

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

David W. Redding¹, Alex L. Pigot¹, Ellie E. Dyer¹, Çağan H. Şekercioğlu^{2,3}, Salit Kark⁴ & Tim M. Blackburn^{1,5*}

Human-mediated translocation of species to areas beyond their natural distribution (which results in ‘alien’ populations¹) is a key signature of the Anthropocene², and is a primary global driver of biodiversity loss and environmental change³. Stemming the tide of invasions requires understanding why some species fail to establish alien populations, and others succeed. To achieve this, we need to integrate the effects of features of the introduction site, the species introduced and the specific introduction event. Determining which, if any, location-level factors affect the success of establishment has proven difficult, owing 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 introduction events of alien birds⁴ to show that environmental conditions at the introduction location, notably climatic suitability and the presence of other groups of alien species, are the primary determinants of successful establishment. Species-level traits and the size of the founding population (propagule pressure) exert secondary, but important, effects on success. Thus, current trajectories of anthropogenic environmental change will most probably facilitate future incursions by alien species, but predicting future invasions will require the integration of 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 (that is, 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 the establishment of alien species focused 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, behavioural traits¹¹ and ecological traits¹²) have subsequently been shown to explain variation in the successful establishment of alien populations. However, determining which—if any—location-level factors have a general effect on success at a global level, and across large taxonomic groups, has proven challenging for several reasons.

First, many biotic (for example, recipient assemblage composition¹³) and abiotic (for example, climate¹⁴ and 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 organization—at the population (for example, stochastic weather events), species (for example, climatic affinity), community (for example, native species richness) and landscape (for example, habitat composition) levels. Third, how a new environment interacts with a species is dependent on the evolutionary and adaptive history of the species that is introduced¹⁶: a harsh environment for a house sparrow (*Passer domesticus*) may or may not be harsh for the closely related

Eurasian tree sparrow (*Passer 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 (for example, latitude¹⁸ or island versus continent¹⁹) or gross differences between native and alien environments²⁰, and typically ignore spatial autocorrelation²¹. Therefore, we await an integrated analysis of variation in the establishment of alien populations.

Here we undertake a global analysis to identify both the absolute and relative contributions of location, species and event-level processes in predicting the establishment of alien populations. Using birds as a model system, we interrogate data on the success or failure of 4,346 individual introduction events that span 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 factors account for both the mean and temporal variability in the abiotic environment, the suitability of the environment in terms of its similarity to the conditions that are experienced by a species in its native range (termed ‘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 hypothesized 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 approximation²³. This method provides efficient and accurate parameter estimations for complex inferences that incorporate both random and fixed effects, which allows 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; $n = 1,530$, Watanabe–Akaike information criterion (w AIC) = 892.96, area under receiver operating curve = 0.75; see ‘Statistical modelling outline’ in Methods for further details). This result was robust to the precise way in which introduction events were defined (Extended Data Fig. 1) and highlights the fact 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).

¹Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, London, UK. ²Biodiversity and Conservation Ecology Laboratory, Department of Biology, University of Utah, Salt Lake City, UT, USA. ³College of Sciences, Koç University, Istanbul, Turkey. ⁴The Biodiversity Research Group, The School of Biological Sciences, The University of Queensland, Brisbane, Queensland, Australia. ⁵Institute of Zoology, Zoological Society of London, London, UK. *e-mail: t.blackburn@ucl.ac.uk

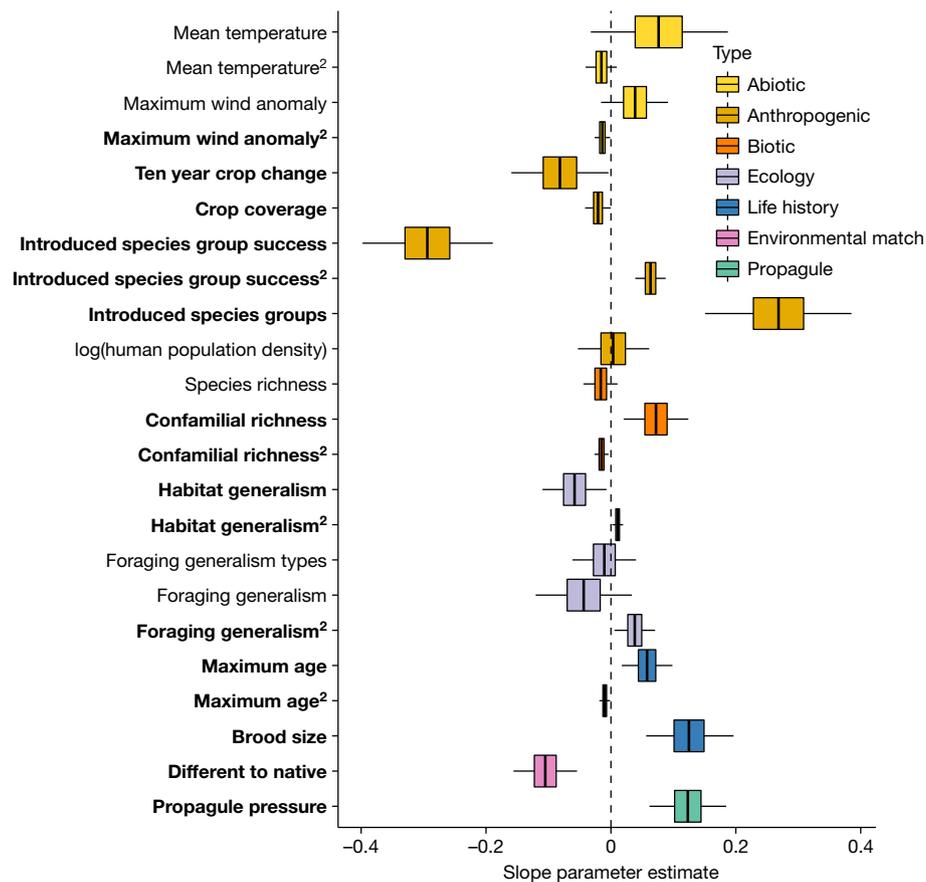


Fig. 1 | Posterior distributions for fixed-effects parameter estimates for the best-fitting model of the success of alien bird establishment. Box plots summarize the posterior marginal distributions for all fixed-effects parameters (β) from a Bayesian regression of the most-conservative subset of the data ($n = 1,530$ introductions). Box widths show the interquartile

range, the mean is represented as a bold vertical line within each box, and whiskers denote the 2.5–97.5th percentiles (that is, 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. For further details, see Extended Data Table 1.

