

1 **Supplementary Information**

2

3 **1. Supplementary Methods**

4 **A. Definitions**

5 We define ‘allopatry’ as complete geographic isolation, without any contact between
6 individuals of (nascent) species, ‘parapatry’ as contact between (nascent) species without
7 substantial geographic overlap, and ‘sympatry’ as substantial overlap in range between species.

8 We define ‘local co-occurrence’ as the occurrence of two species at the same site on the
9 same day. Species that are allopatric have no local co-occurrence by definition. Locally co-
10 occurring species have the opportunity to compete with each other, to learn each others’
11 phenotypes, to breed or attempt to breed with each other, and to pass parasites to each other.
12 These interactions can impact evolutionary processes, even when they occur at low frequencies.

13 ‘Allopatric speciation’ is speciation where divergence proceeds in geographic isolation.
14 ‘Parapatric speciation’ occurs when interactions between diverging populations are continuously
15 (or near-continuously) possible over the course of speciation, but where the ranges of diverging
16 populations do not substantially overlap.

17 We use a functional definition of ‘sympatry’ for analyses, where we define sympatry of
18 species ranges by the fraction of the smaller species range that is part of the larger species range
19 (10% or 20%). More broadly, ‘sympatry’ is substantial overlapping of species ranges. Here, we
20 treat ‘co-existence’ and sympatry equivalently, although other authors {{730 Chesson 2000;}}
21 prefer more stringent definitions of co-existence.

22 **B. Background to methods**

23 *Sampling*

24 *Sister pair co-occurrence and sympatry analyses* To generate a sample of species pairs, we
25 extracted all tips that were sister to a single other tip from the species-level maximum likelihood
26 phylogeny of Burleigh et al {{641 Burleigh 2015;}}. We do not expect that exclusion of species
27 that are sister to a clade of >1 species biased results in any consistent way. Further, although the
28 hypotheses examined in our analyses do not strictly require that only “true” sister species are
29 examined, for terminological consistency we sought to limit the inclusion of sister pairs from the
30 phylogeny that were unlikely to represent each others’ closest living relatives. We therefore
31 retained the sister species pairs from genera with $\geq 75\%$ of the species in the tree. We then

32 checked cases where sister taxa belong to different genera to determine whether their inclusion
33 as sister species was reasonable (see SI Dataset 7).

34

35 *Observational data quality and criteria for inclusion in analyses* We scored co-occurrence states
36 for each species pair by checking the eBird database for instances where both species were
37 reported from the same locality on the same day. We then checked all sister species co-
38 occurrence states against species distribution maps{{689 NatureServe 2014; 639 Anonymous
39 2015;}}, and examined eBird co-occurrence records in detail where there was little or no overlap
40 in range maps. When co-occurrences were attributable exclusively to anthropogenic
41 introductions, species pairs were re-scored as “not co-occurring”. For example, there is only one
42 co-occurrence record of *Tangara seledon* and *T. fastuosa*, which may be attributable to human-
43 aided movements. Range maps for these species indicate allopatric isolation, and thus, this pair
44 was re-scored as “not co-occurring”. Where there was ambiguity regarding the role of
45 anthropogenic analyses in co-occurrence, the species pair was removed from co-occurrence
46 analyses (see SI Dataset 7). Where taxonomic confusion or misidentifications were plausible
47 explanations for discrepancies between observation records and range maps, we removed the
48 species pair from analyses (see SI Dataset 7).

49 To determine which sister pairs locally co-occur within breeding ranges, we began by
50 assuming that locally co-occurring sister pairs co-occur in breeding ranges if neither species is
51 migratory nor makes documented non-breeding season movements as indicated by range maps
52 {{689 NatureServe 2014}}. We then examined the local co-occurrence records of sister pairs for
53 which either species is migratory or makes non-breeding season movements. To qualify as
54 locally co-occurring in the breeding range, species pairs had to co-occur during the breeding
55 season{{639 Anonymous 2015;}} in the breeding range{{689 NatureServe 2014;}}. We
56 checked co-occurrence records for observations that occurred during the breeding season, and as
57 much as possible, we scored pairs conservatively (erring towards non-co-occurrence). For
58 example, pairs for which the only local co-occurrence observations came from the beginning of
59 the breeding season for both species (e.g. early May for north temperate migratory songbird
60 breeders with May-July breeding seasons) were scored as not co-occurring in the breeding range,
61 as these co-occurrences plausibly occur while both species are still migrating. Among locally co-
62 occurring species pairs for which either species is migratory or moves during the non-breeding

63 season ($n=215$ pairs), 31.6% (68 pairs) did not co-occur in the breeding range during the
64 breeding season (see also Fig. 1c).

65 As we include sister pairs in the main analyses where the number of observations n of
66 one or both species can be low (e.g. $10 < n < 100$), we likely underestimated the frequency of co-
67 occurrence across sister species. Given the prior expectation that co-occurrence of sister species
68 should be limited to pairs with longer divergence times, this bias should result in conservative
69 scoring of co-occurrence. In other words, there are sister pairs in the main analysis that likely
70 locally co-occur by our definition, but for which there are no local co-occurrence observations in
71 the eBird data set. Examples of a “no co-occurrence” score in our data set despite apparent
72 overlapping distributions include *Guarouba guarouba* - *Diopsittaca nobilis* and *Montifringilla*
73 *adamsi* - *Montifringilla nivalis*. See additionally the Results subsection *Sensitivity analyses:*
74 *minimum observation threshold*, SI Tables S11-S14, and SI Table S17, which detail sensitivity
75 analyses with alternate minimum observation thresholds (20 or 50 observations as minimum
76 thresholds instead of 10).

77

78 *Our approach to transitions to secondary contact compared with previous efforts* Our definition
79 of local co-occurrence is appealing in several respects. First, whereas the quantification and
80 categorization of geographic overlap are somewhat arbitrary and complex, our definition of local
81 co-occurrence is simple and straightforward. Secondly, even limited co-occurrence may be
82 important during speciation. Several recent studies have lumped parapatry together with
83 allopatry in analyzing species ranges during or following speciation {{645 Pigot 2013; 751 Pigot
84 2015; 648 Weir 2011;}}, but in parapatry, species have the opportunity to interact, and even
85 mate, with each other. Thus, even very limited co-occurrence can be biologically important, and
86 our approach allows us to examine the significance of limited co-occurrence. Third, there is little
87 bias in the choice of taxa in our data set, which may allow our study to be more generalizable
88 over birds as a whole. For example, we do not exclude marine or island species. Lastly, we also
89 use divergence time estimates from two different time-calibrated trees for all birds. Using
90 multiple time estimates lends robustness to our conclusions, and the use of time-calibrated trees
91 including all analyzed species pairs yields divergence time estimates estimated under a single
92 time-calibrated framework across species pairs {{751 Pigot 2015;}}.

93

94 *Sampling for sympatry analyses* To examine the transition from secondary contact to sympatry,
95 we limited the data set to sister pairs with some evidence of local co-occurrence. We reasoned as
96 follows. Some sister pairs occur in complete isolation and do not have the opportunity to interact.
97 These pairs may have diverged in ways that would allow sympatry, or may have other aspects of
98 their biology that would allow sympatry, but we cannot know if they would achieve sympatry on
99 secondary contact because we do not have evidence that they interact. In analyses attempting to
100 determine which factors could speed up or slow the transition to sympatry following divergence,
101 these pairs do not provide evidence in either direction. The noise associated with these pairs
102 could prevent discovery of biologically meaningful signal, and is thus best removed from
103 analyses seeking to understand the transition from secondary contact to sympatry. In the main
104 analyses, removing species pairs that do not co-occur resulted in species pair data sets with $n =$
105 440 locally co-occurring pairs.

106

107 *Sampling for body mass disparity analyses* We examined the distribution of body mass disparity
108 values (see *Body mass* section below) to remove outliers that may reflect biases or errors. We
109 subsequently removed three sister pairs from body mass disparity analyses, as their disparity
110 values either appeared distorted by reporting of only one sex's mass in extremely dimorphic
111 species (*Psarocolius atrovirens* – *P. angustifrons*) or because data for one of the two species
112 appeared incorrect by a large magnitude when compared to a second body mass data
113 source{{639 Anonymous 2015;}} (*Cyanolyca argentigula* – *C. pumilo* and *Milvus milvus* – *M.*
114 *migrans*).

115

116 **C. Taxonomy and taxonomic reconciliation**

117 The two phylogenetic frameworks that we used for analyses are based on different
118 taxonomies, and these taxonomies conflicted with the other data sources in some cases. The
119 phylogeny used in main analyses{{641 Burleigh,J.Gordon 2015;}} adopts the taxonomy of
120 Clements (v6.6{{434 Clements 2011;}}). The alternate phylogeny used in sensitivity analyses
121 {{646 Jetz,W. 2012;}} adopts the BirdLife International taxonomy (v3,
122 <http://www.birdlife.org/datazone/info/taxonomy>), with a small number of modifications (see the
123 Supplementary Information of the Jetz et al. study{{646 Jetz 2012;}}). We reconciled these
124 taxonomies by comparing the species lists, and treating different names equivalently where the

125 designated taxa appeared equivalent (e.g. where taxonomies differed in genus name but the
126 species epithet was the same, or where there were differences in spelling). We checked
127 taxonomic histories in AviBase (<http://avibase.bsc-eoc.org/avibase.jsp>) to make these
128 comparisons. In some cases we had to remove species or species pairs from analyses. This is
129 because taxonomies differ in classifying some taxa as subspecies or species, and in these cases it
130 was not always possible to keep these taxa in analyses (e.g. where pairs of taxa comprise a
131 species pair in the “Burleigh” phylogeny, but are subspecies of a single species in the “Jetz”
132 phylogeny, represented by a single tip in the phylogeny).

133 Similar to the Burleigh *et al.* phylogeny{{641 Burleigh,J.Gordon 2015;}}, eBird and the
134 body mass data set follow the Clements taxonomy{{434 Clements 2011;}}. Similarly to the Jetz
135 *et al.* {{646 Jetz 2012;}} phylogeny, the range maps we used to estimate sympatry follow the
136 BirdLife International taxonomy (<http://www.birdlife.org/datazone/info/taxonomy>). Taxonomic
137 confusion may have resulted in a small number of mismatches between data sets and
138 phylogenetic taxonomy, though the direction of any resulting biases in our analyses is unclear.

139

140 **D. Phylogenetic framework**

141 For our main analyses, we use a fossil dating of the supermatrix phylogeny with 6,714 species as
142 tips{{641 Burleigh,J.Gordon 2015;}}. Of the two global, species-level phylogenies of birds with
143 similar levels of species sampling, there are several advantages to the phylogeny we use for the
144 main analyses. This phylogeny was built without topological constraints, and the molecular
145 branch lengths were estimated from the same alignment used to infer the phylogeny{{731
146 Gatesy 2004;}}. This property of the phylogeny is desirable, as the relative branch lengths across
147 the phylogeny must be informative for analyses that rely on divergence time estimates for
148 species pairs. The alternate global phylogeny{{646 Jetz 2012;}} uses a topological constraint
149 based on previous reconstructions of the deeper nodes in the avian phylogeny {{724 Hackett
150 2008;}}. Interestingly, the Burleigh *et al.* analyses largely recapitulate the topology of the
151 Hackett *et al.* phylogeny{{724 Hackett 2008;}} that forms the constraint for deep nodes in the
152 Jetz *et al.* analyses. To account for uncertainty in divergence times in our analyses, we use
153 divergence times from the Jetz *et al.* analyses and found that our results using the Burleigh tree
154 are generally consistent with results from this alternate phylogenetic reconstruction and dating
155 method (see Sensitivity Analyses section below).

