Supplementary Information Appendix

for the article:

Maria H. Hällfors, Laura H. Antão, Malcolm Itter, Aleksi Lehikoinen, Tanja Lindholm , Tomas Roslin, Marjo Saastamoinen: Shifts in the timing and duration of breeding for 73 boreal bird species over four decades. *Proceedings of the National Academy of Sciences of the United States of America* www.pnas.org/cgi/doi/10.1073/pnas.1913579117 (2020)

Figures S1- S10

Table S1

Texts S1-S4

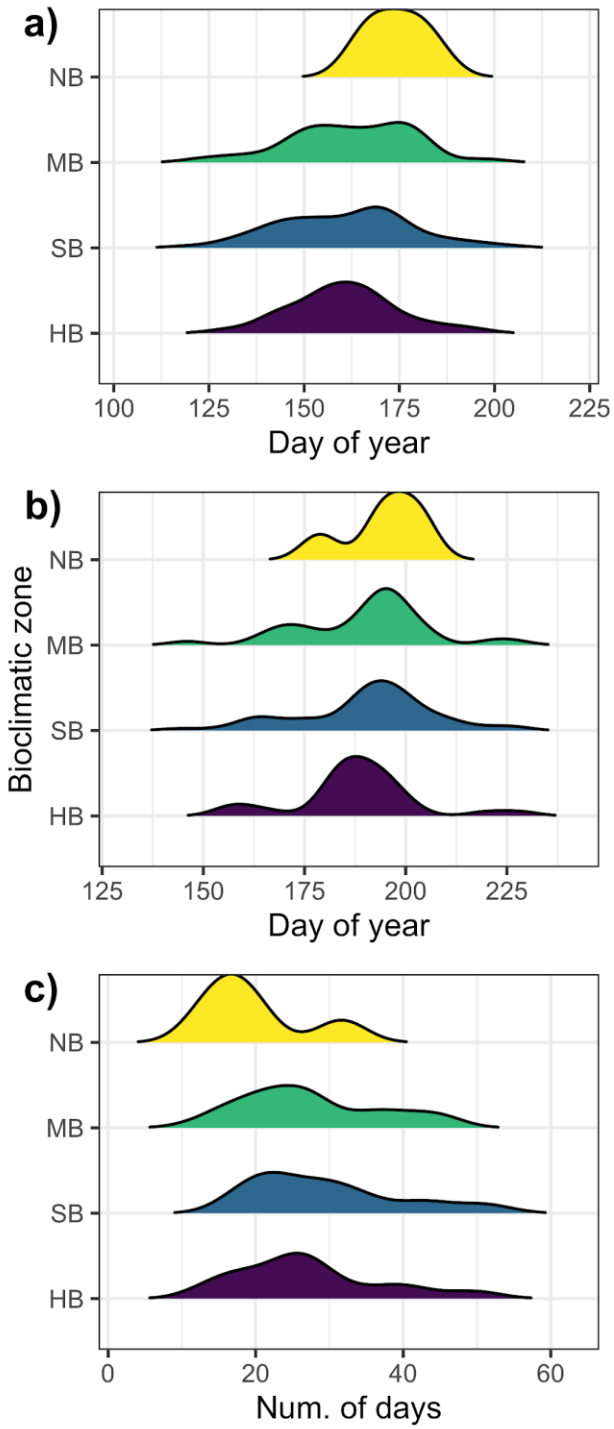

Figure S1. Distribution of species-level posterior mean estimates of the average beginning (a), end (b), and duration (c) of the breeding period by bioclimatic zone. HB= Hemiboreal; SB= South Boreal; MB= Middle Boreal; NB= North Boreal. Note the different scales on the x-axis: metrics for beginning and end of breeding period are in day of year, while duration is in absolute number of days.

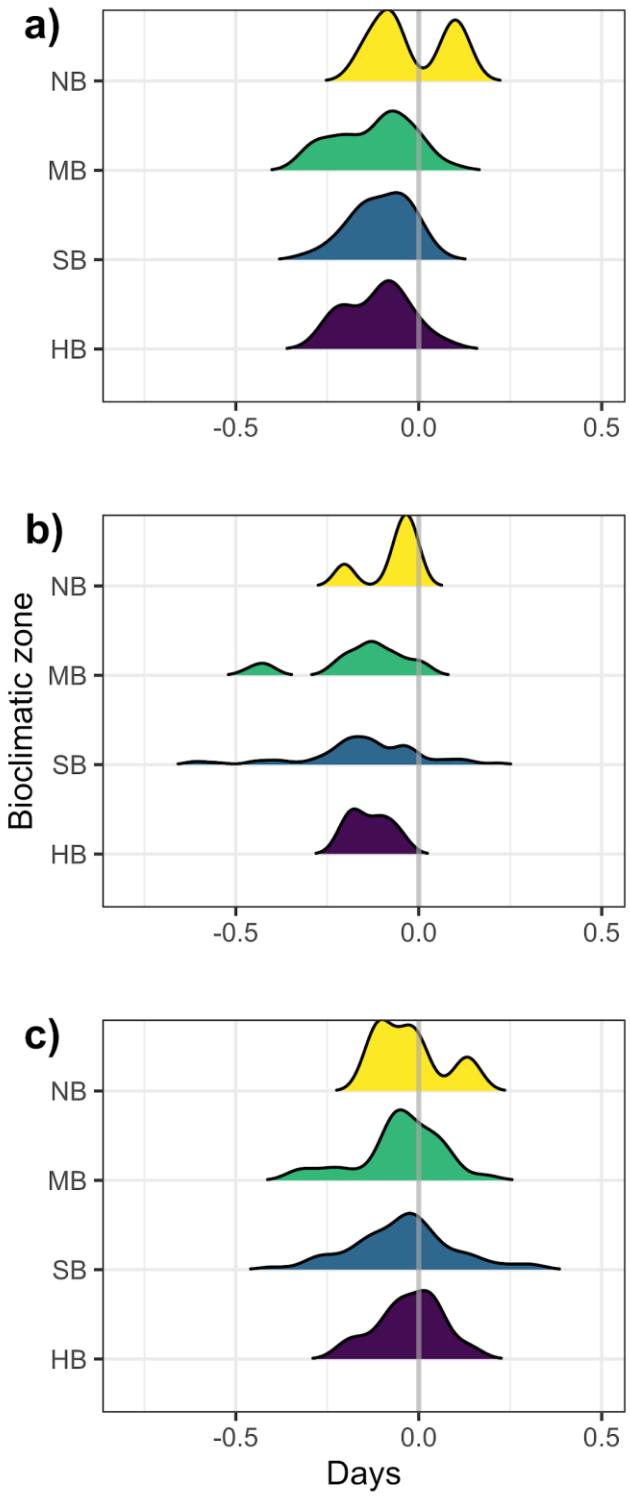

Figure S2. Distribution of species-level posterior mean annual shifts in the beginning (a), end (b), and duration (c) of the breeding period by bioclimatic zone. HB= Hemiboreal; SB= South Boreal; MB= Middle Boreal; NB= North Boreal.