A strong anthropogenic determinant of establishment success is the number of alien taxonomic groups that are already established at a location at the time of introduction. The positive effect of the number of alien groups that have been introduced is broadly consistent with the invasion meltdown hypothesis¹⁷, which states that the ecological disruptions that are caused by (or enable) earlier invasions facilitate further successful introductions. This result is not simply indexing anthropogenic environmental disturbance; although crop coverage and human population density were included in the best-fitting model, they did not have a strong and consistent global signal for the successful establishment of alien populations (Fig. 1, Extended Data Fig. 2). This may be due to historical patterns of introductions being mainly restricted to areas that are already disturbed¹². In fact, our analysis shows that less-disturbed areas have higher rates of establishment success, with rapid agricultural land conversion not only causing native species declines² but also negatively affecting alien species—at least during 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 the successful establishment of alien populations is highest in locations in which environmental conditions are more similar to those in the species' native range (environmental match, Figs. 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 that

are naturally experienced by each introduced species. Environmental extremes are also important²⁴, with establishment success reduced by the occurrence of historical storm events in the period immediately after introduction. Anecdotal evidence had previously suggested that extreme weather was a cause of specific establishment failures (for example, 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 the species richness of native birds on the successful establishment of aliens; this switched to a hump-shaped relationship (Fig. 1) when considering only the most closely related species, which are presumably similar in ecological terms. These results help to clarify previous contradictory findings by showing that, although overall native biodiversity may inhibit invasions, this effect is relatively weak compared to other extrinsic and intrinsic factors, and it may be partially masked by the tendency for locations with some closely related species to be more environmentally suitable for aliens introduced there, and thus be more susceptible to establishment (that is, the 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, in birds larger brood sizes promote

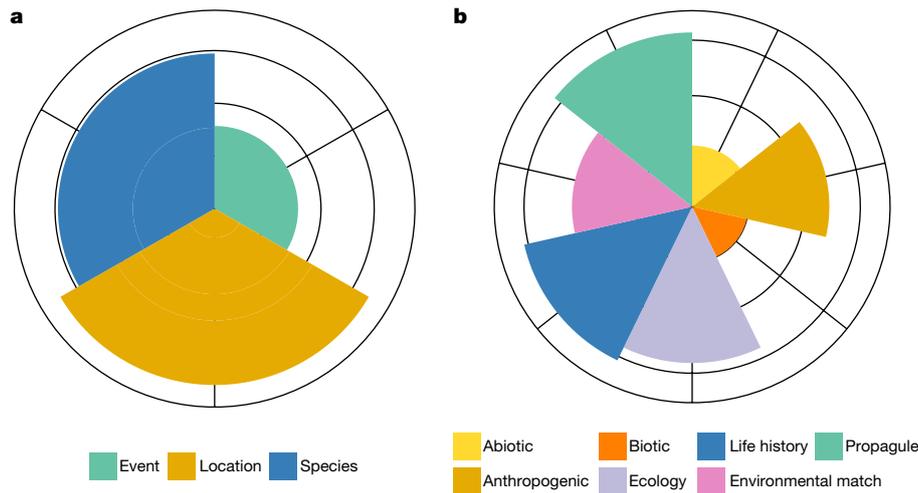


Fig. 2 | Relative effect size of different categories of predictors in the best-fitting model of the success of alien bird establishment. Each wedge represents the sum of the change in w_{AIC} for the fixed effects in each category when added to a Bayesian regression of establishment

establishment, whereas lifespan showed a hump-shaped relationship with invasion success (Fig. 1, Extended Data Fig. 2), which confirms previous evidence of a trade-off between the benefits of fast and slow life histories¹⁰. Although species with fast life histories can quickly gain a 'foothold' at a new location through rapid population growth, slower life histories provide resilience against demographic and environmental variation, which enables alien populations to better ride out extreme conditions^{27,28}. In our model, there is also evidence that foraging

success versus failure ($n = 1,530$ introductions). **a**, Variables classified into location-, species- and event-level categories. **b**, Sub-categories within the broad levels shown in **a** ($n = 1,530$ introduction events).

generalism and habitat-use generalism may, together, increase establishment success. Life-history variables are, in general, strongly phylogenetically conserved (for example, brood size, $\lambda = 0.96$; Fig. 3), which implies that related species may have similar rates of establishment success. However, globally, establishment success has a much weaker phylogenetic signal ($\lambda = 0.4$; Fig. 3), owing to the fact that phylogenetically conserved traits are overwhelmed by the combined spatial effects of the local environment and propagule pressure, which tend to exhibit

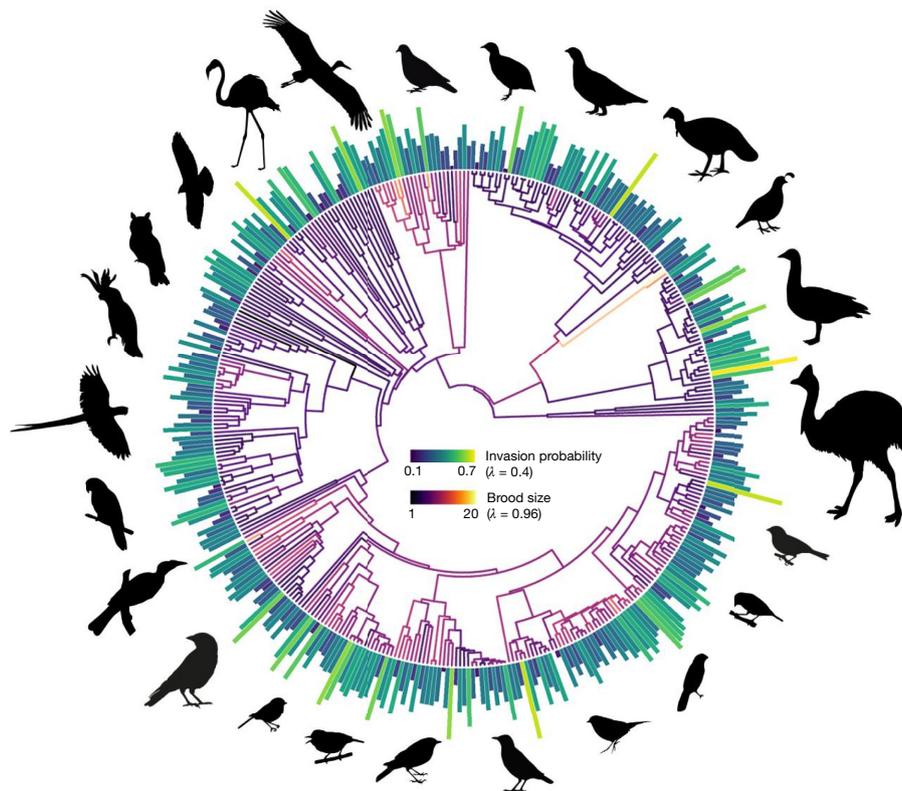


Fig. 3 | Phylogenetic patterns of invasion probability across alien birds. This figure shows 358 species with the highest-quality information on introduction events. Blue, green and yellow outer bars show the mean establishment potential of a species across all 1° grid cells beyond its native range, 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 indicating lower brood sizes. Silhouettes (from <http://phylopic.org/>) show the approximate location of avian taxa.

little phylogenetic signal. The inherently idiosyncratic nature of these effects with regard to the identity of the species introduced (the 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.

