

Published in final edited form as:

Nature. 2019 June 06; 571(7763): 103–106. doi:10.1038/s41586-019-1292-2.

Location-level processes drive the establishment of alien bird populations worldwide

David Redding¹, Alex L. Pigot¹, Ellie E. Dyer¹, Çağrı H. Akçaya^{2,3}, Salit Kark⁴, and Tim M. Blackburn^{1,5}

¹Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, Gower Street, London, WC1E 6BT, United Kingdom

²University of Utah, Department of Biology, Biodiversity and Conservation Ecology Lab, 257 South 1400 East, Salt Lake City, UT 84112-0840, USA

³Koç University, College of Sciences, Rumelifeneri, Sariyer 34450, Istanbul, Turkey

⁴The Biodiversity Research Group, The School of Biological Sciences, The University of Queensland, Brisbane, QLD, 4072 Australia

⁵Institute of Zoology, Zoological Society of London, Regent's Park, London, NW1 4RY, United Kingdom

Abstract

Human-mediated translocation of species to areas beyond their natural distribution (here termed aliens¹) is a key signature of the Anthropocene² and a primary driver of global biodiversity loss and environmental change³. Stemming the tide of invasions requires understanding why some species fail to establish alien populations, while others succeed. To achieve this, we need to integrate the impact of features of the introduction site, the species introduced, and the specific introduction event. However, determining which, if any, location-level factors affect establishment success has proved difficult due to the multiple spatial, temporal and phylogenetic axes along which environmental variation may influence population survival. Here, we apply Bayesian hierarchical regression analysis to a global spatially and temporally explicit database of alien bird introduction events⁴ to show that environmental conditions at the introduction location, notably climatic suitability and the presence of other alien species groups are the primary determinants of establishment success. Species-level traits and founding population size (propagule pressure) exert

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence should be directed to T.M.B. t.blackburn@ucl.ac.uk.

Data Availability Statement

All data generated or analysed during this study are included in this published article (and its supplementary information files).

Code Availability Statement

Code used to calculate the final model is included in this published article (and its supplementary information files).

Author Contributions D.W.R., A.P., T.M.B. developed the overall study design. E.E.D. and C.H.S. oversaw initial data collation. D.W.R. & A.P. carried out the modelling and data processing with assistance from T.M.B. All authors contributed to writing the manuscript.

Reprints and permissions information is available at www.nature.com/reprints. All data and code are included with this submission.

The authors declare no competing interests.

secondary, but still important, effects on success. Thus, current trajectories of anthropogenic environmental change will most likely facilitate future incursions by alien species, but predicting future invasions will require integrating multiple location, species, and event-level characteristics.

Globally, alien species are accumulating at ever-increasing rates⁵, mainly driven by growing trade and transport connectivity⁶. Once an alien species is established (i.e. self-sustaining) in a new location, the economic and environmental costs of eradicating it or controlling its spread are often prohibitive³. Understanding the processes that facilitate or inhibit the initial establishment of alien species is therefore a critical step in limiting the future threat of biological invasions. Most early attempts to predict alien species establishment focussed on the characteristics of the introduced species or the introduction location⁷, but with limited success⁸, and did not consider the key role of idiosyncratic “event-level” factors, notably propagule pressure⁹. Some species-level traits (life history¹⁰, behavioural¹¹ and ecological¹²) have subsequently been shown to explain variation in alien establishment success. However, determining which, if any, location-level factors affect success generally at a global level and across large taxonomic groups has proved challenging, for several reasons.

First, many different biotic (e.g. recipient assemblage composition¹³) and abiotic (e.g. climate¹⁴, disturbance¹⁵) factors may be important. Second, these factors vary across both space and time, and drive differences in susceptibility at a range of levels of biological organisation – population (e.g. stochastic weather events), species (e.g. climatic affinity), community (e.g. native species richness), and landscape (e.g. habitat composition). Third, how a new environment interacts with a species is dependent on the evolutionary and adaptive history of the species introduced¹⁶: a harsh environment for a house sparrow (*Passer domesticus*) may or may not be harsh for the closely related Eurasian tree sparrow (*P. montanus*), and vice versa. Fourth, alien introductions happen in synergy with other major anthropogenic environmental changes such as increasing human population density, agricultural land conversion, and the presence of other alien species¹⁷. Yet, despite this apparent complexity, many previous analyses have treated location-level variables in a relatively simplistic way, considering either only coarse features of locations (e.g. latitude¹⁸, island versus continent¹⁹) or gross differences between native and alien environments²⁰, and typically ignore spatial autocorrelation²¹. Therefore, we still await an integrated analysis of variation in alien establishment.

Here, we undertake a global analysis to identify both the absolute and relative contributions of location, species, and event-level processes in predicting alien establishment. Using birds as a model system, we interrogate data on the success or failure of 4,346 individual introduction events spanning 708 species and, crucially, include information on propagule pressure, the key event-level driver of establishment⁹. To assess the specific influence of location, we consider a wide array of abiotic, biotic and anthropogenic factors. These account for both the mean and temporal variability in the abiotic environment, the suitability of the environment in terms of its similarity to conditions experienced by a species in its native range (‘environmental match’), metrics of human disturbance, and the characteristics of recipient biological communities, including both their diversity and their phylogenetic

similarity to each introduced species. Finally, we incorporate aspects of species' life history, behaviour and ecology that have previously been hypothesised to explain establishment success in alien birds. Features of introduction events are not random with regard to the identity, relatedness and characteristics of the species introduced¹⁶, their spatial location of origin and introduction⁴, nor to propagule pressure²², and so we undertake this analysis using Bayesian hierarchical regressions, inferred using Integrated Nested Laplace Approximation (INLA)²³. This method provides efficient and accurate parameter estimation for complex inferences incorporating both random and fixed effects, allowing us to control for spatial and temporal non-independence in the abiotic and biotic features of locations, and for taxonomic non-independence in species traits.

At a global scale, combinations of location-, species- and event-level variables are selected as important terms across all fitted models, including the best fitting model of avian establishment success (Fig. 1, Extended Data Table 1, $n = 1530$, $wAIC = 892.96$, $AUC = 0.75$ – for definitions see methods). This result was robust to the precise way in which introduction events were defined (Extended Data Fig. 1) and highlights that alien establishment cannot be adequately explained by characteristics of the environment, the species, or the specific introduction event in isolation. The most strongly supported individual determinant of establishment is the environment of the recipient location (Fig. 2a). Within this category, anthropogenic features, followed by climatic suitability, have the greatest influence on establishment success (Fig. 2b).

A strong anthropogenic determinant of establishment success is the number of alien taxonomic groups already established at a location at the time of introduction. The positive effect of the number of alien groups introduced is broadly consistent with the invasion meltdown hypothesis¹⁷, whereby ecological disruptions caused by, or enabling, earlier invasions facilitate further successful introductions. This result is not simply indexing anthropogenic environmental disturbance; crop coverage and human population density, while included in the best fitting model, did not have a strong and consistent global signal for alien establishment success (Fig. 1, Extended Data Fig. 2). This may be due to historical patterns of introductions being mainly restricted to already disturbed areas¹². In fact, our analysis shows that *less* disturbed areas have higher establishment success rates, with rapid agricultural land-conversion not only causing native species declines², but also negatively impacting alien species, at least in the early stages of the invasion process.

Previous evidence has suggested that species are more likely to establish when they are pre-adapted to local climatic conditions¹⁶ and our analysis confirms this hypothesis. We found that alien establishment success is highest in locations where environmental conditions are more similar to those in the species' native range ('environmental match', Fig. 1-2), albeit with the proviso that average conditions across the range are relatively crude measures of climatic preferences. Our analysis also suggests a hump-shaped effect of mean annual temperature on establishment (Fig. 1). This relationship implies a "Goldilocks effect", such that locations with intermediate conditions are more amenable to establishment than those that are too hot or too cold, regardless of the conditions naturally experienced by each species. Environmental extremes are also important²⁴, with establishment success reduced by the occurrence of historical storm events in the period immediately following

introduction. Anecdotal evidence had previously suggested extreme weather as a cause of specific establishment failures (e.g., the house crow (*Corvus splendens*) on Mauritius²⁵), and our spatiotemporal analysis identifies this as a general effect in the global record of avian introductions.