156 **E. Fossil calibration**

157 The phylogenetic tree from Burleigh *et al.* (2015) has molecular branch lengths. In order to
158 make the tree ultrametric, with branch lengths representing time, we performed a penalized
159 likelihood analysis with r8s v. 1.71 {{760 Sanderson 2003;}}. r8s was chosen because a Bayesian
160 divergence time estimation analysis, for example using BEAST {{182 Drummond 2007;}}, was
161 not computationally feasible for this large dataset. Twenty well-vetted fossil observations with
162 minimum and maximum dates from throughout the avian tree were selected to calibrate the
163 branch lengths (SI Dataset 1). We determined the optimal smoothing parameter for r8s using a
164 cross-validation analysis in which we fixed the age of crown Psittacopasserae to 60 million years
165 ago (mya), midway between its minimum (53.5 mya) and maximum (66.5 mya) ages, and
166 examined how closely different smoothing parameters estimated the fossil-constrained node ages
167 when they were not constrained. The optimal smoothing parameter was 3.2. In r8s analyses that
168 did not constrain the age of the root of the avian tree, estimates of the ages of the root and
169 earliest divergences were unrealistically ancient. Yet, an optimal age constraint for the avian tree
170 is not obvious. We fixed the age of the root to 110 mya. Although this makes the age of the
171 earliest divergences more credible, it makes little difference in the ages of the sister taxa in the
172 tree.

173 **F. Ecological divergence of sister pairs**

174 *Body mass* Body size impacts many aspects of organismal biology {{734 Calder 1983; 735 Peters
175 1986;}}, and its divergence in nascent species should have profound effects on their interactions.
176 Theory suggests that body size differences between species should reduce competition {{733
177 Wilson 1975;}}. Such differences could also play a role in the reduction of reproductive
178 interference if they are associated with increased pre- or post-zygotic reproductive
179 isolation {{737 Streelman 2003;}}. The establishment of sympatry in recently diverged taxa is
180 frequently associated with substantial divergence in body size {{741 Huxley 1942; 736 Abbott
181 1977; 315 Schluter 2000;}}, suggesting that body mass divergence indeed facilitates coexistence.

182 In this study, we used body mass as a proxy for body size, with values sourced from a
183 published global data set {{661 Dunning 2008;}} and an update
184 (<https://ag.purdue.edu/fnr/Documents/WeightBookUpdate.pdf>). We incorporated body mass
185 disparity for species pairs by taking the difference in log mean of the two species, resulting in a
186 unitless measure of divergence. In cases where different values were reported for males and

187 females, we took the mean of these values to represent a species-level mass measurement before
188 calculating body mass disparity for the species pair.

189

190 *Beak morphology* Divergence in beak morphology has long been thought to play a central role in
191 the ecological niche divergence of bird species via its association with foraging differences{{739
192 Hutchinson 1959; 726 Schoener 1965;}}. Beak morphology differences, inclusive of size, are
193 thought to contribute to enabling co-existence{{740 Grant 1968;}}, and are thus considered to
194 correspond to differences in the α -niche{{645 Pigot 2013;}}. Beak morphology differences
195 could impact the probability of sympatry, and could also impact the probability of breeding local
196 co-occurrence (versus the absence of co-occurrence) through its effects on facilitating sympatry.
197 Here we use three measures of beak morphology – culmen length, beak depth, and beak width –
198 to represent beak morphology as a proxy of foraging differences between species. The mean
199 number of individuals per species for these measurements was $4.72 \pm 4.62SD$ across 1,884 species
200 sampled. We sought to represent sexes evenly for each species as much as possible for these
201 measurements. Mean values for culmen length, beak depth, and beak width were then taken for
202 each species (males, females, and unassigned individuals were averaged together). The mean
203 proportion of individuals identified to sex across species samples was $.898 \pm .234SD$. In species
204 for which at least one individual was identified by sex (1,833 species), the mean proportion male
205 was $.534 \pm .176SD$. In species for which all sampled individuals were identified to sex (1,487
206 species), the mean proportion male was $.526 \pm .130SD$.

207 We performed a phylogenetic Principal Components Analysis (PPCA){{304 Revell
208 2009;}} to account for collinearity and phylogenetic effects in the divergence of beak
209 morphology. From this analysis, we obtained scores along three principal components for each
210 species (Supplementary Table 1, Database 4).

211 We then included beak morphology in analyses of the establishment of secondary contact
212 (including any contact and breeding contact) and the establishment of sympatry in two ways: 1)
213 we used a single distance metric to represent beak divergence, and 2) we included the divergence
214 along each of the PC axes separately. In the PPCA, the first principal component loads heavily
215 on all three linear measurements, indicating that it predominantly reflects bill size differences. In
216 order to extract a single distance metric from the PC space that adequately incorporated bill
217 shape aspects, we first re-scaled each of the principal components to unit variance, then

218 calculated Euclidean distance in this transformed PC space. The re-scaling of PC space increases
219 the relative importance of bill shape differences in the Euclidean distance metric so that the
220 metric is not overly dominated by size differences, and reflects shape differences in addition to
221 size. In the alternate sets of analyses, we include species pair differences in each of the three
222 principal components as separate predictors.

223 Both the Euclidean distance in beak morphology space and the beak PC1 are likely to be
224 correlated with body size, which could result in spurious correlations of co-occurrence or
225 sympatry with these variables when true correlations exist with body size differences instead. As
226 all our GLM include body mass disparity as a predictor, we have controlled for body size
227 differences in our analyses, such that support for the importance of beak morphology differences
228 in our multi-model inferences is independent of the role of body size⁴³.

229

230 **H. The role of dispersal in transitions to secondary contact and sympatry**

231 As the speed of range expansion may be associated with dispersal capacity in birds, elevated
232 dispersal capacity may shorten waiting times from geographic isolation to secondary
233 contact{{751 Pigot 2015;}}. If elevated dispersal results in rapid re-establishment of contact
234 following the initial molecular or phenotypic divergence of populations, high levels of gene flow
235 may result, slowing or preventing further divergence and speciation{{745 Weeks 2014; 647
236 Claramunt 2012;}}. However, where rapid ecological or social divergence occurs in taxa with
237 elevated dispersal, population merging may be prevented following rapid transitions to
238 secondary contact, with speciation proceeding despite contact{{746 Grant 2014}}.

239 For incipient species that do not merge following secondary contact, elevated dispersal could
240 facilitate more rapid transitions to sympatry by promoting range expansions that might otherwise
241 be slowed by ecological and reproductive interactions. When these interactions inhibit the range
242 expansions of closely related species, they may do so by impacting potential colonizers, who
243 occur in small numbers or at small densities during colonization{{417 Case 2005;}}. The
244 inhibition of colonization that occurs as a function of interactions at species borders can occur as
245 a priority effect, and not because a resident species (or resident nascent species) is necessarily
246 ecologically superior to the colonizer{{417 Case 2005;}}. Species with elevated dispersal
247 capacities may more easily surmount the challenges faced during colonization of a sister species
248 range because fatter dispersal tails may result in higher numbers or greater frequencies of the

249 colonizing species within the range of the resident, increasing the probability of establishment by
250 decreasing the probabilities of extirpation through stochastic or Allee effects{{747 Lockwood
251 2005; 685 Stephens 1999;}}. Alternatively, populations found within the ranges of sister species
252 may sometimes be sink populations with negative growth rates, buoyed by dispersal from non-
253 sympatric populations. Such source-sink dynamics could yield sympatry where the species at
254 lower density is sustained only because of dispersal ability and not because the two species have
255 diverged sufficiently in ecological and/or reproductive traits.

256 Direct data on dispersal distances or dispersal kernels are not available for most
257 species{{742 Jønsson, Knud Andreas 2016; 743 Koenig 1996;}}. However, the wing shapes of
258 birds bear variation that is indicative of dispersive capacity variation – birds with higher wing
259 aspect ratios have higher natal dispersal distances{{738 Dawideit, Britta A 2009;}} and reduced
260 molecular variation across ranges{{748 Burney 2009;}}. In this study, we assess the potential
261 contribution of dispersal capacity to the timing and outcomes of secondary contact by including
262 the hand-wing index, a metric representing wing aspect ratio, as a proxy for dispersal capacity, in
263 analyses of local co-occurrence and sympatry in sister pairs. As indicated in the Methods section,
264 the hand-wing index is calculated as:

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

265 where WL (wing length) is the standard length of the closed wing (the length of the longest
266 primary), and SL (secondary length) is the distance from the carpal joint to the tip of the first
267 secondary feather. Higher values of the hand-wing index (HWI) are correlated with greater
268 dispersal distances, such that we hypothesized that species with higher HWI should have shorter
269 waiting times to secondary contact and higher probabilities of sympatry given contact. We use
270 hand-wing index measurements made on 1,889 species, with a mean of $4.72 \pm 4.62SD$
271 measurements per species.

272

273 **Data set organization**

274 **Dataset 1** is the species pair dataset for GLMs of local co-occurrence and breeding local
275 co-occurrence probability with divergence time as the only predictor ($n = 952$ pairs). It
276 comprises species pairs that met the following criteria: 1) the pair have a sister relationship in the
277 phylogeny we use for our main analyses{{641 Burleigh 2015;}}; 2) species-level sampling

278 within the phylogeny was $\geq 75\%$; 3) there were no additional reasons to suspect the sister pair
279 was not a “true” sister pair; 4) taxonomic instability is unlikely to have caused erroneous co-
280 occurrence observations in the eBird dataset; 5) the pair’s co-occurrence score is not ambiguous
281 (where ambiguity arises because only questionable co-occurrence records exist but co-
282 occurrence is feasible according to range maps); 6) the pair’s co-occurrence is not the result of
283 human introduction; 7) the pair is not a formerly sympatric pair for which there is no longer
284 contact because of human-caused extirpations; 8) the minimum number of species observations
285 in the eBird dataset (downloaded October 2013) was ≥ 10 (see SI Dataset 7).

286 **Dataset 2** is the species pair dataset for GLMs of local co-occurrence and breeding local
287 co-occurrence probability designed to examine the factors associated with increased or decreased
288 waiting times to secondary contact. It is the subset of Dataset 1 for which information was
289 available for all traits (body mass, beak morphology, and hand-wing index; $n = 848$ pairs).

290 **Dataset 3** is the species pair dataset for GLMs examining the probability of sympatry
291 versus parapatry in species pairs for whom there is evidence of breeding season local co-
292 occurrence and a full complement of trait data ($n = 440$ pairs).

293 **Dataset 4** is the set of morphological measurements by species (culmen length, beak
294 depth, beak width, hand-wing index; $n = 1,881$ species).

295 **Dataset 5** was used for analyses of body mass divergence. It is the subset of the species
296 pairs in Dataset 1 for which body mass estimates were available for both species ($n = 872$ pairs).

297 **Dataset 6** is the list of calibrations (fossil and basal node) used to date the Burleigh et al.
298 2015 phylogeny.

299 **Dataset 7** is the list of all sister species from the Burleigh et al. 2015 maximum
300 likelihood phylogeny that were excluded from analyses, along with the reasons for exclusion.

301

302 **Analyses**

303 *General approach to GLM* We chose to use generalized linear models (GLM) to analyze
304 transitions to secondary contact and sympatry because of their flexibility in accommodating
305 continuous and discrete predictors {{749 Bolker 2009;}}, the ease of examining a large number
306 of candidate models, including many predictors and their interactions {{667 Calcagno, V.;}}, and
307 because we could control for divergence time in assessing the contribution of other factors in
308 affecting the probability of local co-occurrence and sympatry. We expected that, as species

309 ranges begin in isolation under allopatric speciation, and move in a finite space, the probability
310 of secondary contact and sympatry necessarily increases with time.

311 In generating and assessing GLM, we included both main effects and pairwise interaction
312 terms. As we considered as many as eight predictors in our models, the number of possible
313 models to consider was very large. Instead of considering all possible models for each analysis,
314 we used a genetic algorithm in the R package `glmulti` (Calcagno, V.;) to search model
315 space for the best models. The genetic algorithm is so called because processes analogous to
316 population genetic processes (referred to as mutation, recombination, immigration) are employed
317 in searching model space. When implementing the genetic algorithm, a population of models is
318 maintained. New candidate models are generated through modifications to the model population
319 (mutation and recombination), and via the introduction of random models to the population
320 (immigration). The genetic algorithm is an efficient way to search model space in lieu of
321 exhaustively fitting all possible models.

322 Within `glmulti`, we fit candidate models using the R function `glm` while maintaining a
323 confidence set (population) size of 300 models. We constrained model space by limiting the
324 number of terms that could enter models to 10 and by only considering ‘marginal’ models (those
325 models where any term included in an interaction is also included as a main effect – (Venables 1997;
326 Grafen 2002;)). Model selection was performed using AICc scores.

327
328 *Beak morphology in GLM* We incorporated aspects of beak morphology in two ways for GLMs
329 analyzing which factors predict the probability of co-occurrence and sympatry. We began with a
330 phylogenetic principal components analysis on culmen length, beak depth, and beak width. For
331 some sets of models, we included as a predictor the Euclidean distance between sister pairs in
332 beak morphology PC space. However, the Euclidean distance can be dominated by the first
333 principal component, which explained much of the variation but can largely reflect body size
334 differences. Thus, we first scaled each of the three PCs to unit variance, such that PC2 and PC3
335 were better represented in Euclidean distances. As this Euclidean distance will still incorporate
336 body size differences, we controlled for body size differences in all GLM routines by including
337 log body mass disparity as a separate predictor (Freckleton 2002;). In the alternate
338 modeling routines that included beak morphology, species pair differences along each of the

339 three PC axes were included as separate predictors, in the same way body mass disparity for each
340 pair was included in GLMs (see also **Ecological divergence of sister pairs**, above).

341

342 *Simulations of range dynamics* The relationship of the probability of co-occurrence and
343 divergence time in sister pairs is surprising, in that we did not recover the near-zero intercept
344 expected under allopatric speciation{{657 Barraclough 2000; 645 Pigot 2013;}}. However, with
345 sufficiently short delays in the establishment of secondary contact, this signature of geographic
346 isolation could be lost. It was thus necessary to examine the minimal rate of transition to
347 secondary contact that would be required to yield an intercept equivalent to our empirical
348 estimate of the intercept in a generalized linear model with divergence time as the sole predictor.

349 We simulated range dynamics with two possible states: isolation and contact, assuming
350 all species pairs begin in isolation. We make the simplifying assumption that a single rate
351 governs the range dynamics, i.e. transitions into and out of contact, across all species pairs, with
352 stochasticity responsible for the variation in this process among pairs. While this assumption
353 sacrifices some realism, especially as we find that there are biological factors that explain
354 variation in the probability of local co-occurrence when controlling for divergence time, the
355 process is still noisy in nature, such that a simple stochastic model may yet be highly informative
356 regarding the timescale of these processes. Moreover, Pigot and Tobias (2013) found that a
357 model with a single rate governing transitions to sympatry (instead of contact, as we analyze
358 here) had better relative support than a model with rates that varied as a function of
359 morphological differences. Thus, the simple stochastic model may be sufficient to characterize
360 geographic range transitions{{645 Pigot 2013;}}.

361 To simulate transitions into and out of secondary contact, we considered the set of
362 divergence times for all sister pairs in our analyses. We began by setting time $t = 0$,
363 corresponding to the outset of divergence for each sister pair. For a given set of two parameters
364 representing 1) the rate of transition into secondary contact from isolation, σ , and 2) from
365 secondary contact back into isolation, ε , we calculated the waiting time to the next transition
366 event. The type of event (transition into contact or transition out of contact) is chosen after
367 calculating the probability corresponding to each based on their relative rates, and then applied
368 pseudorandomly to one of the sister pairs. When the time t reaches the time equal to the
369 divergence time for a sister pair, the sister pair's state (secondary contact or isolation) is

370 recorded, and the pair removed from the set of pairs for which a transition can take place. At the
371 end of the simulation, each sister pair is either in contact or in isolation.

372 We took the set of states, and the associated divergence time for each sister pair, and used
373 a generalized linear model to calculate an intercept of the relationship between the probability of
374 co-occurrence (secondary contact) and divergence time in a fashion identical to our empirical
375 analysis. We considered forward transition rates (into contact from isolation) from 0.1 to 0.8
376 transitions per million years, with increments of 0.1 transitions per million years. We then
377 considered values of ϵ as multiples of σ (.005, .01, .05, .1, .2, and .5 times σ). We ran 5 replicate
378 simulations for each parameter combination, and found the maximum intercept from each
379 replicate set. We present the maximum intercept for each σ across all ϵ (Extended Data Fig. 4).
380 We performed a sensitivity analysis on divergence time estimates by performing the same
381 analysis using the alternate set of divergence time estimates for sister pairs from the Jetz *et al.*
382 analyses{{646 Jetz 2012;}} (Extended Data Fig. 5).

383
384 *Trait divergence models* We fit four evolutionary models{{662 Uyeda 2011;}} to sister pair
385 divergence of mean body mass and beak morphology PC1 to examine relative support. The first
386 of these models, a model of constrained evolution, corresponds to bounded evolution
387 independent of time. In this model, body mass divergence is a random draw from a normal
388 distribution, and the expected variance σ_p in body mass disparity among species pairs is
389 equivalent at all divergence times. In the remaining three models, this time-independent model is
390 included as a component, and an additional time-varying component is added. In the gradual
391 evolution model, the time-varying component exists as a Brownian motion process, where the
392 expected variance in body mass divergence increases linearly with time (the expected variance at
393 time $t = \sigma_{BM}^2 t$). In the single pulse model, the time-varying component of evolutionary
394 divergence is modeled as an instantaneous pulse of evolution, where the magnitude of the pulse
395 is drawn from a normal distribution $N(0, \sigma_D^2)$ and the waiting times to a pulse are exponentially
396 distributed, with rate parameter λ . In the multiple pulse model, the number of pulses is drawn
397 from a Poisson process with rate parameter λt , and the magnitude of each pulse is again a draw
398 from a normal distribution $N(0, \sigma_D^2)$.

399

400 **Sensitivity analyses**

401 *Alternate divergence times* In our main analyses (GLM on the sister pair data set, Markov model
402 simulations of the transition to local co-occurrence, and mode analyses on body mass
403 divergence), we used divergence time estimates for species pairs from the fossil calibration of a
404 recent supermatrix molecular phylogeny {{641 Burleigh,J.Gordon 2015;}}. There is inherent
405 uncertainty and error in the dating of divergence times, however, and when possible, analyses
406 that depend on estimated dates from phylogenetic analyses should test their robustness to
407 alternate date estimates. To examine the robustness of our analyses to divergence time estimate
408 methods, we extracted divergence time estimates for the same set of species pairs in our main
409 analyses from an alternate species-level phylogenetic analysis{{646 Jetz,W. 2012;}}. There are
410 two types of phylogenetic analyses in the well-known Jetz et al. analyses, both of which use a
411 ‘backbone’ phylogeny as a constraint on deeper phylogenetic relationships {{724 Hackett
412 2008;}}. There are phylogenetic analyses in which only those species for which molecular data
413 were obtained are included, and there are additional analyses where species without molecular
414 data supporting their position are added to the phylogeny, subject to taxonomic constraints. In
415 our sensitivity analyses, we use divergence time estimates from only the former, as we expect
416 there is the potential to be misled from the latter analysis{{732 Rabosky 2015;}}, in part because
417 the topological constraints imposed for species without molecular data iteratively placed in
418 resolved positions on the phylogeny in some cases may not be justifiable{{641
419 Burleigh,J.Gordon 2015;}}. We replicated sister pair GLM analyses, the Markov model
420 simulations of the transition to local co-occurrence, and mode analyses on body mass
421 divergence, and show that all of our major results and conclusions are robust to these alternate
422 divergence time estimates for species pairs. These sensitivity analyses required that both species
423 from each species pair appear in the ‘molecular data-only’ Jetz *et al.* analyses, and thus the
424 sample size of these analyses is reduced. As the Jetz *et al.* analyses are Bayesian analyses, the
425 phylogenetic results are pseudo-posterior distributions of phylogenies, where each phylogeny has
426 a different divergence time estimate for a given species pair. We therefore take as our estimated
427 divergence times the means of divergence time estimates across 100 phylogenetic trees from the
428 posterior distribution of the molecular-only analyses.

429 We took an additional step to examine the sensitivity of our analyses to divergence time
430 estimates by repeating GLMs of the probability of local co-occurrence and breeding local co-
431 occurrence with divergence time as the sole predictor for: 1) divergence time estimates from

432 each of 100 bootstraps of the Burleigh tree; and 2) divergence time estimates from each of the
433 10,000 pseudo-posterior samples of the Jetz *et al.* analyses.

434

435 **Potential biases**

436 As our analyses use divergence time estimates that are based on molecular branch lengths, then
437 calibrated using the fossil record, it is possible that more recent divergence times will be
438 comparatively over-estimated relative to more ancient divergence times. This effect arises
439 because coalescent times for molecular variation pre-date species splitting by an amount of time
440 that is independent of the age of the split{{750 Edwards, ScottV 2000;}}. Recent divergence
441 times are disproportionately impacted by this bias, as the differences between molecular branch
442 lengths and species splitting times are a greater fraction of the total divergence time for recent
443 splits. Fossil calibration may reduce this bias slightly, but the effect is likely to remain. The
444 bias's effect in our analyses should be that we have underestimated the frequency of early
445 secondary contact, which suggests that the intercept of the relationship between the probability
446 of co-occurrence and divergence time should be higher, and the slope flatter. This bias should
447 work in the same direction as our likely under-estimation of the frequency of species pairs that
448 locally co-occur, arising from weak observation records for some included taxa. Furthermore, we
449 are likely to have inferred a longer delay to mass divergence pulses than if we had access to true
450 species divergence times. Overall, transitions from allopatry to secondary contact, and from
451 secondary contact to sympatry, and additionally the divergence in body size, are likely to be
452 faster than what we present here if the main bias in our divergence time estimates stems from the
453 difference between coalescent times and species splits.

454 Because we use a species-level phylogeny (see Burleigh *et al.* 2015), we are assessing co-
455 occurrence of recognized species. A similar analysis that included subspecies, races, ecomorphs,
456 geographic variants, phylogroups, or other forms below the species level may add resolution at
457 low ages of evolutionary divergence, and could reveal non-linearities in the relationship between
458 divergence time and the probability of co-occurrence. We note, though, that there is
459 hybridization among some of the sister pairs included in the analysis. This observation indicates
460 that we are examining a spectrum of stages in the speciation process, including what amounts to
461 incomplete speciation for some species concepts.

462 For latitudinal analyses, biases may arise from the unevenness of systematic and
463 taxonomic efforts across latitude. Less effort has been made in bird systematics and taxonomy
464 per species at lower latitudes, which may result in fewer species-level splits in tropical species
465 complexes. If species complexes at lower latitudes are under-split relative to taxa at higher
466 latitudes, we may fail to capture similar spectra of speciation stages at different latitudes. It is
467 possible that this bias could lead to the underrepresentation of low-latitude allopatric species
468 pairs. This bias may be counteracted by the undersampling of species observations at low
469 latitudes, which should introduce an opposing bias in local co-occurrence analyses, because we
470 have failed to report true co-occurrence of tropical species pairs.

471

472 **2. Supplementary Tables**

473

474 **Table S1:** Loadings for the phylogenetic Principal Components Analysis (phylogenetic PCA) on
475 mean beak measurements for 1,881 species.

	PC1	PC2	PC3
Culmen length	-0.986	0.163	0.013
depth	-0.973	-0.206	0.106
width	-0.978	-0.134	-0.162

476

477

478

479

480

481

482

483

484

485

486

487

488

489 **Table S2-S3:** Multi-model inference of predictor effects for breeding range co-occurrence versus
 490 absence of such co-occurrence in bird sister species ($n = 849$ species pairs). Generalized linear
 491 models were implemented with binomial errors and a logit link. Predictors were: divergence
 492 time, mass disparity, latitude, log mean hand-wing index, and either Euclidean distance in beak
 493 PC morphospace (SI Table S2) or each of the disparities along the beak PC axes (SI Table S3).
 494 Coefficient estimates, unconditional variance, confidence interval and relative importance for all
 495 predictors with importance >0 among models with $\Delta AIC < 2$. The number of models with ΔAIC
 496 < 2 is in parentheses following model parameters.

497

498 **Table S2: Burleigh divergence times and Euclidean beak distances (24 models)**

Predictor	Estimate	Variance	Conf.	
			interval	Importance
latitude x migration	6.67E-01	3.41E-02	3.62E-01	1
DT	3.19E-01	1.19E-02	2.14E-01	1
mass disparity	1.27E-01	1.41E-02	2.33E-01	1
latitude	6.45E-02	9.56E-03	1.92E-01	1
(Intercept)	3.28E-02	8.61E-03	1.82E-01	1
migration	-4.15E-01	4.34E-02	4.09E-01	1
DT x migration	-3.22E-01	3.61E-02	3.73E-01	9.31E-01
mass disparity x migration	-2.80E-01	4.69E-02	4.25E-01	8.44E-01
mean log HWI	6.29E-02	5.75E-03	1.49E-01	6.88E-01
DT x mass disparity	-5.58E-02	4.38E-03	1.30E-01	5.81E-01
Euc beak dist	-8.06E-03	1.43E-03	7.42E-02	3.15E-01
Euc beak dist x migration	1.04E-01	2.98E-02	3.39E-01	2.53E-01
latitude x mass disparity	9.47E-03	4.58E-04	4.20E-02	1.12E-01
latitude x mean log HWI	2.59E-03	5.49E-05	1.45E-02	6.40E-02
Euc beak dist x DT	-3.37E-03	5.20E-05	1.42E-02	4.71E-02
mean log HWI x migration	2.97E-03	8.99E-05	1.86E-02	3.22E-02
Euc beak dist x mean log HWI	3.76E-03	6.34E-05	1.56E-02	3.08E-02
DT x latitude	-7.37E-04	1.06E-05	6.40E-03	2.98E-02
mean log HWI x mass	1.63E-04	4.56E-06	4.19E-03	2.86E-02

disparity
DT x mean log HWI 9.01E-05 5.11E-06 4.44E-03 2.85E-02

499

500

501 **Table S3: Burleigh divergence times and individual beak PCs (8 models), see legend above**

502 **Table S2**

Predictor	Estimate	Variance	Conf.	
			interval	Importance
latitude x migration	6.62E-01	3.41E-02	3.63E-01	1
DT	2.71E-01	1.38E-02	2.30E-01	1
mass disparity	1.41E-01	1.35E-02	2.28E-01	1
beak PC2 disparity	6.46E-02	1.06E-02	2.02E-01	1
(Intercept)	5.83E-02	8.43E-03	1.80E-01	1
latitude	5.53E-02	9.61E-03	1.92E-01	1
beak PC2 disparity x mass	-1.25E-01	2.58E-03	9.98E-02	1
disparity				
migration	-4.19E-01	4.22E-02	4.03E-01	1
mass disparity x migration	-3.06E-01	5.20E-02	4.48E-01	8.10E-01
beak PC2 disparity x migration	1.85E-01	4.03E-02	3.94E-01	6.49E-01
DT x migration	-2.12E-01	4.72E-02	4.26E-01	6.16E-01
mean log HWI	2.90E-02	2.64E-03	1.01E-01	3.44E-01
beak PC1 disparity	4.30E-02	5.63E-03	1.47E-01	2.03E-01

503

504

505

506

507

508

509

510

511

512 **Table S4-S5:** Multi-model inference of predictor effects for sympatry versus parapatry of bird
513 sister species that locally co-occur in breeding ranges ($n = 441$ species pairs). Generalized linear
514 models were implemented with binomial errors and a logit link. Predictors were: divergence
515 time, mass disparity, latitude, migration status (if either species is migratory or has non-breeding
516 season movements), log mean hand-wing index, and either Euclidean distance in beak PC
517 morphospace (SI Table S4) or each of the species disparities along the beak PC axes (SI Table
518 S5). Coefficient estimates, unconditional variance, confidence interval and importance, for all
519 predictors with importance >0.5 among models with $\Delta AIC < 2$. The number of models with
520 $\Delta AIC < 2$ is in parentheses following the specifications of the four model sets.

521

522 **Table S4: Sympatry versus parapatry GLM with Euclidean beak distances (8 models)**

Predictor	Estimate	Variance	Conf.	
			interval	Importance
(Intercept)	7.61E-01	1.33E-02	2.27E-01	1
DT	7.21E-01	2.43E-02	3.06E-01	1
mass disparity	1.30E-01	2.11E-02	2.85E-01	1
mean log HWI	5.25E-02	1.33E-02	2.27E-01	1
DT x mean log HWI	-3.22E-01	1.93E-02	2.73E-01	1
Euc beak dist	-9.19E-03	3.23E-03	1.12E-01	3.74E-01
Euc beak dist x DT	9.40E-02	2.37E-02	3.03E-01	2.83E-01
Euc beak dist x mass disparity	-1.52E-02	1.09E-03	6.47E-02	1.07E-01
migration	-8.75E-03	1.09E-03	6.49E-02	9.76E-02
mean log HWI x mass disparity	3.66E-03	2.63E-04	3.19E-02	9.52E-02
DT x mass disparity	2.73E-03	2.24E-04	2.94E-02	9.36E-02
latitude	2.34E-04	9.57E-05	1.92E-02	9.13E-02

523

524

525

526

527

528 **Table S5: Sympatry versus parapatry GLMs with individual beak PCs (10 models), see**
 529 **legend above Table S4**

Predictor	Estimate	Variance	Conf.	
			interval	Importance
(Intercept)	7.61E-01	1.47E-02	2.38E-01	1
DT	7.38E-01	2.64E-02	3.19E-01	1
beak PC1 difference x DT	6.38E-01	5.70E-02	4.69E-01	1
mass disparity	1.38E-01	3.73E-02	3.80E-01	1
mean log HWI	5.98E-02	1.36E-02	2.29E-01	1
beak PC1 difference	2.11E-02	3.08E-02	3.45E-01	1
DT x mean log HWI	-3.77E-01	2.03E-02	2.80E-01	1
DT x mass disparity	-3.12E-01	4.02E-02	3.94E-01	9.05E-01
beak PC3 difference	-1.13E-02	1.02E-03	6.28E-02	1.68E-01
beak PC2 difference	2.06E-02	1.88E-03	8.52E-02	1.34E-01
beak PC1 difference x mass disparity	-1.86E-02	1.54E-03	7.71E-02	1.26E-01
beak PC1 difference x beak PC3 difference	1.81E-02	1.42E-03	7.40E-02	9.00E-02
mean log HWI x mass disparity	4.98E-03	2.55E-04	3.14E-02	7.71E-02
beak PC1 difference x mean log HWI	3.51E-03	2.10E-04	2.85E-02	7.26E-02
migration	-4.95E-03	5.24E-04	4.50E-02	7.15E-02
latitude	6.95E-04	6.51E-05	1.59E-02	6.91E-02

530

531

532

533

534

535

536

537

538 **Table S6-S7: Sensitivity analyses for divergence time estimates** Multi-model inference of
 539 predictor effects for breeding range co-occurrence versus absence of such co-occurrence, in bird
 540 sister species ($n = 774$ species pairs). Generalized linear models were implemented with
 541 binomial errors and a logit link. Predictors were: divergence time (alternate estimates from Jetz
 542 *et al.* 2012), mass disparity, latitude, log mean hand-wing index, and either Euclidean distance in
 543 beak PC morphospace (SI Table S6) or each of the disparities along the beak PC axes (SI Table
 544 S7). Coefficient estimates, unconditional variance, confidence interval and relative importance
 545 for all predictors with importance >0 among models with $\Delta AIC < 2$. The number of models with
 546 $\Delta AIC < 2$ is in parentheses following model parameters.

547

548 **Table S6: Jetz divergence times and Euclidean beak distances (17 models)**

Predictor	Estimate	Variance	Conf.	
			interval	Importance
latitude x migration	6.57E-01	3.63E-02	3.74E-01	1
mass disparity	2.18E-01	2.09E-02	2.84E-01	1
DT	2.02E-01	1.78E-02	2.62E-01	1
latitude	4.02E-03	1.14E-02	2.10E-01	1
(Intercept)	-1.06E-01	1.63E-02	2.51E-01	1
migration	-3.35E-01	5.21E-02	4.48E-01	1
mass disparity x migration	-4.17E-01	4.29E-02	4.07E-01	1
mean log HWI	6.90E-02	6.40E-03	1.57E-01	7.73E-01
DT x mass disparity	-8.16E-02	5.99E-03	1.52E-01	7.15E-01
latitude x mass disparity	1.78E-02	1.37E-03	7.27E-02	1.79E-01
Euc beak dist	1.13E-02	7.50E-04	5.37E-02	1.49E-01
DT x migration	-2.80E-02	4.05E-03	1.25E-01	1.12E-01
mean log HWI x mass disparity	-2.57E-03	5.06E-05	1.40E-02	5.29E-02
Euc beak dist x migration	1.33E-02	8.40E-04	5.69E-02	4.75E-02
Euc beak dist x DT	-4.69E-03	9.90E-05	1.95E-02	4.75E-02
Euc beak dist x mass disparity	-3.88E-03	6.62E-05	1.60E-02	4.64E-02

latitude x mean log HWI	7.97E-04	1.54E-05	7.71E-03	4.45E-02
DT x latitude	-1.03E-03	3.86E-05	1.22E-02	4.38E-02
DT x mean log HWI	-8.57E-04	2.93E-05	1.06E-02	4.38E-02
mean log HWI x migration	1.48E-03	9.70E-05	1.93E-02	4.37E-02

549

550

551 **Table S7: Jetz divergence times and individual beak PCs (5 models), see legend above**

552 **Table S6**

Predictor	Estimate	Variance	Conf.	
			interval	Importance
latitude x migration	6.61E-01	3.40E-02	3.62E-01	1
DT	2.77E-01	1.30E-02	2.24E-01	1
beak PC2 difference	9.21E-02	7.73E-03	1.73E-01	1
latitude	6.00E-02	9.60E-03	1.92E-01	1
(Intercept)	5.78E-02	8.39E-03	1.80E-01	1
mass disparity	3.96E-02	1.66E-02	2.53E-01	1
beak PC2 difference x mass disparity	-1.31E-01	2.53E-03	9.88E-02	1
migration	-4.06E-01	4.12E-02	3.98E-01	1
beak PC1 difference	1.70E-01	1.43E-02	2.35E-01	8.65E-01
DT x migration	-2.56E-01	4.12E-02	3.99E-01	7.90E-01
mass disparity x migration	-2.07E-01	4.67E-02	4.24E-01	6.28E-01
mean log HWI	1.43E-02	9.02E-04	5.89E-02	1.86E-01

553

554

555

556

557

558

559

560

561

562 **Table S8-S9: Sensitivity analyses for divergence time estimates** Multi-model inference of
 563 predictor effects for sympatry versus parapatry of bird sister species with breeding range local
 564 co-occurrence ($n = 382$ species pairs). Generalized linear models were implemented with
 565 binomial errors and a logit link. Predictors were: divergence time (Jetz), mass disparity, latitude,
 566 migration status (if either species is migratory or has non-breeding season movements), log mean
 567 hand-wing index, and either Euclidean distance in beak PC morphospace (SI Table S8) or each
 568 of the disparities along the beak PC axes (SI Table S9). Coefficient estimates, unconditional
 569 variance, confidence interval and importance, for all predictors with importance >0.5 among
 570 models with $\Delta AIC < 2$. The number of models with $\Delta AIC < 2$ is in parentheses following the
 571 specifications of the four model sets.

572

573 **Table S8: Jetz divergence times and Euclidean beak distance (7 models)**

Predictor	Estimate	Variance	Conf.	
			interval	Importance
DT	1.00E+00	9.03E-02	5.91E-01	1
mass disparity	5.75E-01	4.37E-02	4.11E-01	1
(Intercept)	4.10E-01	4.17E-02	4.02E-01	1
mean log HWI	2.75E-01	3.89E-02	3.88E-01	1
DT x mass disparity	-3.61E-01	1.79E-02	2.63E-01	1
DT x mean log HWI	-3.97E-01	5.57E-02	4.64E-01	9.00E-01
mean log HWI x mass disparity	7.04E-02	1.34E-02	2.27E-01	3.38E-01
migration	-2.24E-02	5.33E-03	1.44E-01	2.02E-01
Euc beak dist	4.33E-03	3.44E-04	3.64E-02	1.00E-01
latitude	-1.58E-03	1.51E-04	2.42E-02	9.66E-02

574

575 **Table S9: Jetz divergence times and individual beak PCs (13 models), see legend above**
 576 **Table S8**

Predictor	Estimate	Variance	Conf.	
			interval	Importance

DT	1.20E+00	1.08E-01	6.45E-01	1
beak PC1 difference x DT	1.17E+00	2.14E-01	9.10E-01	1
mass disparity	7.34E-01	6.66E-02	5.08E-01	1
(Intercept)	2.99E-01	4.68E-02	4.25E-01	1
mean log HWI	2.85E-01	3.42E-02	3.64E-01	1
DT x mean log HWI	-4.91E-01	4.65E-02	4.24E-01	1
beak PC1 difference	-6.31E-01	9.68E-02	6.12E-01	1
DT x mass disparity	-7.83E-01	4.42E-02	4.14E-01	1
beak PC2 difference	2.72E-01	3.04E-02	3.43E-01	9.35E-01
migration	-2.97E-02	5.01E-03	1.39E-01	1.33E-01
mean log HWI x mass disparity	1.90E-02	1.62E-03	7.90E-02	1.01E-01
beak PC3 difference	-1.36E-02	8.82E-04	5.84E-02	9.19E-02
beakPC2 difference x mean log HWI	-1.45E-02	1.03E-03	6.30E-02	8.77E-02
beak PC1 difference x mass_disparity	-1.41E-02	9.73E-04	6.13E-02	8.67E-02
beak PC1 difference x mean log HWI	1.27E-02	8.69E-04	5.80E-02	7.89E-02
mean log HWI x migration	3.54E-02	5.19E-03	1.42E-01	7.62E-02
beak PC1 difference x beak PC2 difference	1.07E-02	7.53E-04	5.40E-02	6.71E-02
latitude	-1.58E-03	5.83E-05	1.50E-02	5.17E-02
beak PC2 difference x mass disparity	2.22E-03	1.94E-04	2.74E-02	5.09E-02
beak PC2 difference x DT	3.85E-05	2.59E-04	3.16E-02	4.99E-02

577

578

579

580

581

582 **Table S10:** Parameter estimates and AIC for log body mass divergence models (Uyeda et al.
 583 2011). The parameter λ is a rate parameter; thus its inverse (reported in millions of years) may be
 584 interpreted as the average waiting time in the two pulse models.

Model	σ_p	σ_{BM}^2	σ_D^2	$1/\lambda$	AIC
<i>Burleigh divergence times (n=872 species pairs)</i>					
White noise	0.368				707.8
Brownian motion	0.051	0.010			553.8
Single pulse	0.016		0.025	0.664	-345.0
Multiple pulse	0.042		0.226	17.508	452.2
<i>Jetz divergence times (n=790 species pairs)</i>					
White noise	0.357				614.4
Brownian motion	0.051	0.013			445.9
Single pulse	0.015		0.034	0.592	-190.0
Multiple pulse	0.034		0.152	8.442	350.1

585
 586

587 **Table S11:** Parameter estimates and AIC for beak PC1 divergence models (Uyeda et al. 2011).
 588 The parameter λ is a rate parameter; thus its inverse (reported in millions of years) may be
 589 interpreted as the average waiting time in the two pulse models.

Model	σ_p	σ_{BM}^2	σ_D^2	$1/\lambda$	AIC
<i>Burleigh divergence times (n=926 species pairs)</i>					
White noise	0.277				249.5
Brownian motion	0.036	0.005			106.2
Single pulse	0.012		0.015	0.559	-755.7
Multiple pulse	0.031		0.122	19.524	32.5
<i>Jetz divergence times (n=839 species pairs)</i>					
White noise	0.271				194.4
Brownian motion	0.035	0.007			55.3
Single pulse	0.011		0.020	0.553	-563.9

Multiple pulse

0.021

0.091

8.774

-50.7

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619 **Table S12-S13: Sensitivity analyses of the threshold for the minimum number of eBird**
620 **observations per sister pair** Multi-model inference of predictor effects for breeding range co-
621 occurrence versus absence of such co-occurrence, in bird sister species, with a threshold
622 minimum number of observations of 20 ($n = 802$ species pairs). Generalized linear models were
623 implemented with binomial errors and a logit link. Predictors were: divergence time, mass
624 disparity, latitude, log mean hand-wing index, and either Euclidean distance in beak PC
625 morphospace (SI Table S11) or each of the disparities along the beak PC axes (SI Table S12).
626 Coefficient estimates, unconditional variance, confidence interval and relative importance for all
627 predictors with importance >0 among models with $\Delta AIC < 2$. The number of models with ΔAIC
628 < 2 is in parentheses following model parameters.

629

630 **Table S12: Threshold for minimum observation number equal to 20, and Euclidean beak**
631 **distances (10 models)**

632

Predictor	Estimate	Variance	Conf.	
			interval	Importance
mass disparity	1.19E+00	4.94E-01	1.38E+00	1
DT	9.26E-02	7.59E-04	5.41E-02	1
latitude x migration	4.64E-02	1.81E-04	2.64E-02	1
latitude	2.09E-03	6.82E-05	1.62E-02	1
(Intercept)	-5.91E-01	8.69E-02	5.79E-01	1
migration	-6.27E-01	2.50E-01	9.81E-01	1
DT x mass disparity	-9.24E-02	2.55E-03	9.92E-02	9.34E-01
mass disparity x migration	-1.47E+00	7.72E-01	1.73E+00	9.13E-01
DT x migration	-6.18E-02	1.56E-03	7.76E-02	9.02E-01
Euc beak dist	8.99E-03	7.77E-03	1.73E-01	2.72E-01
Euc beak dist x migration	2.04E-01	1.32E-01	7.14E-01	1.99E-01
mean log HWI	4.01E-03	9.11E-04	5.93E-02	1.90E-01
latitude x mass disparity	4.47E-03	8.83E-05	1.85E-02	1.84E-01
Euc beak dist x DT	-2.79E-03	3.27E-05	1.12E-02	6.65E-02
DT x latitude	2.52E-06	6.09E-09	1.53E-04	6.56E-02

633
634
635
636

Table S13: Threshold for minimum observation number equal to 20, with individual beak PCs as predictors (4 models), see legend above Table S11

Predictor	Estimate	Variance	Conf. interval	Importance
beak PC2 difference x migration	1.75E+00	8.63E-01	1.82E+00	1
mass disparity	1.00E+00	3.82E-01	1.21E+00	1
beak PC2 difference	6.27E-01	3.48E-01	1.16E+00	1
DT	5.52E-02	5.78E-04	4.72E-02	1
latitude x migration	4.55E-02	1.84E-04	2.66E-02	1
latitude	2.47E-03	5.41E-05	1.44E-02	1
(Intercept)	-5.17E-01	4.94E-02	4.36E-01	1
migration	-1.01E+00	2.30E-01	9.42E-01	1
beak PC2 difference x mass disparity	-2.41E+00	9.17E-01	1.88E+00	1
mass disparity x migration	-1.53E+00	8.75E-01	1.84E+00	8.52E-01
DT x migration	-3.62E-02	1.89E-03	8.54E-02	5.04E-01
Beak PC1 difference	4.61E-01	3.61E-01	1.18E+00	4.75E-01

637
638
639
640
641
642
643
644
645
646
647

648 **Table S14-S15: Sensitivity analyses of the threshold for the minimum number of eBird**
649 **observations per sister pair** Multi-model inference of predictor effects for breeding range co-
650 occurrence versus absence of such co-occurrence, in bird sister species, with a threshold
651 minimum number of observations of 50 ($n = 689$ species pairs). Generalized linear models were
652 implemented with binomial errors and a logit link. Predictors were: divergence time, mass
653 disparity, latitude, log mean hand-wing index, and either Euclidean distance in beak PC
654 morphospace (SI Table S13) or each of the disparities along the beak PC axes (SI Table S14).
655 Coefficient estimates, unconditional variance, confidence interval and relative importance for all
656 predictors with importance >0 among models with $\Delta AIC < 2$. The number of models with ΔAIC
657 < 2 is in parentheses following model parameters.

658
659 **Table S14: Threshold for minimum observation number equal to 50, and Euclidean beak**
660 **distances (17 models)**

661

Predictor	Estimate	Variance	Conf.	
			interval	Importance
mass disparity	9.28E-01	9.46E-01	1.91E+00	1
latitude x migration	5.22E-02	2.01E-04	2.79E-02	1
latitude	-1.37E-03	1.19E-04	2.14E-02	1
(Intercept)	-3.23E-01	2.06E-01	8.92E-01	1
migration	-9.23E-01	1.96E-01	8.69E-01	1
mass disparity x migration	-1.99E+00	6.70E-01	1.61E+00	1
DT	4.68E-02	7.86E-04	5.50E-02	9.44E-01
DT x mass disparity	-5.62E-02	3.75E-03	1.20E-01	6.50E-01
mean log HWI	2.69E-02	9.73E-03	1.94E-01	4.14E-01
Euc beak dist	1.96E-02	6.63E-03	1.60E-01	2.32E-01
mean log HWI x mass disparity	5.75E-02	1.67E-02	2.54E-01	1.17E-01
latitude x mass disparity	1.55E-03	1.73E-05	8.17E-03	9.47E-02
Euc beak dist x migration	7.33E-02	2.34E-02	3.00E-01	9.11E-02
DT x latitude	2.76E-05	1.68E-08	2.54E-04	8.39E-02

latitude x mean log HWI	5.91E-04	1.83E-06	2.66E-03	5.83E-02
DT x migration	-6.85E-04	6.34E-06	4.95E-03	4.44E-02

662
663
664
665

Table S15: Threshold for minimum observation number equal to 50, with individual beak PCs as predictors (20 models), see legend above Table S14

Predictor	Estimate	Variance	Conf.	
			interval	Importance
mass disparity	1.00E+00	8.08E-01	1.77E+00	1
latitude x migration	5.22E-02	2.01E-04	2.78E-02	1
DT	4.83E-02	7.07E-04	5.22E-02	1
latitude	-1.53E-03	1.18E-04	2.13E-02	1
(Intercept)	-4.37E-01	2.33E-01	9.47E-01	1
migration	-9.47E-01	1.94E-01	8.65E-01	1
mass disparity x migration	-1.96E+00	6.44E-01	1.58E+00	1
DT x mass disparity	-6.40E-02	4.51E-03	1.32E-01	6.65E-01
mean log HWI	5.94E-02	1.78E-02	2.62E-01	5.98E-01
beak PC1 difference	1.29E-01	1.31E-01	7.11E-01	3.06E-01
beak PC2 difference	2.55E-02	2.19E-02	2.90E-01	2.48E-01
beak PC2 difference x migration	2.23E-01	1.81E-01	8.34E-01	1.33E-01
beak PC1 difference x DT	1.12E-02	5.84E-04	4.75E-02	9.51E-02
latitude x mean log HWI	6.85E-04	2.52E-06	3.12E-03	7.25E-02
mean log HWI x mass disparity	2.06E-02	2.29E-03	9.39E-02	3.87E-02
latitude x mass disparity	6.09E-04	2.67E-06	3.21E-03	3.52E-02
beak PC3 difference x mass disparity	3.51E-02	5.89E-03	1.51E-01	3.23E-02
beak PC3 difference	-1.47E-02	1.05E-03	6.36E-02	3.23E-02
DT x migration	-4.53E-04	2.81E-06	3.29E-03	2.94E-02

666

667 **Table S16-S17: Sensitivity analyses for the scoring of sympatry (sympatry scored for pairs**
668 **when overlap of smaller range is >20%)** Multi-model inference of predictor effects for
669 sympatry versus parapatry of bird sister species with breeding range local co-occurrence ($n = 441$
670 pairs). Generalized linear models were implemented with binomial errors and a logit link.
671 Predictors were: divergence time, mass disparity, latitude, migration status (if either species is
672 migratory or has non-breeding season movements), log mean hand-wing index, and either
673 Euclidean distance in beak PC morphospace (SI Table S15) or each of the species disparities
674 along the beak PC axes (SI Table S16). Coefficient estimates, unconditional variance, confidence
675 interval and importance, for all predictors with importance >0.5 among models with $\Delta AIC < 2$.
676 The number of models with $\Delta AIC < 2$ is in parentheses following the specifications of the four
677 model sets.

678

679 **Table S16: Sympatry at >20%, Euclidean beak distances (7 models)**

Predictor	Estimate	Variance	Conf.	
			interval	Importance
DT	6.32E-01	1.91E-02	2.72E-01	1
(Intercept)	2.78E-01	1.18E-02	2.14E-01	1
mass disparity	1.84E-01	1.69E-02	2.56E-01	1
mean log HWI	1.38E-01	1.23E-02	2.18E-01	1
DT x mean log HWI	-2.73E-01	1.46E-02	2.38E-01	1
Euc beak distance	-4.96E-02	7.51E-03	1.70E-01	3.47E-01
Euc beak distance x DT	4.62E-02	7.03E-03	1.65E-01	2.09E-01
latitude	4.45E-03	2.25E-04	2.95E-02	1.04E-01
migration	8.47E-03	1.04E-03	6.34E-02	1.01E-01
DT x mass disparity	1.70E-03	1.62E-04	2.51E-02	9.58E-02
mean log HWI x mass disparity	2.72E-04	1.27E-04	2.22E-02	9.47E-02

680

681

682

683

684 **Table S17: Sympatry at >20%, and individual beak PCs (10 models), see legend above**
 685 **Table S16**

Predictor	Estimate	Variance	Conf.	
			interval	Importance
DT	6.58E-01	1.99E-02	2.77E-01	1
(Intercept)	2.46E-01	1.31E-02	2.25E-01	1
mean log HWI	1.68E-01	1.32E-02	2.26E-01	1
mass disparity	1.53E-01	1.73E-02	2.58E-01	1
beak PC3 difference	-2.65E-01	1.78E-02	2.62E-01	1
DT x mean log HWI	-3.15E-01	1.53E-02	2.43E-01	1
beak PC2 difference	2.09E-01	2.35E-02	3.01E-01	9.17E-01
beak PC3 difference x DT	1.93E-01	2.33E-02	3.00E-01	8.69E-01
migration	1.05E-02	4.73E-03	1.35E-01	2.52E-01
beak PC2 difference x migration	-9.51E-02	2.74E-02	3.25E-01	2.10E-01
beak PC2 difference x beak PC3 difference	7.54E-03	2.70E-04	3.23E-02	9.51E-02
beak PC2 difference x mean log HWI	-1.32E-02	7.71E-04	5.46E-02	9.34E-02
beak PC3 difference x migration	2.94E-02	3.83E-03	1.22E-01	7.62E-02
beak PC3 difference x mass disparity	5.10E-03	1.79E-04	2.63E-02	5.68E-02
beak PC2 difference x mass disparity	-4.57E-03	1.43E-04	2.35E-02	5.40E-02
beak PC2 difference x DT	1.89E-03	6.64E-05	1.60E-02	4.46E-02
beak PC1 difference	-1.79E-03	7.11E-05	1.66E-02	4.40E-02
mean log HWI x mass disparity	-1.06E-03	4.00E-05	1.24E-02	4.33E-02
latitude	-3.20E-04	2.17E-05	9.16E-03	4.25E-02

687
688
689
690
691

Table S18: The minimum number of observations from each species pair predicts the probability of local co-occurrence, with a threshold minimum number of observations equal to: a) 10 ($n = 849$ pairs); and b) 50 ($n = 689$ pairs).

a.

Predictor	Estimate	Std. Error	z value	p
(Intercept)	-1.8331	.2490	-7.362	1.81e-13
\log_{10} (minimum # observations)	0.9369	.1058	8.859	< 2e-16

692
693

b.

Predictor	Estimate	Std. Error	z value	p
(Intercept)	-2.1119	2.1711E-05	-5.823	5.79e-09
\log_{10} (minimum # observations)	1.0377	0.1433	7.239	4.51e-13

694
695

Supplementary Results and Discussion

Sensitivity analyses: Alternate divergence time estimates

696
697
698 *Local co-occurrence* Using the Jetz et al. divergence time point estimates for species pair
699 analyses (mean divergence times from 100 pseudo-posterior samples), we find remarkably
700 similar estimates for the probability of local co-occurrence and breeding local co-occurrence
701 across divergence times (Extended Data Figs. 2, 3). Analyses using divergence times from the
702 Jetz *et al.* analyses resulted in slightly higher intercepts (any co-occurrence: 0.546, breeding co-
703 occurrence: 0.468) than those using divergence times from the Burleigh *et al.* tree (any co-
704 occurrence: 0.503, breeding co-occurrence: 0.434). The relationship with divergence time had a
705 similarly flat slope (Extended Data Fig. 3).

706 GLMs using 100 bootstraps of the Burleigh tree and 10,000 pseudo-posterior samples from the
707 Jetz *et al.* analyses show limited sensitivity to divergence time estimates within each
708 phylogenetic framework, but suggest that repeating analyses across phylogenies constructed
709 using different methods captures more variation from phylogenetic uncertainty than examining
710 uncertainty within a single phylogenetic analysis (via e.g. bootstraps or sets of trees from a
711 Bayesian posterior, Extended Data Fig. 3).

712

713 *Factors affecting the probability of breeding co-occurrence* In a sensitivity analysis assessing
714 whether alternate divergence time estimates yield similar inferences regarding the role of

715 latitude, migration, divergence in ecological traits, and dispersal in the probability of breeding
716 local co-occurrence, multi-model inferences from GLMs were extremely similar (compare SI
717 Tables S2 and S6, and SI Tables S3 and S7). Multi-model inferences from GLMs where
718 Euclidean beak distance was a potential predictor were extremely similar whether we used
719 divergence time estimates from the Burleigh *et al* or Jetz *et al.* analyses {{646 Jetz,W. 2012;}}.
720 In these model sets, divergence time, mass disparity, latitude, migration, and mean log hand-
721 wing index were all important predictors, as were interactions between latitude and migration,
722 and mass disparity and migration. Each model set had a third interaction with importance $>.6$,
723 which differed in the model sets dependent on the divergence time estimates used (divergence
724 time \times migration in models using our divergence times, divergence time \times mass disparity in
725 models using the Jetz *et al.* estimates). The broad agreement in variable importance from these
726 model sets suggests that this analysis is robust to different approaches to estimating divergence
727 time.

728 In model sets where differences in each beak morphology PC were potential predictors,
729 the main difference between the sensitivity analysis and our main analysis is that the sensitivity
730 analysis with Jetz *et al.* divergence times revealed greater support for the inclusion of beak PC1
731 differences as a main effect (importance = .87). This effect receives less support, although it
732 receives some, in our main analyses (importance = .20). Additionally, the sensitivity analysis did
733 not recover support for one interaction effect (PC2 differences \times migration) that had support in
734 our main analyses. However, there was support in the sensitivity analysis for the main effects of
735 the variables included in this interaction (PC2 differences and migration), and each variable was
736 included in other interaction effects in the sensitivity analysis. Overall, this sensitivity analysis
737 found extremely similar support for the importance of the predictors with support in the main
738 analyses (divergence time, mass disparity, latitude, migration, beak PC2 differences, and the
739 following interactions: mass disparity \times migration, divergence time \times migration, and latitude \times
740 migration), indicating again that our results are robust to approaches to estimating divergence
741 time on phylogenies.

742

743 *Factors affecting the probability of parapatry vs. sympatry* In a sensitivity analysis assessing the
744 impact of using alternate divergence time estimates on GLMs built to discover which factors are
745 associated with the probability of parapatry vs. sympatry for species pairs in contact, our results

746 from the sensitivity analysis were again extremely similar to the main results (compare SI Tables
747 S4 and S8, and SI Tables S5 and S9). In analyses incorporating Euclidean beak distance as a
748 potential predictor, the sensitivity analysis found substantial support for a single interaction
749 effect (divergence time \times mass disparity, importance = 1) that had little support (importance =
750 .09) in the main analyses. However, the sensitivity analysis confirmed support for all the
751 predictors with strong support (importance $>$ 0.6) in the main analyses: divergence time, mass
752 disparity, mean log HWI, and the interaction effect divergence time \times mean log HWI. In model
753 sets where differences in each beak morphology PC were potential predictors, support among
754 potential predictors was again strikingly similar. The main difference between our main analysis
755 and the sensitivity analysis was that the sensitivity analysis found substantial support
756 (importance = 0.93) for the main effect of beak PC2 differences, whereas our main analysis
757 found less support (importance = 0.13) for the main effect of beak PC2 differences. Overall, the
758 sensitivity analysis again confirmed support for the predictors that were found to be important in
759 our main analyses.

760

761 *Simulations of range dynamics* In the main analyses, we used Markov simulations of range
762 dynamics to aid in the interpretation of GLM results that suggested a high probability of co-
763 occurrence at very early stages following initial divergence in bird sister pairs. We sought to
764 estimate the approximate minimum rate of secondary contact establishment required to yield the
765 unexpectedly high intercept we found in our initial analyses of the probability of breeding local
766 co-occurrence with divergence time. Using our divergence time estimates both for the analysis of
767 co-occurrence probability and for simulations, we found an approximate minimum rate of 0.3
768 transitions per million years (see main text). Using the alternate set of divergence time estimates
769 from the Jetz et al. pseudo-posterior for sensitivity analyses, we recover a higher intercept of the
770 relationship between breeding local co-occurrence probability and divergence time (0.468 versus
771 0.434 – see *Local co-occurrence* section above). We thus similarly employed range dynamics
772 simulations to estimate an approximate minimum rate of transition to secondary contact that
773 could yield an intercept = 0.468 for the relationship between probability of breeding local co-
774 occurrence and divergence time, with the alternate divergence time estimate set. We find a
775 higher estimate of the approximate minimum rate of transition to secondary contact, σ : 0.45
776 transitions from isolation to contact per million years. We performed simulations of range

777 dynamics with $\sigma = 0.45$ and $\varepsilon = 0.2 \times \sigma$ or $\varepsilon = 0.5 \times \sigma$ (we recovered intercepts ≥ 0.468 at both
778 these values of ε), and we report the approximate minimum percentage of (nascent) species pairs
779 that come into contact by particular points in time by taking the minimum of the estimates for
780 different values of ε . At a forward transition rate of 0.45 transitions from isolation to contact per
781 million years (and a reverse rate $\varepsilon = 0.225$ transitions from contact to isolation), $\sim 4.2\%$ of species
782 pairs come into contact before 100,000 years following initial divergence, $\sim 34.3\%$ of species
783 pairs come into contact before 1 million years following initial divergence, and 50% of species
784 pairs come into contact before 1.92 million years following initial divergence. These estimates
785 suggest that early contact is even more prevalent than the main results we report, and thus our
786 conclusions with respect to the timescale of the transition to secondary contact, and the
787 importance of traits in determining these outcomes, are robust. We may even be underestimating
788 the importance of these early interactions.

789

790 *Trait divergence models* The body mass and beak divergence analyses indicate that a single-
791 pulse model for body mass divergence had far more relative support than the constrained (time-
792 independent) model, the gradual (Brownian motion) model, and the multiple-pulse model.
793 Importantly, the mean waiting times to body mass and beak morphology divergence pulses for
794 single pulse models are $\sim 665,000$ and $560,000$ years, respectively. In combination with our
795 analysis of the timescale of transition to secondary contact, it is then likely that most species
796 pairs have undergone a pulse of divergence evolution in ecological traits before coming into
797 secondary contact. This result is consistent with the findings of Tobias and colleagues { {687
798 Tobias 2014; } }, who found that ecological trait divergence frequently occurs prior to sympatry
799 in the suboscine bird family Furnariidae, and is of sufficient magnitude that there is little
800 selection for divergence (character displacement) upon the transition to sympatry.

801 In the sensitivity analysis using the alternate divergence time estimates, the single pulse models
802 similarly have the most relative support of the four models by far (smallest ΔAIC for body mass
803 = 489.8, SI Table S10; smallest ΔAIC for beak PC1 = 513.2, SI Table S11). The average waiting
804 times, calculated as $1/\lambda$ (the inverse of the rate parameter λ), are just $\sim 592,000$ and $553,000$
805 years old respectively. Thus, using the alternate divergence times, we again find support for a
806 single pulse model of body mass divergence with a short average waiting time to body mass
807 divergence relative to the timescale of transitions into secondary contact. Indeed, the results of

808 this sensitivity analysis suggest that a pulse of body mass divergence may be a typical
809 component of the speciation process.

810

811 *Sensitivity analyses: minimum observation threshold* We conducted sensitivity analyses for the
812 minimum number of eBird observations for species pairs to be included in local co-occurrence
813 analyses. In the main results, we use a minimum value of 10 for either species in the species pair.
814 As noted in the Methods section of the main text, when this threshold for inclusion of species
815 pairs is applied, the minimum number of observations of a species within species pairs remains a
816 strong predictor of the probability of local co-occurrence in the data set, suggesting that the
817 threshold is low – i.e. conservative with respect to providing evidence for co-occurrence (SI
818 Table S17). Thus, in sensitivity analyses, we implemented minimum observation values of 20
819 and 50. When comparing GLM with divergence time as the only predictor, it is evident that the
820 probability of local co-occurrence increases at lower divergence times with the minimum
821 observation threshold. The intercept of the relationship, for example, is progressively higher with
822 increasing minimum observation values (compare Fig. 1a and 1b with Extended Data Fig. 7).
823 Meanwhile, the slope of the relationship decreases, such that there is less variation in the
824 probability of local co-occurrence with divergence time (compare Fig. 1a and 1b with Extended
825 Data Fig. 7).

826 Higher minimum values likely exclude species outside of North America, especially
827 tropical species, and more rare species. Thus, the differences in our results in these sensitivity
828 analyses of the threshold minimum number of observations may be relevant with respect to
829 sampling bias, or to real biological differences in local co-occurrence probabilities (e.g. with
830 latitude). However, we note that at a threshold of 10 observations, we are much more likely to be
831 underestimating than overestimating the probability of local co-occurrence early in the speciation
832 process. Thus, our main conclusions are robust to the minimum observation threshold.
833 Furthermore, the results suggest that it is possible that as observational evidence accrues, we will
834 find that the probability of local co-occurrence is generally higher than we have estimated at low
835 divergence times.

836 *Sensitivity analysis: minimum threshold for sympatry* We performed sensitivity analyses as a
837 check on the threshold for minimum overlap for ranges to be considered sympatric. We
838 distinguish sympatric from parapatric ranges to analyze the transition to sympatry using GLMs.

839 In the main text, we used a threshold of 10% overlap of the smaller species range, and in
840 sensitivity analyses we use a threshold of 20%^{751 Pigot 2015;}. For multi-model inferences
841 that incorporated beak morphology differences by including Euclidean distance in beak
842 morphology PC space as a possible predictor, the results were nearly identical for the two
843 thresholds (compare SI Tables S4 and S15). All predictors with importance > 0.6 in our main
844 analysis received similar support in the sensitivity analysis model set. For multi-model
845 inferences that incorporated beak morphology differences by allowing differences along each of
846 the three beak morphology principal components to be predictors, the model sets were similar in
847 finding support for effects of divergence time, mass disparity, the mean of the log hand-wing
848 index, and an interaction between divergence time and the mean of the log hand-wing index.
849 However, support for beak morphology differences as predictors varied between these model
850 sets (compare SI Tables S5 and S16). In the main analyses, PC1 and an interaction between PC1
851 and divergence time had support as predictors (importance = 1 for both), whereas these
852 predictors had less support in the sensitivity analyses (importance of beak PC1 difference = .044,
853 importance of the beak PC1 difference × divergence time = 0). Conversely, there was support for
854 beak PC2 (importance = .92) and PC3 differences (importance = .10) and the beak PC3
855 difference × divergence time interaction (importance = .87) in the sensitivity analysis, whereas in
856 the main analysis there was little support for these three predictors (importance scores .13, .17,
857 and 0 respectively). It thus appears that we find some limited support for some aspects of beak
858 divergence as predictors of sympatry in these models, but the sensitivity analysis indicates that it
859 is unclear which aspect of beak divergence may be most important.

860

861 **Supplementary References**

- 862 1. Chesson, P. General theory of competitive coexistence in spatially-varying environments.
863 *Theor. Popul. Biol.* 58, 211-237 (2000).
- 864 2. Burleigh, J. G., Kimball, R. T. & Braun, E. L. Building the avian tree of life using a large-
865 scale, sparse supermatrix. *Mol. Phylogenet. Evol.* **84**, 53-63 (2015).
- 866 3. BirdLife International and NatureServe. in *Bird species distribution maps of the world*
867 (BirdLife International and NatureServe, Cambridge, UK and Arlington, USA, 2014).
- 868 4. *Handbook of the Birds of the World Alive* (eds del Hoyo, J., Elliott, A., Sargatal, J., Christie,
869 D. A. & de Juana, E.) (Lynx Edicions, Barcelona, 2015).
- 870 5. Fritz, S. A. & Purvis, A. Selectivity in mammalian extinction risk and threat types: a new
871 measure of phylogenetic signal strength in binary traits. *Conserv. Biol.* **24**, 1042-1051 (2010).
- 872 6. Pigot, A. L. & Tobias, J. A. Species interactions constrain geographic range expansion over
873 evolutionary time. *Ecol. Lett.* **16**, 330-338 (2013).

- 874 7. Pigot, A. L. & Tobias, J. A. Dispersal and the transition to sympatry in vertebrates.
875 *Proceedings of the Royal Society of London B: Biological Sciences* **282**, 20141929 (2015).
- 876 8. Weir, J. T. & Price, T. D. Limits to Speciation Inferred from Times to Secondary Sympatry
877 and Ages of Hybridizing Species along a Latitudinal Gradient. *Am. Nat.* **177**, 462-469 (2011).
- 878 9. <http://www.birds.cornell.edu/clementschecklist/>. Accessed September 9, 2016.
- 879 10. Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. The global diversity of
880 birds in space and time. *Nature* **491**, 444-448 (2012).
- 881 11. Gatesy, J., Baker, R. & Hayashi, C. Inconsistencies in arguments for the supertree approach:
882 Supermatrices versus supertrees of Crocodylia. *Syst. Biol.* **53**, 342-355 (2004).
- 883 12. Hackett, S. J. *et al.* A phylogenomic study of birds reveals their evolutionary history. *Science*
884 **320**, 1763-1768 (2008).
- 885 13. Sanderson, M. J. R8s: Inferring Absolute Rates of Molecular Evolution and Divergence
886 Times in the Absence of a Molecular Clock. *Bioinformatics* **19**, 301-302 (2003).
- 887 14. Drummond, A. J. & Rambaut, A. BEAST: Bayesian Evolutionary Analysis by Sampling
888 Trees. *BMC Evolutionary Biology* **7** (2007).
- 889 15. Calder, W. A. Ecological Scaling: Mammals and Birds. *Annu. Rev. Ecol. Syst.* **14**, 213-230
890 (1983).
- 891 16. Peters, R. H. in *The ecological implications of body size* (Cambridge University Press, 1986).
- 892 17. Wilson, D. S. The Adequacy of Body Size as a Niche Difference. *Am. Nat.* **109**, 769-784
893 (1975).
- 894 18. Streelman, J. T. & Danley, P. D. The stages of vertebrate evolutionary radiation. *Trends in*
895 *Ecology & Evolution* **18**, 126-131 (2003).
- 896 19. Huxley, J. Evolution. The Modern Synthesis. *Evolution. The Modern Synthesis.* (1942).
- 897 20. Abbott, I., Abbott, L. K. & Grant, P. R. Comparative Ecology of Galapagos Ground Finches
898 (Geospiza-Gould) - Evaluation of Importance of Floristic Diversity and Interspecific
899 Competition. *Ecol. Monogr.* **47**, 151-184 (1977).
- 900 21. Schluter, D. in *The ecology of adaptive radiation* (Oxford University Press, Oxford, UK,
901 2000).
- 902 22. Dunning, J. B. in *Body masses of birds of the world* (Taylor and Francis Group, Boca Raton,
903 Florida, 2008).
- 904 23. Hutchinson, G. E. Homage to Santa Rosalia or why are there so many kinds of animals? *Am.*
905 *Nat.* **93**, 145-159 (1959).
- 906 24. Schoener, T. W. The Evolution of Bill Size Differences among Sympatric Congeneric
907 Species of Birds. *Evolution* **19**, 189-213 (1965).
- 908 25. Grant, P. R. Bill size, body size, and the ecological adaptations of bird species to competitive
909 situations on islands. *Syst. Zool.* **17**, 319-338 (1968).
- 910 26. Revell, L. J. Size-Correction and Principal Components for Interspecific Comparative
911 Studies. *Evolution* **63**, 3258-3268 (2009).
- 912 27. Weeks, B. C. & Claramunt, S. Dispersal has inhibited avian diversification in Australasian
913 archipelagoes. *Proc. Biol. Sci.* **281**, 20141257 (2014).
- 914 28. Claramunt, S., Derryberry, E. P., Remsen, J. V., Jr. & Brumfield, R. T. High dispersal ability
915 inhibits speciation in a continental radiation of passerine birds. *Proceedings of the Royal Society*
916 *B-Biological Sciences* **279**, 1567-1574 (2012).
- 917 29. Grant, P. R. & Grant, B. R. in *40 years of evolution: Darwin's finches on Daphne Major*
918 *island* (Princeton University Press, 2014).

- 919 30. Case, T. J., Holt, R. D., McPeck, M. A. & Keitt, T. H. The community context of species'
920 borders: ecological and evolutionary perspectives. *Oikos* **108**, 28-46 (2005).
- 921 31. Lockwood, J. L., Cassey, P. & Blackburn, T. The role of propagule pressure in explaining
922 species invasions. *Trends in Ecology & Evolution* **20**, 223-228 (2005).
- 923 32. Stephens, P. A., Sutherland, W. J. & Freckleton, R. P. What is the Allee effect? *Oikos* **87**,
924 185-190 (1999).
- 925 33. Jönsson, K. A. *et al.* Tracking Animal Dispersal: From Individual Movement to Community
926 Assembly and Global Range Dynamics. *Trends in Ecology & Evolution* (2016).
- 927 34. Koenig, W. D., Van Vuren, D. & Hooge, P. N. Detectability, philopatry, and the distribution
928 of dispersal distances in vertebrates. *Trends in Ecology & Evolution* **11**, 514-517 (1996).
- 929 35. Dawideit, B. A., Phillimore, A. B., Laube, I., Leisler, B. & Böhning-Gaese, K.
930 Ecomorphological predictors of natal dispersal distances in birds. *J. Anim. Ecol.* **78**, 388-395
931 (2009).
- 932 36. Burney, C. W. & Brumfield, R. T. Ecology predicts levels of genetic differentiation in
933 Neotropical birds. *Am. Nat.* **174**, 358-368 (2009).
- 934 37. Bolker, B. M. *et al.* Generalized linear mixed models: a practical guide for ecology and
935 evolution. *Trends in ecology & evolution* **24**, 127-135 (2009).
- 936 38. <http://CRAN.R-project.org/package=glmulti>.
- 937 39. Venables, W. N. & Ripley, B. D. in *Modern applied statistics with S-PLUS* (Springer, New
938 York, 1997).
- 939 40. Grafen, A. & Hails, R. in *Modern statistics for the life sciences* (Oxford University Press,
940 Oxford, 2002).
- 941 41. Freckleton, R. P. On the misuse of residuals in ecology: regression of residuals vs. multiple
942 regression. *Journal of Animal Ecology* **71.3**, 542-543-545 (2002).
- 943 42. Barraclough, T. G. & Vogler, A. P. Detecting the geographical pattern of speciation from
944 species-level phylogenies. *Am. Nat.* **155**, 419-434 (2000).
- 945 43. Rabosky, D. L. No substitute for real data: A cautionary note on the use of phylogenies from
946 birth?death polytomy resolvers for downstream comparative analyses. *Evolution* **69**, 3207-3216
947 (2015).
- 948 44. Edwards, S. & Beerli, P. Perspective: gene divergence, population divergence, and the
949 variance in coalescence time in phylogeographic studies. *Evolution* **54**, 1839-1854 (2000).
- 950 45. Tobias, J. A. *et al.* Species coexistence and the dynamics of phenotypic evolution in adaptive
951 radiation. *Nature* **506**, 359-+ (2014).

952

953 **Supplementary Code**

954 **1. R8s block for time calibration of Burleigh et al. (2015) maximum**

955 **likelihood phylogeny**

956 Begin r8s;

957 BLFORMAT lengths=persite nsites=32383 ultrametric=no;

958 collapse;

959 MRCA clade1 Casuariiformes_Dromaiidae_Dromaius_novaehollandiae
960 Casuariiformes_Casuariidae_Casuarius_casuarius;
961 MRCA clade2 Anseriformes_Anatidae_Amazonetta_brasiliensis
962 Anseriformes_Anseranatidae_Anseranas_semipalmata;
963 MRCA clade3 Galliformes_Phasianidae_Lagopus_lagopus
964 Galliformes_Odontophoridae_Callipepla_gambelii;
965 MRCA clade5 Podicipediformes_Podicipedidae_Aechmophorus_occidentalis
966 Phoenicopteriformes_Phoenicopteridae_Phoenicopus_chilensis;
967 MRCA clade6 Steatornithiformes_Steatornithidae_Steatornis_caripensis
968 Nyctibiiformes_Nyctibiidae_Nyctibius_leucopterus;
969 MRCA clade7 Apodiformes_Apodidae_Apus_apus
970 Apodiformes_Hemiprocnidae_Hemiproctne_comata;
971 MRCA clade8 Gruiformes_Gruidae_Grus_grus
972 Gruiformes_Rallidae_Rallus_limicola;
973 MRCA clade9 Charadriiformes_Alcidae_Alle_alle
974 Charadriiformes_Laridae_Sterna_forsteri;
975 MRCA clade10 Phaethontiformes_Phaethontidae_Phaethon_aethereus
976 Podicipediformes_Podicipedidae_Podiceps_auritus;
977 MRCA clade11 Suliformes_Fregatidae_Fregata_magnificens
978 Suliformes_Sulidae_Sula_granti;
979 MRCA clade12 Suliformes_Phalacrocoracidae_Phalacrocorax_pelagicus
980 Suliformes_Anhingidae_Anhinga_anhinga;
981 MRCA clade13 Sphenisciformes_Spheniscidae_Eudyptes_pachyrhynchus
982 Procellariiformes_Pelecanoididae_Pelecanoides_urinatrix;
983 MRCA clade14 Sphenisciformes_Spheniscidae_Spheniscus_mendiculus
984 Sphenisciformes_Spheniscidae_Eudyptula_minor;

985 MRCA clade15 Pelecaniformes_Threskiornithidae_Platalea_ajaja
986 Pelecaniformes_Ardeidae_Botaurus_lentiginosus;
987 MRCA clade16 Galbuliformes_Galbulidae_Galbula_ruficauda
988 Piciformes_Picidae_Celeus_flavescens;
989 MRCA clade17 Coraciiformes_Coraciidae_Eurystomus_orientalis
990 Coraciiformes_Meropidae_Merops_orientalis;
991 MRCA clade18 Coraciiformes_Todidae_Todus_multicolor
992 Coraciiformes_Alcedinidae_Alcedo_meninting;
993 MRCA clade19 Falconiformes_Falconidae_Falco_tinnunculus
994 Falconiformes_Falconidae_Polihierax_semitorquatus;
995 MRCA clade20 Phaethontiformes_Phaethontidae_Phaethon_aethereus
996 Passeriformes-oscines_Thraupidae_Hemispingus_atropileus;
997 MRCA clade21 Passeriformes-oscines_Thraupidae_Hemispingus_atropileus
998 Passeriformes-suboscines_Pipridae_Corapipo_altera;
999 MRCA root Tinamiformes_Tinamidae_Crypturellus_undulatus Passeriformes-
1000 suboscines_Thamnophilidae_Terenura_humeralis;
1001 constrain taxon=clade1 min_age=24.5 max_age=58.7;
1002 constrain taxon=clade2 min_age=53.5 max_age=86.3;
1003 constrain taxon=clade3 min_age=24 max_age=51.81;
1004 constrain taxon=clade5 min_age=46.6 max_age=61.6;
1005 constrain taxon=clade6 min_age=51.81 max_age=66.5;
1006 constrain taxon=clade7 min_age=51 max_age=66.5;
1007 constrain taxon=clade8 min_age=53.9 max_age=66.5;
1008 constrain taxon=clade9 min_age=20.44 max_age=47.8;
1009 constrain taxon=clade10 min_age=56 max_age=72.1;
1010 constrain taxon=clade11 min_age=51.81 max_age=66.5;
1011 constrain taxon=clade12 min_age=24.52 max_age=51.81;

```
1012 constrain taxon=clade13 min_age=60.5 max_age=72.1;
1013 constrain taxon=clade14 min_age=9 max_age=27;
1014 constrain taxon=clade15 min_age=53.9 max_age=66.5;
1015 constrain taxon=clade16 min_age=31 max_age=58.5;
1016 constrain taxon=clade17 min_age=51.81 max_age=66.5;
1017 constrain taxon=clade18 min_age=31 max_age=55;
1018 constrain taxon=clade19 min_age=16 max_age=47;
1019 constrain taxon=clade20 min_age=53.5 max_age=66.5;
1020 constrain taxon=clade21 min_age=26 max_age=55;
1021 constrain taxon=root max_age=110;
1022 DIVTIME method=PL algorithm=tn;
1023 set penalty=add smoothing=10 minDurfactor=0.00001;
1024 describe plot=chrono_description;
1025 end;
```

1026

1027 2. A custom script to measure breeding range overlap for a set of sister species
1028 pairs (where the first species of each pair is in the vector data\$Sp1, and the
1029 second species of each pair is in vector data\$Sp2) based on
1030 NatureServe/IUCN range map polygons, run using R (3.2.2) and a set of
1031 libraries for spatial statistics. The script outputs a vector of overlap values
1032 where the overlap value is the fraction of the smaller range overlapped by
1033 the larger range.

```
1034 library(sp)
1035 library(raster)
1036 library(rgdal)
1037 library(rgeos)
1038 library(maptools)
1039 for (i in 1:length(data$Sp1)){
1040   sh1<-readOGR(dsn=".", layer=data$Sp1[i])
```

```

1041 r<-sh1@data$SEASONAL==1 | sh1@data$SEASONAL==2
1042 sh1_br<-sh1[,]
1043 a<-sh1_br@data$Shape_Area
1044 a<-sum(a)
1045 sh1_br<-unionSpatialPolygons(sh1_br, IDs =
1046     seq(1,length(sh1_br@polygons)))
1047 sh2<-readOGR(dsn=".", layer=data$Sp2[i])
1048 q<-sh2@data$SEASONAL==1 | sh2@data$SEASONAL==2
1049 sh2_br<-sh2[,]
1050 b<-sh2_br@data$Shape_Area
1051 b<-sum(b)
1052 sh2_br<-unionSpatialPolygons(sh2_br, IDs =
1053     seq(1,length(sh2_br@polygons)))
1054 c<-min(a,b)
1055 int<-intersect(sh1_br, sh2_br)
1056 if (is.null(int)) {
1057     prop_symp[i]<-0
1058 } else {
1059     d<-c()
1060     for (j in 1:length(int@polygons)){
1061         e<-c()
1062         for (k in 1:length(int@polygons[[j]]@Polygons)){
1063             e<-c(e,int@polygons[[j]]@Polygons[[k]]@area)
1064         }
1065         d<-c(d,e)
1066     }
1067     prop_symp[i]<-sum(d)/c
1068 }
1069 }
1070
1071

```