Finally, we confirm the strong general role of propagule pressure which, consistent 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 owing to stochastic and Allee effects, whereas the success of larger populations³⁰ depends on the species- and location-level effects that we identify here. Our analysis highlights the key role of the presence of other groups of alien species in establishment success (which suggests that locations that are already hotspots for introductions are especially susceptible to accumulating alien species), but also shows 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 exacerbated depending on the precise combinations of species and sites at which the introductions are occurring. Our analyses confirm the urgent need for enhanced management programmes to prevent or mitigate the negative effect of these invasions.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1292-2>.

Received: 1 October 2018; Accepted: 17 May 2019;

Published online 19 June 2019.

- Blackburn, T. M. et al. A proposed unified framework for biological invasions. *Trends Ecol. Evol.* **26**, 333–339 (2011).
- Lewis, S. L. & Maslin, M. A. Defining the Anthropocene. *Nature* **519**, 171–180 (2015).
- Pimentel, D. *Biological Invasions: Economic and Environmental Costs of Alien Plant, Animal, and Microbe Species* (CRC, Boca Raton, 2011).
- Dyer, E. E., Redding, D. W. & Blackburn, T. M. The global avian invasions atlas, a database of alien bird distributions worldwide. *Sci. Data* **4**, 170041 (2017).
- Seebens, H. et al. No saturation in the accumulation of alien species worldwide. *Nat. Commun.* **8**, 14435 (2017).
- Banks, N. C., Paini, D. R., Bayliss, K. L. & Hodda, M. The role of global trade and transport network topology in the human-mediated dispersal of alien species. *Ecol. Lett.* **18**, 188–199 (2015).
- Drake, J. A. et al. *Biological Invasions: A Global Perspective* (John Wiley & Sons, Chichester, 1989).
- Crawley, M. in *Colonization, Succession And Stability. 26th Symposium of the British Ecological Society* (eds Gray, A. J. et al.) 429–453 (Blackwell Scientific Publications, Oxford, 1986).
- Lockwood, J. L., Cassey, P. & Blackburn, T. The role of propagule pressure in explaining species invasions. *Trends Ecol. Evol.* **20**, 223–228 (2005).
- Sol, D. et al. Unraveling the life history of successful invaders. *Science* **337**, 580–583 (2012).
- Sol, D., Griffin, A. S., Bartomeus, I. & Boyce, H. Exploring or avoiding novel food resources? The novelty conflict in an invasive bird. *PLoS ONE* **6**, e19535 (2011).
- Sol, D., González-Lagos, C., Lapidra, O. & Díaz, M. in *Ecology and Conservation of Birds in Urban Environments* (eds Murgui, E. & Hedblom, M.) 75–89 (Springer, 2017).
- Duncan, R. & Forsyth, D. in *Conceptual Ecology and Invasion Biology: Reciprocal Approaches to Nature* (eds Cadotte, M. W. et al.) 405–421 (Springer, 2006).
- Peterson, A. T. Predicting the geography of species' invasions via ecological niche modeling. *Q. Rev. Biol.* **78**, 419–433 (2003).
- Wagner, V. et al. Alien plant invasions in European woodlands. *Divers. Distrib.* **23**, 969–981 (2017).
- Duncan, R. P., Blackburn, T. M. & Sol, D. The ecology of bird introductions. *Annu. Rev. Ecol. Syst.* **34**, 71–98 (2003).
- Jeschke, J. M. et al. Support for major hypotheses in invasion biology is uneven and declining. *NeoBiota* **14**, 1–20 (2012).
- Pyšek, P. & Richardson, D. M. The biogeography of naturalization in alien plants. *J. Biogeogr.* **33**, 2040–2050 (2006).
- Jeschke, J. M. Across islands and continents, mammals are more successful invaders than birds. *Divers. Distrib.* **14**, 913–916 (2008).
- 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* **39**, 665–675 (2016).
- Case, T. J. Global patterns in the establishment and distribution of exotic birds. *Biol. Conserv.* **78**, 69–96 (1996).
- Cassey, P., Blackburn, T. M., Sol, D., Duncan, R. P. & Lockwood, J. L. Global patterns of introduction effort and establishment success in birds. *Proc. R. Soc. Lond. B* **271**, S405–S408 (2004).
- Rue, H., Martino, S. & Chopin, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. B* **71**, 319–392 (2009).
- Ameca y Juárez, E. I. A., Mace, G. M., Cowlishaw, G. & Pettorelli, N. Natural population die-offs: causes and consequences for terrestrial mammals. *Trends Ecol. Evol.* **27**, 272–277 (2012).
- Cheke, A. Seafaring behaviour in house crows *Corvus splendens*—a precursor to ship-assisted dispersal? *Phelsuma* **16**, 65–68 (2008).
- Fridley, J. D. et al. The invasion paradox: reconciling pattern and process in species invasions. *Ecology* **88**, 3–17 (2007).
- Allen, W. L., Street, S. E. & Capellini, I. Fast life history traits promote invasion success in amphibians and reptiles. *Ecol. Lett.* **20**, 222–230 (2017).
- Sæther, B.-E. et al. Life-history variation predicts the effects of demographic stochasticity on avian population dynamics. *Am. Nat.* **164**, 793–802 (2004).
- Hayes, K. R. & Barry, S. C. Are there any consistent predictors of invasion success? *Biol. Inv.* **10**, 483–506 (2008).
- Duncan, R. P., Blackburn, T. M., Rossinelli, S. & Bacher, S. Quantifying invasion risk: the relationship between establishment probability and founding population size. *Methods Ecol. Evol.* **5**, 1255–1263 (2014).
- Dyer, E. E. et al. The global distribution and drivers of alien bird species richness. *PLoS Biol.* **15**, e2000942 (2017).
- Simberloff, D. & Von Holle, B. Positive interactions of nonindigenous species: invasional meltdown? *Biol. Inv.* **1**, 21–32 (1999).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

