Supporting Information

Climate matching drives spread rate but not establishment success in recent unintentional bird introductions

Pedro Abellán, José L. Tella, Martina Carrete, Laura Cardador and José D. Anadón

doi:10.1073/pnas.1704815114

E-mail: pabellan@qc.cuny.edu

SI Materials and methods

Dataset. Data on introduced birds in the Iberian Peninsula were obtained from a comprehensive database of exotic birds in Spain and Portugal (1), which compiles records of exotic species observed in the wild in both countries from 1912 to 2012. This dataset, which is based on a systematic review of national peer-reviewed journals, national and regional bird atlases, books, regional ornithological yearbooks and monographs, websites that compile bird observations or photographs of birds in Spain and Portugal, as well as our own data and unpublished observations from a number of ornithologists or researchers (see dataset details in ref. 1), includes over 11,200 records of exotic birds in the Iberian Peninsula. The location of each record was assigned to 0.05×0.05 arc degree cells (approx. 5×5 km). When coordinates were not provided in the literature source, the record was georeferenced from the locality description using Google Earth software. Those records lacking sufficient information for adequate georeferencing (16% of the total of records) were not further used for spatial analyses.

According to this database, a total of 335 exotic birds have been recorded by observers in the wild in the Iberian Peninsula during this time period, although a large number corresponds to species only observed occasionally (often a single record). To avoid a bias toward anecdotic introductions, we focused only on those species with at least five georeferenced records. Obligate brood parasites (*Vidua* genus) and domestic varieties were also excluded. Our final dataset consisted of 107 bird species, and all analyses were subsequently restricted to these species (see Dataset S1). Established species were those that have established self-sustaining populations or, at least, whose reproduction in the wild has been regularly verified (n=26; see ref.1). It should be noted that this classification might be biased towards non-established species since recently introduced species were classified as non- established, although they can currently or in a near future establish breeding populations, but not enough time has passed to allow it (1).

Only some of those species that successfully establish self-sustaining populations are able to spread into locations away from the point of introduction, where individuals survive and reproduce, becoming invasive (2). For the subset of established species, we estimated the increase through time in the number of occupied 5x5km cells, as an estimate of the rate of spatial spread (Fig. S2). Although the increase in the number of new occupied cells per year may not be just the result of a population spread process, but can also be partially influenced by the release/escape of new individuals (as it seems to be the case of *Cyanoliseus patagonus*, for which there is little evidence of reproduction but of many escape events, Fig. S2), it provides a reliable estimate to further investigate factors explaining differences in spread rate among species. In the case of *Threskiornis aethiopicus*, the two Spanish breeding nuclei were eradicated and current spread comes from the nearby French invasive populations; the rest of species show clear spread from single or multiple introduction sites, with little contribution of new escapes/releases (Authors unpublished information, Fig. S2). Additionally, we also calculated the degree of invasion or spread by quantifying the number of occupied 5x5km cells in the study area as a measure of invaded range size.

Event-level factors. We considered several factors related to introduction history, propagule pressure and origin of species. Years since first introduction (i.e. the number of years since the species was first recorded as introduced relative to 2012) was used as a variable reflecting introduction history as well as the number of years with records and the range of years with records in our dataset. Because these three variables were highly correlated (Pearson correlation coefficients r > 0.75) we retained just the time since first introduction. Propagule pressure was estimated as the total number of live birds reported by CITES that have been legally traded from each of the native ranges to Spain and Portugal from 1976 (the first year for which CITES have been compiling records) to 2012 (www.cites.org). Although this measure is more accurate for species listed in CITES appendixes (i.e., parrots), it is the only information available for all species included in the study and represents the minimum number of individuals traded, so results are conservative. Furthermore, because importation data does not include those birds that have been bred in captivity in Spain and Portugal which could also escape or be released into the wild, we also used the number of introduction (or escape) events as a proxy of propagule pressure. For each species, the number of introduction events was estimated using graph theory from the geographic locations of their records. The *igraph* package (3) in R (4) was used to obtain a network in which any two nodes (georeferenced records) were deemed connected by an edge if they were separated by a geographic distance lower than 100 km. The number of isolated or non-connected sub-networks present in the whole graph was assimilated to the number of introduction events. The assumption here was that those records more distant from each other than this distance threshold represent (or come from) independent introduction events. Alternative thresholds (50 and 150 km) were also considered, providing highly correlated values ($r > 0.73$). Finally, for the most commonly traded cage species (Psittaciformes and Passeriformes), we used information of their abundance in the pet market (i.e. market availability) from Carrete & Tella (5) as an additional surrogate of propagule pressure for this subset of species. It should be noted that under our approach "event-level factors" refer to the global colonization or invasion process in the study area, and not to each one of the introduction events.

On the other hand, Carrete & Tella (5) have shown that, among exotic cage bird species, those that were caught in the wild and traded on the pet market were more successful invaders than those that were bred in captivity over a long period of time. The reason is that the ability to cope with important aspects such as predators in new environments seems to have been lost in captive-bred species (6, 7). We here also tested this hypothesis. For this purpose, passerines (songbirds) and parrots (orders Passeriformes and Psittaciformes, respectively), which numerically encompass most of pet or cage birds (5), were classified as wild-caught or captive-bred species according to their main origin (see classification in ref. 5).

Location-level factors. We assessed the degree to which the introduced region resembles the species' native range (i.e. climate matching between the region of introduction and origin of the species). Methods for quantifying the environmental niche and estimating niche differences typically rely on either direct comparisons of species–environment relationships in environmental space (usually using ordination techniques) or model predictions (8). We adopted here both approaches to assess differences in native and introduced environmental niches for each species. Vector range maps were obtained from the BirdLife International's World Bird Database (9). For each species, we excluded the areas occupied only during the non-breeding season or during migration (10). Georeferenced native range occurrences were obtained from GBIF (Global Biodiversity Information Facility, www.gbif.org). Species' records with lower resolution than that of climatic layers (see below) were discarded. Our final dataset comprised over 1,200,000 record locations for 107 species.

First, we measured niche overlap between native and invaded ranges using the same approach as Broennimann et al. (8) that allows calculating niche overlap while correcting for differences in the availability of environments between study areas. This involves (1) the application of multivariate analyses to summarize environmental variables in a one- or two-dimensional environmental space, (2) calculation of the density of occurrences across the environmental space, and (3) measurement of niche overlap across the environmental space. A principal components analysis (PCA) was calibrated using global climate conditions from the 19 bioclimatic variables of the Worldclim database (11) at a 5 arc minutes spatial resolution. The first two PCA axes, which accounted for 77.6% of the variation in the data, were then used to create a gridded climate space of 100 × 100 square cells, in which each cell corresponds to a unique set of global climate conditions. The first axis mostly correlates with temperature and the second with precipitation variables. Then, we used a kernel density function to convert native occurrences of each species and the available climate in the native range (i.e., background conditions) into densities in order to correct for environmental availability (8). Background conditions were defined by those spatial grid cells within the species' breeding range, as obtained from BirdLife database, with a buffer area of 400 km. The same approach was used for all grid cells within the Iberian Peninsula, representing the region of introduction. Two metrics of niche overlap, Schoener's D and Hellinger's distance (I), as proposed by Warren et al. (12), were calculated from the occupancies in the environmental space depicted by the two first axes of the PCA (8, 12). Both indices indicate the overall match between two niches over the whole climatic space and varies between 0 (no overlap at all) and 1 (complete overlap). In niche comparisons, both species' native range and Iberian conditions were projected along the axes of the PCA calibrated on the global climate space. Thus, the environmental space remained constant allowing a more robust comparison across species. It should be noted here that our measure of niche similarity refers to the whole target region (i.e., the Iberian Peninsula). Nevertheless, using the records for each species in the invaded region (i.e. the 5x5km cells in which the species has been recorded) provided niche overlap measures highly correlated with those using all grid cells of the Iberian Peninsula (r = 0.91) for the subset of species (n = 89) that had enough records to perform the approach of Broennimann et al. (8).