The extent to which communities differ in their biotic resistance to introduced species has remained controversial, with studies variously reporting positive, negative or no effects of local species richness on patterns of establishment²⁶. Overall, we found that the biotic environment had a relatively weak effect on establishment compared to the other location-, species- and event-level factors. Nevertheless, accounting for these other factors revealed a potential negative effect of native bird species richness on alien establishment success, with this switching to a hump-shaped relationship (Fig. 1) when considering only the most closely related and presumably ecologically similar species. These results help clarify previous contradictory findings, by showing that while overall native biodiversity may inhibit invasions, (i) this effect is relatively weak compared to other extrinsic and intrinsic factors, and (ii) it may be partially masked by the tendency for locations with many closely related species to be more environmentally suitable, and thus be more susceptible to establishment (i.e. biotic acceptance hypothesis¹⁷).

In addition to environmental factors, features of the species' life history and ecology are strongly supported as determinants of establishment success. In particular, larger brood sizes promote establishment, while lifespan showed a hump-shaped relationship with invasion success (Fig. 1, Extended Data Fig. 2), confirming previous evidence of a trade-off between the benefits of fast and slow life histories¹⁰. While species with fast life histories can gain a quick 'foothold' at a new location through rapid population growth, slower life histories give resilience against demographic and environmental variation, allowing alien populations to be better able to ride out extreme conditions^{27,28}. In our model, there is also evidence that foraging specialism and habitat-use generalism may, taken together, increase establishment success. Life history variables are generally strongly phylogenetically conserved (e.g. brood size, $\lambda = 0.96$, Fig. 3), implying that related species could have similar rates of establishment success. However, globally, establishment success has a much weaker phylogenetic signal ($\lambda = 0.4$; Fig. 3), due to phylogenetically conserved traits being overwhelmed by the combined spatial effects of the local environment and propagule pressure, all of which tend to exhibit little phylogenetic signal. The inherently idiosyncratic nature of these effects with regard to the identity of the species introduced (Spearman ρ between predictions based on life history and the final model is 0.64) explains why it has proven difficult to identify consistent life history predictors of establishment in isolation²⁹.

Lastly, we confirm the strong general role of propagule pressure which, in line with previous work on alien birds³⁰, is best represented by an asymptotic log-term (Fig. 1, Extended Data Fig. 2): small founding populations are likely to fail due to stochastic and Allee effects, while the success of larger populations³⁰ depends instead on the species- and location-level effects we identify here. Our analysis highlights the key role of the presence of other alien species groups, suggesting that locations that are already hotspots for introductions are especially susceptible to accumulating alien species, but also show that alien species are more likely to establish when they are pre-adapted to local climatic conditions. Growth in

global trade means that an ever-growing number of species are being introduced to novel locations^{4,31}, and the environmental matches of ever more species are being tested against new environments. These trajectories will facilitate future incursions by alien species, exhibiting features of an invasion meltdown³², which, as we show, could be further exacerbated depending on precise combinations of species and sites where the introductions are occurring. Our analyses confirm the urgent need for enhanced management programs to prevent or mitigate the negative impact of these invasions.

Methods

Alien introduction events

We collated all the records from the GAVIA database of global bird introductions^{4,31} that contained geo-referenced introduction events at known specific (e.g. sub-national and below) point locations (i.e. the locations where the species was recorded as having escaped or been released) and at specific dates, excluding those records from GAVIA that related to spread after an introduction event (Supplement Information Data 1). This process initially resulted in 5834 records, with accompanying spatial polygon data created by drawing around the smallest geographical unit to which the introduction event could be reasonably attributed⁴. While some records were at a specific location a (e.g. a single building address, park or harbour), a small minority of event records could only be assigned at a coarser spatial scale (e.g. city or county). We note there is some geographical bias in the records with most of the introductions occurring in the Australasian region (19% of records), followed by the Palearctic (18%), Oceanic (17%) and the Nearctic (16%), with fewest records in the Afrotropics (12%), Neotropics (10%), IndoMalay (8%), and Antarctica (<1%). However, introduction events occur predominately within regions and high-profile historical routes between continents (e.g. Europe to Australia) are relatively rare given the huge increase in introductions in recent decades⁵ through, for example, accidental transport and wildlife trade³¹ (Extended Data Figure 3).

Using data from both the original, and external sources, we thoroughly checked all introduction records for potential errors and removed any records that were possibly dubious – usually due to misreported dates or locations across multiple references. This resulted in $n = 4346$ unique introduction event records (Supplementary Information Data 1). We then used text information, again from the original source, to categorise the introduction events as either known specific introduction events (“introductions”, $n = 1784$) first sightings (“sightings”, $n = 584$), or as having no clear designation (“unknown”, $n = 1978$). Finally, we noted if any records were part of a chronological sequence of introduction events involving a single species at a single location e.g. Eurasian skylark (*Alauda arvensis*) imported and released on six separate occasions in the Barrabool Hills, VIC, Australia from 1852-1880. From these we created four data subsets of decreasing size but increasing specificity: “All Records” contained all the records in the database ($n = 4346$), “Intro. & Unk.” contained records that are known introductions and records that have no detailed description ($n = 3762$), and “Introductions” contained all records that are specified as detailed introduction events ($n = 1784$) and “Last Introduction” ($n = 1530$). This final data subset contained known introduction events, but with events that were part of a chronological sequence of

introduction events collapsed into a single record, summarised using the date of the last introduction events and the cumulative propagule pressure across events ($n = 1530$).

Based on prior hypotheses of avian establishment success reported in the literature (Supplementary Information Data 2), we collated information on a wide-range of covariates that could reasonably impact upon establishment success. We categorised these covariates into the three categories of establishment determinants, as defined by Duncan et al. 16 – location-, species- and event-level factors –further distinguishing between different types of species- and location-level factors as below:

Event-level Factors - Propagule pressure

We extracted from the original reference source, where available, a numerical estimate of propagule pressure (founding population size), measuring both the number of introduction events per record (propagule number⁹) and the number of individual birds that were introduced at each event (propagule size⁹). For a minority ($n = 67$ or $\sim 0.01\%$) of records that only had descriptive text regarding the number of individuals introduced, we translated any common terms according to the following rules. When describing individuals released: “few”=3, “several”=5, “some”=10, “small numbers”=10, “many”=25, “flock”=25, “large numbers”=100, “shipment”=200, “mass”=250, “great numbers”=250. When describing propagule number: “repeated”=5, “several”=5, “releases”=2, “numerous”=10, “many”=10, “frequent”=10. We decided on these numbers by summarising, where available, records that contained both these descriptive qualifiers *and* a numerical figure for number introduced. To calculate propagule pressure, that is the relative size of the introduction effort⁹, we used the recorded number of individuals introduced. When this data type was missing we added in the median propagule size (5 individuals introduced), or, if the number of discrete introduction events were available, we used the median propagule size multiplied by propagule number.

Species-level Factors - Life History Traits

For each species, we assembled data from published sources on a number of life-history traits previously linked to establishment success in birds, including mean clutch size, number of clutches per year, age at first breeding (months) and maximum lifespan (years)^{10,33,34}. We additionally included data on mean adult body mass (g)³⁵. Species for which data could not be collected (clutch size [11%], number of clutches per year [48%], age at first breeding [66%], lifespan [52%]) were assigned the mean value of the lowest inclusive taxonomic rank (i.e. genus, family, order) for which data were available. This approach is justified because most of variance in avian traits, as calculated from our data, occurs at taxonomic levels above the genus (clutch size [91%], number of clutches per year [70%], age at first breeding [83%], lifespan [62%]). We also include a previously used measure called brood value, which is expressed as $\log_{10}(1/[(\text{broods per year}) \times (\text{reproductive lifespan})])$ ¹⁰ and represents investment in future over current reproduction.

Species-level Factors - Behavioural Traits

For each species, we assembled data on relative brain size, quantified as the residuals from a least squares regression of brain size on body size (both log-transformed)³⁶. Relative brain

size provides a metric of behavioural flexibility that has been shown to relate to establishment success in birds³⁷. Species with missing data (72%) were assigned the mean value of the lowest inclusive taxonomic rank (i.e. genus, family, order) for which data were available. As for life history traits, most of the variance in brain size (93%) occurs at taxonomic levels above the genus.

Species-level Factors - Ecological Traits

Data on species-specific diets and foraging strategies came from Kissling et al.³⁸: for both of these variables a total value of 100 was divided between categories to represent the percentage of time a species spends feeding on a particular food type or foraging in any particular location. Habitat use data for each species were extracted from the IUCN Red List database⁴⁰. For each ecological variable (diet, foraging and habitat), we calculated two measures of generalism, using the total number of different categories used and Simpson's diversity measure $D39$.