METHODS

Alien introduction events. We collated all of the records from the GAVIA database of global bird introductions^{4,31} that contained geo-referenced introduction events at known specific (for example, sub-national and below) point locations (that is, the locations at which 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 5,834 records, with accompanying spatial polygon data created by drawing around the smallest geographical unit to which the introduction event could be reasonably attributed⁴. Some records were at a specific location (for example, a single building address, park or harbour), and a small minority of event records could only be assigned at a coarser spatial scale (for example, city or county). We note there is some geographical bias in the records: most of the introductions occur in the Australasian region (19% of records), followed by the Palaearctic (18%), Oceanic (17%) and the Nearctic (16%), with the fewest records being from the Afrotropic (12%), Neotropic (10%), Indomalayan (8%) and Antarctic (<1%) realms. However, introduction events occur predominately within regions and high-profile historical routes between continents (for example, Europe to Australia) are relatively rare given the huge increase in introductions in recent decades⁵; for example, through accidental transport and the trade in wildlife³¹ (Extended Data Fig. 3).

Using data from the original source in GAVIA as well as external sources, we thoroughly checked all introduction records for potential errors and removed any records that were possibly dubious—usually owing to misreported dates or locations across multiple references. This resulted in $n = 4,346$ unique introduction-event records (Supplementary Data 1). We then used text information (again from the original source) to categorize the introduction events as either known, specific introduction events (introductions, $n = 1,784$), first sightings (sightings, $n = 584$) or as having no clear designation (unknown, $n = 1,978$). Finally, we noted whether any records were part of a chronological sequence of introduction events that involved a single species at a single location; for example, the Eurasian skylark (*Alauda arvensis*) was imported and released on 6 separate occasions in the Barrabool Hills (Victoria, Australia) from 1852 to 1880. From these we created four data subsets of decreasing size but increasing specificity: ‘All Records’ contained all the records in the database ($n = 4,346$); ‘Intro. & Unk.’ contained records that are known introductions as well as records that have no detailed description ($n = 3,762$); ‘Introductions’ contained all records that are specified as detailed introduction events ($n = 1,784$); and ‘Last Introduction’ ($n = 1,530$). This final subset of the data contained known introduction events, but with events that were part of a chronological sequence of introduction events collapsed into a single record, summarized using the date of the last introduction event and the cumulative propagule pressure across events. Owing to the exhaustive nature of the data collection process that we used, all available data were used in each component analysis and no statistical methods were used to predetermine sample size.

Based on previous hypotheses of avian establishment success reported in the literature (Supplementary Data 2), we collated information on a wide range of covariates that could reasonably affect establishment success. We categorized these covariates into the three categories of establishment determinants, as previously defined¹⁶—location-, species- and event-level factors—and further distinguished between different types of species- and location-level factors as described 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 and ‘great numbers’ = 250. When describing propagule number: ‘repeated’ = 5, ‘several’ = 5, ‘releases’ = 2, ‘numerous’ = 10, ‘many’ = 10 and ‘frequent’ = 10. We decided on these numbers by summarizing, 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 (five 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 that have previously been 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 in grams³⁵. Species for which data could not be collected (clutch size (11%), number of clutches

per year (48%), age at first breeding (66%) and lifespan (52%)) were assigned the mean value of the lowest inclusive taxonomic rank (that is, genus, family or order) for which data were available. This approach is justified because most of the variance in avian traits, as calculated from our data, occurs at taxonomic levels above that of genus (clutch size (91%), number of clutches per year (70%), age at first breeding (83%) and lifespan (62%)). We also include a previously used measure known as 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. **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 previously 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 (that is, genus, family or 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.

Ecological traits. Data on species-specific diets and foraging strategies came from a previous publication³⁸: for both of these variables, a total value of 100 was divided between categories to represent the percentage of time that 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 database³⁹. 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 D^{40} .

Location-level factors. Abiotic environment. Global geophysical data (altitude above sea level and slope percentage) were downloaded in pre-processed geoTiff format at 1-km grid scale⁴¹. A third variable, altitude variance, was computed with the R function `aggregate` (raster package⁴²) using the variance of the altitude values of 3×3 grid cells, such that all 9 cells had the same final value. Bioclimatic 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-second (~ 1 -km) resolution. They consisted of mean annual temperature, annual variation in temperature (temperature seasonality⁴³), mean annual rainfall and annual variation in temperature (precipitation seasonality⁴³). Abiotic data for islands came from an island-specific dataset⁴⁴ and had the same variables as for large terrestrial land masses, but were represented as a spatially referenced spreadsheet that contained 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 dataset. We also used a measure of remoteness (again giving continents a value of zero), using the ‘SLMP’ column of the dataset.

Historical climate data (1850–2007) were downloaded as 6×4 -km 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 dataset⁴⁵. Historical spatiotemporal land-cover data (1700–2007) were downloaded as global ESRI format ascii data grids at ~ 5 -km resolution, consisting of the proportion cover of primary and secondary habitats, from the Harmonised Land Use dataset⁴⁶. To reduce collinearity in a regression (that is, the dummy variable trap) the ‘other’-category data were 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 that intersected the introduction-event polygon (R function `extract`⁴²). For very specific introduction events, this would be the single cell in which the event polygon was located; for less-specific data for the introduction event, there was sometimes more than one grid cell that overlapped the polygon. For the spatiotemporal data, extractions for each record were undertaken only on the temporally nearest data layers. For example, for the historical climate data, a maximum anomaly value at that location over the 10 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 represent a yearly rather than a 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 and 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-cover types, we may not be able to resolve their specific effects on invasion success. For each land-cover type noted above, an additional variable was constructed; this was the gradient of change in land cover in each grid cell over the ten years before each introduction event. This was calculated 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.

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 ($\sim 110 \times 110$ km) in a Behrmann projection. These maps show the extent of occurrence for each species, and so are relatively crude depictions of the area that is 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; and BIO15, precipitation seasonality)⁴⁵. Using each of these input variables (to capture, as a single variable, the environmental match between the introduction site and the environmental conditions 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 environment', 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 of measuring native preferences, as fine-scale habitat variation within the range may act to bias the mean value; however, finer-scale data are not available for all introduced species.

