Tempo and timing of ecological trait divergence associated with transitions to coexistence in birds

Jay P. McEntee[1,2], Joseph A. Tobias[3], & J. Gordon Burleigh[1]

Affiliations:

[1]Biology Department, University of Florida, PO Box 118525, 220 Bartram Hall, Gainesville, FL 32611-8525, USA

[2]Ecology and Evolutionary Biology Department, University of Arizona, PO Box 210088, Biological Sciences West Room 310, 1041 E. Lowell St., Tucson, Arizona 85721, USA

[3]Department of Life Sciences, Imperial College London, Silwood Park, Ascot, Berkshire, SL5 7PY, UK

Number of words in the summary paragraph (252), in the manuscript as a whole, excluding summary paragraph, references, methods, and captions (1,780)

Number of words in Methods Summary (360) and in Methods for online version of the article (2,162)

Number of references in main text (50)

Number of figures (3) and tables (0)

Name and complete mailing address of the person to whom correspondence should be sent:

Jay McEntee, Biology Department, University of Florida, PO Box 118525, 220 Bartram Hall, Gainesville, FL 32611-8525, USA Email: jaymcentee@ufl.edu

Summary paragraph

Speciation in vertebrates is often viewed as a three-stage process beginning with an allopatric phase (geographic isolation), followed by secondary contact, and finally the transition to coexistence in overlapping geographical ranges (sympatry)[1-4]. In some forms of this model, the delay in establishing secondary sympatry is due to the slow divergence in ecological traits[5], where such divergence reduces competition[6] and/or reproductive interference [3]. However, we know little about the general tempo and timing of ecological trait divergence for allopatrically speciating pairs, and how these factors impact transitions from allopatry to sympatry. Here, we combine divergence time estimates, trait measurements, and geographic range data for 952 avian sister species pairs worldwide to examine the tempo and timing of ecological trait divergence, and how such divergence may impact the three-stage speciation process. Our analyses indicate that sister pair divergences in body mass and beak morphology, important ecological traits, are better explained by a pulse-and-stasis evolution model than a gradual divergence model. For sister pairs in secondary contact, body mass divergence and beak divergence are associated

with earlier transitions to sympatry. Our evidence suggests that the contribution of trait divergence to the transition to sympatry stems from pulses of trait divergence early in the speciation process, with a limited contribution from gradual trait divergence. Incorporating early trait divergence pulses of varying magnitude into the three-stage speciation model can explain a perplexing set of observations in bird speciation: prolonged mutual exclusion in some older species pairs, marked geographic divergence early in speciation, and instances of rapid sympatry[7].

TEXT

The relative importance and role of gradual versus pulsed trait evolution in generating diversity continues to be part of a long-lasting debate in evolutionary biology[8-10]. Pulses have been described and modeled as near-instantaneous jumps separated by relatively long periods of relative stasis, while gradual evolution often has been modeled as a random walk. While both gradual and pulsed evolution almost certainly occur, a central question in evolutionary biology is which has greater importance in the generation of broad patterns of diversity. A challenge in this debate is that support for different models likely varies with the temporal scale of analysis. At the coarse temporal scales typical of phylogenetic studies of extant taxa, near-instantaneous pulsed evolution interspersed with relative stasis may be difficult to characterize because the signal of pulses can blur into gradualism. Meanwhile, pulses observed over short microevolutionary (population-level) scales may not contribute strongly to the trait variation evident at macroevolutionary levels[11, 12].

Many proponents of the pulsed evolution model have argued that evolutionary change is concentrated at speciation. As this hypothesis predicts that evolutionary change depends on the number of speciation events instead of clade age in phylogenies, studies have tested this hypothesis on phylogenetic trees, finding some support for speciation-associated pulses[13-16]. However, an issue with parsing evolutionary change between cladogenesis (speciation) and anagenesis (within-lineage evolution) from phylogenetic trees is that speciation events are represented as instantaneous events in the tree. Speciation may instead comprise a series of processes that vary in duration[1] – sometimes referred to as the speciation continuum[17]. Change in ecological traits may be gradual or pulsed within this period, and could occur early[18] or late[5] in the process.

In the allopatric speciation model thought to represent the most common path to speciation in vertebrates, there are three phases: allopatry (geographic isolation), secondary contact, and sympatry (coexistence over substantial area[19]). If speciation is a non-instantaneous process with these phases, we can ask about the mode of trait divergence over the course of speciation. Even when evolutionary

change appears concentrated at cladogenesis (speciation) in phylogenetic studies, this pattern could result from relatively higher rates of gradual change over the duration of speciation, or it could result from qualitatively different near-instantaneous pulses that break up periods of stasis. Interestingly, divergence during the three-stage allopatric speciation model is often depicted as[7] or implicitly assumed to be[20] a continuous, gradual process. However, authors have additionally noted a decoupling of phenotypic and molecular divergence rates over the timescales relevant to speciation processes[7], suggesting that phenotypic divergence may not be gradual over the course of speciation. If pulse-and-stasis tempos, instead of gradualism, predominate even on the relatively short timescale of speciation, there are consequences for how we view the three-stage allopatric speciation process. For example, as we expect that sympatry must wait for ecological and/or reproductive trait divergence to accrue[21], the gradual trait divergence model of allopatric speciation suggests that rapid sympatry should generally be possible only in those clades with the fastest rates of gradual divergence. If pulses generally occur early during the speciation process, rapid sympatry may be expected to follow pulses of larger magnitude instead of being a product of high background rates of gradual divergence. What is needed is an approach that allows us to examine evidence for the mode and timing of divergence across the allopatric speciation process for traits that play a role in mediating transitions through this process.

Here we leverage the variation in phenotypic divergence, divergence times, and geographic stages of speciation among 952 bird species pairs to assess the divergence mode of traits associated with the establishment of sympatry. As we were interested in making inferences that are general across birds, we extracted species pairs from a global phylogeny of birds with 6,714 species as tips[22] (see Supplementary Information). Competition[6] and reproductive interference[23, 24] can prevent the coexistence of incipient and recent species [4], thus variation in trait divergence should predict which species pairs in secondary contact are sympatric (coexisting) versus parapatric (having abutting distributions). Ecological trait divergence should be especially important in mediating sympatry, as it can ease both competition and reproductive interference [18, 25]. We found that divergence in ecological traits is associated with coexistence, as disparity in both body mass (a composite ecological trait) and bill morphology (a foraging-associated trait) predicted sympatry for species pairs in secondary contact (n = 441) across a set of analyses (relative importance = 1 and estimated coefficients > 0 for both trait divergence predictors in multi-model generalized linear modeling approach; Fig. 1, Supplementary Tables 4-5, 8-9, 15-16; see Methods). These analyses further accounted for variation among species pairs in divergence time, latitude, migratory behavior, and dispersal (via a morphological proxy).

Our results are in accordance with previous findings in birds (body mass[26], beak morphology[20],) in suggesting that ecological trait divergence facilitates coexistence. However, the association of trait divergence and sympatry was not as consistent across variants of our analyses for beak morphology as it was for body mass (Fig. 2, Supplementary Tables 4-5, 8-9, 15-16). Body size differences therefore appear to be paramount, with divergence in the foraging trait (beak) being secondary. As divergence time was also an important predictor (relative importance = 1 in multi-modeling framework, Fig. 1) of sympatry for sister pairs in secondary contact, our results suggest that trait divergence and divergence time are decoupled to some degree, in that ecological trait divergence predicts the outcome of secondary contact while controlling for the role of divergence time.

To better understand the role of trait divergence in transitions through three-stage allopatric speciation, we sought to understand the tempo and timing of trait divergence relative to the establishment of secondary contact and sympatry. Thus, we explored a set of evolutionary models to understand the mode of ecological trait divergence at the timescale of speciation. Different modes of divergence (gradual versus pulsed) leave different signatures across sets of species pair divergence times and trait disparities. We considered four divergence models[12] to explain body mass disparities and the component (phylogenetic PC1[27]) of beak morphology associated with sympatry in the main analysis. One model is a pure white noise model in which trait disparity is independent of sister pair age, and the remaining three models combine a white noise component, representing bounded evolution on short timescales, with alternate longer timescale components. The longer-timescale components are either gradual (Brownian motion) or pulsed evolution (where disparity arises through a single or multiple pulses), with the pulse rate determined by a Poisson process. Using AIC to compare models for both traits, we find strongest relative support for models with a single divergence pulse (Supplementary Tables 10-11).

In the preferred single pulse models for trait divergence, the mean estimated waiting times to pulses are ~665,000 years for body mass and ~560,000 years for beaks (Fig. 3 for body mass, Supplementary Tables 10-11). Comparing these estimates with the timing of secondary contact and sympatry in allopatric speciation, it is evident that trait divergence pulses largely precede sympatry (Fig. 3). Divergence pulses may precede even secondary contact in most cases (Extended Data Fig. 1); however, we find that secondary contact is frequent in young sister pairs. Indeed, across species pairs, the estimated intercept for the linear relationship between divergence time and the probability of breeding co-occurrence is .434 (Extended Data Fig. 2), which requires an approximate minimum rate of transition to secondary contact from allopatry of 0.3 per million years (Extended Data Fig. 3,

Supplementary Information). While our evidence suggesting that sympatry follows (via ecological sorting) rather than causes (via character displacement[26]) trait divergence is in line with a recent study in minimally dispersive birds[20], our results suggest that ecological sorting of previously diverged taxa is a more general pattern across bird diversity, despite strong dispersal in many taxa.

Additionally, under the single pulse model preferred here, an early divergence pulse is followed by stasis. The magnitude of infrequent pulses may therefore have a central importance in downstream processes. For incipient species separated by low-magnitude pulses of ecological traits, little substantial additional divergence is likely to accrue via gradual evolution. Hence, whereas under gradual divergence models the eventual establishment of sympatry is expected to be facilitated by the continuous divergence of traits, stasis instead of gradualism may be the dominant evolutionary mode through time over the course of the three-stage allopatric speciation model. As a consequence, species pairs differentiated by low-magnitude pulses of ecological trait divergence may be more subject to prolonged periods of mutual exclusion. As mutual exclusion prohibits range expansion[3, 7], minimally ecologically differentiated incipient species may generally have smaller range sizes and higher probabilities of extinction[28]. These results then suggest that the extinction of minimally ecologically differentiated incipient (or recent) species may be an important dynamic in shaping diversity. At the other end of the divergence spectrum, instances of strong ecological trait divergence coupled with rapid sympatry are evident in nature[29]. While such cases are often treated in the literature as exceptional instances of atypical speciation, such events are an expected, if relatively low-frequency, outcome of an allopatric speciation model with divergence pulses.

We note that gradual divergence models can account for prolonged mutual exclusion between highly similar species when divergence is slow. However, gradual divergence cannot simultaneously account for highly divergent taxon pairs with rapid sympatry unless brief bursts of exceptionally fast gradual divergence (mimicking pulses) are invoked. A general model for bird speciation that instead incorporates a pulse-and-stasis mode for traits helps to rectify observations from nature that may otherwise appear to arise from fundamentally different processes. Under the pulse-and-stasis model of allopatric speciation we advocate here, instances of rapid sympatry coupled with strong trait divergence are expected, as are instances of prolonged mutual exclusion between highly similar species.

We suggest that our results point to an important question in speciation: what is the relevance and role of pulses of phenotypic evolution in speciation? We propose that our results are consistent with the 'Ephemeral Divergence' model[30] in which 'successful' speciation can translate the origin of forms below the species level into their long-term persistence. Here we interpret the establishment of

sympatry as a measure of speciation *success* [5]. Transitions to sympatry indicate that range expansion has been achieved by one or both members of a species pair. Such range expansions may decrease the probability of extinction and better the prospects of subsequent geographic speciation[28] for incipient species. In our species pair data set, parapatric ranges are indicative of mutual exclusion, with range expansion prevented. Species in these circumstances may be more subject to near-term extinction than those that have expanded ranges into sympatry. Our evidence then suggests that ecological trait divergence pulses of large magnitude may disproportionately contribute to the emergent patterns of diversity at longer time-scales.

METHODS SUMMARY

Using a sister species pair data set of 952 pairs sampled from across avian diversity, we examined 1) the timing and outcomes of secondary contact, and 2) the mode and timing of ecological trait divergence. We determined whether sister pairs had established secondary contact by scoring local co-occurrence from site-based inventories[31] and whether sister pairs had established sympatry by scoring range overlap from range maps[32]. Using multi-model inference of generalized linear models (GLM), we analyzed how variation in environment (latitude), life history (migratory versus sedentary behavior), ecological trait differences (divergence in body mass and beak morphology), and a proxy for dispersal capacity (hand-wing index) explain variation in the timing of secondary contact and sympatry. Sympatry was defined by range overlap of >10% of the smaller of the two species' ranges (>20% in sensitivity analyses), with overlap calculated using spatial data libraries in R. Latitude was calculated from sets of occurrence records[33]. Migratory behavior information was sourced from the natural history literature[34], and body mass data was compiled from the literature [35, 36]. Beak morphology and a morphological index for dispersal capacity (hand-wing index) were measured from museum specimens. To aid in the interpretation of an initial result indicating that the intercept of the relationship between the probability of local co-occurrence and divergence time is unexpectedly high, we used simulations of range dynamics that model transitions into and out of secondary contact from an initially allopatric configuration[37]. We analyzed the mode and timing of ecological trait divergence by fitting evolutionary models incorporating short-term and long-term processes, reflecting bounded evolution over short periods and gradual or pulsed divergence over the longer periods[12]. We estimated divergence times using a novel set of 20 fossil calibrations for a maximum likelihood phylogeny of 6,714 bird species. We show that our results are robust to different approaches to phylogenetic inference and dating by performing parallel analyses with an alternate time-calibrated phylogeny[38]. Analyses and data

visualizations were performed in R using a range of packages, including 'glmulti' for generalized linear model generation and selection; 'phytools' and 'diversitree' for phylogenetic comparative analyses; 'rgdal', 'rgeos', and 'raster' for GIS analyses; and 'maxLik' for likelihood searches of evolutionary divergence models.

Acknowledgements

We are grateful to numerous data collectors who contributed to eBird, GenBank, the CRC bird body mass data set (see Supplementary Information for further details). Catherine Sheard, Monte Neate-Clegg and Sam Jones measured morphometric traits from museum specimens, and Dan Ksepka, Rebecca Kimball, and Ed Braun helped with dating the phylogenetic tree. Aaron Ragsdale and Gleb Zhelezov assisted with likelihood calculations for the body mass divergence analyses; Alex Pigot helped with models of range dynamics. This work was supported by the National Science Foundation (DEB-1208428 to J.G.B.) and Natural Environment Research Council (NE/ I028068/1 to J.A.T.).

Author contributions J.G.B. and J.P.M. conceived and designed the study; J.G.B. performed dating analyses and assembled phylogenetic, occurrence, and body mass information; J.A.T. provided morphometric data; J.P.M. integrated data sets, and designed and performed statistical analyses with significant input from J.G.B.; J.P.M. produced figures and tables; J.P.M. wrote the manuscript, with significant input from J.A.T. and J.G.B.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to jaymcentee@ufl.edu.

References

1. Mayr, E. in *Systematics and the Origin of Species* (Columbia University Press, 1942).
2. Price, T. D. *et al*. Niche filling slows the diversification of Himalayan songbirds. *Nature* **509**, 222-+ (2014).
3. Weir, J. T. & Price, T. D. Limits to speciation inferred from times to secondary sympatry and ages of hybridizing species along a latitudinal gradient. *Am. Nat.* **177**, 462-469 (2011).
4. Pigot, A. L. & Tobias, J. A. Species interactions constrain geographic range expansion over evolutionary time. *Ecol. Lett.* **16**, 330-338 (2013).
5. Rundell, R. J. & Price, T. D. Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology & Evolution* **24**, 394-399 (2009).

6. MacArthur, R. & Levins, R. The limiting similarity, convergence, and divergence of coexisting species. *Am. Nat.*, 377-385 (1967).
7. Price, T. in *Speciation in birds* (Roberts and Company, Greenwood, Village, Colorado, 2008).
8. Gould, N. E. J. Punctuated equilibria: an alternative to phyletic gradualism. (1972).
9. Gould, S. J. & Eldredge, N. Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* **3**, 115-151 (1977).
10. Pennell, M. W., Harmon, L. J. & Uyeda, J. C. Is there room for punctuated equilibrium in macroevolution? *Trends in ecology & evolution* **29**, 23-32 (2014).
11. Estes, S. & Arnold, S. J. Resolving the paradox of stasis: Models with stabilizing selection explain evolutionary divergence on all timescales. *Am. Nat.* **169**, 227-244 (2007).
12. Uyeda, J. C., Hansen, T. F., Arnold, S. J. & Pienaar, J. The million-year wait for macroevolutionary bursts. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 15908-15913 (2011).
13. Eldredge, N. The allopatric model and phylogeny in Paleozoic invertebrates. *Evolution* **25**, 156-167 (1971).
14. Pagel, M., Venditti, C. & Meade, A. Large punctuational contribution of speciation to evolutionary divergence at the molecular level. *Science* **314**, 119-121 (2006).
15. Bokma, F. Time, species, and separating their effects on trait variance in clades. *Syst. Biol.* **59**, 602-607 (2010).
16. Rabosky, D. L. & Adams, D. C. Rates of morphological evolution are correlated with species richness in salamanders. *Evolution* **66**, 1807-1818 (2012).
17. Pereira, R. J. & Wake, D. B. Ring species as demonstrations of the continuum of species formation. *Mol. Ecol.* **24**, 5312-5314 (2015).
18. Rundle, H. D. & Nosil, P. Ecological speciation. *Ecol. Lett.* **8**, 336-352 (2005).
19. Lack, D. Ecological aspects of species-formation in passerine birds. *Ibis* **86**, 260-286 (1944).
20. Tobias, J. A. *et al*. Species coexistence and the dynamics of phenotypic evolution in adaptive radiation. *Nature* **506**, 359-+ (2014).
21. Mayr, E. in *Animal Species and Evolution* (Belknap Press of Harvard University Press, Cambridge, Massachusetts, 1963).
22. Burleigh, J. G., Kimball, R. T. & Braun, E. L. Building the avian tree of life using a large-scale, sparse supermatrix. *Mol. Phylogenet. Evol.* **84**, 53-63 (2015).
23. Hochkirch, A., Groening, J. & Buecker, A. Sympatry with the devil: reproductive interference could hamper species coexistence. *J. Anim. Ecol.* **76**, 633-642 (2007).
24. Groening, J. & Hochkirch, A. Reproductive interference between animal species. *Q. Rev. Biol.* **83**, 257-282 (2008).
25. Podos, J. Correlated evolution of morphology and vocal signal structure in Darwin's finches. *Nature (London)* **409**, 185-188 (2001).
26. Bothwell, E., Montgomerie, R., Lougheed, S. C. & Martin, P. R. Closely related species of birds differ more in body size when their ranges overlap-in warm, but not cool, climates. *Evolution* **69**, 1701-12 (2015).
27. Revell, L. J. Size-correction and principal components for interspecific comparative studies. *Evolution* **63**, 3258-3268 (2009).
28. Rosenzweig, M. L. in *Species diversity in space and time* (Cambridge University Press, 1995).
29. Schluter, D. in *The ecology of adaptive radiation* (Oxford University Press, Oxford, UK, 2000).
30. Futuyma, D. J. Evolutionary constraint and ecological consequences. *Evolution* **64**, 1865-1884 (2010).
31. Sullivan, B. L. *et al*. The eBird enterprise: An integrated approach to development and application of citizen science. *Biol. Conserv.* **169**, 31-40 (2014).

32. BirdLife International and NatureServe. in *Bird species distribution maps of the world* (BirdLife International and Natureserve, Cambridge, UK and Arlington, USA, 2014).
33. Sullivan, B. L. *et al*. eBird: A citizen-based bird observation network in the biological sciences. *Biol. Conserv.* **142**, 2282-2292 (2009).
34. *Handbook of the Birds of the World Alive* (eds del Hoyo, J., Elliott, A., Sargatal, J., Christie, D. A. & de Juana, E.) (Lynx Edicions, Barcelona, 2015).
35. Dunning, J. B. in *Body masses of birds of the world* (Taylor and Francis Group, Boca Raton, Florida, 2008).
36. https://ag.purdue.edu/fnr/Documents/WeightBookUpdate.pdf.
37. Pigot, A. L. & Tobias, J. A. Dispersal and the transition to sympatry in vertebrates. *Proceedings of the Royal Society of London B: Biological Sciences* **282**, 20141929 (2015).
38. 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).

METHODS

Sister pairs We initially selected sister species pairs (2,076 initial pairs) from the maximum likelihood topology of Burleigh *et al*.'s[22] avian supermatrix phylogenetic tree (hereafter "Burleigh tree"), which contains 6,714 species of the ~10,500 bird species in the world (http://www.birds.cornell.edu/clementschecklist/). This selection excluded species that are sister to clades of >1 species, which generated a manageable sample for downstream data quality checks while still allowing for adequate sampling for statistical analysis. To exclude pairs that may not represent "true" sister species, we reduced this data set by retaining only those sister pairs belonging to genera with ≥75% species-level sampling in the Burleigh tree (excluding 763 pairs), then examined evidence for sister relationships for species belonging to different genera and removed species pairs unlikely to be true sister species. The final lists of included (SI Dataset 1) and excluded species pairs (SI Dataset 7) are presented in the Supplementary Information, with exclusion criteria specified for all excluded pairs.

Divergence times The divergence time estimates from the Burleigh tree were obtained from a penalized likelihood analysis implemented in r8s[39], using the maximum likelihood topology and molecular branch lengths from the Burleigh et al. supermatrix tree[22]. We used twenty carefully vetted fossil calibrations across avian diversity, and we further constrained the root of the tree to a maximum age of 110 mya (although the age constraint of the root made little difference to the estimated sister pair divergence times, Extended Data Fig. 4). A list of the fossil calibrations (Supplementary Information Dataset 6) and a command block for the r8s analyses are available in the Supplementary Information. We performed sensitivity analyses using alternate sets of divergence time estimates, both from bootstrap analysis of the Burleigh et al. tree[22] and from an independent phylogenetic and dating analysis[38] (see Supplementary Information and e.g. Extended Data Figs. 5, 6, and 10).

Ecological trait measurements *Body mass* Divergence in body size may be a strong contributor to ecological divergence between species, which could reduce competition between species[29, 40]. In this study, we use body mass as a proxy for body size. Body mass data were compiled from a published global dataset[35] and a recent update[36]. If multiple body mass values were reported, we took the mean, and if male and female body masses were reported separately, we used the mean of the male and female mean body mass values. Our body mass disparity scores were calculated as the difference between species in natural log of mean body mass [12].

Beak morphology To quantify similarity in foraging ecology among sister species, we focused on three beak measurements (culmen length, beak depth, and beak width) associated with foraging niche and prey item selection [4, 41, 42]. Culmen length was measured as the distance from the distal part of the nostril to the beak tip. Beak depth and beak width were both measured at the distal part of the nostril. All beak measurements were made to the nearest 0.1mm. See Supplementary Information for further details and rationale. To account for colinearity between beak measurements, we performed a phylogenetic Principal Components Analysis (phylogenetic PCA[27]; see Supplementary Table 1 for PC loadings) before analyzing the role of beak morphology divergence in transitions to secondary contact and sympatry.

Secondary contact For a given species pair, local co-occurrence was defined as the occurrence of both species on the same day at the same reported coordinates. All sister species pairs for which local co-occurrence has been documented were considered to have established secondary contact. We estimated co-occurrence using ~178 million bird species observation records stored in the eBird observational record database[33, 43]. Because co-occurrence is unlikely to be reported for species with very few observations, we excluded sister pairs where at least one species had <10 eBird sightings reported. This threshold is low, as even after this filtering, the minimum number of observations strongly predicts the probability of species pair local co-occurrence in our data set (GLM with the log of the minimum observations as sole predictor: coefficient estimate = $3.8 \times 10^{-4} \pm 8.5 \times 10^{-5}$ SE, see Supplementary Information). Consequently, we conducted sensitivity analyses adopting minima of 20 and 50 observations (see Extended Data Fig. 7; Supplementary Tables 11-14). We also checked observational evidence for co-occurrence, discounting cases likely attributable to anthropogenic introductions, and excluding cases potentially based on misidentifications or taxonomic confusion (see included and excluded species pairs in Datasets 1 and 7).

Dispersal We hypothesized that secondary contact may occur more quickly for highly dispersive taxa because greater dispersal capacities should result in faster range expansions in nascent species[44]. Some

authors have suggested that this rapid resumption of contact in highly dispersive taxa may serve to slow or reverse the speciation process by permitting gene flow, instead of accelerating the eventual transition to sympatry[21]. However, a recent study indicated that increased dispersal capacity is associated with faster transition rates from non-sympatry (either allopatry or parapatry) to sympatry in vertebrates[37]. Our study attempts to separate the transitions from allopatry to secondary contact, and secondary contact to sympatry, to examine whether there are separate effects of dispersal capacity for either transition.

As it is difficult to measure dispersal capacity directly, here we use a proxy for flight performance that is in turn correlated with natal dispersal distance[45]. Wings with high aspect ratio are associated with efficient long-distance flight. Thus, to quantify dispersal capacity we used an index of wing shape related to the aspect ratio of the wing[44]:

$$Hand-wing\ index = 100 \times \frac{WL - SL}{WL}$$

where WL (wing length) is the standard length of the closed wing, and SL (secondary length) is the distance from the carpal joint to the tip of the first secondary feather (both measurements were made to the nearest mm).

Geographic configurations We calculated percent breeding range overlap from range polygons[32] with a custom R script, using the R libraries rgdal, rgeos, maptools, and raster. Species pairs with a range overlap >10% of the smaller range[37] were scored as sympatric. Sensitivity analyses were performed where overlap of >20% was scored as sympatric (Supplementary Tables 16-17).

Analyses

Secondary contact and sympatry We examine the probability of local co-occurrence, breeding range local co-occurrence, and sympatry versus parapatry using GLM with binomial error distributions, implemented in R[46]. We distinguished evidence for breeding range local co-occurrence from other local co-occurrence observations by checking the dates and localities of co-occurrence records against breeding phenology[34] and breeding range maps[32]. Our criteria for evidence of breeding range local co-occurrence were that co-occurrences had to take place during known breeding seasons of both species[34], within the known breeding range of one of the two species[32].

In analyses of local co-occurrence and breeding local co-occurrence for sister pairs, we began by predicting the probability of co-occurrence with divergence time as the only predictor (Fig. 1a-b and Extended Data Fig. 2). We subsequently performed a model generation and selection routine (the genetic algorithm of R package glmulti[47], see Supplementary Information) to examine which among a set of phenotypic measures and an environmental variable (latitude) best predict local co-occurrence or sympatry while accounting for the effects of divergence time. Our model generation routine permitted

all pairwise interactions between predictors to enter the model, under the constraint that all models were marginal. We included as predictors the between-species disparity in two ecologically important traits that have also been associated with mate choice in birds: body mass[48] and beak morphology[49]. We incorporated disparity in beak morphology between sister species in one of two ways: either as the Euclidean distance between species in PC space (following scaling of all PC's to unit variance) or by including species differences along each PC axis as separate predictors. To account for differences among sister species pairs in dispersal ability, we include the average hand-wing index[44] of the sister pair in GLM. We further included divergence time, latitude (average of the two median observational latitudes for each species from eBird), and migratory status (if either member of a species pair is migratory to any degree according to species range maps, the pair is scored as migratory; otherwise they are scored as non-migratory). We report support for all predictors entering the set of local co-occurrence models with $\Delta AIC < 2$ (Supplementary Tables 2-3). All continuous variables were scaled and centered, such that estimated slope magnitudes for individual variables are meaningful in relation to one another.

For GLM examining the probability of sympatry versus parapatry, we first limited the sister species data set to those pairs that locally co-occur in breeding ranges. This restriction removes clades for which sister interactions are not the primary determinant of breeding sympatry, and focuses the analysis instead on those taxa that interact to some degree[50]. We then categorized the ranges of locally co-occurring sister species as parapatric or sympatric by calculating overlap from species range maps[32], using a cutoff of 10% range overlap of the smaller range[37] (as any range overlap cutoff to designate sympatry is arbitrary, we performed sensitivity analyses using a cutoff of 20%, see Supplementary Information). The response variable in GLM is the geographic configuration: parapatric (interacting but without substantial range overlap) versus sympatric. We again used a genetic algorithm (see Supplementary Information) to generate model variants and performed model selection using the R package glmulti[47].

To check the sensitivity of our results to the divergence time estimates we used, we repeated all GLM analyses using mean divergence times for our species pairs from 100 samples of the pseudo-posterior distribution of phylogenies of Bayesian species-level phylogenetic analysis for birds[38] (SI Tables S6-S9). For analyses examining the probability of local co-occurrence (and breeding local co-occurrence) with divergence time, we additionally performed exhaustive sensitivity analyses employing divergence time estimates from 100 bootstraps of the Burleigh tree and for each of the 10,000 psuedo-posterior samples from the Jetz et al. analyses.

Tempo of body mass divergence

In our GLM analyses, we find that increased body mass disparity and beak morphology (phylogenetic PC1) are associated with increased probability of sympatry among pairs that have come into contact (Fig. 1, Supplementary Tables 4-5, 8-9, 15-16). This result occurs while controlling for divergence time, suggesting that increased body mass and beak morphology disparity accelerate the establishment of sympatry. But the result raises the issue of how body mass disparity accrues in sister pairs of birds. If body mass accrual is gradual, then there should be little potential for body mass divergence to impact transitions from secondary contact to sympatry shortly following divergence. However, if body mass disparity more typically accrues in rapid pulses, such pulses could affect transitions to sympatry shortly following divergence if they frequently occur before secondary contact. To examine the tempo and timing of body mass and beak PC1 divergence then, we investigated the relative support for four models of divergence for the species pairs from the full dataset for which body mass or beak morphology data were available (n = 872 species pairs for body mass, n = 926 species pairs for beak morphology). We fit models of time-independent bounded evolution and three models that include a bounded evolution component on shorter timescales and an additional component for longer timescales[12]. These longer-timescale components are a gradual evolution model (Brownian motion) and two forms of pulsed divergence: a single pulse model where a single instantaneous displacement occurs following a waiting time sampled from an exponential distribution, and a multiple pulse model where the expected number of displacements for a given divergence time is determined by a Poisson process. We examine relative support for these models using AIC from likelihood calculations performed in R.

Simulations of range dynamics To aid in the interpretation of the intercept of local co-occurrence probability that we calculated from empirical data, we present stochastic simulations of the process of secondary contact establishment. We used these simulations to place an approximate lower bound on the rate of secondary contact establishment. To perform this estimation, we simulated the establishment of secondary contact using a simple model[4, 37], in which sister pairs can be in one of two states: co-occurring and not co-occurring. We simulated transitions into and out of contact over a set of possible rates from 0.1 to 0.8 per million years, in which the forward rate (rate of transition from isolation to contact, σ) is always greater than or equal to the reverse rate (rate of transition out of contact, ε). The forward and reverse rates are constant (the Constant-Rate model of[4], but with reversion to isolation included), and the variation in rates among species arises only from stochasticity. Reverse rates were simulated at .005, .01, .05, .1, .2, and .5 times each of the forward rates. We present the maximum intercept calculated across all reverse rates (ε) for each simulated forward rate (σ) (Extended Data

Figure S4). To calculate the approximate percentage of species pairs coming into secondary contact by given points in time following divergence (100,000 years, 1 million years), we simulated range dynamics with σ=0.3, and ε=.15 (corresponding to the minimal σ that yielded intercept > 0.434, and the value of ε that yielded the highest intercept for σ=0.3).

Figure 1 | Factors associated with the establishment of secondary contact and coexistence in birds The relative importance of predictors in generalized linear models of the probability of **a, c**, breeding season secondary contact and **b, d,** sympatry for two predictor sets. Relative importance is calculated as the proportion of the summed model weights for all models in the candidate set (those with ΔAIC<2) including the predictor. The relative importance can be interpreted as the overall support for each predictor as a key factor in predicting the probability of secondary contact or sympatry across the candidate set[47]. Predictors with green bars are positively associated with the probability of co-occurrence or sympatry in the superior model of the candidate set when controlling for other variables (i.e. predictions are made with other variables set to their means, except divergence time which is set to its median and migration which is set to non-migratory; see Supplementary Tables 2–5 and 6-9 for multi-model slope estimates for each predictor). Migratory behavior is associated with increased probability of breeding co-occurrence, controlling for divergence time. The identity of all pairwise interactions with relative importance >0.6 are indicated by the numbers to the right of each variable's importance score.

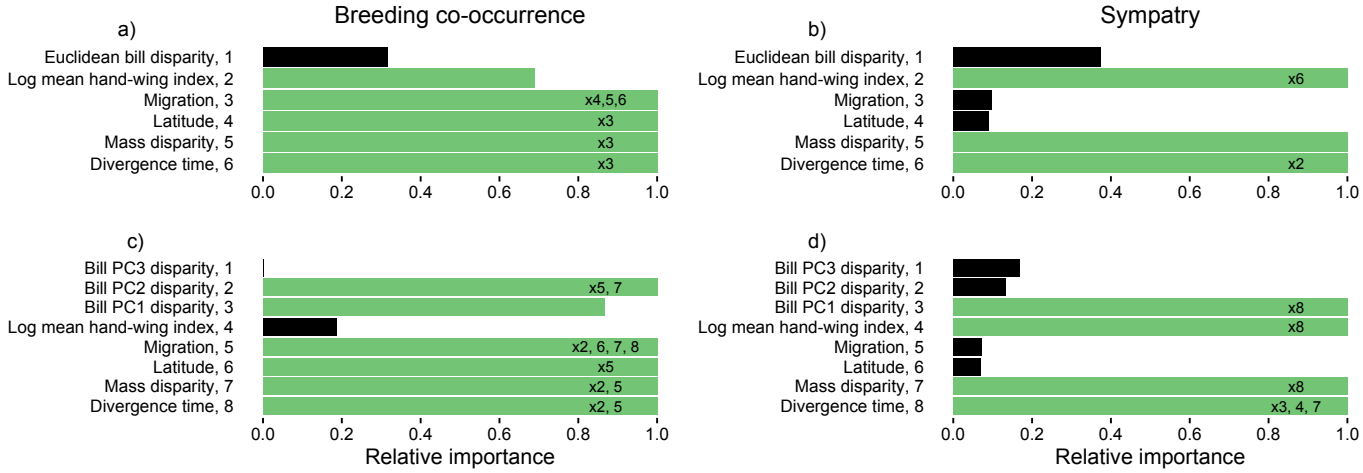


Figure 2 | Body mass divergence tempo for avian sister species Stochastic pulsed models of evolution (either a single pulse or multiple pulse model) fit species pair log body mass disparities and divergence times (n = 872 species pairs) better than a gradual evolution model (Brownian motion) and a

white noise model (disparity is independent of divergence time). For any given time slice along the x-axis, the probability density follows a normal distribution (most apparent in the white noise model where the probability density distribution is the same at all time slices). In the remaining models, variance increases with time. Probability densities are normalized within each time slice, such that relative probability density can be assessed within each time slice (and not across time slices). The single pulse model is the best fit by AIC (ΔAIC over the multiple pulse model: 797), with a fitted mean waiting time to the pulse of 665,000 years, suggesting that the majority of species pairs in the data set have incurred a pulse of body mass divergence (see Figure 3). Assuming this model accurately reflects evolutionary process, species pairs with larger-magnitude pulses of body mass divergence are associated with higher probabilities of sympatry given secondary contact. The empirical data points are plotted only on the white noise model, as it is difficult to examine the other fitted models when points are plotted.

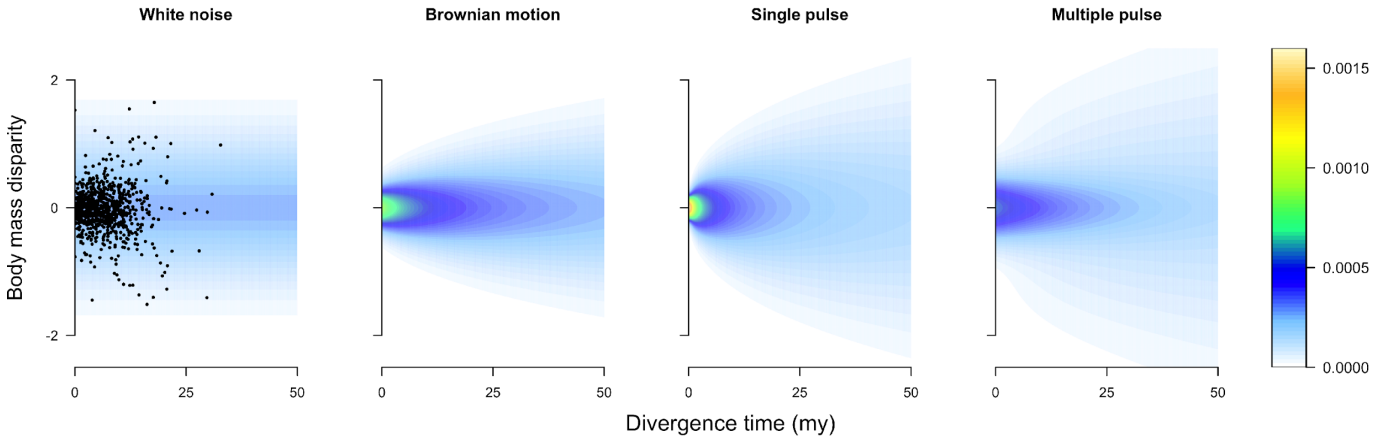
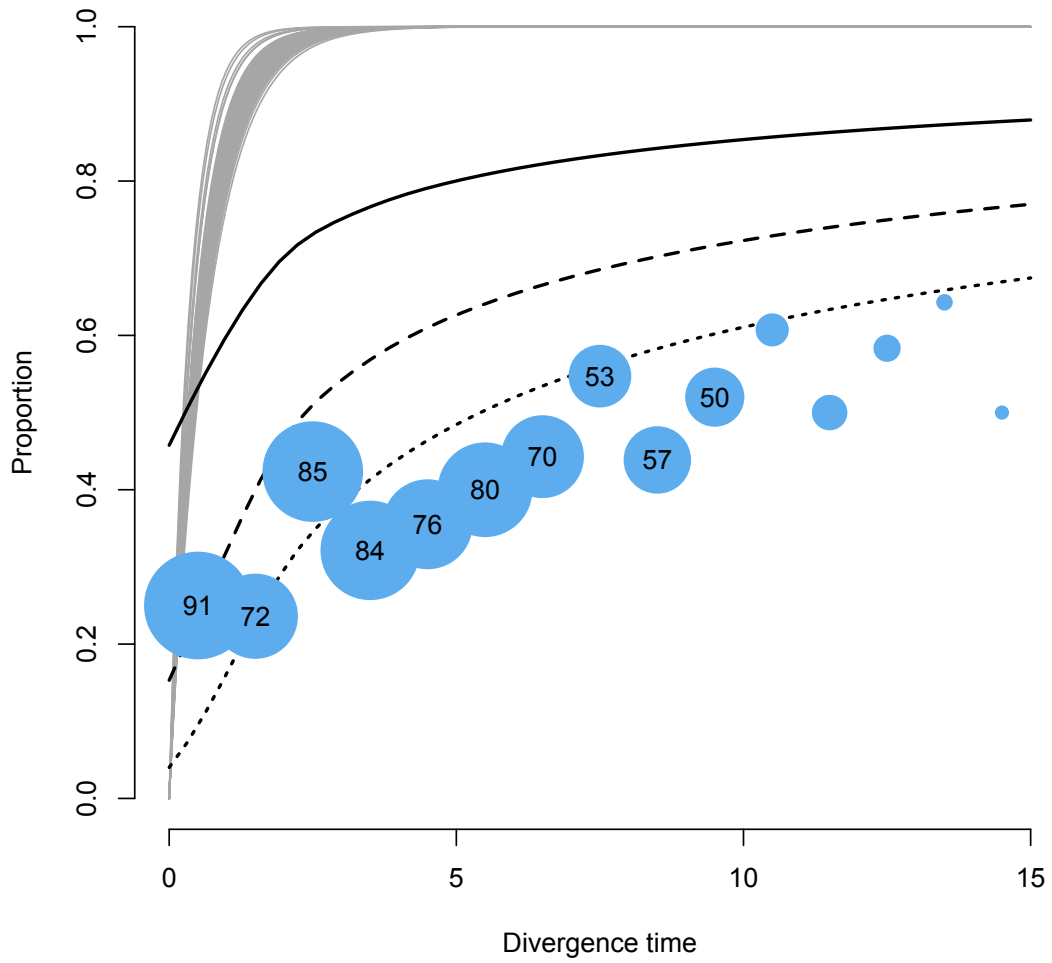


Figure 3 | Pulsed body mass divergence and sympatry The establishment of sympatry occurs relatively slowly compared to body mass divergence pulses in the preferred single pulse model. The blue circles are the proportion of sympatric sister pairs for 1-million-year divergence time intervals. Circle sizes represent sample sizes (*n*) within each 1-million-year interval, with *n* shown where ≥ 50. The gray lines are the cumulative probability distributions (CPD) for having incurred a pulse of body mass divergence, with each of the 100 lines representing the CPD from a fitted single pulse model using divergence times from a dated bootstrap phylogeny from a bootstrap analysis of the Burleigh phylogenetic tree[22]. The variation among the presented lines is indicative of the effect of phylogenetic uncertainty. The black curves indicate the predicted proportion of sister pairs with body mass divergence greater than 10% (solid), 20% (dashed), and 30% (dotted) based on the single pulse model

fitted with divergence times from the maximum likelihood Burleigh phylogenetic tree[22]. Note the rapid increase in these proportions at low divergence time and their subsequent leveling off.

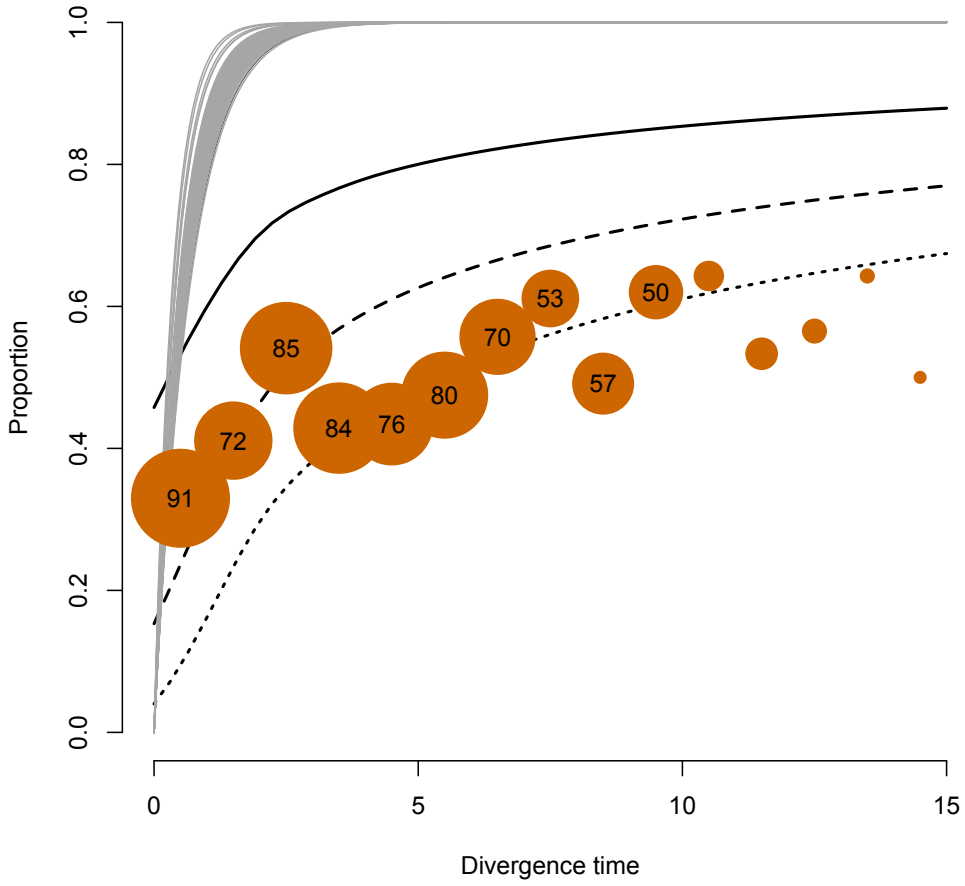


39. Sanderson, M. R8s: Inferring Absolute Rates of Molecular Evolution and Divergence Times in the Absence of a Molecular Clock. *Bioinformatics* **19**, 301-302 (2003).
40. Wilson, D. S. The Adequacy of Body Size as a Niche Difference. *Am. Nat.* **109**, 769-784 (1975).
41. Schoener, T. W. The Evolution of Bill Size Differences among Sympatric Congeneric Species of Birds. *Evolution* **19**, 189-213 (1965).
42. Miles, D. B. & Ricklefs, R. E. The Correlation between Ecology and Morphology in Deciduous Forest Passerine Birds. *Ecology* **65**, 1629-1640 (1984).
43. http://www.ebird.org.

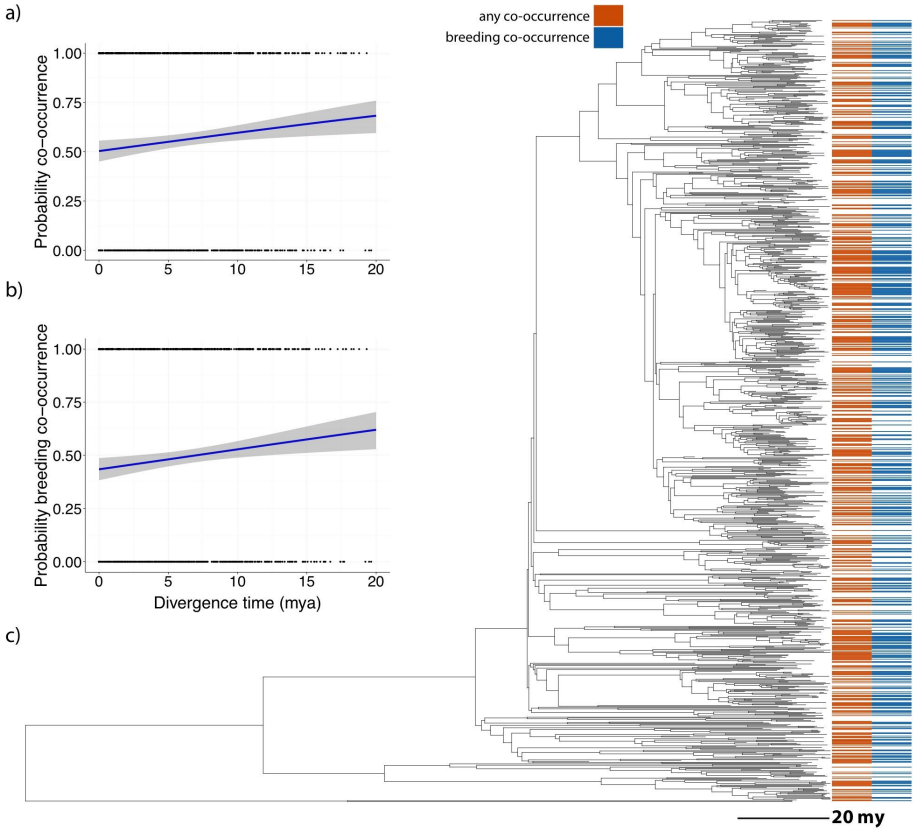
502 44. Claramunt, S., Derryberry, E. P., Remsen, J. V.,Jr. & Brumfield, R. T. High dispersal ability
503 inhibits speciation in a continental radiation of passerine birds. *Proceedings of the Royal Society B-*
504 *Biological Science***s 27**9, 1567-1574 (2012).
505 45. Dawideit, B. A., Phillimore, A. B., Laube, I., Leisler, B. & Böhning-Gaese, K. Ecomorphological
506 predictors of natal dispersal distances in birds. *J. Anim. Ecol***. 7**8, 388-395 (2009).
507 46. R Core Team. R: A language and environment for statistical computing. (2012).
508 47. http://CRAN.R-project.org/package=glmulti.
509 48. Yasukawa, K. Male quality and female choice of mate in the red-winged blackbird (Agelaius
510 phoeniceus). *Ecolog***y 6**2, 922-929 (1981).
511 49. Grant, P. R. & Grant, B. R. Hybridization, sexual imprinting, and mate choice. *Am. Nat***. 14**9, 1-28
512 (1997).
513 50. Lovette, I. J. & Hochachka, W. M. Simultaneous effects of phylogenetic niche conservatism and
514 competition on avian community structure. *Ecolog***y 8**7, S14-S28 (2006).
515

Extended Data Figures

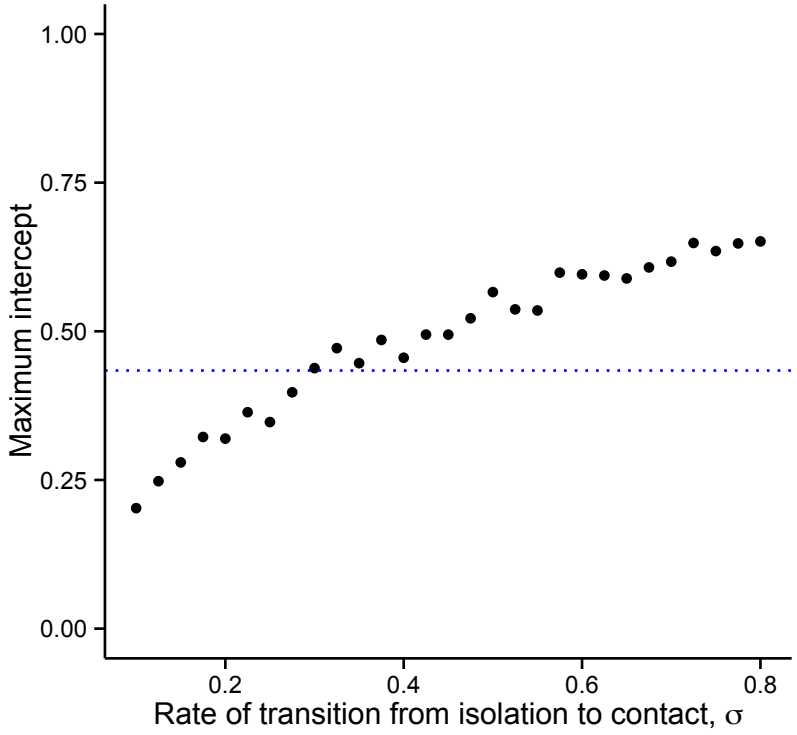
Extended Data Figure 1 | Pulsed body mass divergence and secondary contact The pulsed model of body mass divergence supported in our analyses suggests that most change in body mass occurs close to the outset of speciation. At least some divergence pulses are likely to pre-date secondary contact. The orange circles are the proportion of sister pairs that co-occur during the breeding season for 1-million-year divergence time intervals. Circle sizes represent sample sizes (n) within each 1-million-year interval, with n shown where ≥ 50. The gray lines are the cumulative probability distributions (CPD) for having incurred a pulse of body mass divergence, with each of the 100 lines representing the CPD from a fitted single pulse model using divergence times from a dated bootstrap phylogeny from a bootstrap analysis of the Burleigh phylogeny{{641 Burleigh 2015;}}. The variation among the presented lines is indicative of the role of phylogenetic uncertainty. The black curves indicate the predicted proportion of sister pairs with body mass divergence greater than 10% (solid), 20% (dashed), and 30% (dotted) based on the single pulse model fitted with divergence times from the maximum likelihood Burleigh phylogeny{{641 Burleigh 2015;}}. Note the rapid increase in these proportions at low divergence time, and their subsequent leveling off.



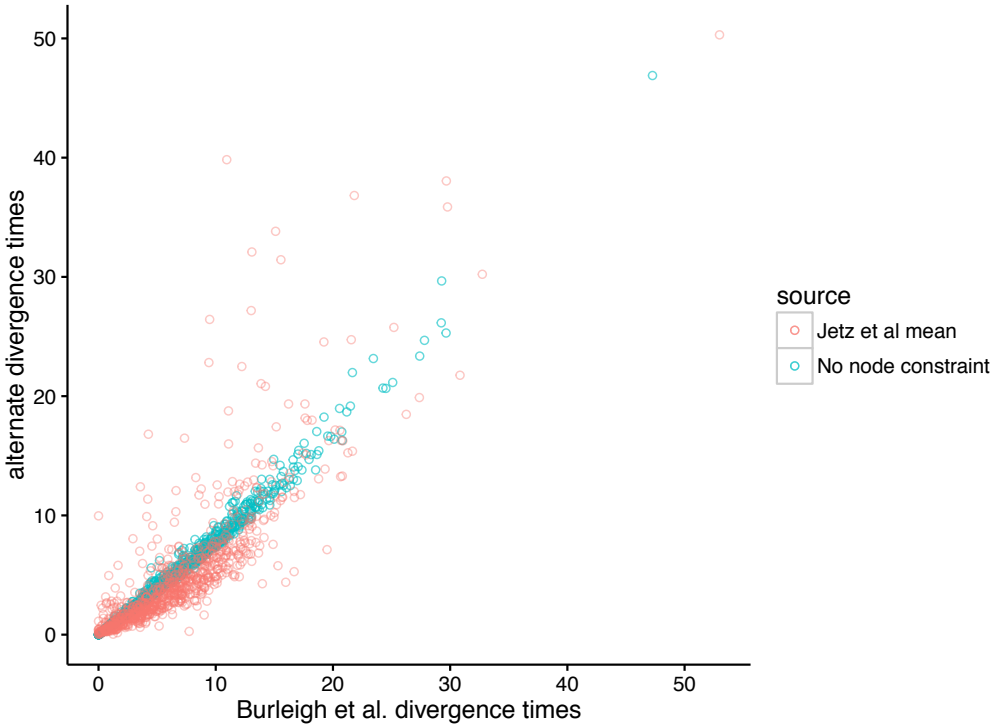
Extended Data Figure 2 | Patterns of co-occurrence across avian sister species. a, predicted probability of local co-occurrence and **b**, predicted probability of breeding range local co-occurrence for sister species of birds as a function of divergence time. Divergence time estimates come from a dating analysis of the Burleigh *et al.* supermatrix phylogeny{{641 Burleigh 2015;}}. Points at 0 and 1 on the y-axis are data points from species pairs with "no co-occurrence" or "local co-occurrence" respectively. The non-zero intercepts are not consistent with long-duration (>1 million year) phases of complete geographic isolation being necessary at the outset of speciation across birds{{657 Barraclough,T.G. 2000;}}. **c**, tree showing evolutionary relationships among 952 pairs of sister species included in local co-occurrence analyses, with co-occurrence treated as a binary variable for each pair. Sister pairs with breeding range local co-occurrence are a subset of those with any local co-occurrence. The tree is derived from the Burleigh et al. supermatrix phylogeny{{641 Burleigh 2015;}}, pruned to the sister pair dataset. Terminal branches are further pruned such that tips represent the most recent common ancestors of sister species pairs. There is little phylogenetic clustering of co-occurrence of either type (see Text).



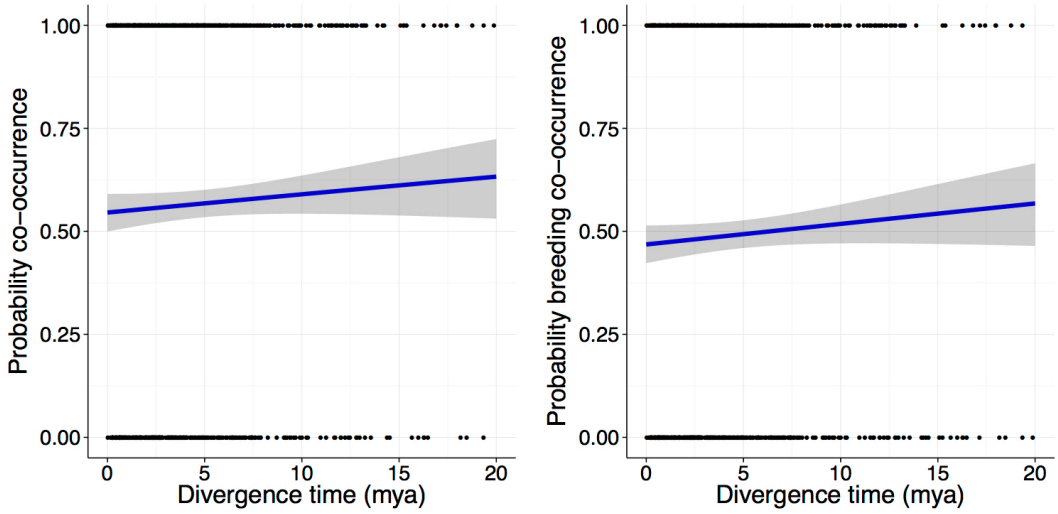
Extended Data Figure 3 | Estimating the minimum rate of transition to secondary contact from allopatry. Maximum intercepts estimated from generalized linear models of simulated species pair configurations (co-occurring or not co-occurring) from a simple Markov model of range dynamics. Simulations were accomplished via a Gillespie algorithm implementation of the Markov model, with the set of divergence times equal to the empirical divergence time estimates (see *Fossil calibration* section in Supplementary Information). Maximum intercepts are the maximum across all reverse transition rates (see Methods and Supplementary Information), with five replicates per reverse transition rate, ε, for each forward rate, σ.



Extended Data Figure 4 | Correlation of alternate divergence time estimates with those used in the main text Plots of two alternate divergence time estimates for bird species pairs (mean divergence times across 100 pseudo-posterior samples from the Jetz et al. analyses, and estimates from a r8s analysis of the Burleigh tree with no time constraint imposed on the basal node) against the estimated divergence times used for analyses presented in the main text (from r8s analysis of the Burleigh tree, implementing a maximum node age of 110 million years for the basal node). The alternate divergence times are highly correlated with the Burleigh tree divergence times used in the main text, for which the maximum basal node age was 110 million years (r = .823 for the Jetz et al. estimates, and r = .991 for the no-constraint estimates from the Burleigh tree). Note that though ancient nodes are older when there is no maximum constraint on the basal node of the Burleigh et al. phylogeny, divergence time estimates for sister pairs are slightly younger.

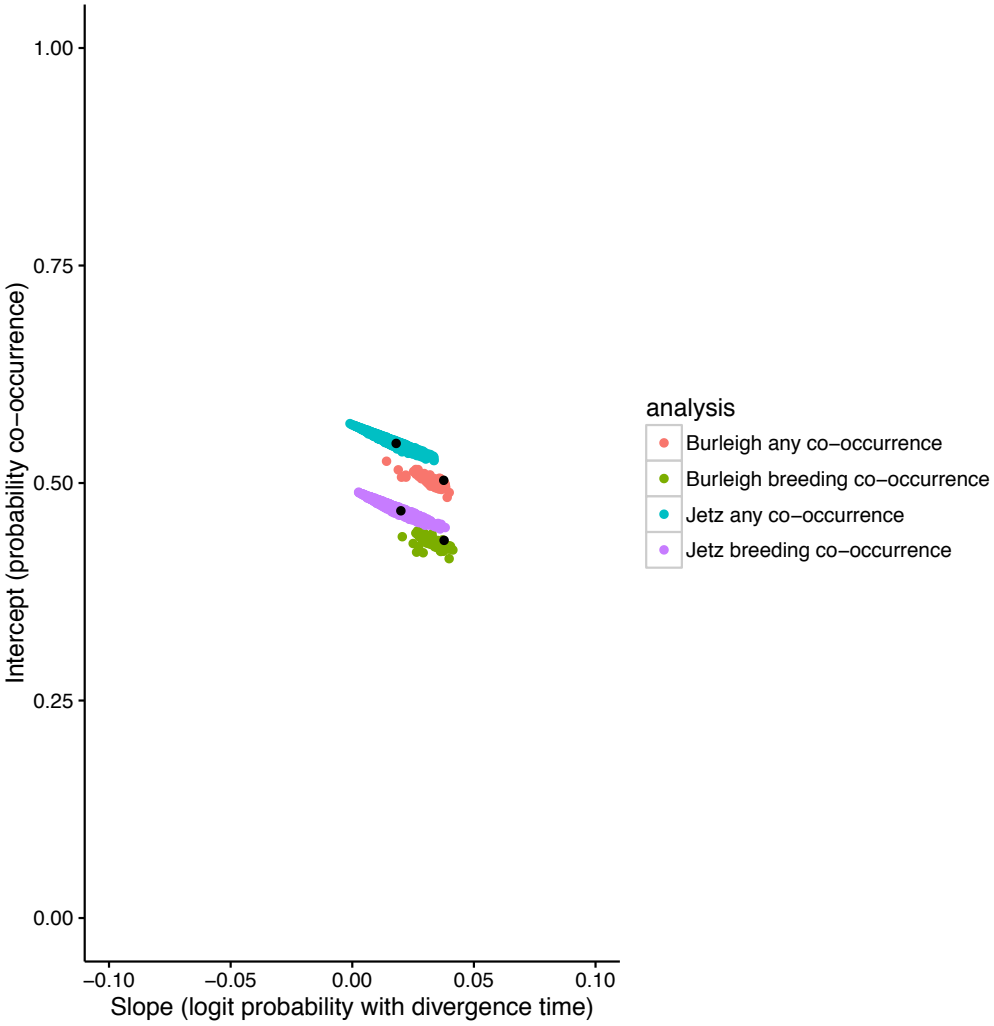


Extended Data Figure 5 | Sensitivity analyses of the relationship of local co-occurrence and divergence time, using alternate divergence time estimates. Predicted a) probability of local co-occurrence and b) probability of breeding range local co-occurrence for sister species of birds as a function of divergence time. Divergence times are the means across 100 phylogenies from the posterior distribution of the Jetz et al. (2012) analysis. Points at 0 and 1 are individual data points from species pairs with "no co-occurrence" or "local co-occurrence." These analyses indicate that the intercept is not near 0, counter the expectation stemming from extended allopatry as the dominant mode of bird speciation (Barraclough and Vogler 2000).

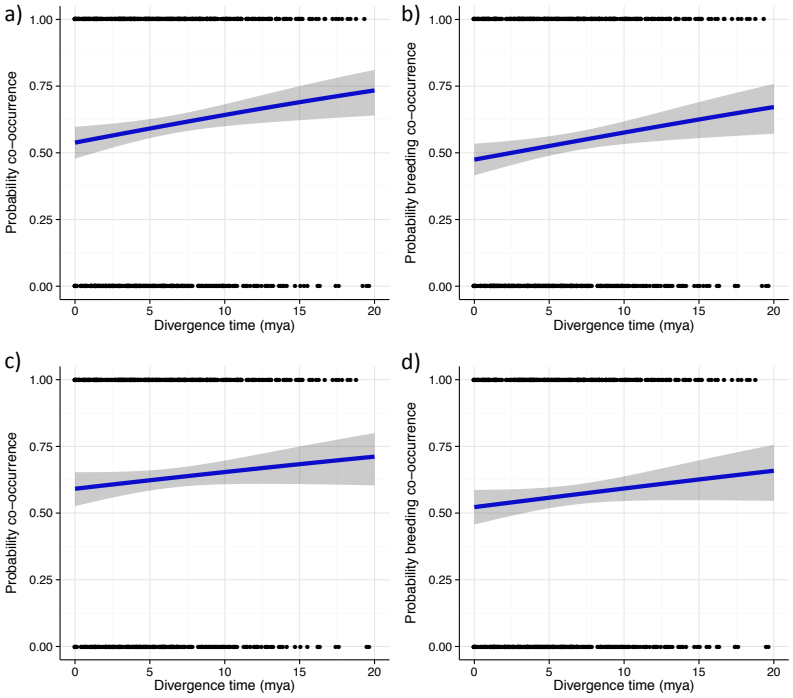


Extended Data Figure 6| Sensitivity analyses examining the probability of local co-occurrence with divergence time

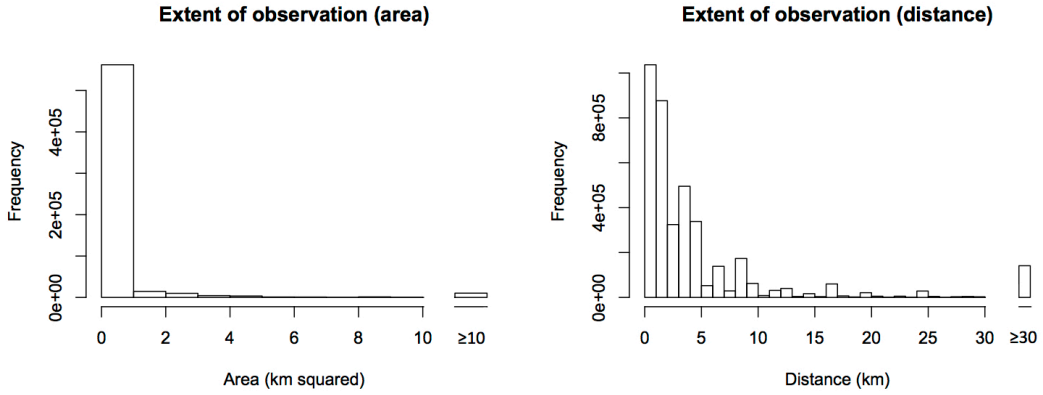
Sensitivity analyses for the relationship of the probability of co-occurrence with divergence time among bird sister pairs. Clouds of points represent the results of individual regressions of the probability of co-occurrence on divergence time, with divergence time estimates from 100 bootstrap samples of the Burleigh et al. (2015) phylogenetic analyses or from 10,000 samples from the pseudo-posterior of the Jetz et al. (2012) analyses. Black dots within each cloud represent either the results of analyses reported in the main text from the maximum likelihood phylogeny of Burleigh et al. (2015) or sensitivity analyses using the mean divergence time estimate for each sister pair from the Jetz et al. analysis pseudo-posterior. Note that there is little overall variation in the estimated slope and intercept for the probability of any co-occurrence or of breeding co-occurrence within and between phylogenetic analyses.



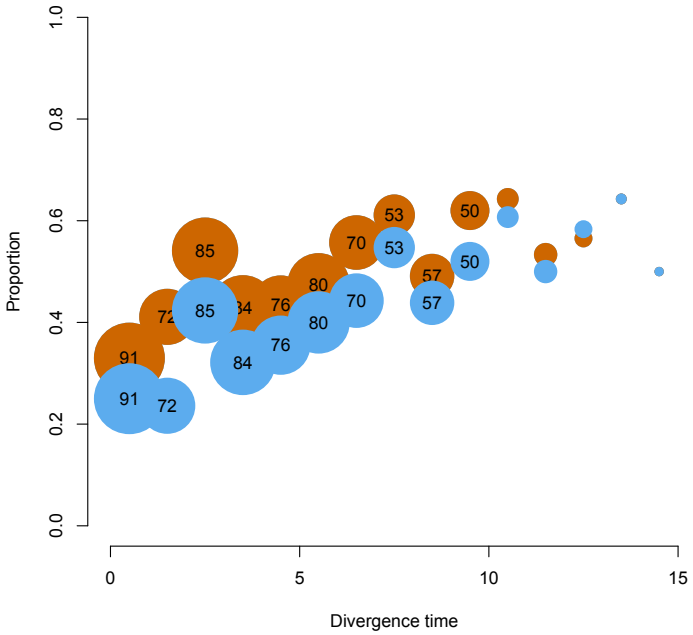
Extended Data Fig. 7 | The relationship between probability of local co-occurrence and divergence time for different thresholds of the minimum number of observations. Probability of any local co-occurrence with divergence time; minimum number of species observations equal to **a,** 20, and **c,** 50**.** Probability of breeding season local co-occurrence with divergence time; minimum number of species observations equal to **b,** 20, and **d,** 50. Note that the intercept of the relationship becomes higher with increasing observation number minima, and that the slopes of the relationship become slightly flatter (compare with Fig. 1a, b).



Extended Data Figure 8 | Distributions of areas and distances covered by site-based inventories. Histograms of observation areas (n = 608,855) and distances (n = 3,918,920) reported for eBird checklists, which correspond to observations given for a single reported locality.



Extended Data Figure 9 | Comparison of secondary contact and sympatry establishment across divergence times in sister pairs. Proportion of breeding season co-occurring (orange) and breeding season sympatric (blue) sister species in 1-million year intervals. Circle sizes represent sample sizes (n) within each 1-million-year interval, with n shown where ≥ 50.



Extended Data Figure 10 | Sensitivity analyses for the estimation of the minimum rate of transition to secondary contact from allopatry. Maximum intercepts estimated from generalized linear models of simulated species pair configurations (co-occurring or not co-occurring) from a simple Markov model of range dynamics. Simulations were accomplished via a Gillespie algorithm implementation of the Markov model, with the set of divergence times equal to the empirical divergence time estimates (sensitivity analysis with Jetz *et al.* (2012) divergence times). Maximum intercepts are the maximum across all reverse transition rates (see Methods), with five replicates per reverse transition rate, ε, for each forward rate, σ. We infer higher rates of transition from allopatry to secondary contact, and thus shorter average waiting times (see Supplementary Information).