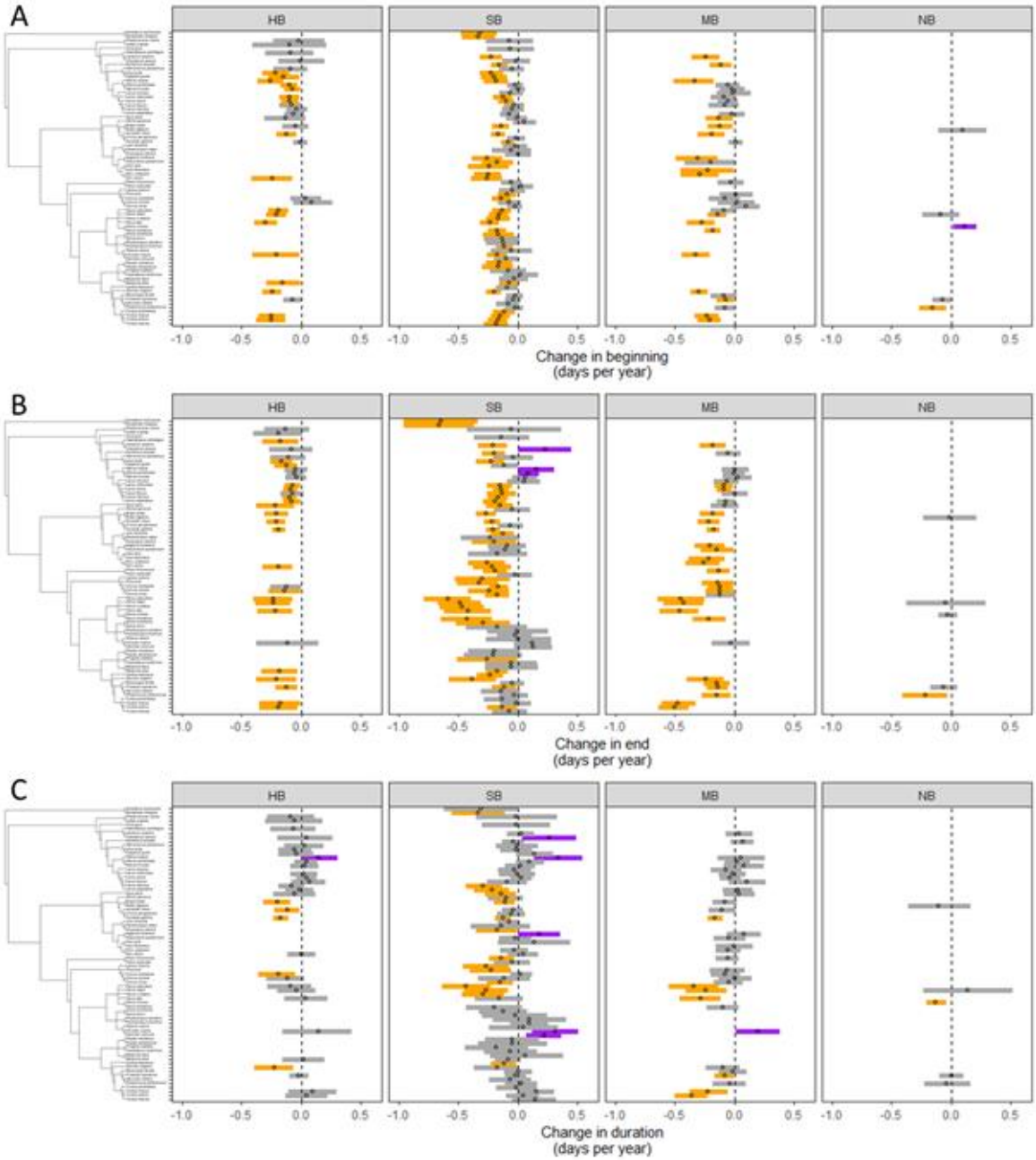

Figure S3. Summary of posterior distributions of species-specific shifts in the (a) beginning, (b) end, and (c) duration of the breeding period within each bioclimatic zone - HB= Hemiboreal; SB= South Boreal; MB= Middle Boreal; NB= North Boreal. Points represent the posterior mean with lines indicating with lower and upper 2.5 quantiles. Species with evidence of advancing or contracting their breeding periods are highlighted in orange, and species delaying or expanding are highlighted in purple. Species are ordered by their phylogeny (left panel) .