An alternative measure of niche climatic matching was calculated as the distance in the environmental space between the centroid of species' scores along PCA axes (an estimate of the centre of the distribution for each species along an axis) and the centroid of Iberian conditions (scores along PCA axes). Niche metrics were calculated using the 'ecospat' R package (13).

Second, we also used species distribution models (SDMs) to measure niche overlap between native and invaded ranges. In this case, niche overlap metrics were calculated using the environmental space as defined by the gradient of predicted probabilities obtained with SDMs (8) using the 'ecospat' library. SDMs were calibrated using occurrence data in each native range and a set of bioclimatic variables as predictors, using the Maxent modelling algorithm (14) as implemented in the R package 'dismo' (15). The Maxent algorithm has been shown to perform better than other correlative methods that use presence and background data (16). While it is generally accepted that the potential range distribution of invasive species is best addressed using its global coverage (17), we only used native information for model calibration because the focus of this study was not to produce the best-fit model in the invasive range, but to determine the environmental similarity between the two native and the invasive ranges. As climatic predictors we used 7 bioclimatic variables (obtained from Worldclim database) commonly used in avian distribution modelling to denote bioclimatic controls (e.g. (18)) representing average and extreme climatic conditions: annual mean temperature, temperature seasonality, maximum temperature of the warmest month, minimum temperature of the coldest month, annual precipitation, precipitation of the driest month and precipitation seasonality. Ten replicate niche models were then constructed for each species within its native range and, in each model, all native occurrence records were partitioned randomly into a calibration set (training set = 70% of all records) and a testing set (validation set = 30%). We selected 500 iterations for model convergence and employed the regularization procedure to prevent overfitting.

Potential bias in presence records was addressed by creating bias grids for use in Maxent modeling, following procedures outlined by Elith et al. (19), which has been shown to provide more accurate model predictions (20). For each species, we produced a bias grid using information on surveyed grids retrieved from GBIF for all species belonging to the same genus. We derived kernel density maps of sampled grids and rescaled it from 1 to 20 to reduce numeric disparities (19), so cell values reflected sampling effort and gave a weight to random background data used for modeling. Species' breeding range, as obtained from BirdLife database (21), with a buffer area of 400 km, were used as background area. Performance of Maxent models was assessed using the Continuous Boyce Index (22), which was especially designed for presence only models. It is obtained by plotting the relationship between the predicted number of evaluation points (P) and the number of points expected by a random distribution (E) along the gradient of habitat suitability (HS) scores. The index, defined as the Spearman's r between P/E and HS, ranges from 0 (random) to 1 (perfect prediction), with negative values indicating models that predict worse than random. Results of Maxent models are summarized in Dataset 1.

Because results using the different measures of climatic matching were qualitatively congruent, we only report the outputs for the measure of niche overlap between native and invaded ranges using Schoener's D from the PCA procedure for further analyses, while results for the SDM approach, I index and centroid distances are provided in the

Table S4. Additionally, we explored the role of temperature alone in climate matching, as it could be argued that temperature should represent the main direct climatic constraint on the invasion success. Precipitation is likely to be more related to food supply which is often also affected by other factors (e.g. urbanization) and could mask the role of climate-matching as some exotic birds may establish in human settlements. Because the first PCA axis was mostly correlated with temperature (e.g. $r = 0.96$ with minimum temperature of the coldest month), we calculated niche overlap along this axis following the approach proposed by Broennimann et al. (8) and outlined above. We found that niche overlap measures considering only temperature were highly correlated with those measures using both PCA axes when using the whole target region ($r = 0.9$, $p < 0.001$) and when using the records for each species in the invaded region ($r = 0.76$, $p < 0.01$) for the subset of species that had enough records to perform the approach of Broennimann et al. (8) (n = 89).

Species-level factors We considered several factors related to species traits:

Relative brain size. Brain mass data were obtained from different literature sources (see reference list in Dataset S1). We compiled brain size information for a total of 1357 bird species (both species introduced and species that have never been introduced), including 74 of those in our dataset of exotic birds. We used actual brain mass, where available, but we also included cranial endocast measures converted to mass by multiplying the reported value by the density of fresh brain tissue (1.036 g/ml) (ref. 23 and references therein). To control for the allometric effect of body size on brain size, we used the residuals of log-log regressions against body mass (relative brain size, hereafter). Body mass was obtained from the same sources as was brain mass and the residuals were taken from a log-log regression on all available species. The relationship between brain size and body mass was strongly positive and linear (linear regression, R2 = 0.84; slope ± SE, 0.580±0.007; P < 0.0001), and the residuals were unrelated with body mass (R2 = 0.0007). Then, for those species in our dataset of exotic birds for which brain mass was not available (n=33), we used the average brain residual of the species from the same genus (24).

Life history strategies. We collected information for a set of life-history traits, namely: clutch size, number of broods per year, fecundity, egg mass, incubation period, fledgling period, lifespan and age at first breeding (sources detailed in Supplementary Information) to estimate the fast-slow continuum of life-history strategies of the different species. Following Bielby et al. (25; see also ref 24), we used a factor analysis to simplify the pattern of covariation among traits by positing latent variables underlying the data. In order to avoid potential biases in the estimates, information for both species introduced and species that have never been introduced was compiled. A total of 253 species, for which information was available for all the eight traits, were used in the factor analysis, including 52 of those in our dataset of exotic birds. The confounding effect of body size (26) was removed by regressing life-history variables on body size after log transformation, using ordinary least squares, and computed residuals for use in the factor analysis. The first factor was retained as an estimate of the fast-slow continuum (44% of variance explained). However, because clutch size (i.e. the residuals of log-log regression against body mass) was highly correlated with this derived variable (r = 0.91) and was available for all of the target species, we used it as surrogate of the fast-slow continuum in order to maximize the number of species included in the analysis.

Recently, a wide cross-species comparison in birds (24) suggested that rather than a fast or slow strategy, a bet-hedging life history characterized by delayed reproduction and longer lifespan is linked to invasion. In order to test this hypothesis, and following Sol et al. (24), we computed for each species its brood value (27), which accounts for the ability of species to prioritize current survival over future reproduction, expressed as log10(1/[number of broods per year x reproductive life span]). For species for which either or both of these parameters were unavailable (n = 25), values were extrapolated from the mean for congeners (24). Variation in brood value was poorly correlated with the fast-slow continuum, so providing a different life-history trait. Furthermore, as an additional proxy of life history strategy, we also explored body mass (mean of male and female masses), as obtained from different sources (21, 28).

Niche breadth. For each species, an estimate of niche breadth was calculated using the area of the PCA envelope surrounding the native distribution points in the global PCA climate space (see above) after excluding the 5% of most extreme values. Additionally, geographic range size in native areas was also used as a proxy of niche breadth (see e.g. ref. 29). Range size data were obtained from BirdLife International (21).

Migratory status. Species were classified as migratory (i.e. species for which a substantial proportion of the global or regional population makes regular or seasonal cyclical movements beyond the breeding range, with predictable timing and destinations) or non-migratory, according to the information provided by BirdLife International (21).

Modeling invasion success. To test the link between the different predictors and establishment success, we conducted logistic regressions, in which the outcome of the introduction was the dependent variable, taking a value of 0 when the species failed in establishing self-sustaining populations and 1 when it succeeded. Most predictor variables were log transformed to improve compliance with normality, and all the continuous predictors were standardized to allow comparisons among estimates.