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 (in millions of years) that separate an introduced alien species from its closest relative in the recipient community. Recipient communities were designated as those bird species that had ranges that overlapped with any of the introduction polygon—though we note that not all species in this sample would be interacting if they used very different habitats. Phylogenetic distances were calculated as the mean across 100 phylogenies sampled at random from the posterior distribution of trees from a previous phylogeny⁴⁷ with the Hackett backbone. These variables thus quantify the overall species richness of the location of introduction (metric (i)), and the location's richness relative to the phylogenetic position of the introduced species (metrics (ii), (iii) and (iv)).

Anthropogenic environment. To determine the role of human disturbance and urbanization¹² 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 whether any alien species were present at a given site before 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, and 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 spatiotemporal knowledge currently available; 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 before 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, as implemented in the R package R-INLA⁴⁸ (Supplementary Information, code 1). We used this method because it provides accurate parameter (for example, 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 w AIC⁴⁹ and the conditional predictive ordinate (CPO). w AIC 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, w AIC provides a method to penalize 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 (for example, 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 w AIC tends to $\sum \ln(\text{CPO})$ under ideal circumstances⁵¹ (in our case, Spearman correlation ~ 0.99), we henceforth report only w AIC as a proxy for CPO.

We model the 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 the central limit theorem given our large numbers of trials. This was due to the computational efficiency of using 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' and 'failed') to 0. For a large proportion of records, the success or failure of an avian introduction was unknown ($n = 2,234$). 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; <https://www.gbif.org/>). 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 'observation date' associated 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 past 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 past 10 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 degree, 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 w AIC 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 visualizing 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 Fig. 4a, b) and spatially mapping the same values (Extended Data Fig. 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 used an AUC approach (AUC 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, and an AUC value close to 1 indicates a perfect predictive ability.

Analysis protocol. We first used the most-conservative dataset for analysis—which contained known introductions events ($n = 1,530$ records) (Supplementary 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 (Extended Data Table 1). For each additional step of complexity, we recorded the change (Δ) in the w AIC value⁴⁹ and only included random effects that increased the fit of the model by more than two w AIC units. We used uninformative priors in all cases except for the spatial term, for which 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 owing to the large number of covariates and high collinearity between them), we then added all explanatory variables (Supplementary Data 2) into a regression model in a

stepwise manner and, after each step, assessed model fit using the $wAIC$ value. So that the effect sizes of the different covariates could be better compared, each explanatory variable was standardized 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 $\Delta_wAIC > 2$ to select better models⁵³. 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 situations in which 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 whether 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 subsets of data. 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 ($n = 19,561$ 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 probably make up the vast majority of future invasions⁵. When predicting, we set the random-effects introduction year as 2015 to be as close to the present day as possible and propagule pressure as 150, the lowest value after the threshold beyond which the number introduced has limited effect (Extended Data Fig. 2). All predicted values for which the confidence was low (such that the 95% confidence intervals for the grid cell estimate covered 0 and 1), were 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 `phylosig`⁵⁴) 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.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All data generated or analysed during this study are included with the paper and its Supplementary Information.

Code availability

Code used to calculate the final model is included in Supplementary Information.

33. Myhrvold, N. P. et al. An amniote life-history database to perform comparative analyses with birds, mammals, and reptiles. *Ecology* **96**, 3109 (2015).
34. Tacutu, R. et al. Human Ageing Genomic Resources: integrated databases and tools for the biology and genetics of ageing. *Nucleic Acids Res.* **41**, D1027–D1033 (2013).
35. Wilman, H. et al. EltonTraits 1.0: species-level foraging attributes of the world's birds and mammals. *Ecology* **95**, 2027 (2014).
36. Sayol, F. et al. Environmental variation and the evolution of large brains in birds. *Nat. Commun.* **7**, 13971 (2016).

37. Sol, D., Duncan, R. P., Blackburn, T. M., Cassey, P. & Lefebvre, L. Big brains, enhanced cognition, and response of birds to novel environments. *Proc. Natl Acad. Sci. USA* **102**, 5460–5465 (2005).
38. Kissling, W. D., Sekercioglu, C. H. & Jetz, W. Bird dietary guild richness across latitudes, environments and biogeographic regions. *Glob. Ecol. Biogeogr.* **21**, 328–340 (2012).
39. BirdLife International. *IUCN Red List Bird Habitat Classifications* <http://www.iucnredlist.org> (2017).
40. Simpson, E. H. Measurement of diversity. *Nature* **163**, 688 (1949).
41. Hastings, D. A. et al. (eds) *The Global Land One-Kilometer Base Elevation (Globe) Digital Elevation Model*. Version 1.0 <http://www.ngdc.noaa.gov/mgg/topo/globe.html> (National Oceanic and Atmospheric Administration, National Geophysical Data Center, 1999) (2015).
42. Hijmans, R. J. et al. *raster: Geographic Data Analysis and Modeling*. <https://cran.r-project.org/web/packages/raster/index.html> (2016).
43. Hijmans, R. et al. *WorldClim*. Version 1.3 (Univ. California, Berkeley, 2005).
44. Weigelt, P., Jetz, W. & Kreft, H. Bioclimatic and physical characterization of the world's islands. *Proc. Natl Acad. Sci. USA* **110**, 15307–15312 (2013).
45. Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. & Jones, P. D. Uncertainty estimates in regional and global observed temperature changes: a new data set from 1850. *J. Geophys. Res. Atmos.* **111**, D12106 (2006).
46. Hurtt, G. C. 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. *Clim. Change* **109**, 117 (2011).
47. Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. The global diversity of birds in space and time. *Nature* **491**, 444–448 (2012).
48. Blangiardo, M., Cameletti, M., Baio, G. & Rue, H. Spatial and spatio-temporal models with R-INLA. *Spat. Spatiotemporal Epidemiol.* **7**, 39–55 (2013).
49. Watanabe, S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* **11**, 3571–3594 (2010).
50. Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. The deviance information criterion: 12 years on. *J. R. Stat. Soc. B* **76**, 485–493 (2014).
51. Gelman, A., Hwang, J. & Vehtari, A. Understanding predictive information criteria for Bayesian models. *Stat. Comput.* **24**, 997–1016 (2014).
52. Sullivan, B. L. et al. eBird: A citizen-based bird observation network in the biological sciences. *Biol. Conserv.* **142**, 2282–2292 (2009).
53. Raftery, A. E. Bayesian model selection in social research. *Sociol. Methodol.* **25**, 111–163 (1995).
54. Revell, L. J. `phytools`: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).