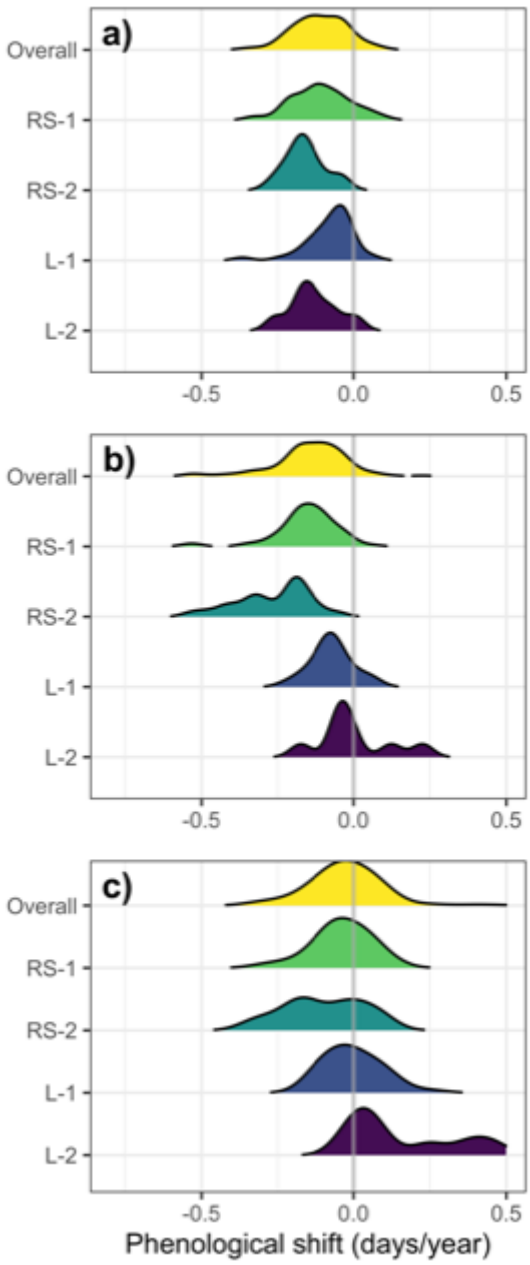

Figure S4. Summary of lower-bound re-analysis shifts in species-specific breeding periods over the study period (1975-2017). Distribution of species-specific posterior mean annual shifts in the beginning (a), end (b), and duration (c) of the breeding period based on re-analysis applying the lower-bound of 95 percent bootstrapped confidence intervals for each phenological metric. Species-specific shifts are summarized overall and by trait group (RS: resident/short-distance migrant; L: long-distance migrant; 1: one brood per season; 2: two or more broods per season). Negative values indicate an advancement of the beginning and end of breeding or a contraction of the breeding period, while positive values indicate the opposite.

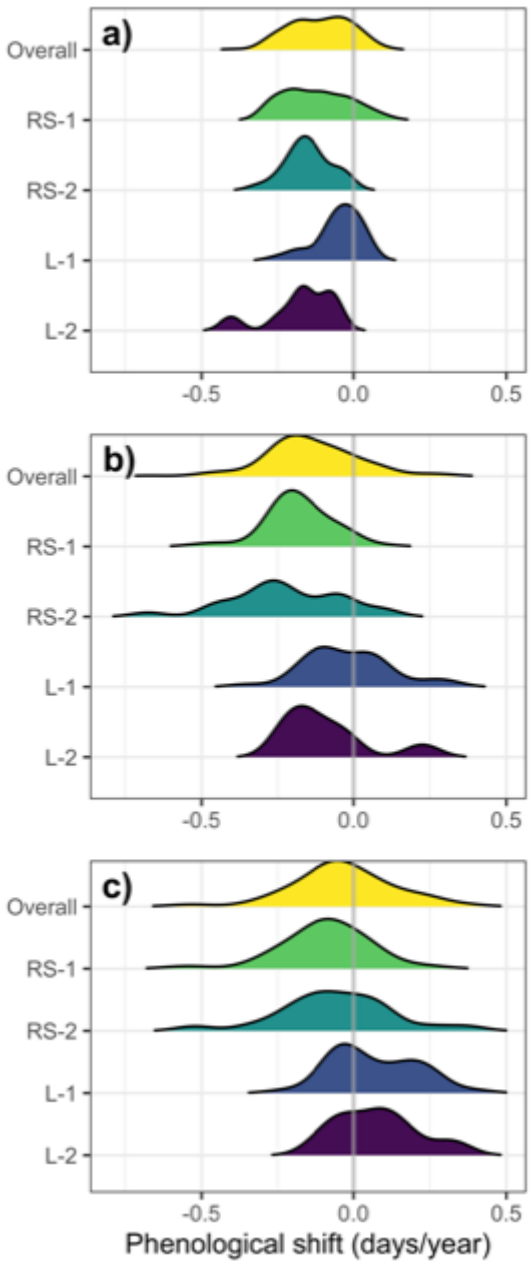

Figure S5. Summary of upper-bound re-analysis shifts in species-specific breeding periods over the study period (1975-2017). Distribution of species-specific posterior mean annual shifts in the beginning (a), end (b), and duration (c) of the breeding period based on re-analysis applying the upper-bound of 95 percent bootstrapped confidence intervals for each phenological metric. Species-specific shifts are summarized overall and by trait group (RS: resident/short-distance migrant; L: long-distance migrant; 1: one brood per season; 2: two or more broods per season). Negative values indicate an advancement of the beginning and end of breeding or a contraction of the breeding period, while positive values indicate the opposite.

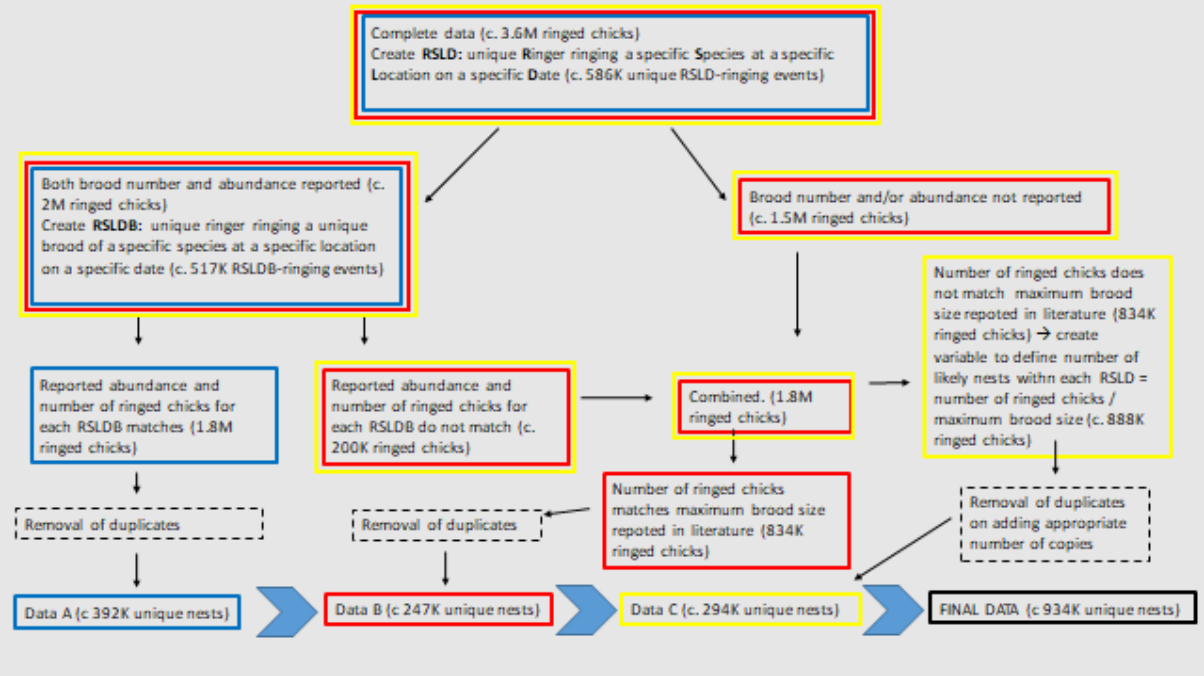

Figure S6. Workflow for defining unique 934 087 ringing events per nest to obtain the complete data on 166 species (See SI Appendix Text S1 for description). For the purposes of this analysis, we additionally only included records after 1975 and species with enough observations to confidently estimate the phenology metrics and their shifts over time (see *Species inclusion data* in Methods in the main manuscript).