Species can generally not be considered independent data points because closely related ones tend to share many morphological, physiological and ecological traits due to their common evolutionary history (30). A previous study (1) has shown that those exotic taxa that have succeed in establishing viable breeding populations in the wild in Spain and Portugal are a non-random subset of all introduced birds in terms of phylogeny, so we performed our analyses in a phylogenetic comparative framework. For comparison, we also performed non-phylogenetic logistic regressions, which assume all species values are independent. Dated phylogenies of all extant bird species have been recently built (31) and sets of 10,000 pseudo-posterior samples of these phylogenies are available for download at http://birdtree.org/. In order to account for phylogenetic uncertainty, we subsampled 1,000 trees (see (31) for further details on phylogeny construction). We began with separate univariate analyses to test the effect of each predictor variable on establishment. We used phylogenetic logistic regression (PLR) (32), as implemented in the R package ‘phylolm’ (33), to assess the relationship between single predictor variables and establishment success. By this approach we were able to assess the influence that each variable had on the probability of establishment independently of the phylogenetic relationships among species. Then, we examined the combined influence of predictor variables on establishment success in phylogenetic multiple logistic regression. We used a multi-model approach based on Akaike’s Information Criterion adjusted for small sample sizes (AICc) to evaluate the parameter estimates and the relative importance of predictor variables in a likelihood-based framework (34). We identified the best model based on AICc and calculated the relative importance of each predictor variable as the sum of the AICc weights of all models that included this variable in the set of most likely models (ΔAICc < 4). We conducted this approach for each of the 1,000 phylogenetic trees, and we report the mean and the central range that contains 95% of parameter estimates. Goodness-of-model fit was evaluated by calculating the explained deviance (D^2) (35). It is not currently possible to obtain D^2 for PLRs (32) so we relied on the results of non-phylogenetic logistic regression.

We also tested the link between the different predictors and the rate of spatial spread in the study area for the subset of established species using phylogenetic generalized least squares (PGLS) models. PGLS models were constructed with the R library ‘nlme’ (36) assuming a Brownian motion model of evolution and with the rate of spatial spread (increment in the number of occupied 5x5km cells over time, log-transformed) as response variable (for comparison, we also performed non-phylogenetic ordinary least squares). Similarly, we also tested how the different factors explained the size of the invaded range (number of occupied 5x5km cells in the study area, also log-transformed). As for PLR, we fitted univariate PGLS models and all possible PGLS multivariate models from predictor variables to identify the most likely models and to calculate Akaike weights and variable importance based on AICc and model averaging.

To evaluate the relative independent effect of the explanatory variables, we performed a hierarchical partitioning (37) on the subsets of variables selected in the final best PLR and PGLS models. To do so we used the ‘hier.part’ package in R (38). Briefly, this analysis splits the variation explained by each predictor into independent and joint effects, allowing assessing the importance of the common effect among variables, and thus accounting for their collinearity. A 1,000-randomization procedure was carried out to test the statistical significance of the independent effects of each predictor (37). Because the ‘hier.part’ analysis does not support PLR nor PGLS models, we relied on the results of non-phylogenetic models.

References

1. Abellán P, Carrete M, Anadón JD, Cardador L, Tella JL (2016) Non-random patterns and temporal trends (1912-2012) in the transport, introduction and establishment of exotic birds in Spain and Portugal. *Divers Distrib* 22(3):263–273.
2. Blackburn TM, et al. (2011) A proposed unified framework for biological invasions. *Trends Ecol Evol* 26(7):333–339.
3. Csardi G, Nepusz T (2006) The igraph software package for complex network research. *InterJournal, Complex Syst* 1695:1–9.
4. Team RDC (2011) *R: a language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna).
5. Carrete M, Tella J (2008) Wild-bird trade and exotic invasions: a new link of conservation concern? *Front Ecol Environ* 6(4):207–211.
6. Cabezas S, Carrete M, Tella JL, Marchant TA, Bortolotti GR (2013) Differences in acute stress responses between wild-caught and captive-bred birds: a physiological mechanism contributing to current avian invasions? *Biol Invasions* 15(3):521–527.
7. Carrete M, Tella JL (2015) Rapid loss of antipredatory behaviour in captive-bred birds is linked to current avian invasions. *Sci Rep* 5:18274.
8. Broennimann O, et al. (2012) Measuring ecological niche overlap from occurrence and spatial environmental data. *Glob Ecol Biogeogr* 21(4):481–497.
9. BirdLife International and NatureServe (2011) *Bird species distribution maps of the world* (BirdLife International, Cambridge, UK and NatureServe, Arlington, USA).
10. Strubbe D, Beauchard O, Matthysen E (2015) Niche conservatism among non-native vertebrates in Europe and North America. *Ecography (Cop)* 38(3):321–329.
11. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25(15):1965–1978.
12. Warren DL, Glor RE, Turelli M (2008) Environmental Niche Equivalency Versus Conservatism: Quantitative Approaches To Niche Evolution. *Evolution (N Y)* 62(11):2868–2883.
13. Broenniman O., Petitpierre B, Ran C (2014) ecospat: Spatial ecology miscellaneous methods. R package version 1.0.
14. Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecol Modell* 190(3):231–259.
15. Hijmans RJ, Phillips S, Leathwick J, Elith J (2013) dismo: Species distribution modeling. R package version 0.9-1.
16. Ferrier S, et al. (2006) Novel methods improve prediction of species ' distributions from occurrence data. 2(January).
17. Broennimann O, Guisan A (2008) Predicting current and future biological invasions: both native and invaded ranges matter. *Biol Lett* 4(5).
18. Cardador L, Carrete M, Gallardo B, Tella JL (2016) Combining trade data and niche modelling improves predictions of the origin and distribution of non-native European populations of a globally invasive species. *J Biogeogr* 43(5):967–978.
19. Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. *Methods Ecol Evol* 1(4):330–342.
20. Hertzog LR, Besnard A, Jay-Robert P (2014) Field validation shows bias-corrected pseudo-absence selection is the best method for predictive species-distribution modelling. *Divers Distrib* 20(12):1403–1413.
21. BirdLife International (2012) Species factsheets. Available at: http://www.birdlife.org [Accessed February 1, 2016].
22. Hirzel AH, Le Lay G, Helfer V, Randin C, Guisan A (2006) Evaluating the ability of habitat suitability models to predict species presences. *Ecol Modell* 199(2):142–152.
23. Iwaniuk AN, Nelson JE (2002) Can endocranial volume be used as an estimate of brain size in birds? *Can J Zool* 80(1):16–23.
24. Sol D, et al. (2012) Unraveling the Life History of Successful Invaders. *Science (80-)* 337(6094):580–583.
25. Bielby J, et al. (2007) The fast-slow continuum in mammalian life history: an empirical reevaluation. *Am Nat* 169(6):748–57.
26. Jeschke JM, Kokko H (2009) The roles of body size and phylogeny in fast and slow life histories. *Evol Ecol* 23(6):867–878.
27. Bókony V, et al. (2009) Stress Response and the Value of Reproduction: Are Birds Prudent Parents? *Am Nat* 173(5):589–598.
28. Dunning J (2008) *CRC Handbook of Avian Body Masses* (CRC Press, Boca Raton). 2nd editio.
29. Blackburn TM, Duncan RP (2001) Determinants of establishment success in introduced birds. *Nature* 414(6860):195–197.
30. Felsenstein J (1985) Confidence-Limits on Phylogenies - an Approach Using the Bootstrap. *Evolution (N Y)* 39(4):783–791.
31. Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO (2012) The global diversity of birds in space and time. *Nature* 491(7424):444–448.
32. Ives AR, Garland T (2010) Phylogenetic Logistic Regression for Binary Dependent Variables. *Syst Biol* 59(1):9–26.

33. Ho LST, Ané C (2013) Phylolm: Phylogenetic Linear Regression. R package version 3.1-102.
34. Burnham KP, Anderson DR (2003) *Model Selection and Multimodel Inference* (Springer, New York). 2nd Ed.
35. Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecol Modell* 135(2):147–186.
36. Pinheiro J, Bates D, DebRoy S, Sarkar D (2016) Nlme: linear and nonlinear mixed effects models. R package version 3.1-102.
37. Mac Nally R (2002) Multiple regression and inference in ecology and conservation biology: further comments on identifying important predictor variables. *Biodivers Conserv* 11(8):1397–1401.
38. Walsh C, Mac Nally R (2005) hier.part: hierarchical partitioning. R package version 1.0-1.

Fig. S1. Hierarchical partitioning showing independent effects of different factors on spatial spread in established species. The left and right panel show variance explained by the subset of variables selected in the best PGLS models on the degree of spatial spread (number of occupied 5x5km cells) for all established species and for the subset of cage species, respectively. The asterisk denotes significance at the *P* < 0.05 level.