Location-level Factors - Abiotic environment

Global geophysical data (altitude above sea level and slope %) were downloaded in pre-processed geoTiff format at 1km grid scale⁴¹. A third variable, 'altitude variance', was computed with the R function *aggregate* (*raster* package⁴²) using the variance of the altitude values of 3x3 grid cells, such that all 9 cells had the same final value. Bio-climatic data in the form of global averages from 1960-2000 were restricted to large terrestrial land masses, and were downloaded as four 'ESRI' format *ascii* data grids, at 30 arc-seconds (~1 km) resolution. They consisted of mean annual temperature, annual variation in temperature ('temperature seasonality'⁴³), mean annual rainfall and annual variation in temperature ('precipitation seasonality'⁴³). Abiotic data for islands came from an island-specific dataset⁴⁴ and had the same variables as for large terrestrial land masses, but represented as a spatially-referenced spreadsheet containing data on climate and physical characteristics of the majority of the world's islands. For islands that were not represented in the gridded bioclimatic data, we identified missing values for the above bioclimatic and altitude data, and by matching the island name in the GAVIA data with the island name in the bioclimatic dataset, we were able to extract the mean annual temperature, annual variation in temperature, mean annual rainfall, annual variation in temperature and altitude. For islands, we also included distance to continent (giving non-island records a value of zero) represented by the "dist" column from that data set. We also employed a measure of remoteness, again giving continents a value of zero, using the "SLMP" column of the data set.

Historical climate data (1850-2007) were downloaded as 6x4km netCDF grids for six main variables: sea surface temperature (SST), air temperature (A), U-wind (Uwind), V-wind (Vwind), sea-level pressure (SLP) and cloudiness (CLDC) from the HADCRUT3 data set⁴⁵. Historical spatio-temporal land-cover data (1700-2007) were downloaded as global 'ESRI' format *ascii* data grids at ~5 km resolution, consisting of proportion cover of primary and secondary habitats, from the Harmonised Land Use Dataset⁴⁶. To reduce collinearity in a regression (i.e. the dummy variable trap) the 'other' category data was not included in the

analysis, such that the addition of all the land-use categories in each grid cell did not sum to one.

These environmental data were extracted at each record location by calculating the mean grid cell values intersecting the introduction event polygon (R function *extract*⁴²). For very specific introduction events, this would be the single cell the event polygon was located in, but for less specific introduction event data there was sometimes more than one grid cell that overlapped the polygon. For the spatio-temporal data, extractions for each record were only undertaken on the temporally nearest data layers. For example, for the historical climate data, a maximum anomaly value at that location over the ten years (120 months) post-introduction was used, with records outside the dataset time period (<5% of the total records) being designated missing using “NA”. For the land-cover data, which was a yearly rather than monthly dataset, records were matched by introduction year to the specific global land-cover layer so that the contemporary (at the time of introduction) percentage cover for five land-use types (Primary, Secondary, Cropland, Pasture, Urban) could be calculated. Records earlier than the starting year of the land-cover dataset (<1% of total records) were designated missing in each land-use covariate using “NA”. We note that there is uneven sampling here, with most historical introductions occurring in human modified landscapes: for instance, in forested areas only 8% of the introduction polygon is designated as ‘Primary’ at the time of introduction. This means for some specific land-types we may not be able to resolve their specific impacts on invasion success. For each land-cover type noted above, an additional variable was constructed which was the gradient of change in land-cover in each grid cell over the 10 years prior to each introduction event. This was calculated by using a linear regression (R function *lm*) of land-cover proportion explained by year, and taking the slope (β) as the value of change per cell.

Location-level Factors - Environmental Match

Range maps for species’ native distributions were downloaded from Birdlife International and NatureServe (www.birdlife.org) and extracted onto an equal area grid (~110 x 110 km) in a Behrmann projection. These maps show species’ extent of occurrence, and so are relatively crude depictions of the area occupied by the species, but are nevertheless commonly used for analyses of this type. We quantified species’ environmental preferences using the mean and standard deviation of climate conditions across grid cells in their native distributions on the basis of four climatic indices from the WorldClim dataset (BIO1, mean annual temperature; BIO4, temperature seasonality; BIO12, annual precipitation; BIO15, precipitation seasonality)⁴³. Using each of these input variables, to capture as a single variable the environmental match between the introduction site and those being experienced in the species’ native range, for each introduction event we calculated the distance, in measurement space, between the Euclidean distance from mean values taken from the grid cells at the introduction site (sources defined in ‘abiotic’ above) to the mean values from the native range of the introduced species. For each climatic axis, we divided the distance by the standard deviation of native climatic values, as some species have very large ranges with a corresponding wide range of acceptable native climatic values. We note that this measure is a relatively coarse way to measure native preferences, as fine scale habitat variation within

the range may act to bias the mean value, but that finer scale data are not available for all introduced species.

Location-level Factors - Biotic Environment

To test whether interactions with native resident species may influence establishment success, for each combination of introduced alien species and grid cell, we calculated four metrics of community diversity and structure: i) the richness of all native resident species, ii) the number of native species in the same genus or, iii) family, and iv) the nearest taxon index, representing the phylogenetic branch length (millions of years) separating an introduced alien species from its closest relative in the recipient community. Recipient communities were designated as those species whose ranges overlapped with any of the introduction polygon, though we note that not all species in this sample would be interacting if they utilised very different habitats. Phylogenetic distances were calculated as the mean across 100 phylogenies sampled at random from the posterior distribution of trees from the Jetz et al. phylogeny with the Hackett backbone⁴⁷. These variables thus quantify the overall species richness of the location of introduction (i), and the location's richness relative to the phylogenetic position of the introduced species (ii – iv).

Location-level Factors - Anthropogenic Environment

To determine the role of human disturbance and urbanisation¹² in facilitating the establishment of invasive species we captured the spatial variation and prior ten-year change in human population density, urban land, crop and pastoral land coverage, from the Harmonised Land Use Dataset⁴⁶ using the same methods as for abiotic environment variables.

We also tested whether establishment probability is related to the prior presence of other alien species. We determined if any alien species were present at a given site prior to each introduction event using published data⁵ on first records for a number of groups (algae, amphibians, arachnids, arthropods, bacteria and protozoans, birds, bryophytes, crustaceans, fishes, fungi, insects, invertebrates, mammals, molluscs, reptiles, plants, viruses), recorded at the level of country or major island group. For larger countries this value would be less accurate, but most of our data are for islands (66% of introduction records). Furthermore, these data represent the best spatio-temporal knowledge currently available and further work using site-level data would be able to examine these relationships in more detail. For each introduction event, we create a binary variable with a value of 1 for each group if any species from that group was present at least one year prior to the bird introduction event and 0 if no species from that group were present.

Statistical Modelling Outline

We modelled the establishment success or failure of bird introductions using a Bayesian hierarchical regression inferred using Integrated Nested Laplace Approximation (INLA) as implemented in the R package R-INLA⁴⁸ (Supplementary Information Code 1). We used this method as it provides accurate parameter (e.g. β) estimates for complex regressions incorporating both spatial and non-spatial random and fixed effects with very low computational overheads²³. We evaluated the model fit for covariate choice via the

Watanabe–Akaike Information Criterion ($wAIC$)⁴⁹ and the Conditional Predictive Ordinate (CPO). $wAIC$ is a criterion for model comparison and is an extension of the Akaike Information Criterion (AIC), but is widely-applicable to Bayesian inference techniques and offers clearer interpretation than other options⁵⁰. Similar to AIC, $wAIC$ provides a method to penalise the ability of the model to fit the observed data by the number of parameters used to create the underlying model. This value is more suitable for a Bayesian framework as it integrates across the whole posterior distribution rather than relying on summary statistics, e.g. mean of posterior distribution. Similarly, the CPO approximates the ‘gold-standard’⁵¹ Leave-One-Out Cross Validation, and calculates the posterior probability of a model inferred without each data point. The sum of the log CPO scores, therefore, represents an estimator for the log marginal likelihood of the model. Given that $wAIC$ tends to $-\ln(CPO)$ under ideal circumstances⁵¹ (in our case Spearman correlation ~ 0.99), we henceforth report only $wAIC$ as a proxy for CPO.