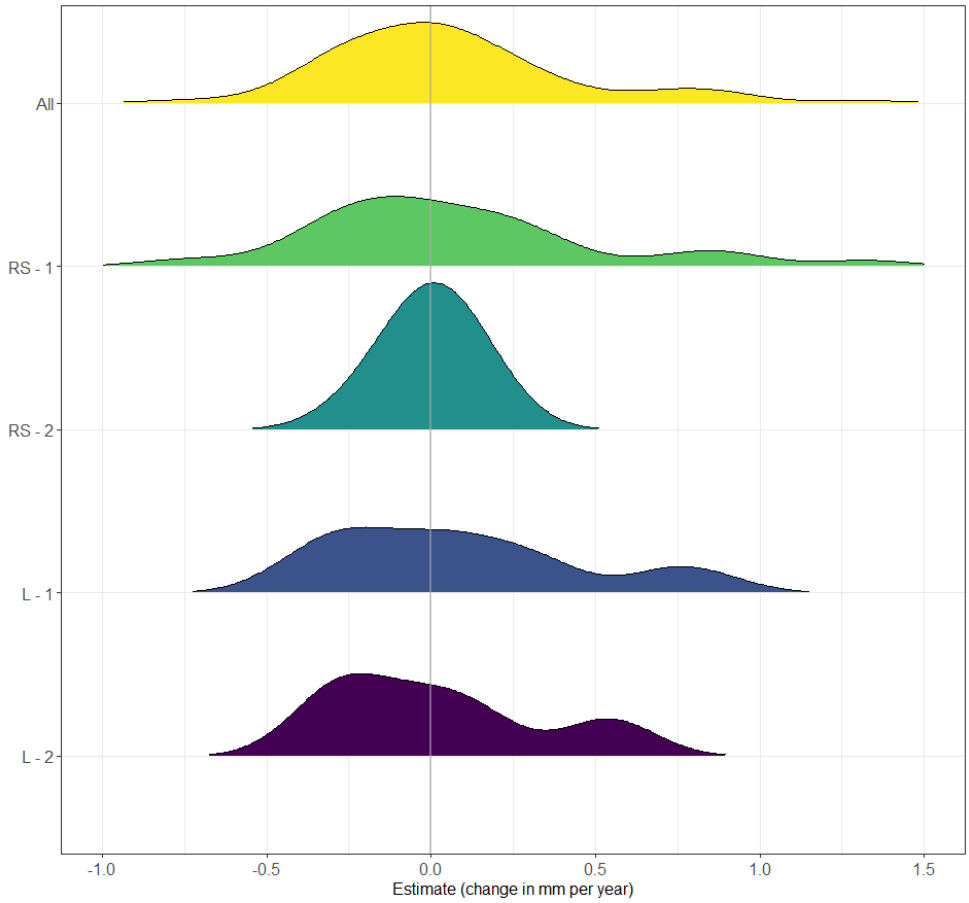

Figure S7. Distribution of estimated changes in wing length (mm) over time for all 73 analyzed species and per trait group (represented by the following number of species: RS-1= 36 species; RS-2= 12 species; L-1=20 species; L-2=5 species). While 52 of the 73 species analyzed showed a significant change in wing length over time based on the available measurements, these changes occurred in both directions and were mostly within a range of +-0.5mm/year, which would be within the range of measurement error. Thus, we conclude that there was no significant evidence to suggest changes in ringer behavior over time.

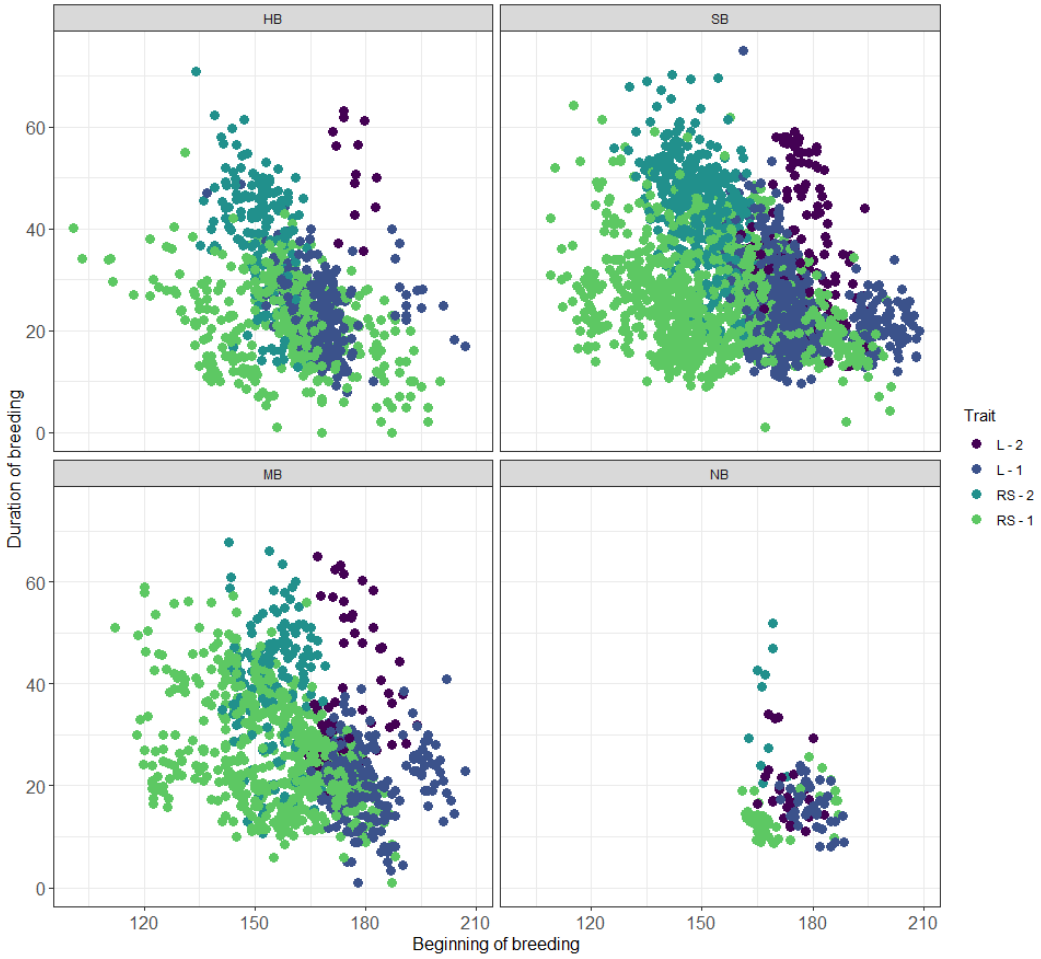

Figure S8. Breeding period duration (number of days) plotted as a function of the beginning of breeding (day of year), per trait group and bioclimatic zone. Each data point represents a unique specie-year combination in a specific bioclimatic zone (HB= Hemi-boreal zone; SB= South Boreal zone; MB= Middle boreal zone; NB= North boreal zone), and colors correspond to trait combinations: green= RS-1 (Resident and short-distance migrant species with one brood); deep turquoise= RS-2 (Resident and short-distance migrant species with two broods); blue= L-1 (Long-distance migrant species with one brood); purple= L-2 (Long-distance migrant species with two broods). Resident and short-distance migrants tend to be ringed earlier in the season compared to long-distance migrants. One-brooded species tend to be ringed for a shorter time during the season (smaller spread along the y-axis) compared to multi-brooded species. The trait groups were represented by the following number of species: RS-1= 36 species; RS-2= 12 species; L-1=20 species; L-2=5 species).

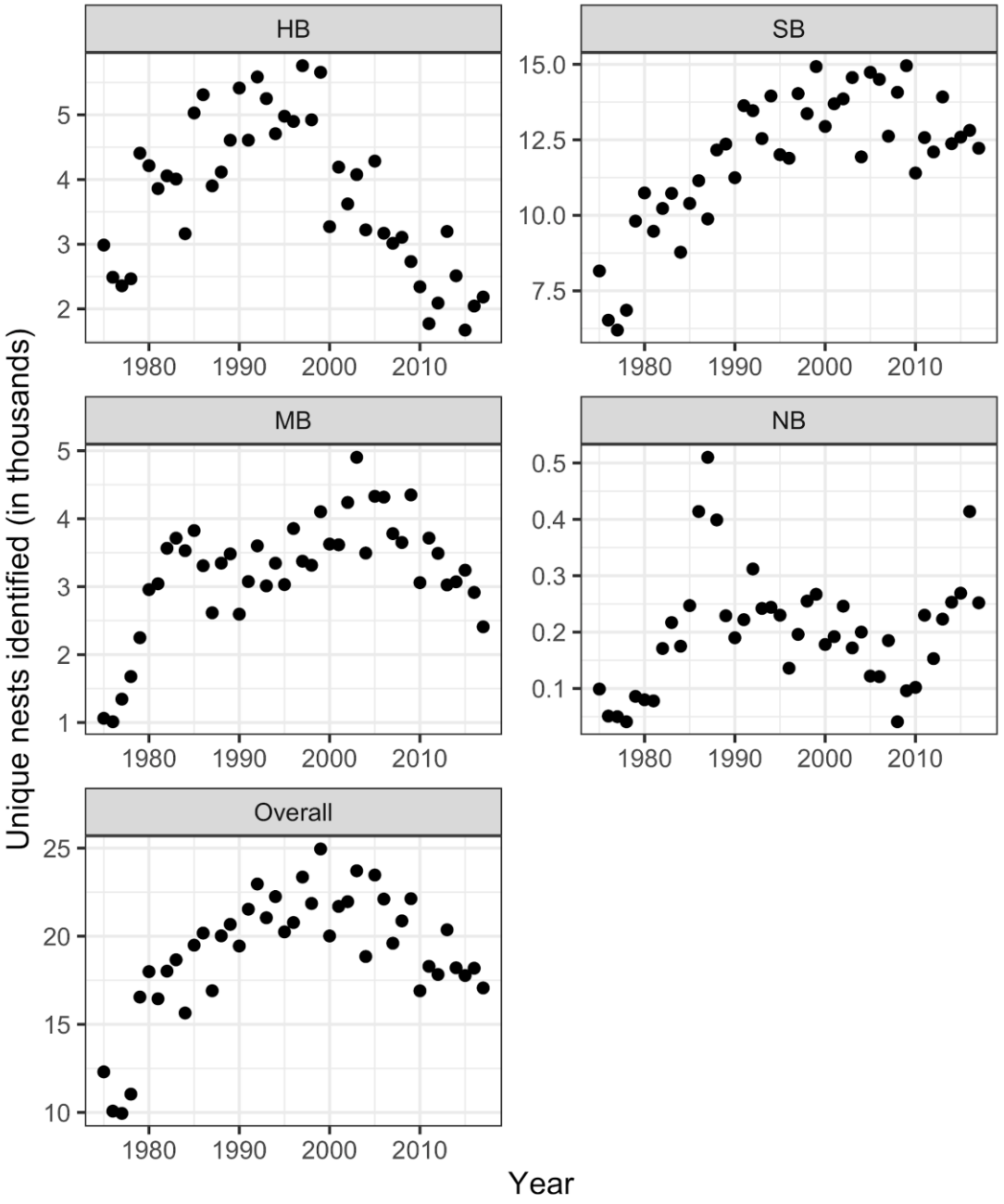

Figure S9. Total number of nest records per year for each bioclimatic zone and overall over the study period (1975-2017).

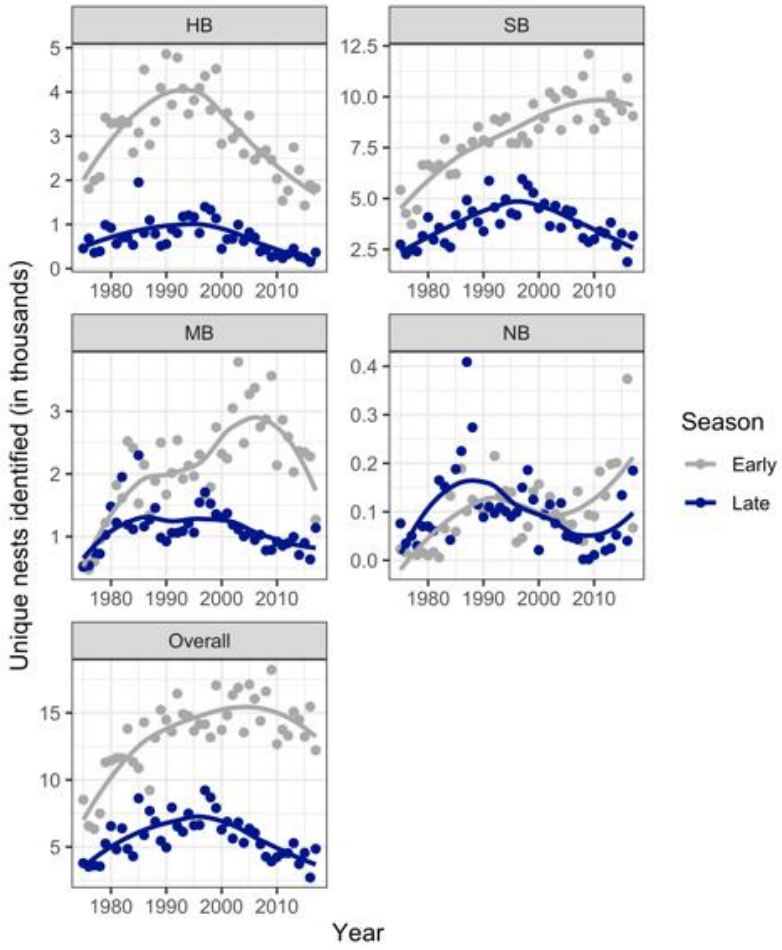

Figure S10. Temporal distribution of the nest records per year comparing the early and late seasons over the study period. In our data set, there are more samples in the beginning of the breeding season than in the end - a pattern common to in most phenological datasets. While there is the potential for this difference in sampling to cause a bias in our results, this should only occur under two conditions: 1) if there are insufficient samples late in the breeding season to accurately identify the end of the breeding; 2) if the relative efforts in the beginning and end of the breeding season change over the study period, such that estimated shifts are potentially confounded with changes in effort. Neither of these conditions are met in our case. Specifically, the number of nests observed in the early beginning (March through June) and late (July - September) part of the breeding season is always in the hundreds to thousands (almost always in the thousands except in the northern boreal zone where sampling efforts are more limited). Further, while there is variation in the number of nests observed in each year over the study period, the proportions of nests observed in the beginning and end of the breeding season are relatively constant over the study period. As such, there are not strong differences in effort levels between the beginning and end of the breeding season across the study period.

Table S1. Summary of the re-analysis of among-species mean breeding period shifts over the study period (1975-2017). Among-species posterior mean estimates are provided overall and by trait group with 95 percent credible intervals in parentheses for re-analyses using lower- and upper-bound of 95 percent bootstrapped confidence intervals of phenological metrics (RS: resident/short-distance migrant; L: long-distance migrant; 1: one brood per season; 2: two or more broods per season). Among-species means for which the credible interval does not include zero (evidence of a breeding period shift) are identified in bold face.

Group	**Among species mean**	
	Lower-bound re-analyis	Upper-bound re-analysis
First		
Overall	**-4.8 (-5.7, -3.8)**	**-4.8 (-5.8, -3.8)**
RS-1	**-4.9 (-6, -3.9)**	**-5 (-6.1, -3.9)**
RS-2	**-5.4 (-6.8, -4)**	**-4.7 (-6.2, -3.2)**
L-1	**-4.2 (-5.3, -3.1)**	**-4.5 (-5.7, -3.3)**
L-2	**-4.2 (-8.3, -0.3)**	-3.7 (-7.8, 0.4)
Last		
Overall	**-6.1 (-7, -5.3)**	**-6.5 (-7.5, -5.6)**
RS-1	**-6.4 (-7.3, -5.5)**	**-7.5 (-8.6, -6.4)**
RS-2	**-7.8 (-9.6, -5.9)**	**-6.7 (-9, -4.4)**
L-1	**-5 (-6, -3.9)**	**-5.3 (-6.6, -4)**
L-2	**-4.8 (-8.9, -0.4)**	-1 (-7.2, 5.5)
Duration		
Overall	**-1.2 (-2, -0.5)**	**-1.5 (-2.5, -0.5)**
RS-1	**-1.2 (-2.1, -0.2)**	**-2.3 (-3.4, -1.1)**
RS-2	**-2.7 (-4.7, -0.7)**	-1.4 (-3.8, 0.9)
L-1	-0.8 (-1.9, 0.3)	-0.5 (-2, 0.9)
L-2	0.6 (-4.3, 5.7)	3.1 (-3.1, 9.3)

Text S1. Processing of raw data

In order to identify single ringing events per nest and day-of-year for each species, we took a number of processing steps with the raw data (also see SI Appendix Fig. S6). First, we aggregated individual ringing events likely done in the same nest. We used a combination of unique **r**inger ID (ringer identification), **S**pecies, **L**ocation (unique longitude and latitude) and **D**ate (unique day, month and year) of ringing to identify unique ringing events per nest (**RSLD** combinations); this yielded 584 354 events which correspond to a ringing event occurring in separate nests by a given ringer on a given day of year for each species in the dataset. However, nestlings in several nests may have been ringed in the same locations (when nests are located very close to each other) and/or more than one ringer may have ringed nestlings cooperatively in the same nest, causing the same nest to occur multiple times in the database. Therefore, we further divided these records into three subsets (A, B and C), based on a brood number (i.e. each ringer's personal chronological sequence of ringed broods) and an abundance value (i.e. number of nestlings observed per brood), when available. Specifically, a unique brood identifier was created based on the combination of **B**rood and the variables described above (**RSLD+B=RSLDB**). For each RSLDB identifier, we tallied the number of ringed nestlings (i.e. number of unique Ring IDs) and compared this to the reported abundance value. If these values matched, each of these unique broods was considered a unique nest and given a unique ID (subset A; 392 421 unique nests). In cases where the tallied number of ringed nestlings and the reported abundance value did not match (either because the ringer had not ringed all nestlings in the brood or because either the brood number or abundance value was incorrectly reported - which of the two was the case we cannot separate) and/or abundance values were missing, we used the tallied number of ringed nestlings per unique RSLD and disregarded the reported abundance value by the ringer, which might or might not be correct. Next, the tallied number of nestlings was compared to the maximum brood size for each species (literature-based estimates; see traits below). When these values matched, the records were assigned a unique nest ID (subset B; 247 399 unique nests). For the remaining records, we calculated the potential number of nests for each RSLD, by dividing the number of ringed nestlings by the maximum brood size for each species. This step was used to calculate the required number of nests to "accommodate" the recorded number of nestling ringed, which were given a unique ID (subset C; 294 284 unique nests). Subsets A, B, and C were then combined, yielding 934 104 records of ringing events in unique nests. Finally, we only included records after 1975, since records before this year were sparse and only represented a small subset of the bird species in the dataset. This processing resulted in 919 713 records of bird ringing events for 166 species, spanning 43 years and four bioclimatic zones in Finland (Available on Dryad [LINK]).