Acknowledgements We thank M. Parnell, V. Franks, F. Spooner, R. Herdson, E. Jones, F. Davis and M. Susko 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.), a Leverhulme Trust grant (RPG-2015-073) (T.M.B., A.L.P. and E.E.D.), and from a King Saud University Distinguished Scientist Research Fellowship (T.M.B., D.W.R. and E.E.D.). D.W.R. is supported by a MRC UKRI/Rutherford Fellowship (MR/R02491X/1). A.L.P. is supported by a Royal Society University Research Fellowship (UF160281).

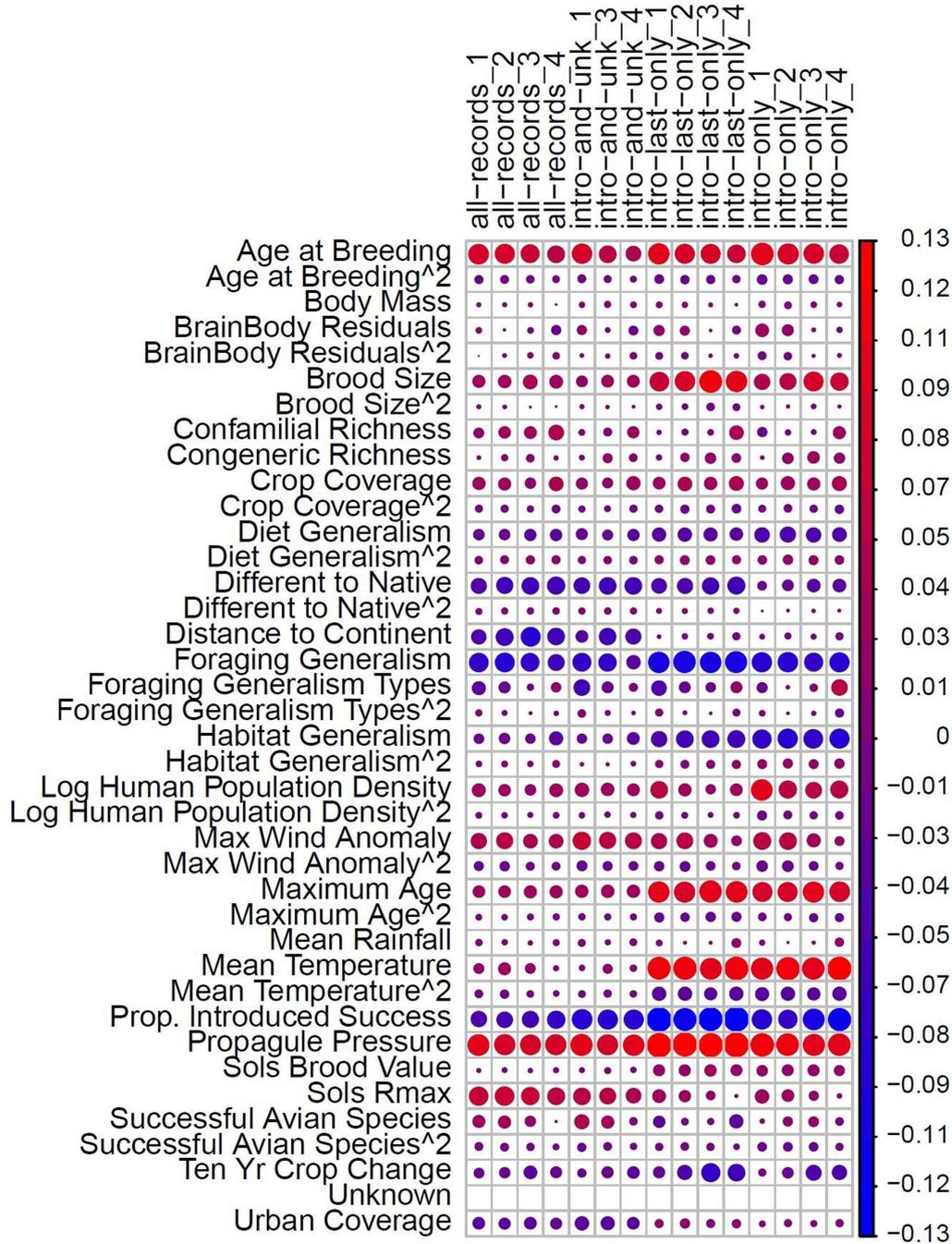
Author contributions D.W.R., A.L.P. and T.M.B. developed the overall study design. E.E.D., S.K. and C.H.S. oversaw initial data collation. D.W.R. and A.L.P. carried out the modelling and data processing, with assistance from T.M.B. All authors contributed to writing the manuscript.

Competing interests The authors declare no competing interests.

Additional information

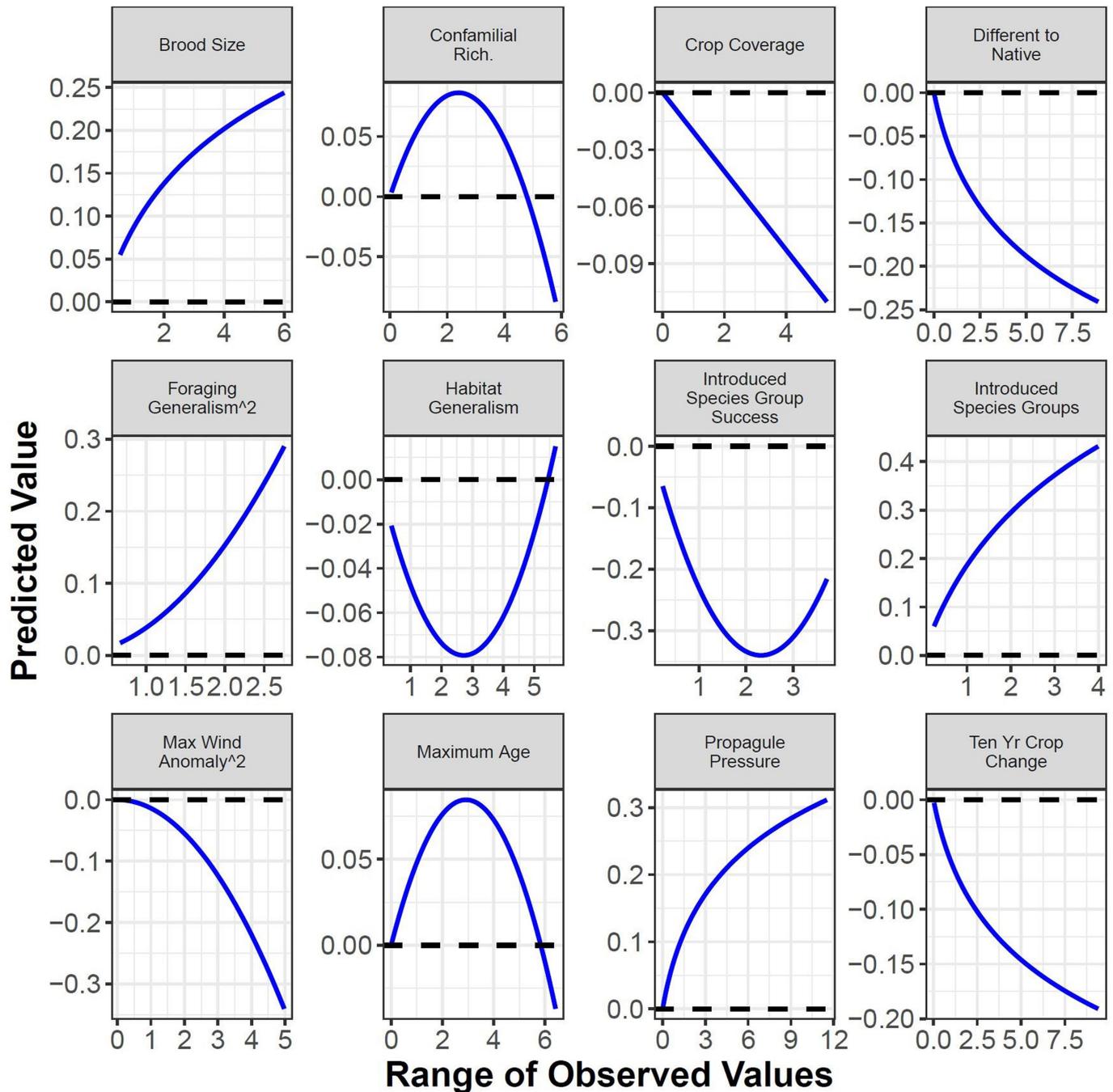
Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1292-2>.