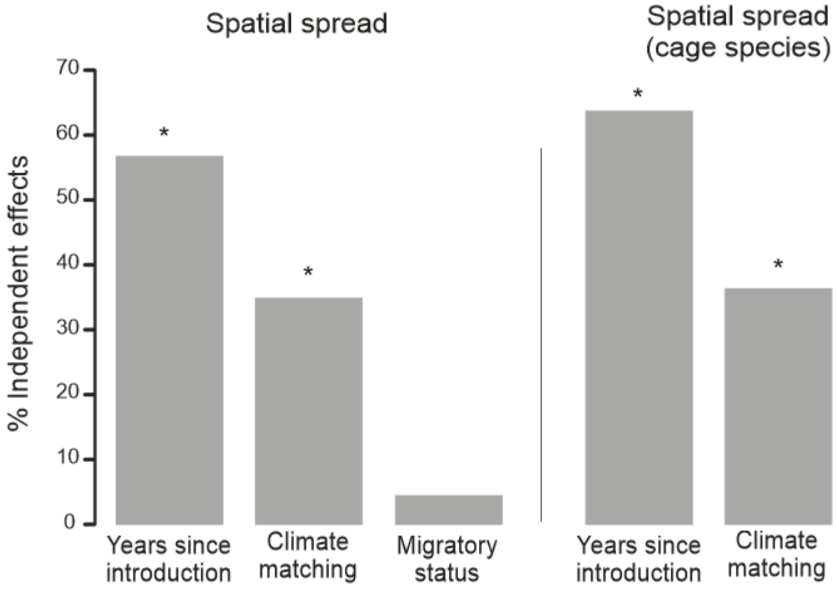
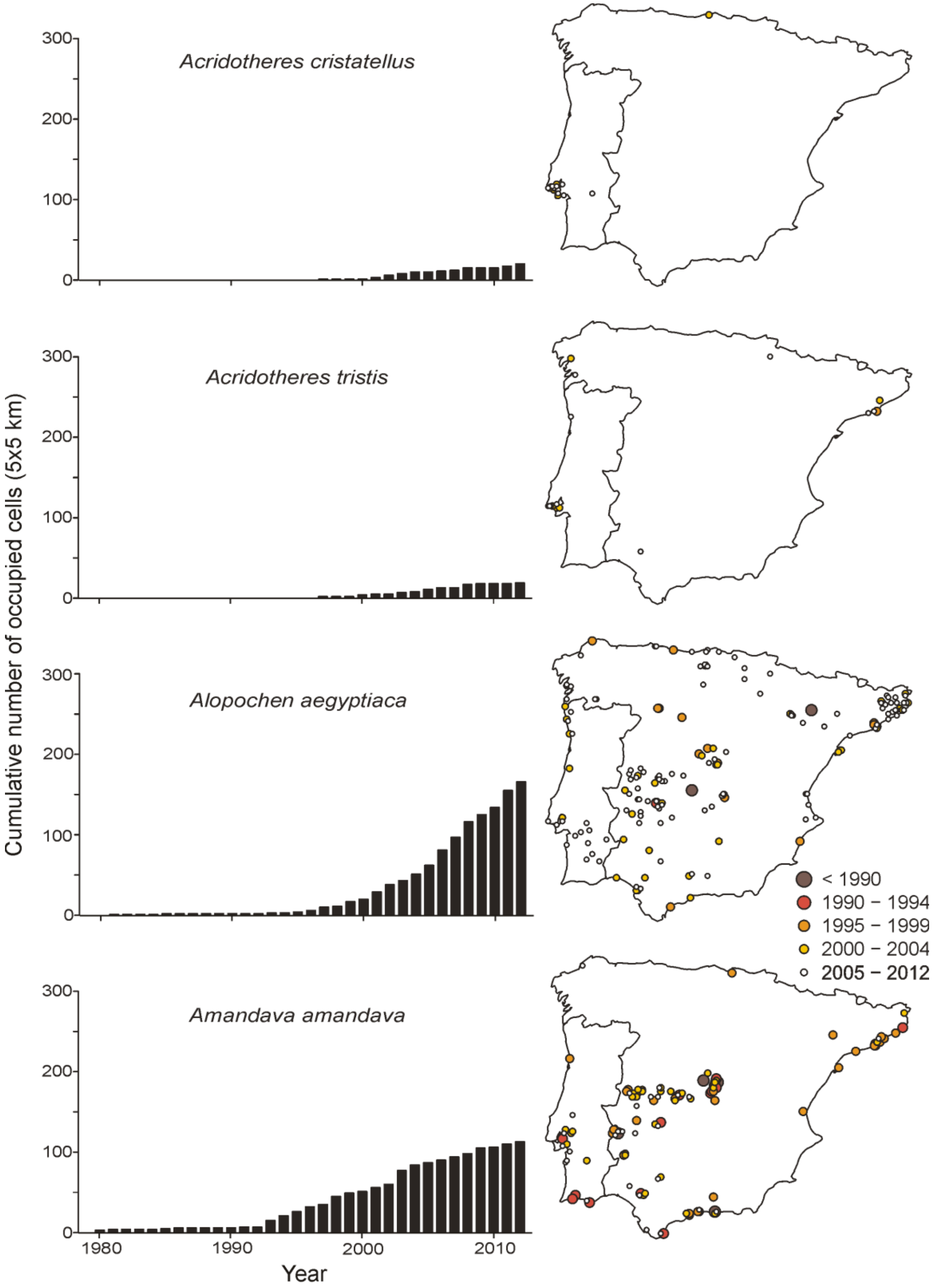
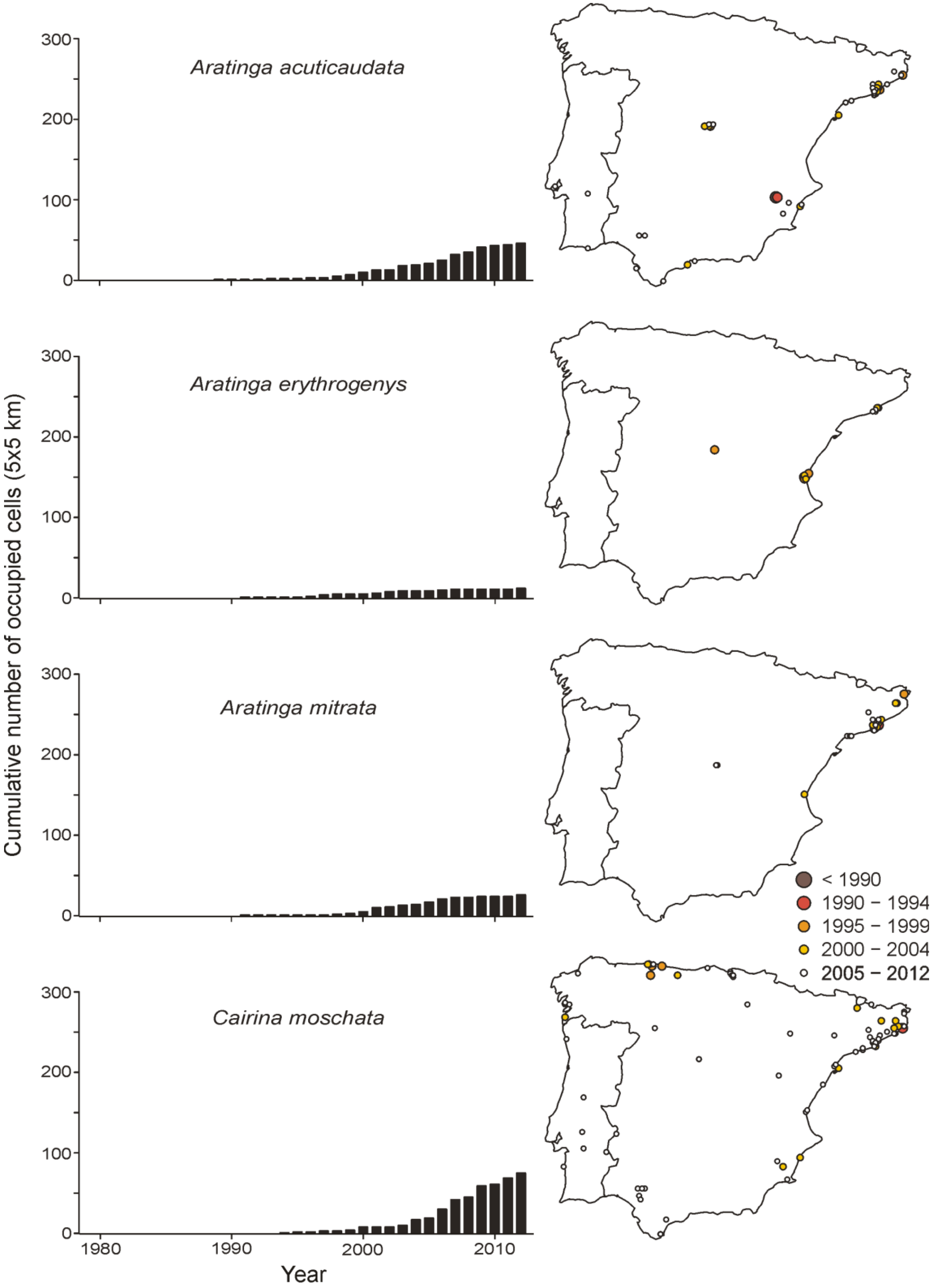
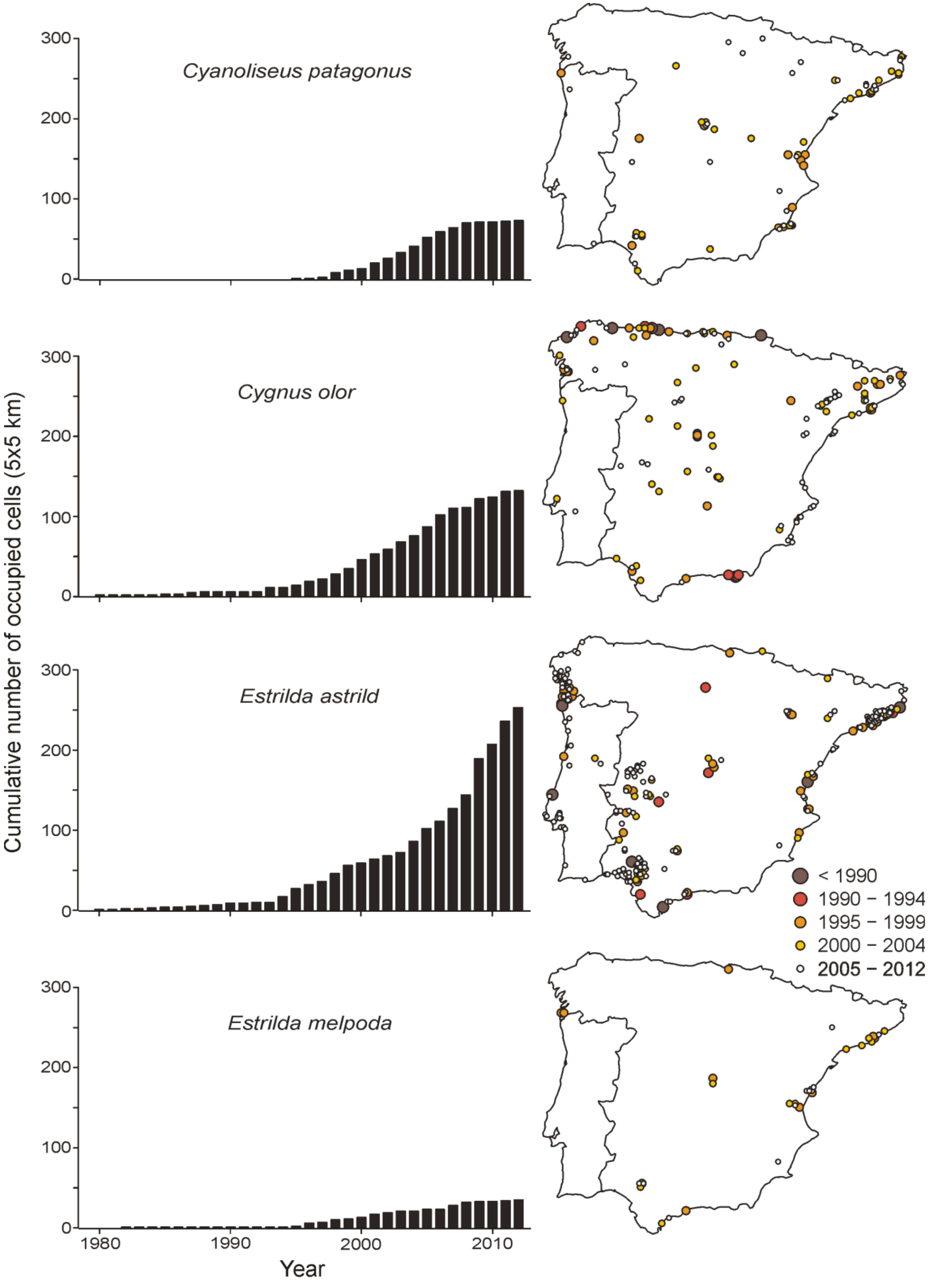


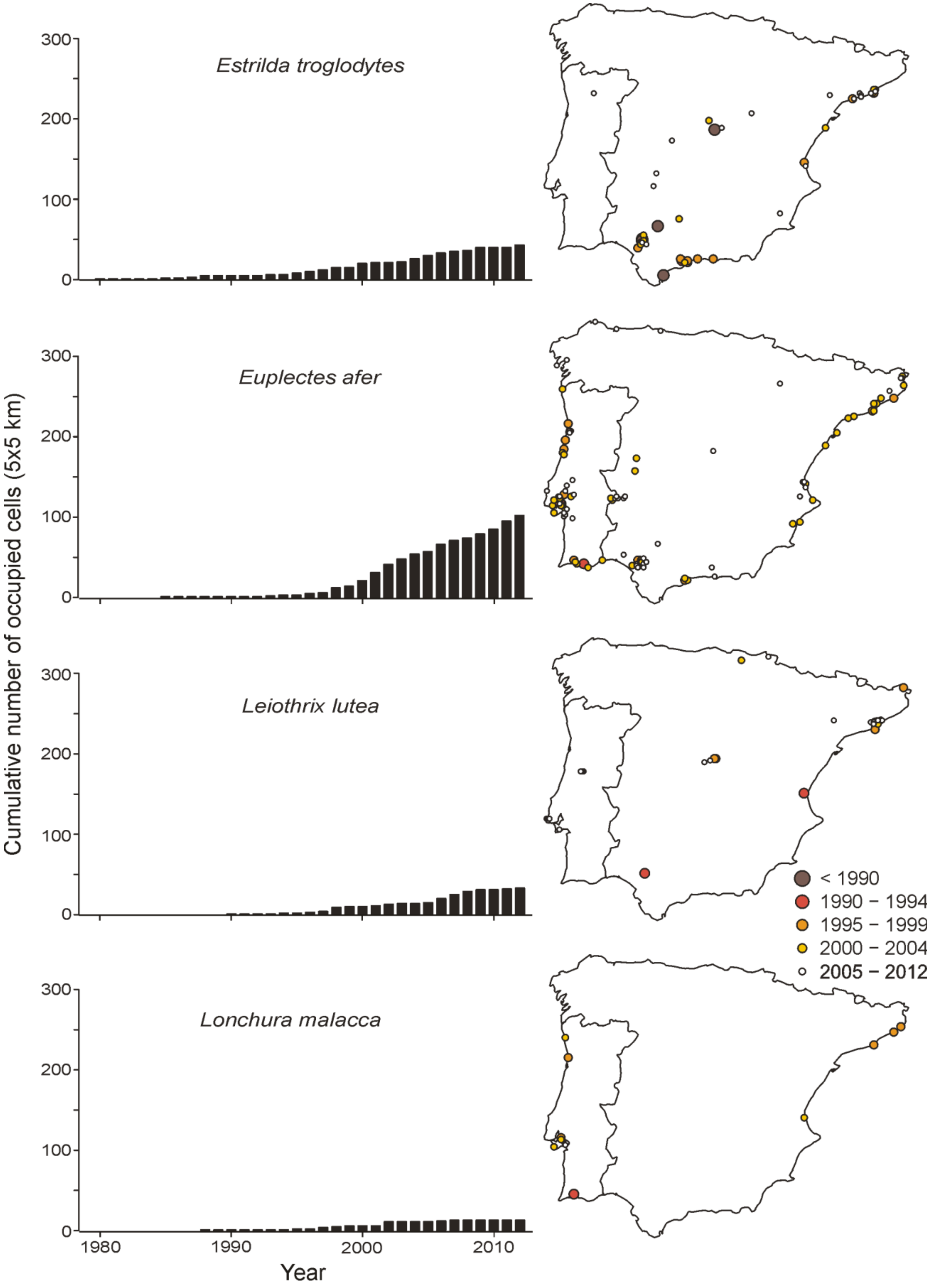
Fig. S2. Temporal evolution of the spatial distribution of exotic bird species established in the Iberian Peninsula. The graph on the left shows the cumulative number of 5x5 km cells over time. The map (right) shows, for each cell, the year in which the species was first recorded. See *SI Materials and methods* for additional information on particular species.

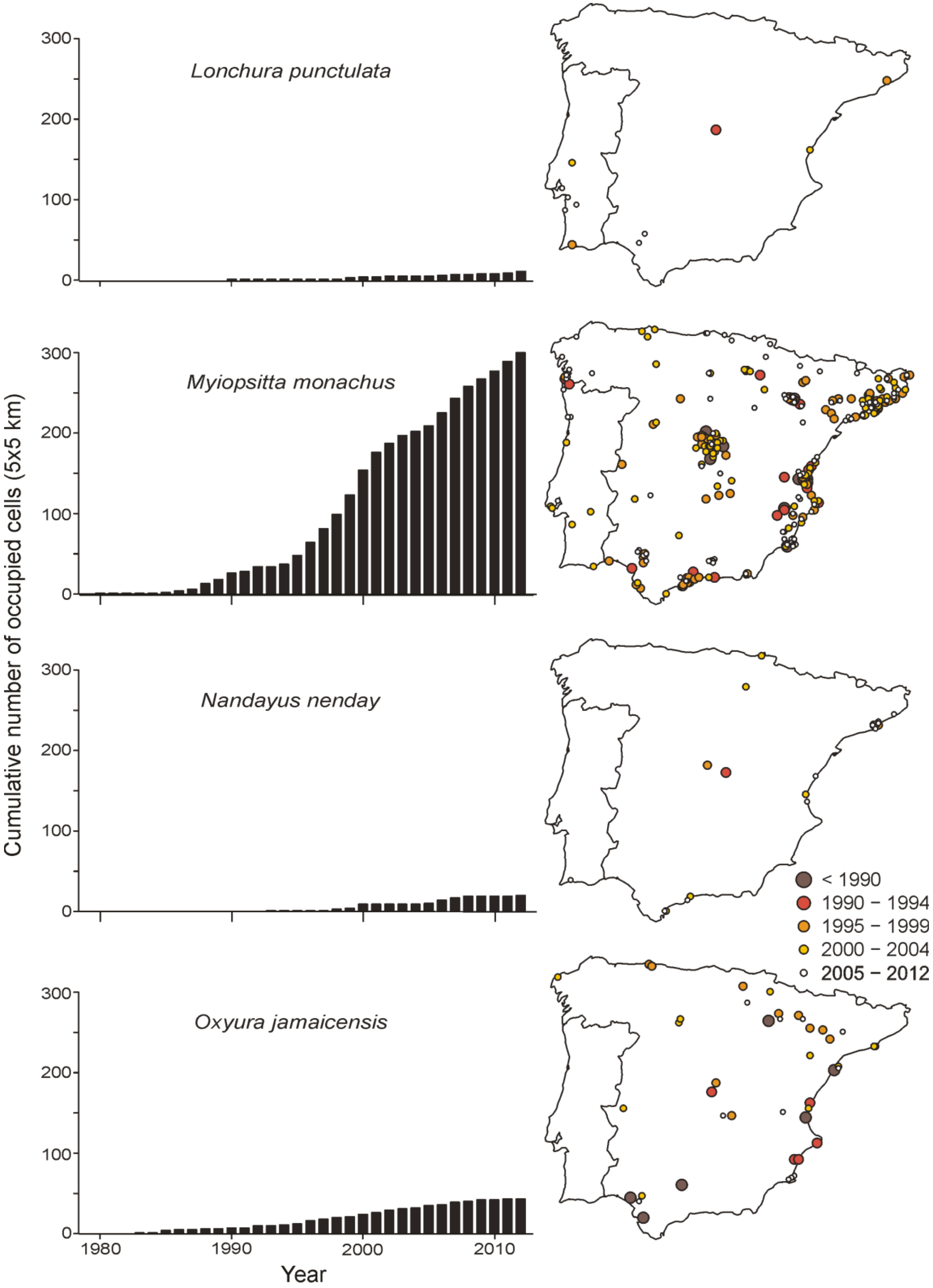




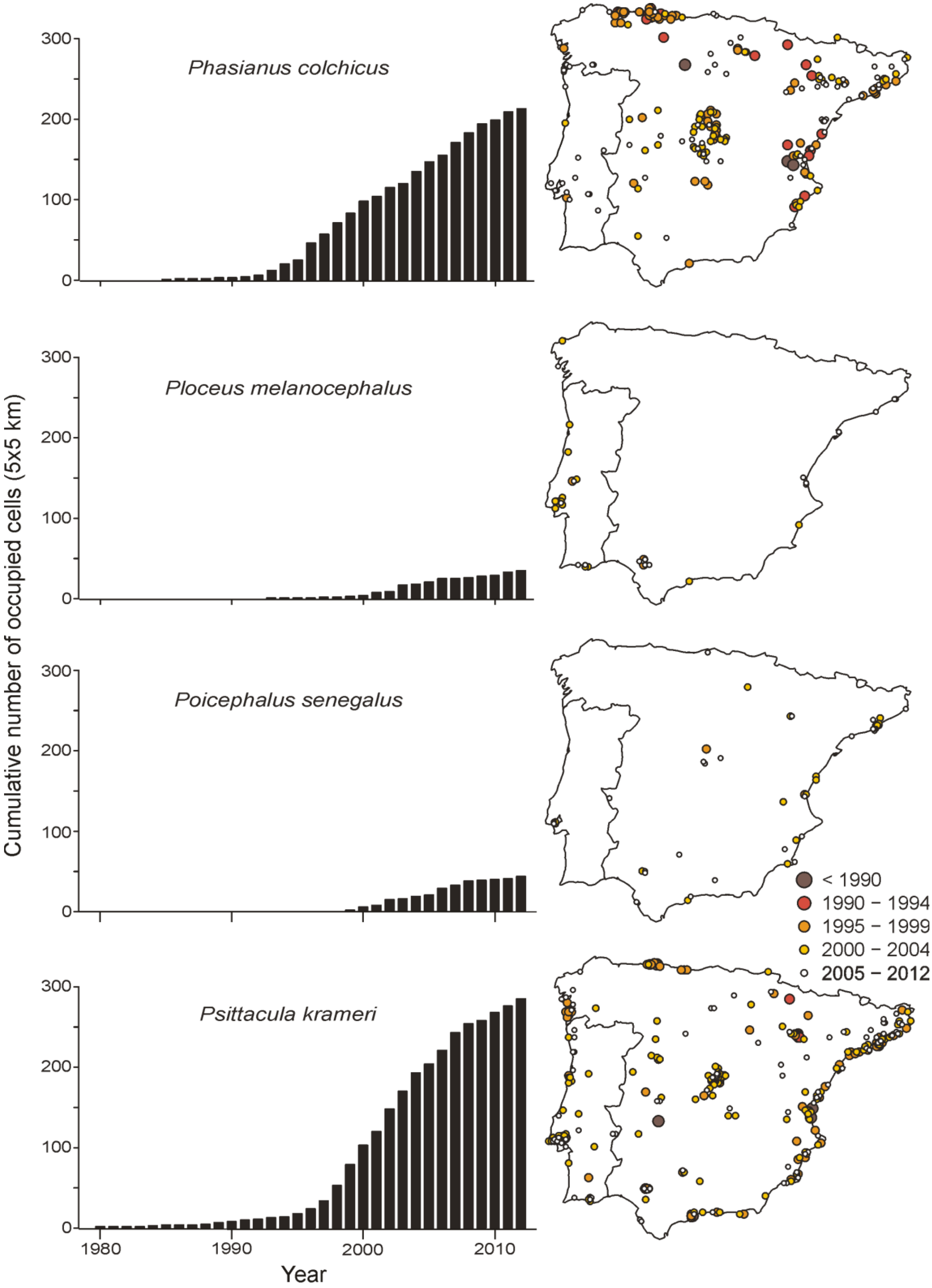
Cumulative number of occupied cells (5x5 km)
Aratinga acuticaudata
Aratinga erythrogenys
Aratinga mitrata
Cairina moschata
300
200
100
0
1980
1990
2000
2010
Year
< 1990
1990 – 1994
1995 – 1999
2000 – 2004
2005 – 2012

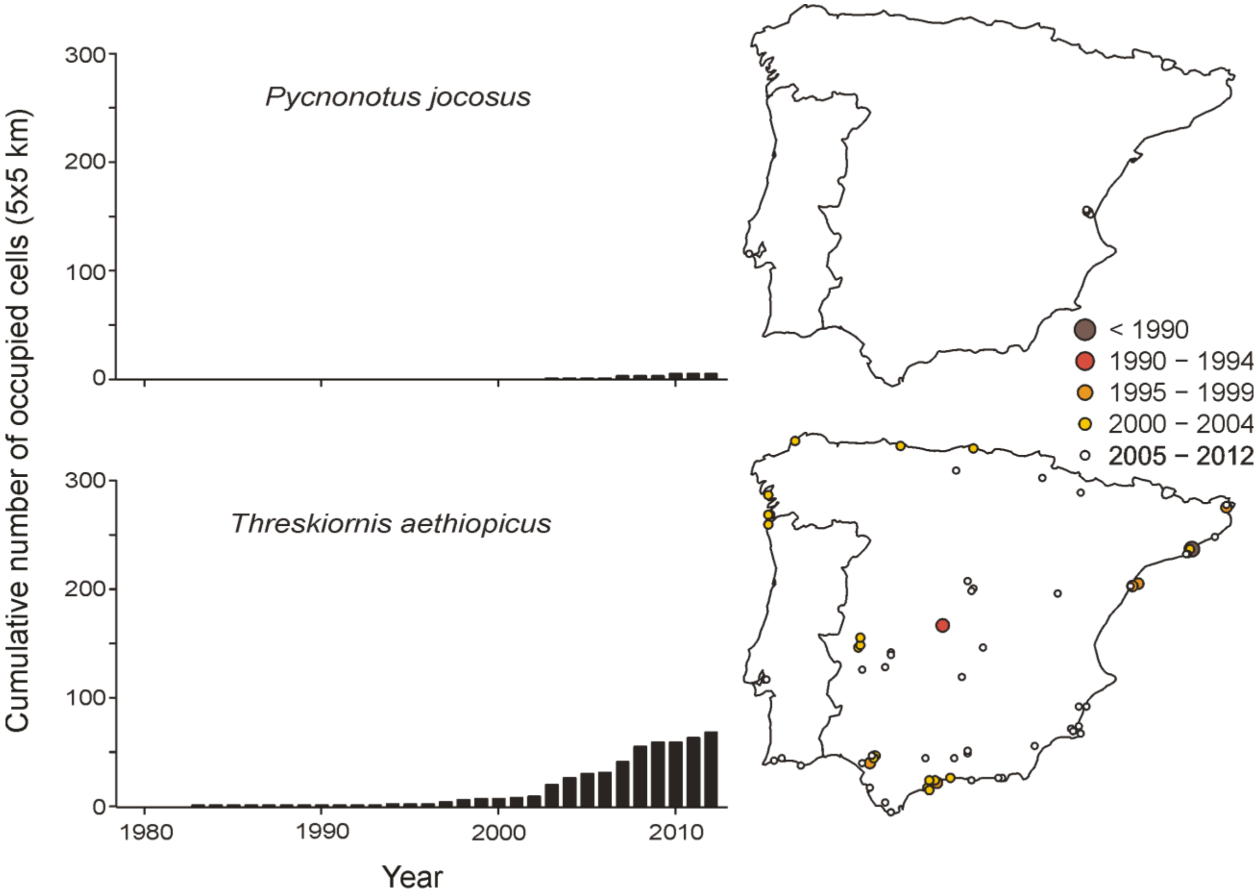

Cyanoliseus patagonus
Cygnus olor
Estrilda astrild
Estrilda melpoda
Cumulative number of occupied cells (5x5 km)
0
100
200
300
1980
1990
2000
2010
Year
< 1990
1990 – 1994
1995 – 1999
2000 – 2004
2005 – 2012


Estrilda troglodytes
Euplectes afer
Leiothrix lutea
Lonchura malacca
Cumulative number of occupied cells (5x5 km)
300
200
100
0
1980
1990
2000
2010
Year
< 1990
1990 − 1994
1995 − 1999
2000 − 2004
2005 − 2012



Lonchura punctulata
Myiopsitta monachus
Nandayus nenday
Oxyura jamaicensis
Cumulative number of occupied cells (5x5 km)
0
100
200
300
1980
1990
2000
2010
Year
< 1990
1990 – 1994
1995 – 1999
2000 – 2004
2005 – 2012


Phasianus colchicus
Ploceus melanocephalus
Poicephalus senegalus
Psittacula krameri
Cumulative number of occupied cells (5x5 km)
300
200
100
0
1980
1990
2000
2010
Year
< 1990
1990 – 1994
1995 – 1999
2000 – 2004
2005 – 2012



Cumulative number of occupied cells (5x5 km)
300
200
100
0
Pycnonotus jocosus
Threskiornis aethiopicus
1980
1990
2000
2010
Year
< 1990
1990 – 1994
1995 – 1999
2000 – 2004
2005 – 2012

Table S1. Results of the non-phylogenetic logistic regressions testing the relationship between the different predictors and establishment success.

	Univariate models		**Best model**		
Variables	**Estimate**	**D^2**	**Estimate**	**D^2**	**Σw_{AIC}**
Event-level factors					
Years from introduction	0.783**	0.094	1.037**	0.29	1.00
Introduction events	0.787**	0.074	0.612		0.54
Number of imported birds	0.627**	0.066	0.881**		1.00
Species-level factors					
Clutch size	0.165	0.004			0.39
Brood value	0.268	0.013	0.549		0.53
Body mass	-0.406	0.026			0.13
Range size	0.077	0.001			0.13
Brain size	0.398	0.025	0.669		0.56
Niche breadth	0.355	0.023	0.869*		1.00
Migrant	-0.673	0.018			0.15
Location-level factors					
Climate matching	-0.155	0.004			0.14

Standardized regression coefficients are showed for univariate models and for the best multivariate model based on Akaike's Information Criterion adjusted for small sample sizes (AICc). Relative importance of variables (ΣwAIC) based on a multi-model information-theory based approach is shown. Variables with strong support (ΣwAIC ≥ 0.8) are shown in bold. * P ≤ 0.05; ** P ≤ 0.01; *** P ≤ 0.0001. Goodness-of-model fit, as evaluated by calculating the explained deviance (D2) is also provided.

Table S2. Results of the ordinary (non-phylogenetic) generalized least squares models testing for the link between the different predictors and the rate of spatial spread in established species.

	Univariate models		**Best model**		
Variables	**Estimate**	**R^2**	**Estimate**	**R^2**	**Σw_{AIC}**
Event-level factors					
Years from introduction	0.363*	0.181	0.253*	0.	**0.82**
Introduction events	0.379*	0.203	0.177		0.39
Number of imported birds	0.137	0.000			0.57
Species-level factors					
Clutch size	0.202	0.067			0.40
Brood value	0.148	0.035			0.19
Body mass	0.192	0.021	0.260		0.75
Range size	0.194	0.022			0.21
Brain size	-0.099	0.016			0.40
Niche breadth	-0.146	0.034			0.17
Migrant	-0.182	0.010	-0.808**		**0.84**
Location-level factors					
Climate matching	0.475**	0.342	0.423**		**1.00**

Standardized regression coefficients are showed for univariate models and for the best multivariate model based on Akaike's Information Criterion adjusted for small sample size (AICc). Relative importance of variables (Σw_{AIC}) based on a multi-model information-theory based approach is shown. Variables with strong support ($\Sigma w_{AIC} \geq 0.8$) are shown in bold. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.0001$. Goodness-of-model fit (R^2) is also provided.

Table S3. Results of the phylogenetic generalized least squares models testing the link between the different predictors and the degree of spatial spread (number of occupied grid cells) of established species.

	Univariate models		**Best model**		
Variables	**Coefficients**	**R^2**	**Coefficients**	**R^2**	**Σw_{AIC}**
Event-level factors					
Years since introduction	**0.351 (0.297, 0.408)**	0.44	**0.237 (0.216, 0.258)**	0.60	**1.00 (1.00, 1.00)**
Introduction events	0.035 (0.000, 0.118)	0.18			0.06 (0.05, 0.09)
Imported birds	0.137 (0.072, 0.256)	0.01			0.08 (0,06, 0.18)
Species-level factors					
Clutch size	0.064 (-0.121, 0.118)	0.02			0.06 (0.05, 0.11)
Brood value	0.001 (-0.048 - 0.054)	0.00			0.06 (0.04, 0.09)
Body mass	-0.313 (-0.450 - -0.151)	0.00			0.07 (0.05, 0.10)
Range size	0.094 (0.052 - 0.157)	0.00			0.08 (0.06, 0.14)
Brain size	-0.264 (-0.370 - -0.187)	0.00			0.10 (0.07, 0.20)
Niche breadth	-0.104 (**-0.127** - -0.063)	0.06			0.07 (0.05, 0.09)
Migrant	-0.157 (**-0.349** - -0.100)	0.00	**-0.263 (-0.330, -0.206)**		**0.91 (0.55, 1.00)**
Location-level factors					
Climate matching	**0.245 (0.221 - 0.268)**	0.28	**0.191 (0.169, 0.207)**		**1.00 (1.00, 1.00)**

Standardized regression coefficients are showed for univariate models and for the best multivariate model based on Akaike's Information Criterion adjusted for small sample sizes (AICc). Relative importance of variables (Σw_{AIC}) based on a multi-model information-theory based approach is shown. Variables with the strongest support ($\Sigma w_{AIC} \geq 0.8$) are shown in bold. Regression coefficients and Σw_{AIC} values represent the median and the central range that contains 95% of values for 1,000 phylogenetic trees. The parameter estimates that are significantly different from zero ($P < 0.05$) are indicated by bold numbers. Goodness-of-model fit (R^2) from ordinary least squares regression is also provided.

Table S4. Results of the phylogenetic regression models testing for the link between the different measures of climate matching and both establishment success (phylogenetic logistic regressions) and spread rate (phylogenetic generalized least squares) in established species.

Climate matching variables	**Estimates**	**P-values**	**D^2/R^2**
Establishment			
PCA.dist	-0.152 (-0.190, -0.033)	0.523 (0.434, 0.867)	0.00
PCA.D	0.136 (-0.075, 0.215)	0.542 (0.335, 0.998)	0.00
PCA.I	0.000 (-0.157, 0.186)	0.627 (0.408, 0.998)	0.00
Maxent.D	-0.367 (-0.426, -0.303)	0.132 (0.093, 0.206)	0.03
Maxent.I	-0.340 (-0.373, -0.309)	0.148 (0.116, 0.185)	0.03
Spread rate			
PCA.dist	-0.429 (-0.542, -0.310)	0.001 (0.000, 0.014)	0.21
PCA.D	0.475 (0.400, 0.545)	0.000 (0.000, 0.000)	0.34
PCA.I	0.470 (0.382, 0.540)	0.000 (0.000, 0.000)	0.27
Maxent.D	0.392 (0.282, 0.529)	0.011 (0.000, 0.049)	0.22
Maxent.I	0.324 (0.385, 0.463)	0.003 (0.000, 0.010)	0.25

Standardized regression coefficients and P values represent the median and the central range that contains 95% of values for 1,000 phylogenetic trees. Goodness-of-model fit, as evaluated by calculating the explained deviance (D^2) from the non-phylogenetic logistic regression or as coefficient of determination (R^2) from ordinary least squares regression, is also provided. Codes: *PCA.dist*, distance in the environmental space between the centroid of species' scores along PCA axes; *PCA.D* and *PCA.I*, Schoener's D and Hellinger's distance (I), respectively, measured on the occupancies in the environmental space depicted by the two first axes of the PCA; *Maxent.D* and *Maxent.I*, Schoener's D and Hellinger's distance (I), respectively, measured on of Maxent output.

Table S5. Results of the phylogenetic logistic regressions testing the relationship between the different predictors and establishment success in cage species (orders Passeriformes and Psittaciformes).

	Univariate models		**Best model**		
Variables	**Coefficients**	**D^2**	**Coefficients**	**D^2**	**Σw_{AIC}**
Event-level factors					
Years since introduction	**0.783 (0.771, 0.798)**	0.11		0.34	0.29 (0.21, 0.48)
Introduction events	**0.794 (0.757, 0.829)**	0.07	**0.961 (0.888, 1.001)**		**0.86** (0.70, **0.92)**
Imported birds	**0.747 (0.719, 0.769)**	0.05			0.51 (0.46, 0.60)
Market availability	-0.026 (-0.026, 0.107)	0.00			0.09 (0.05, 0.18)
Wild-caught/captive-bred	17.381 (16.592, 18.433)	0.07	17.812 (16.841, 18.623)		**1.00 (0.86, 1.00)**
Species-level factors					
Clutch size	0.280 (0.269, 0.345)	0.02			0.42 (0.31, 0.60)
Brood value	0.353 (0.247, 0.356)	0.02			0.09 (0.00, 0.25)
Body mass	-0.235 (-0.306, -0.223)	0.01			0.07 (0.03, 0.13)
Range size	0.093 (0.078, 0.162)	0.00			0.12 (0.02, 0.20)
Brain size	0.055 (0.051, 0.137)	0.00			0.12 (0.04, 0.31)
Niche breadth	0.542 (0.420, 0.593)	0.04	0.657 (0.629, 0.681)		**0.83 (**0.67, **0.96)**
Location-level factors					
Climate matching	-0.042 (-0.048, 0.006)	0.00			0.05 (0.04, 0.08)

Standardized regression coefficients are showed for univariate models and for the best multivariate model based on Akaike's Information Criterion adjusted for small sample sizes (AICc). Relative importance of variables (Σw_{AIC}) based on a multi-model information-theory based approach is shown. Variables with strong support ($\Sigma w_{AIC} \geq 0.8$) are shown in bold. Regression coefficients and Σw_{AIC} values represent the median and the central range that contains 95% of values for 1,000 phylogenetic trees. The parameter estimates that are significantly different from zero ($P < 0.05$) are indicated by bold numbers. Goodness-of-model fit, as evaluated by calculating the explained deviance (D^2) from the non-phylogenetic logistic regression, is also provided.