Text S2 Assessing change in ringer behavior

We performed two sensitivity analyses to evaluate changes in ringer behavior and ringing effort over the study period that could confound the patterns of change in phenology. To assess an effect of ringer behavior, we checked for a potential change in the size of ringed nestlings. This allowed us to evaluate if nestlings were consistently ringed earlier because ringers went out earlier in the season, opposed to general guidelines, and not because of real changes in breeding timing. For those ringing events for which the ringer included wing-length measurements (only reported sporadically; for c. 14% of ringing events for the analyzed 73 species), we used linear regressions to estimate the effect of year on wing- length for each species (see SI Appendix, Fig. S7). While 71% of the species showed a significant change over time, the magnitude of the changes was not biologically meaningful. Additionally, the significant trends appeared in both directions (both increases and decreases), and there were no consistent differences in these trends for the different trait groups we are testing (migratory strategy and number of broods; see SI Appendix, Fig. S7). Thus, we concluded it was unlikely that there was any systematic directional change in ringer behavior over time, and that any trend in size of species is mainly caused by random patterns of negligible biological effect or ringers tending to measure birds of unusual size. Second, to assess potential changes in ringer effort levels across the study period, we plotted the number of unique ringing events (a proxy for ringer effort) in each year within each bioclimatic zone (see SI Appendix, Fig. S9). Despite higher ringer effort levels early in the study period, annual effort levels stabilized in all zones by 1990 and remained relatively constant thereafter. The only exception to this pattern was the hemi-boreal zone, where ringer effort decreased slightly beginning ca. 2000. Even with this decrease, the total number of ringing events in the hemi-boreal zone was at least ~2000 or greater every year over the study period, and thus considered sufficient to accurately estimate the breeding period metrics (see *Phenological Metrics* in the Methods section of the main manuscript). We also note that, while there is variation in the number of nests observed in each year over the study period, the proportions of nests observed in the beginning and end of the breeding season are relatively constant over the study period (see SI Appendix, Fig. S10). As such, there are not strong differences in effort levels between the beginning and end of the breeding season across the study period.

Text S3 Model description

As described in the main article, we adapted the Hierarchical Modeling of Species Communities (HMSC) framework for application to our three phenological metrics for all 73 bird species. Here, we provide additional information on our specific application of the HMSC framework. Further information on the model framework along with worked examples can be found in [1]. Our model notation is as defined in the main article and is adapted from [1] to strengthen the connection to the original model paper and facilitate interpretation.

One key strength of the HMSC framework is it allows species' traits and phylogenetic information to inform species-level responses to model covariates [1]. Specifically, regression coefficients (i.e., fixed effect terms) are assigned exchangeable, Gaussian priors [2, 3], $\boldsymbol{\beta}_{j\cdot} \sim \mathrm{MVN}(\boldsymbol{\mu}_{j\cdot}, \mathbf{V})$, where $\boldsymbol{\beta}_{j\cdot} = \left(\beta_{\mathrm{HB},j}^{(0)}, \ldots, \beta_{\mathrm{NB}}^{(0)}, \beta_{\mathrm{HB},j}^{(1)}, \ldots, \beta_{\mathrm{NB}}^{(1)}\right)'$ is a vector including zone-specific intercepts and annual effects for species j, $\boldsymbol{\mu}_{j\cdot}$ is a vector of the same dimension including among-species mean coefficients, and $\mathbf{V}$ is a covariance matrix capturing variability among the coefficients for species j. Within HMSC, among-species means are modeled as a function of species-level traits, for example $\mu_{jk} = \boldsymbol{t}_j'\boldsymbol{\gamma}_k$, where k indexes the eight species-specific regression coefficients ($k = 1, \ldots, 8$), $\boldsymbol{t}_j$ is a vector containing trait values for species j, and $\boldsymbol{\gamma}_k$ is a vector (of the same dimension as $\boldsymbol{t}_j$) containing unknown coefficient values for each trait for the kth regression coefficient. In our application, trait values correspond to unique combinations of migratory pattern and brood size: (RS-1, RS-2, L-1, L-2), such that $\boldsymbol{\mu}_{j\cdot}$ represents a group mean coefficient for all species with the same trait values as species j.

Phylogenetic information is used to model the dependency among species-specific regression coefficients after accounting for their trait values [1]. Specifically, the combined vector of species-specific regression coefficients is modeled as, $\boldsymbol{\beta}_{\cdot\cdot} \sim \mathrm{MVN}(\boldsymbol{\mu}_{\cdot\cdot}, \mathbf{V} \otimes (\rho\mathbf{C} + (1-\rho)\mathbf{I}))$, where $\mathbf{C}$ is a (73 x 73) phylogenetic covariance matrix for the study species, $\mathbf{I}$ is a 73-dimensional identity matrix, $\otimes$ denotes a Kronecker product, and ρ is a scalar parameter measuring the relative importance of the phylogenetic covariance matrix at describing the covariance of regression coefficients among species. When ρ is close to one, there is evidence that the phylogenetic covariance matrix is meaningful for describing among-species regression coefficient groupings. When ρ is close to zero, species-specific regression coefficients are modeled as independent of one another after accounting for species' traits.

In our specific study, HMSC is a multivariate framework estimating the phenology metric of interest jointly across species. A sampling unit (unique combinations of bioclimatic zone and year) random effect is used to model the covariance in phenology metrics among species after accounting for bioclimatic zone and study year: $\boldsymbol{u}_{it,\cdot} \sim \mathrm{MVN}(\mathbf{0}, \boldsymbol{\Omega})$ where $\boldsymbol{u}_{it,\cdot}$ is a vector containing random effects for all 73 species for zone i and year t and $\boldsymbol{\Omega}$ is a (73 x 73) covariance matrix (accounting for among-

species dependence). Modeling covariance matrices in high-dimensional settings, such as the current application, can be quite challenging and usually requires some form of dimension reduction [4, 5]. HMSC uses a latent factor model to estimate species-specific sampling unit random effects [6]. Specifically, $u_{it,j} = \boldsymbol{\lambda}_j' \boldsymbol{\eta}_{it}$ where $\boldsymbol{\lambda}_j$ is a vector of species-specific factor loadings and $\boldsymbol{\eta}_{it}$ is a identically-sized vector of sampling-unit-specific random factors. The covariance matrix is then calculated as: $\boldsymbol{\Omega} = \boldsymbol{\Lambda}'\boldsymbol{\Lambda}$ where $\boldsymbol{\Lambda} = [\boldsymbol{\lambda}_1|\boldsymbol{\lambda}_2| \ldots |\boldsymbol{\lambda}_{73}]$, a 73-column matrix with each column equivalent to the species-specific factor loadings. Any remaining unexplained variation is modeled using the residual error term, which assumes species are independent (after accounting for species-dependence in the random effect), specifically, $\epsilon_{it,j}$ are independent and identically distributed as $\mathrm{N}\left(0, \sigma_j^2\right)$.

We used Markov chain Monte Carlo (MCMC) simulation to sample from the joint posterior distribution for all model parameters as implemented within the HMSC package (v 3.0-2, [7]) for the R statistical computing environment [8]. Specifically, for each phenological metric, three MCMC chains were run applying the default prior and hyper-parameter values for all model parameters to generate 1,000 posterior samples using a thinning interval of 100 following a burn-in period of 200,000 samples (for the phenological metric *beginning*, the thinning interval and burn-in period were respectively 150 and 300,000). For all metrics, model convergence was assessed visually and using Gelman-Rubin statistics [2].

Two modeling decisions warrant further explanation. First, we chose to model each breeding period metric independently, although we could have included all three metrics within the same model. This was done to improve the interpretability of species covariance estimates. Importantly, preliminary modeling with all three metrics modeled jointly generated nearly identical results as fitting independent models for all three metrics. Second, we chose to use an unstructured, rather than a temporally-structured random effect nested within bioclimatic zone. This was done based on exploratory analysis that indicated low temporal autocorrelation in species-level metrics over time. Further, we did not want temporally-structured random effects to limit our ability to detect temporal shifts in breeding period metrics over time. There was little to no evidence of temporal autocorrelation in the residuals after accounting for the effects of bioclimatic zone and year (mean first order autocorrelation coefficient across all sampling units, species, and metrics equals 0.03 with a 95 percent confidence interval of [0.01, 0.06]).

Text S4 Sensitivity analysis

The lower and upper quantiles of a distribution can be difficult to estimate accurately as they are more sensitive to incomplete observations than measures of centrality, such as the mean. We limited our analysis to only those years in which there were at least 30 unique ringing observations for a given species-zone combination to ensure that we had a sufficient sample size to accurately estimate the phenology metrics (years with less than 30 observations were treated as missing values, but a species-zone combination was only included if there were more than 10 years with sufficient observations; see *Species inclusion criteria* in the Methods section of the main manuscript).

To test the sensitivity of our results and conclusions to uncertainty in derived breeding period metrics, we simulated 95 percent confidence intervals for each metric and species-zone-year combination, based on a non-parametric bootstrap using 10,000 resamples of the original day-of-year observations for each of the 4237 unique combinations of species, zone, and year. We then refit the HMSC model framework to the lower and upper bounds of the bootstrapped confidence intervals for each metric. This resulted in two additional model fits per metric: one applied to the lower bounds of the bootstrapped confidence intervals for all species-zone-year combinations, and one applied to the upper bounds (we refer to these as lower- and upper-bound re-analyses, respectively).

The results of the re-analyses were largely consistent with those presented for the original (non-bootstrapped) data. Mean species-level shifts in each metric were nearly identical to the original data (SI Appendix Figs S4 and S5; Table S1). A total of 20 species (27 percent) exhibited a contracted breeding period under both the lower- and upper-bound re-analyses, compared to 23 species using the original data. The identity of these species differed slightly from the original 23 (15/20 and 17/20 of the species with a contracted breeding period for the lower- and upper-bound re-analyses, respectively, are included in the original 23). Despite these small differences, the majority of species with a contracted breeding period (85 and 95 percent, respectively, for the lower- and upper-bound re-analyses) were resident or short-distance migrants, consistent with our original results. In general, the results of the re-analyses support our main findings: there is a general advancement of the breeding period across study species, with greater advancement in the end relative to the beginning of breeding observed among approximately 30 percent of species, mostly composed of resident and short-distance migratory species. The consistency in results between our original analysis and our re-

analyses using bootstrapped data suggests our results and main conclusions are robust to uncertainty in our derived phenological metrics.

References

1. Ovaskainen O, Tikhonov G, Norberg A, et al (2017) How to make more out of community data? A conceptual framework and its implementation as models and software. Ecology Letters 20:561–576. https://doi.org/10.1111/ele.12757

2. Gelman A Bayesian Data Analysis, Third. Chapman & Hall / CRC Texts in Statistical Science

3. Hobbs NT, Mevin BH (2015) Bayesian models: a statistical primer for ecologists. Princeton University Press

4. Bhattacharya A, Dunson DB (2011) Sparse Bayesian infinite factor models. Biometrika 98:291–306. https://doi.org/10.1093/biomet/asr013

5. Taylor-Rodríguez D, Kaufeld K, Schliep EM, et al (2017) Joint Species Distribution Modeling: Dimension Reduction Using Dirichlet Processes. Bayesian Anal 12:939–967. https://doi.org/10.1214/16-BA1031

6. Ovaskainen O, Abrego N, Halme P, Dunson D (2016) Using latent variable models to identify large networks of species-to-species associations at different spatial scales. Methods in Ecology and Evolution 7:549–555. https://doi.org/10.1111/2041-210X.12501

7. Tikhonov G, Ovaskainen O, Oksanen J, et al Hmsc: Hierarchical Model of Species Communities. R package version 3.0-2. https://CRAN.R-project.org/package=Hmsc

8. R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/