Correspondence and requests for materials should be addressed to T.M.B. **Reprints and permissions information** is available at <http://www.nature.com/reprints>.



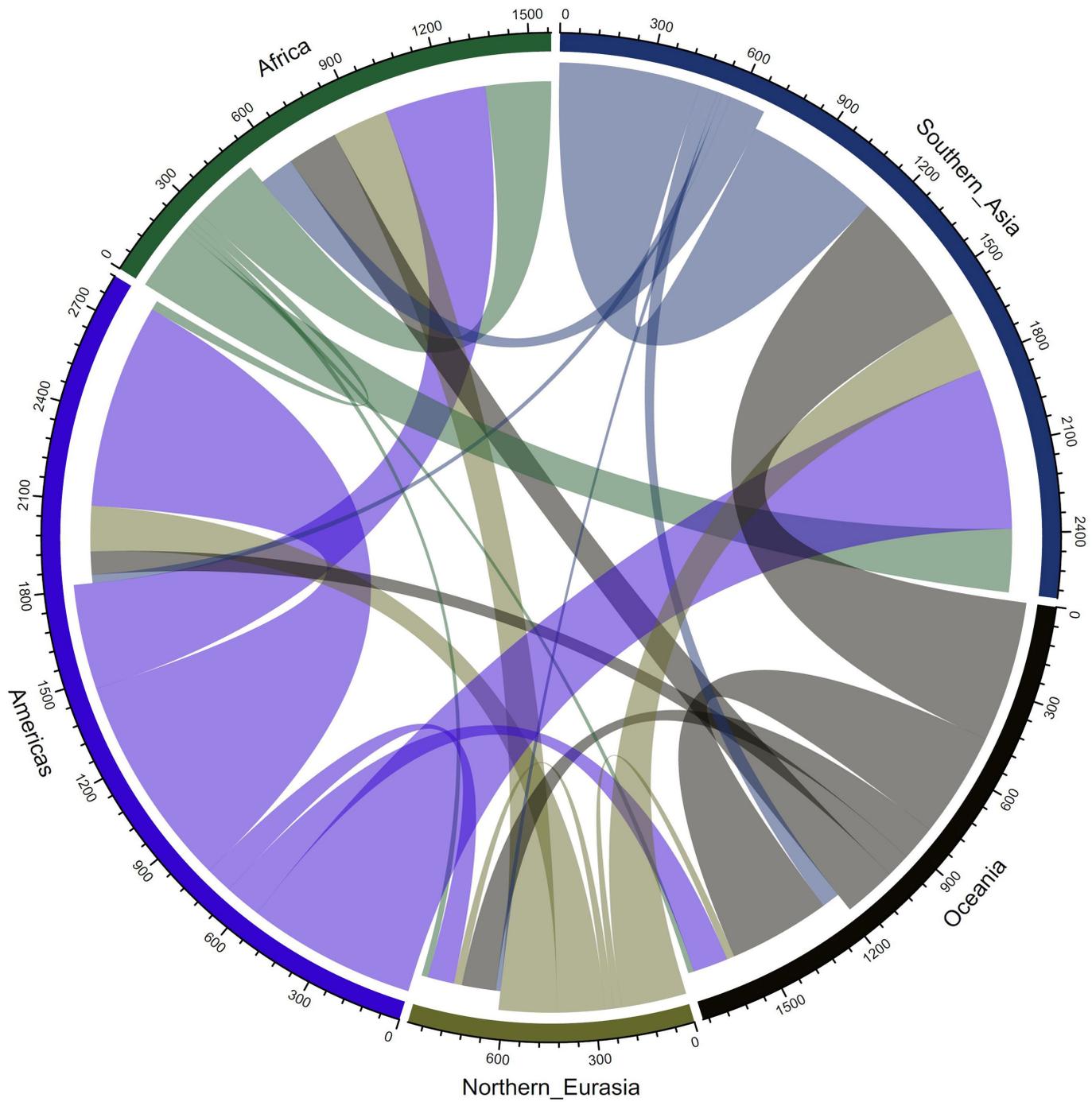
Extended Data Fig. 1 | Sensitivity analysis of slope (β) estimates for the linear terms of a subset of variables across all versions of the input data. Dot size is the size of the β value, with colour representing direction (red, positive; blue, negative). Each row label represents the name of the fixed effect. The column headings represent the subset of data used: 'all-records', all data ($n = 4,346$); 'intro-and-unk', all data, but one record

per species-location event ($n = 3,762$); 'intro-only', detailed introductions only ($n = 1,784$); 'intro-last-only', detailed introductions, but one record per species-location event ($n = 1,530$). The number at end of each column heading indicates the relative size of the buffer used to impute establishment status (Methods).