We model number of establishment successes across the dataset as a binomial random variable (success = 1, failure = 0) and use the normal approximation to the binomial, as expected under central limit theorem given our large numbers of trials. This was due to the computational efficiency of utilising the Gaussian distribution, allowing us to repeat the modelling procedure many times with no loss of predictive accuracy (mean hold-out cross-validated Area Under receiver operating Curve statistic (AUC) for Gaussian = 0.68 ± 0.05 , versus AUC = 0.67 ± 0.06 for the binomial mean). To convert the ‘StatusCat’ column from the GAVIA dataset to the response variable, we recoded the categories of “established” and “breeding” to 1 and the known failure categories (“died out”, “failed”) to 0. For a large proportion of records, the success or failure of an avian introduction was unknown ($n = 2234$). In these cases, we used the introduction event polygon associated with each record to search for the alien species in sightings from eBird⁵² and other sources from within the Global Biodiversity Information Facility (GBIF)⁵³. We downloaded all occurrences from GBIF, within 0.5 degrees of both latitude and longitude of the centre point of each introduction event. Then, using the associated ‘observation date’ with each GBIF record and the ‘mapping date’ associated with each GAVIA record, we counted the total number of individuals seen within the 0.5-degree buffer in the last ten years (2007-2017). We changed unknown values to 1 (succeeded) if there were more than 15 records of the introduced species within 0.5 degree (~ 95 km at the equator) of the introduction site in the last ten years, and 0 (failed) if fewer. To ensure our results were robust to these thresholds, we changed the record density threshold used to identify whether a species is “established” by increasing the buffer size used to capture GBIF observations to 1 and then 1.5 degrees of both latitude and longitude around the introduction point.

To account for known spatial autocorrelation in the input data³¹ we implemented a stochastic partial differential equations model (SPDE) with the hierarchical regression that builds a latent error surface, of user-defined complexity, to account for similarities in more closely located data points⁴⁸. We inferred the regression models using an SPDE term and mesh with varying characteristics to find the range, standard deviation of the range, cut-off and maximum edge values using the $wAIC$ score to determine the best version of the SPDE model. To account for the random effect of phylogenetic non-independence, we included “iid” random effects of species’ taxonomic family and order. To account for temporal

differences in recording accuracy and in methods of introduction³¹ at the same location over time, we included a random-walk auto-correlated random effect⁴⁸ for the year of introduction. To remove effects of very large or small values, for each covariate we capped low values at the 1% quantile and high values as the 99% quantile.

Finally, for each model we assessed all recommended diagnostics to ensure the model was robustly fitted, including plotting and visualising the distribution and probability density of the out-of-sample CPO per data point score (using 10-fold, 10% hold-out, cross validations) (Extended Data Figure 4a,b) and spatially mapping the same values (Extended Data Figure 4c) to check for parts of data that were poorly predicted by the model. Then, using the mean of the posterior distribution of the linear predictor, we employed an AUC approach (Area Under receiver operating Curve score) to calculate the predictive accuracy of each model. This process works by measuring the numbers of correctly and incorrectly labelled predictions across all possible classification threshold values of the binomial response variable. An AUC value equal to or less than 0.5 indicates a predictive ability that is equal to the random expectation, while an AUC value close to 1 indicates a perfect predictive ability.

Analysis protocol

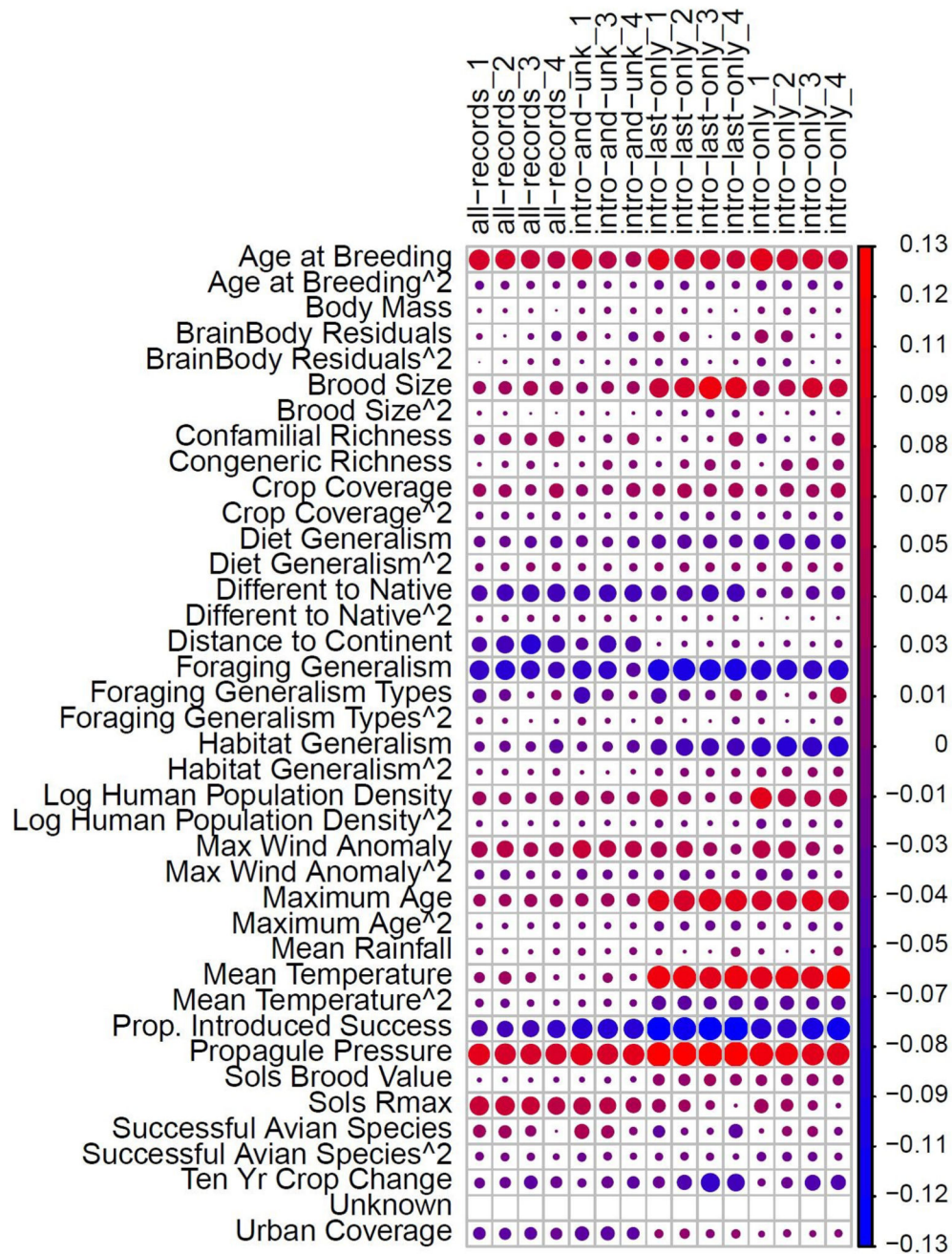
We first used the most conservative dataset for analysis – that which contained known introductions events ($n = 1530$ records) (Supplementary Information Data 3). Then, to ensure our random effect terms were valid in their inclusion, we first fitted a model with just an intercept, then just an intercept and spatial term, and then added in other random effects one at a time (Extend Data Table 1). For each additional step of complexity we recorded the change () in the $wAIC$ value⁴⁹ and only included random effects that increased the fit of the model by more than $2 wAIC$ units. We used uninformative priors in all cases except for the spatial term where we set the priors to a set of reasonable estimates of the range and standard deviation of the range to understand how this specification affected parameter estimates.

To increase the biological interpretability of our models (and due to the large number of covariates and high collinearity between them), we then added all explanatory variables (Supplementary Information Data 2) into a regression model in a step-wise manner and, after each step, assessed model fit using the $wAIC$ value. In order that the effect sizes of the different covariates could be better compared, each explanatory variable was standardised to a mean of zero and standard deviation of one before it was added. At each model choice step we used the standard threshold of $wAIC > 2$ to select better models⁵⁴. When offering steps either forward or backwards, we allowed the choice of either a linear representation of the covariate, the natural logarithm of the term, or a linear and squared term to allow for the situations where a curvilinear relationship fitted better than a linear slope.

To examine whether the model selection process was robust to decisions relating to the database, we ran several additional versions of the stepwise regressions to see if the key variables identified in the main analysis above were still recovered. We first ran a stepwise regression using all the introduction records and then repeated the process with all the other data subsets. We also ran different versions of the stepwise selection with the different buffer sizes for the GBIF missing data interpolation to test the sensitivity of the imputation process.

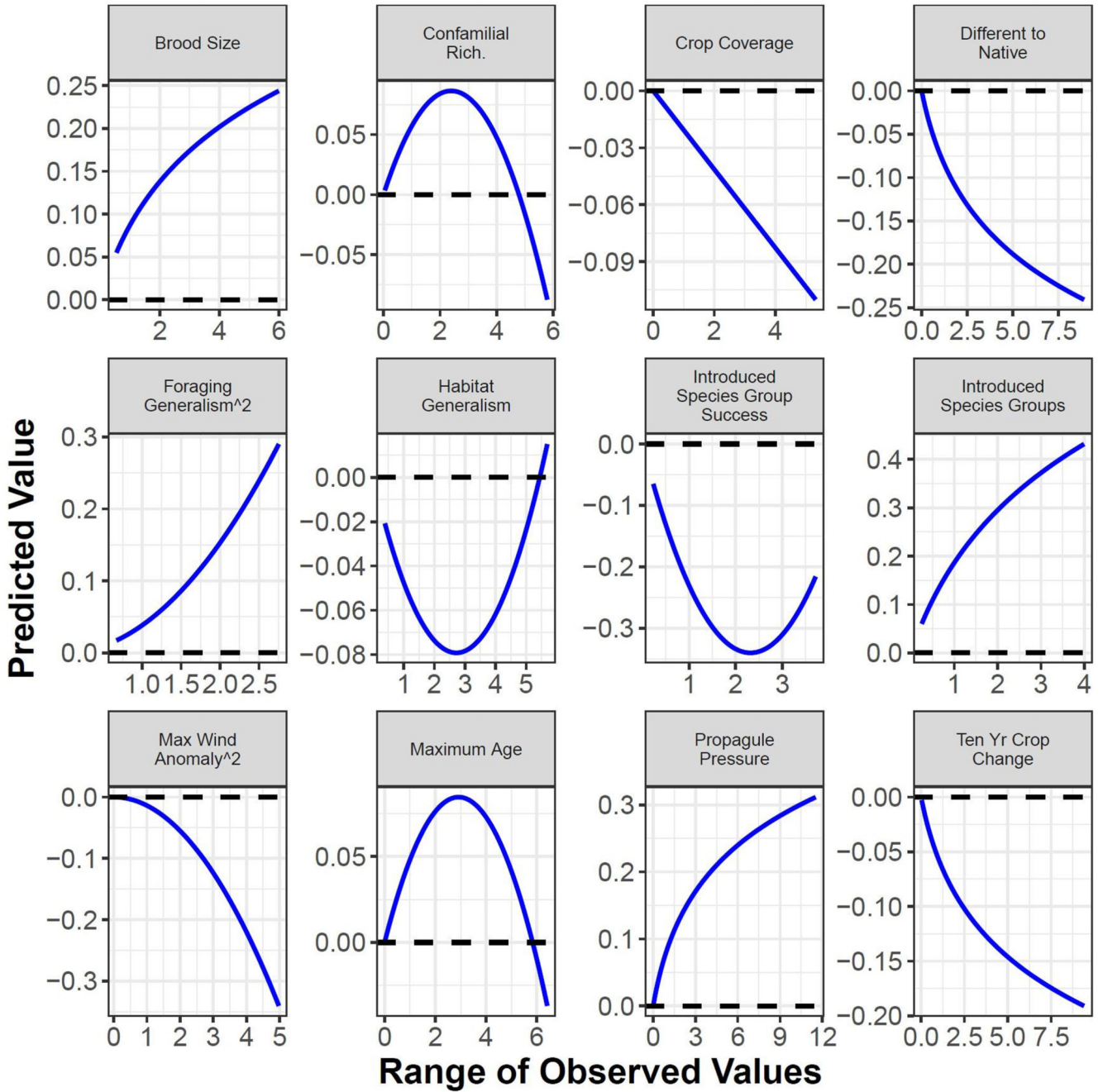
We then used the lowest $wAIC$ model to predict establishment success over a 1-degree grid of points covering all land areas ($n = 19561$ cells) for each introduced alien species ($n = 358$ species). We only used already established alien species because new species are being added to the current pool of aliens at a relatively low rate and this current pool will likely make up the vast majority of future invasions⁵. When predicting, we set the random effects introduction year = 2015 to be as close to the present day as possible and propagule pressure = 150, the lowest value after the threshold beyond which the number introduced has limited impact (Extend Data Fig. 2). All predicted values where the confidence was low, such that the 95% confidence intervals for the grid cell estimate covered 0 and 1, we designated as NA. Using the prediction layers for each species we created two datasets: First, by using a 10% trimmed mean of the probability of success for every species value for each grid cell we were able to determine which areas of the world had the highest ‘establishment potential’ and, therefore, were at risk of this set of introduced species establishing there. Second, by using 10% trimmed means for all values in each of the 384 layers, we were able to create an index of establishment potential per species, which we then mapped on to a recent phylogeny⁴⁷. We calculated the phylogenetic signal of these values using Pagel’s λ (R function *phylosig55*) and used a permutation test to test the probability that these values deviated from 0, indicating a significant relationship between phylogenetic relatedness and species trait values.

Extended Data

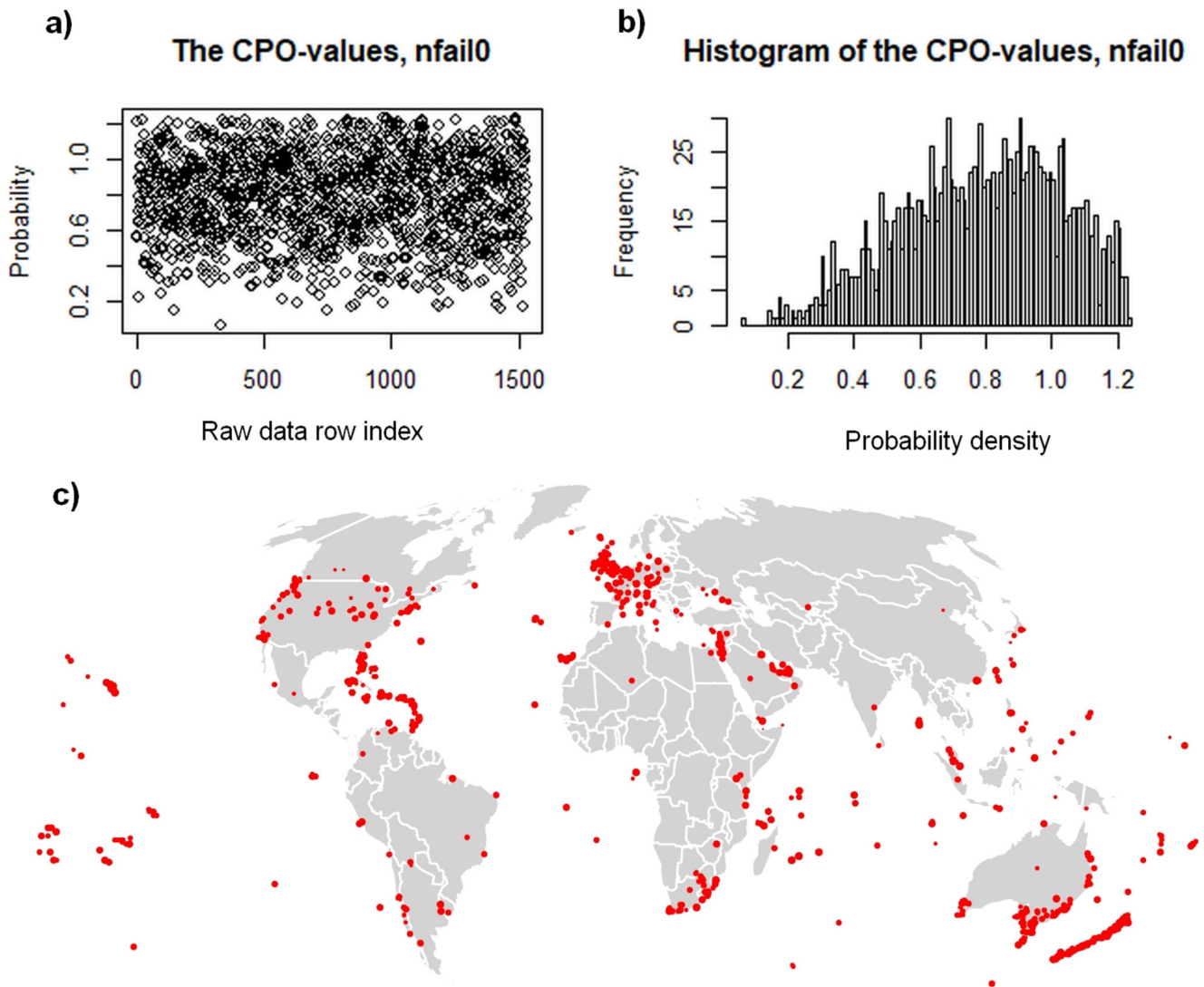


Extended Data Figure 1. Sensitivity analysis of slope (β) estimates for the linear terms of a subset of variables across all versions of the input data.

Dot size is the size of the beta value with colour representing direction with red positive and blue negative. Each row label represents the name of the fixed effect. The column headings represent the data subset used: ‘all-records’ = all data (n = 4346), ‘intro-and-unk.’ = all data but one record per species-location event (n = 3762), ‘intro-only’ = detailed introductions only (n = 1784), ‘intro-last-only’ = detailed introductions but one record per species-location event (n = 1530). The number at end of each column heading indicates relative size of buffer used to impute establishment status (see methods).



Extended Data Figure 2. Approximate shape of fixed effects over the range of observed values. Each panel represents the prediction using β slope estimates from the lowest $wAIC$ model over the known range of values for that given fixed effect (identified by strip title) from the raw data. Only fixed effects whose values were unlikely to include zero are included. All panels from a single Bayesian regression of global avian establishment success ($n = 1530$ introductions).



Extended Data Figure 4. Model diagnostics from the best fitting model.

Panel (a) shows a plot of out-of-sample Conditional Probability Ordinate (CPO) scores for all data points in rank order used in the model; panel (b) the probability density of the CPO scores; (c) map of CPO scores. CPO is the probability of generating each data point in the data set from a posterior fitted without this data point, with each panel allowing visualisation of where in the data the model not be fitting well. All plots from a single Bayesian regression of global avian establishment success ($n = 1530$ introductions).

Extended Data Table 1
All covariates in the best fitting model, from a Bayesian regression of global avian establishment success (n = 1530 introductions).

Each variable (column: Variable name) is assigned a hierarchical category (Category level 1-3), the mean for the posterior distribution for β estimates (Mean β), the standard deviation for the mean value (SD β), and the 25th, 50th and 75th quantiles (25% quant, 50% quant, 75% quant) of the posterior distribution of each β estimate.

Variable name	Category level			Mean β	SD β	25% quant	50% quant	75% quant
	1	2	3					
Intercept				0.234	0.112	0.013	0.234	0.454
Propagule Pressure	Event	Propagule pressure	Propagule pressure	0.123	0.032	0.061	0.123	0.185
Max Wind Anomaly	Location	Abiotic	Max wind anomaly	0.04	0.027	-0.013	0.04	0.092
Max Wind Anomaly^2	Location	Abiotic	Max wind anomaly	-0.013	0.006	-0.025	-0.013	-0.001
Mean Temperature	Location	Abiotic	Mean temperature	0.06	0.055	-0.048	0.059	0.169
Mean Temperature^2	Location	Abiotic	Mean temperature	-0.012	0.013	-0.037	-0.012	0.012
Crop Coverage	Location	Anthropogenic	Crop coverage	-0.022	0.011	-0.043	-0.022	-0.001
Human Popn	Location	Anthropogenic	Human population size	-0.002	0.029	-0.059	-0.003	0.055
Introduced Species Groups	Location	Anthropogenic	Introduced species	0.269	0.06	0.152	0.269	0.387
Intro. Group Success	Location	Anthropogenic	Introduced species	-0.297	0.053	-0.401	-0.297	-0.193
Intro. Group Success^2	Location	Anthropogenic	Introduced species	0.065	0.012	0.04	0.065	0.089
Ten Yr Crop Change	Location	Anthropogenic	Ten Year Crop change	-0.083	0.04	-0.161	-0.083	-0.005
Confamilial Rich.	Location	Biotic	Confamilial richness	0.077	0.027	0.024	0.077	0.129
Confamilial Rich.^2	Location	Biotic	Confamilial richness	-0.016	0.006	-0.027	-0.016	-0.005
Species Rich.	Location	Biotic	Native bird richness	-0.014	0.014	-0.041	-0.014	0.013
Different to Native	Location	Match	Environmental niche difference	-0.105	0.026	-0.156	-0.105	-0.054
Foraging Generalism Cat.	Species	Ecology	Foraging generalism	-0.011	0.026	-0.063	-0.011	0.04
Foraging Generalism	Species	Ecology	Foraging generalism	-0.041	0.04	-0.118	-0.041	0.037
Foraging Generalism^2	Species	Ecology	Foraging generalism	0.037	0.017	0.004	0.037	0.07
Habitat Generalism	Species	Ecology	Habitat generalism	-0.061	0.026	-0.113	-0.061	-0.01
Habitat Generalism^2	Species	Ecology	Habitat generalism	0.011	0.004	0.002	0.011	0.02
Brood size	Species	Life History	Brood size	0.129	0.036	0.059	0.128	0.201
Maximum Age	Species	Life History	Max lifespan	0.056	0.021	0.015	0.056	0.097
Maximum Age^2	Species	Life History	Max lifespan	-0.009	0.004	-0.018	-0.009	-0.001

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors thank Mark Parnell, Victoria Franks, Fiona Spooner, Rebecca Herdson, Elizabeth Jones and Frances Davis for assisting with data collection and map creation. Initial funding for this study was provided by a grant from the Leverhulme Trust (RF/2/RFG/2010/0016) (E.E.D.), with additional support from a UCL IMPACT studentship (10989) (E.E.D.), and from a King Saud University Distinguished Scientist Research Fellowship (T.M.B., D.W.R., E.E.D.). D.W.R. is supported by a MRC UKRI/Rutherford Fellowship (MR/R02491X/1).

References

- Blackburn TM, et al. A proposed unified framework for biological invasions. *Trends Ecol Evol.* 2011; 26:333–339. [PubMed: 21601306]
- Lewis SL, Maslin MA. Defining the anthropocene. *Nature.* 2015; 519:171. [PubMed: 25762280]
- Pimentel, D. Biological invasions: economic and environmental costs of alien plant, animal, and microbe species. CRC Press; Boca Raton: 2011.
- Dyer EE, Redding DW, Blackburn TM. The global avian invasions atlas, a database of alien bird distributions worldwide. *Sci Data.* 2017; 4
- Seebens H, et al. No saturation in the accumulation of alien species worldwide. *Nature Commun.* 2017; 8
- Banks NC, Paine DR, Bayliss KL, Hodda M. The role of global trade and transport network topology in the human-mediated dispersal of alien species. *Ecol Lett.* 2015; 18:188–199. [PubMed: 25529499]
- Drake, JA, , et al. Biological invasions: a global perspective. John Wiley & Sons; Chichester: 1989.
- Crawley, M. Colonization, succession and stability. 26th symposium of the British Ecological Society. Gray, AJ, Crawley, MJ, Edwards, PJ, editors. Blackwell Scientific Publications; Oxford: 1986. 429–453.
- Lockwood JL, Cassey P, Blackburn T. The role of propagule pressure in explaining species invasions. *Trends Ecol Evol.* 2005; 20:223–228. [PubMed: 16701373]
- Sol D, et al. Unraveling the life history of successful invaders. *Science.* 2012; 337:580–583. [PubMed: 22859488]
- Sol D, Griffin AS, Bartomeus I, Boyce H. Exploring or avoiding novel food resources? The novelty conflict in an invasive bird. *PLoS One.* 2011; 6:e19535. [PubMed: 21611168]
- Sol, D, González-Lagos, C, Lapiedra, O, Díaz, M. Ecology and Conservation of Birds in Urban Environments. Murgui, E, Hedblom, M, editors. Springer; 2017. 75–89.
- Duncan, R, Forsyth, D. Conceptual ecology and invasion biology: Reciprocal approaches to nature. Cadotte, MW, McMahon, MN, Fukami, T, editors. 2006. 405–421.
- Peterson AT. Predicting the geography of species' invasions via ecological niche modeling. *Q Rev Biol.* 2003; 78:419–433. [PubMed: 14737826]
- Wagner V, et al. Alien plant invasions in European woodlands. *Divers Distrib.* 2017; 23:969–981.
- Duncan RP, Blackburn TM, Sol D. The ecology of bird introductions. *Annu Rev Ecol Syst.* 2003; 34:71–98.
- Jeschke J, et al. Support for major hypotheses in invasion biology is uneven and declining. *NeoBiota.* 2012; 14:1.
- Pyšek P, Richardson DM. The biogeography of naturalization in alien plants. *J Biogeogr.* 2006; 33:2040–2050.
- Jeschke JM. Across islands and continents, mammals are more successful invaders than birds. *Divers Distrib.* 2008; 14:913–916.

20. Veran S, et al. Modeling spatial expansion of invasive alien species: relative contributions of environmental and anthropogenic factors to the spreading of the harlequin ladybird in France. *Ecography*. 2016; 39:665–675.
21. Case TJ. Global patterns in the establishment and distribution of exotic birds. *Biol Conserv*. 1996; 78:69–96.
22. Cassey P, Blackburn TM, Sol D, Duncan RP, Lockwood JL. Global patterns of introduction effort and establishment success in birds. *Proc R Soc Lond B: Biol Sci*. 2004; 271:S405–S408.
23. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J R Stat Soc B*. 2009; 71:319–392.
24. Ameca y Juárez EIA, Mace GM, Cowlshaw G, Pettorelli N. Natural population die-offs: causes and consequences for terrestrial mammals. *Trends Ecol Evol*. 2012; 27:272–277. [PubMed: 22196829]
25. Cheke A. Seafaring behaviour in House Crows *Corvus splendens*—a precursor to ship-assisted dispersal? *Phelsuma*. 2008; 16:65–68.
26. Fridley J, et al. The invasion paradox: reconciling pattern and process in species invasions. *Ecology*. 2007; 88:3–17. [PubMed: 17489447]
27. Allen WL, Street SE, Capellini I. Fast life history traits promote invasion success in amphibians and reptiles. *Ecol Lett*. 2017; 20:222–230. [PubMed: 28052550]
28. Sæther B-E, et al. Life-history variation predicts the effects of demographic stochasticity on avian population dynamics. *Am Nat*. 2004; 164:793–802. [PubMed: 29641930]
29. Hayes KR, Barry SC. Are there any consistent predictors of invasion success? *Biol Inv*. 2008; 10:483–506.
30. Duncan RP, Blackburn TM, Rossinelli S, Bacher S. Quantifying invasion risk: the relationship between establishment probability and founding population size. *Methods Ecol Evol*. 2014; 5:1255–1263.
31. Dyer EE, et al. The global distribution and drivers of alien bird species richness. *PLOS Biol*. 2017; 15:e2000942. [PubMed: 28081142]
32. Simberloff D, Von Holle B. Positive interactions of nonindigenous species: invasional meltdown? *Biol Inv*. 1999; 1:21–32.
33. Myhrvold NP, et al. An amniote life-history database to perform comparative analyses with birds, mammals, and reptiles. *Ecology*. 2015; 96:3109–3109.
34. Tacutu R, et al. Human Ageing Genomic Resources: integrated databases and tools for the biology and genetics of ageing. *Nucleic acids Res*. 2012; 41:D1027–D1033. [PubMed: 23193293]
35. Wilman H, et al. EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology*. 2014; 95:2027–2027.
36. Sayol F, et al. Environmental variation and the evolution of large brains in birds. *Nature Commun*. 2016; 7
37. Sol D, Duncan RP, Blackburn TM, Cassey P, Lefebvre L. Big brains, enhanced cognition, and response of birds to novel environments. *Proc Natl Acad Sci USA*. 2005; 102:5460–5465. [PubMed: 15784743]
38. Kissling WD, Sekercioglu CH, Jetz W. Bird dietary guild richness across latitudes, environments and biogeographic regions. *Global Ecol Biogeogr*. 2012; 21:328–340.
39. Simpson EH. Measurement of diversity. *Nature*. 1949
40. International, B. IUCN Red List Bird Habitat Classifications. 2017. <http://www.iucnredlist.org>
41. Hengl, T; Kilibarda, M; Carvalho-Ribeiro, E; Reuter, H. Worldgrids—A public repository and a WPS for global environmental layers. 2015. WorldGrids at <http://worldgrids.org/doku.php>
42. Hijmans RJ, et al. Package 'raster'. R package. 2016 accessed 1 October 2016
43. Hijmans, R, , et al. WorldClim, version 1.3. University of California; Berkeley: 2005.
44. Weigelt P, Jetz W, Kreft H. Bioclimatic and physical characterization of the world's islands. *Proc Natl Acad Sci USA*. 2013; 110:15307–15312. [PubMed: 24003123]
45. Brohan P, Kennedy JJ, Harris I, Tett SF, Jones PD. Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J Geophys Res Atmos*. 2006; 111

46. Hurtt GC, et al. Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Climatic Change*. 2011; 109:117.
47. Jetz W, Thomas G, Joy J, Hartmann K, Mooers A. The global diversity of birds in space and time. *Nature*. 2012; 491:444–448. [PubMed: 23123857]
48. Blangiardo M, Cameletti M, Baio G, Rue H. Spatial and spatio-temporal models with R-INLA. *Spat Spatio-temporal Epidem*. 2013; 7:39–55.
49. Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J Machine Learn Res*. 2010; 11:3571–3594.
50. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. The deviance information criterion: 12 years on. *J R Stat Soc B*. 2014; 76:485–493. DOI: 10.1111/rssb.12062
51. Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. *Stat Comp*. 2014; 24:997–1016.
52. Sullivan BL, et al. eBird: A citizen-based bird observation network in the biological sciences. *Biol Conserv*. 2009; 142:2282–2292.
53. Global Biodiversity Information Facility. 2017. <http://www.gbif.org/>
54. Raftery AE. Bayesian model selection in social research. *Sociolog I Methodol*. 1995:111–163.
55. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol*. 2012; 3:217–223.

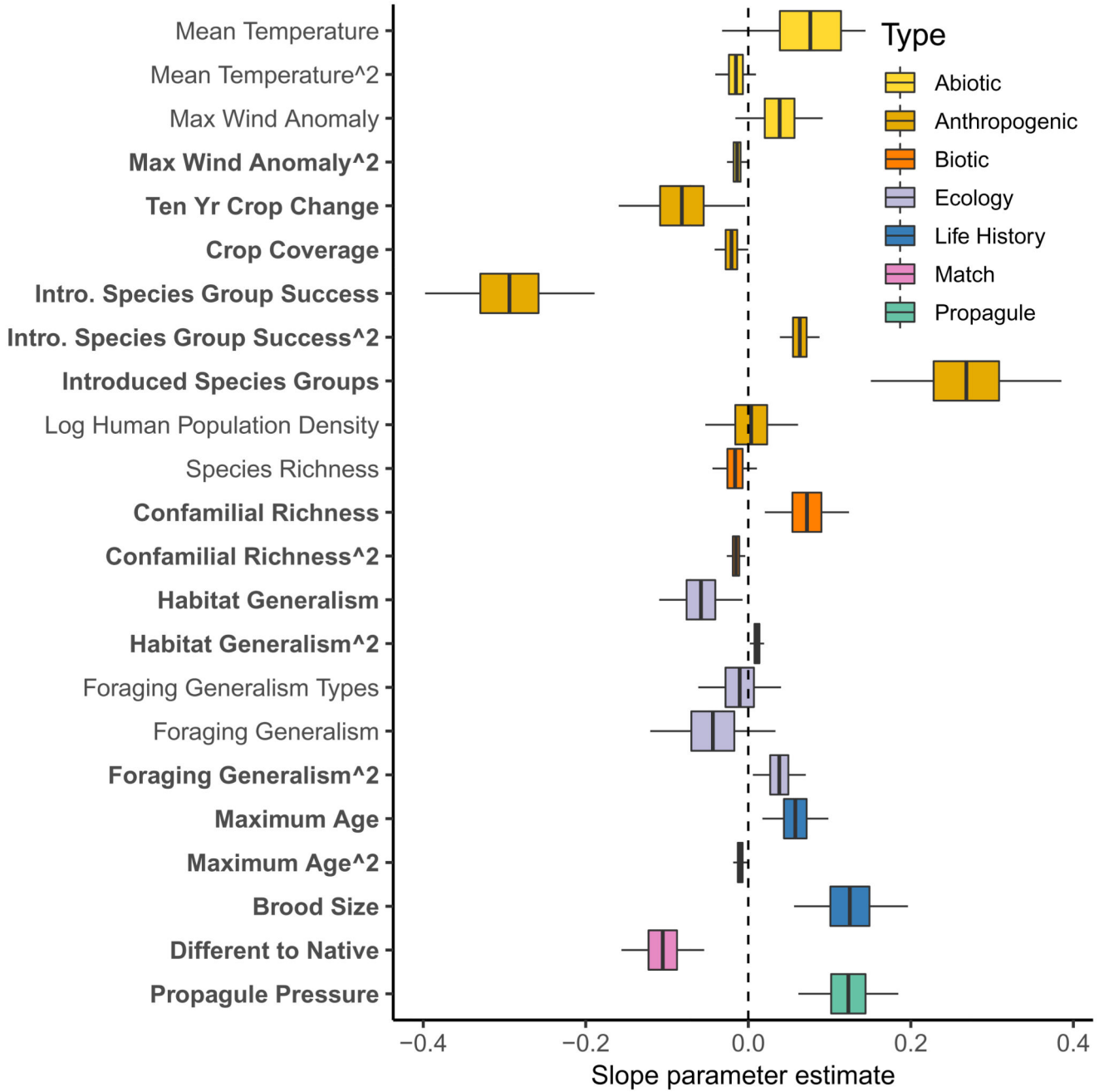
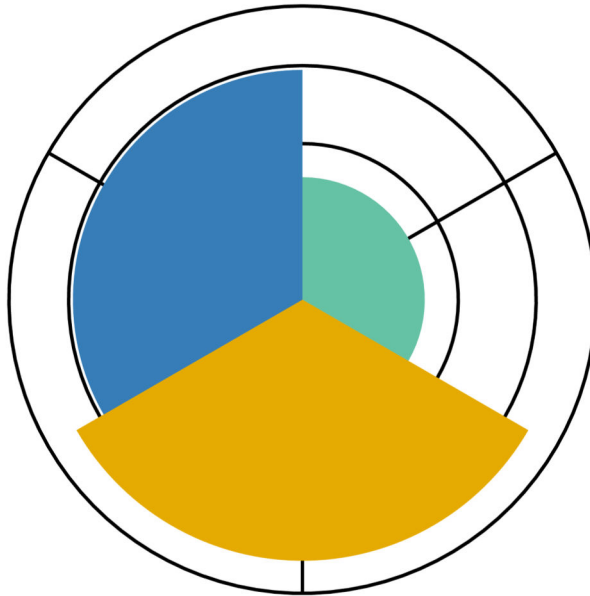
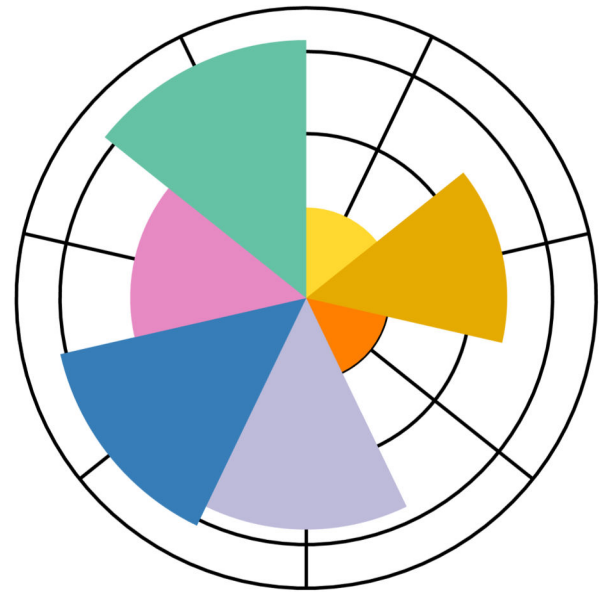


Fig. 1. Posterior distributions for fixed effects parameter estimates for the best fitting model of alien bird establishment success. Boxplots summarise the posterior marginal distributions for all fixed-effects parameters (β) from a Bayesian regression of the most conservative data subset ($n = 1530$ introductions). Box widths show the interquartile range, the mean is represented as a bold vertical line within each box, and whiskers the 2.5th and 97.5th percentiles (i.e. the 95% credibility interval) of the distribution. Colours indicate the fixed effect category, and bold y-axis labels indicate that there is evidence for a non-zero slope for the described data variable. Further details are in Extended Data Table 1.

A

Event Location Species

B

Abiotic Biotic Life Hist. Propagule
Anthrop. Ecology Match

Fig. 2. Relative effect size of different categories of predictors in the best fitting model of alien bird establishment success.

Each wedge represents the sum of the change in $wAIC$ for the fixed effects in each category when added to a Bayesian regression of establishment success versus failure ($n = 1530$ introductions). The left-hand panel (a) presents variables classified into location-, species- and event-level categories, while the right-hand panel (b) presents the sub-categories within those broad levels ($n = 1530$ introduction events).

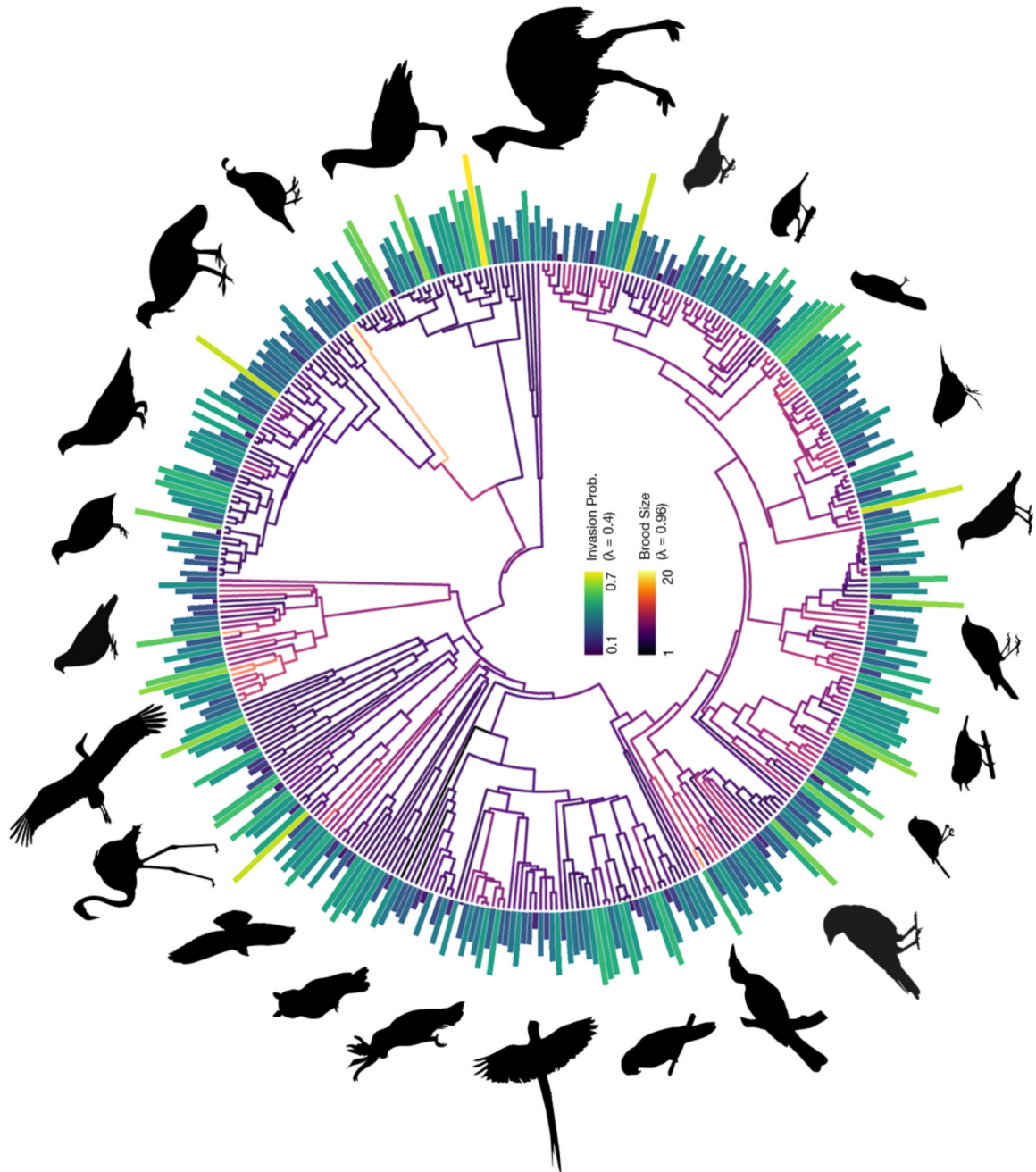


Fig. 3. Phylogenetic patterns of invasion probability across alien birds.

Shows 358 species with the highest quality information on introduction events. Blue-green-yellow outer bars show the mean establishment potential of a species across all 1-degree grid cells beyond its native range, with longer and yellower bars indicating that a species has greater potential to establish outside its native range. Phylogenetic branches are coloured according to brood size, with lighter colours indicating higher brood sizes, and darker

colours lower brood sizes. Silhouettes (from <http://phylopic.org/>) show the approximate location of avian families.