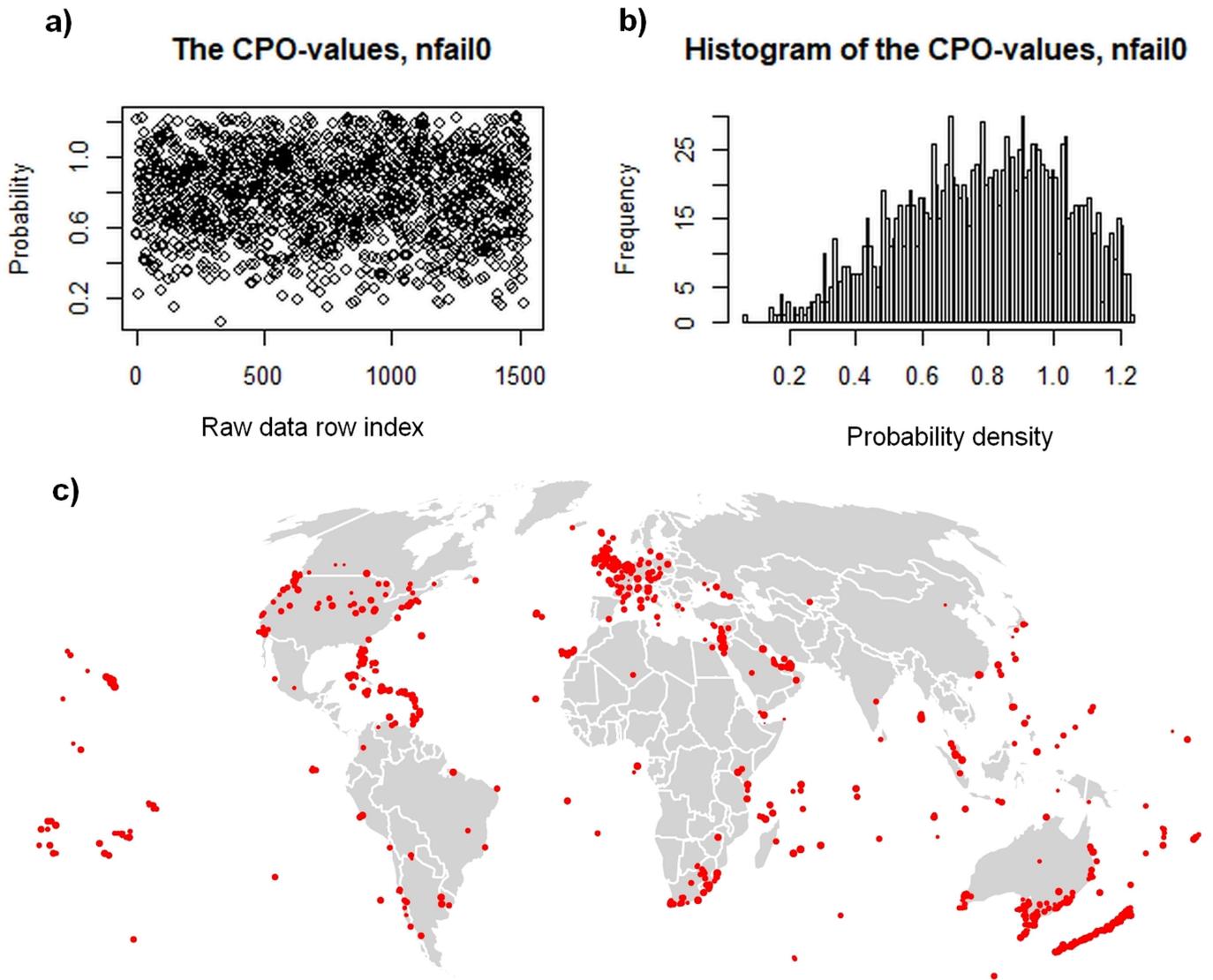
Extended Data Fig. 2 | Approximate shape of fixed effects over the range of observed values. Each panel represents the prediction using β slope estimates from the lowest w AIC model over the known range of values for that given fixed effect (identified by strip title) from the raw

data. Only fixed effects for which the values were unlikely to include zero are included. All panels from a single Bayesian regression of global avian establishment success ($n = 1,530$ introductions).



Extended Data Fig. 3 | Chord diagram showing the directions of origin and introduction location of avian introduction events between regions of the world. The chords near the edge represent introductions to a region;

chords away from the edge show origins of introduction. The width of chord is the relative number of introduction events ($n = 4,346$).



Extended Data Fig. 4 | Model diagnostics from the best-fitting model.
a, Plot of out-of-sample CPO scores for all data points in rank order used in the model. **b,** Probability density of the CPO scores. **c,** Map of CPO scores. CPO is the probability of generating each data point in the

dataset from a posterior fitted without this data point. Each panel allows visualization of where in the data the model does not fit well. All plots from a single Bayesian regression of global establishment success of avian introductions ($n = 1,530$ introductions).

Extended Data Table 1 | All covariates in the best-fitting model, from a Bayesian regression of global avian establishment success

| 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 |

$n = 1,530$ 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 s.d. for the mean value (s.d. β), and the 25th, 50th and 75th quantiles (25% quant, 50% quant and 75% quant) of the posterior distribution of each β estimate.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|-----------------|---|
| Sample size | Literature was exhaustively searched over a 3 year period for as many avian introduction events as possible. The sample size consisted of all records that could be found on this topic during the data collection phase. The sample size was maximised to provide the greatest coverage of data both temporally and spatially and to maximise statistical power. We believe that the dataset represents the majority of the avian introduction data that are available in the scientific literature. |
| Data exclusions | Data that were excluded are marked in the data table and only excluded if accompanying text from the source publication stated that they were dubious records. This meant if the accompanying description data, which contained contextual information about the record taken from the original source literature, contained terms that cast any doubt on the veracity of the record. Such terms were "uncertain", "dubious", "unclear", "possibly" in relation to species identification or the introduction event itself. Also records were not used if the species was considered "vagrant" or "native" to introduction site and therefore not a human introduction. |
| Replication | This is a unique historical dataset of a large majority of documented introduction - replication would only be possible on another taxonomic group but this would take years of effort. |
| Randomization | We randomized the data when performing cross-validation analyses to test the impacts of influential data. |
| Blinding | This is not relevant to the study as there is only one set of data available. |

Reporting for specific materials, systems and methods

Materials & experimental systems

| n/a | Involved in the study |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Unique biological materials |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants |

Methods

| n/a | Involved in the study |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |