Supporting Information

Bartomeus et al. 10.1073/pnas.1218503110

SI Text

Investigation of Potential Bias. We investigated the following potential sources of collecting bias in our data. First, we confirmed that geographical coverage was consistent for all time periods (Fig. S3) and that there was no correlation between collection year, latitude, or longitude (Pearson correlation year-latitude = −0.12, year-longitude = 0.09). Second, we checked for bias in the identity of the bee species collected by different collectors. Such a bias could potentially result in trends in bee species over time, given that different collectors were active during different time periods. The great majority of our collector teams, 1,365 of 1,550, each contributed fewer than 20 specimens to our analysis; thus, the potential for collector bias from any one collector is very small. Only 55 collector teams contributed more than 100 specimens each. To explore the relationship between collector bias and time, we ordinated bee community composition by collector and looked for patterns over time. A nonparametric multidimensional scaling plot (NMDS; stress = 0.11; variance explained = 0.97) of the 55 top collectors as a function of similarity among their collections showed that most collectors are generalists. Even though a minority of collectors was biased toward collecting particular taxa, collector biases did not change significantly over time [Permutational ANOVA (PERMANOVA): F = 1.76; df = 2; P = 0.14; Fig. S5]. Despite this general lack of bias, the genus *Bombus* appears to have been specifically targeted by some collectors (Fig. S5; although excluding the three early collectors who only collected *Bombus* did not qualitatively change our results). For this reason, and also because *Bombus* has been better studied than other genera, we performed all analyses separately for taxa other than *Bombus*, and for *Bombus* alone. NMDS and PERMANOVA were done with the package vegan in R.

Ecological Traits. We compiled information on six ecological traits for the complete set of 187 species for which we ran individual species-level analyses. First, bee body size was measured as the intertegular distance, which is the distance between the two tegulae, small sclerites above the insertion points of the wings. All specimens were measured either in the R.W. laboratory or in the Cornell University Insect Collection. We measured female specimens of all species because these are more abundant in collections and have the prominent role in maintaining populations. For *Bombus*, which varies widely in body size by caste, intertegular distance of both queens and workers were recorded and analyzed separately. We only present the model with *Bombus* workers measurements because workers are far more abundant, but using queens did not alter conclusions with regards to the body size analysis.

Second, data on nest substrate (hole, cavity, soil, stem, wood), sociality (solitary, facultative social, eusocial), parasitism (yes, no) and diet specialization (oligolectic, polylectic), were gathered from the literature for all species analyzed. Voltinism (univoltine, multivoltine), or the number of generations per year, was recorded only when known ($n = 62$) because there is no information available for many bee species. However, note that phenological breadth and the number of generations per year was correlated in our dataset ($r^2 = 0.61$).

Third, we extracted the temporal extent of the flight period, or phenological breadth, of each bee species from our database of 30,138 specimens in our study area (38° to 45° N and −85° to −70° W). Despite the importance of phenological breadth for ecological questions and the large variation among pollinator species in their phenological breadth, quantitative measures for this trait have rarely been reported. Preliminary analysis using species with large sample sizes showed that subsampling 30 independently collected specimens is sufficient to characterize the phenological period, with the range of start and end dates for the estimates based on subsamples being <10 d. Thus, we conducted the phenology analysis only for those species represented by at least 30 independent collecting events. To estimate the length of the flight period, we used the mean 10th and 90th percentiles across 100 subsamples of the data. This truncation helped to remove the influence of sample size and extreme records, thus making our measure more comparable across species.

Fourth, geographical range limits were estimated using our full-specimen database for North America (71,482 specimens) using the maximum and minimum latitude recorded for each species. Our database has poor coverage above 47° N latitude (roughly Québec City, Canada). Hence, the maximum latitude may be underestimated for some species. However, if a bias against collecting bees at the northern latitudes exists, it does not affect the increasing trends reported for species with northern range limits falling within our study area.

Phylogenetic Tree. Sequences for nuclear-coding genes for apoid wasps and bees were downloaded from GenBank in September 2011. The coding regions of 20 genes that were represented in three or more bee tribes were aligned using MUSCLE v. 3.8 (1). Minor adjustments were made by hand using MESQUITE v. 2.73 (2) to retain amino acid coding and to remove introns not identified in some of the GenBank records. Most of the species in our database were not represented in the available GenBank sequences. Therefore, we built a genus-level tree by selecting one sequence per gene per genus (either the longest or, if more than one species was equally long, at random) to form a concatenated matrix of 17,042 bp for genera. The maximum likelihood estimate of the tree was generated using RAx-ML v. 7.2.8-alpha (3) under the GTRCAT model of sequence evolution. Overall, 349 of the 443 world bee genera and 27 apoid wasp genera were included in the final tree (Fig. S6). Genera not represented in our database were pruned from the tree using MESQUITE. When more than one species in a genus was represented in our database, the genus was replaced with a polytomy with near-zero branch lengths.

Because phylogenetic trees for bees are controversial and not fully resolved, we did sensitivity analysis on the use of a phylogenetic tree. Alternative simpler models that used taxonomy (genus nested within family as random effects; dietary breadth estimate = 0.008 ± 0.003, P = 0.006; phenological breadth estimate = 0.00009 ± 0.00005, P = 0.02; body size estimate = −0.003 ± 0.001, P = 0.01; northern latitude estimate = −0.002 ± 0.0004, P < 0.001) instead of a phylogenetic tree produced very similar results (reported in the main text).

1. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.
2. Maddison W., Maddison D. (2010) Mesquite: A modular system for evolutionary analysis. Available at http://mesquiteproject.org/mesquite1.0/mesquite/download/MesquiteManual.pdf. Accessed February 20, 2013.
3. Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688–2690.
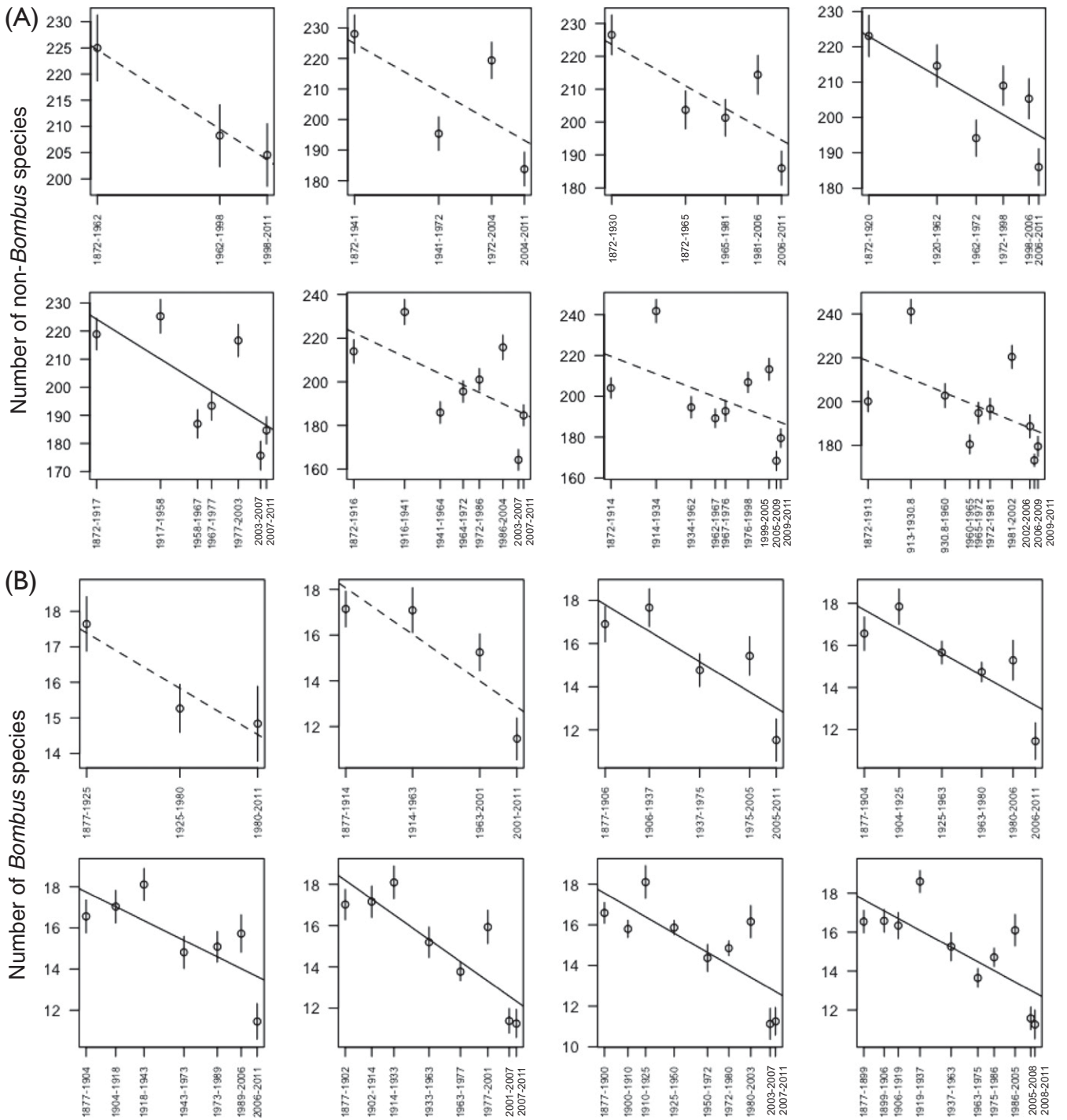
Fig. S1. Sensitivity analysis on the number of bins used to calculate species richness. Data were binned to yield approximately equal numbers of specimen records per bin. Richness calculated after rarefying each time bin to the same number of individuals (± SE), for 3–10 bins. Solid line indicates a significant declining trend as tested by a permutation test. (*A*) Non-*Bombus* species (rarefied to 1,000 specimens per bin) and (*B*) *Bombus* (rarefied to 400 specimens per bin).
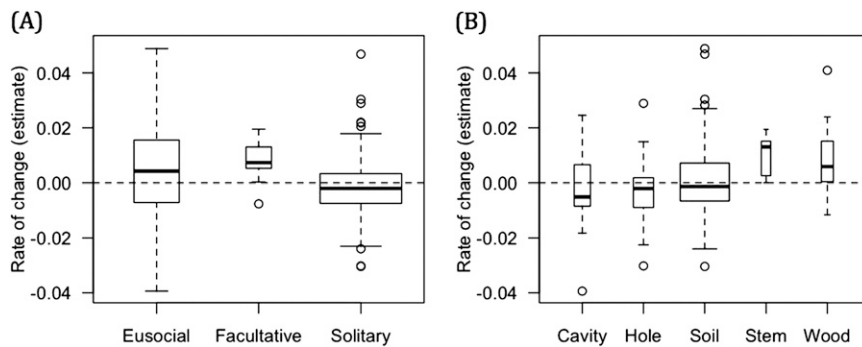
Fig. S2. Estimates of relative change over time, analyzed by life history trait. (*A*) Sociality. (*B*) Nest site location. Boxplot width is proportional to the number of bee species in each analysis category.
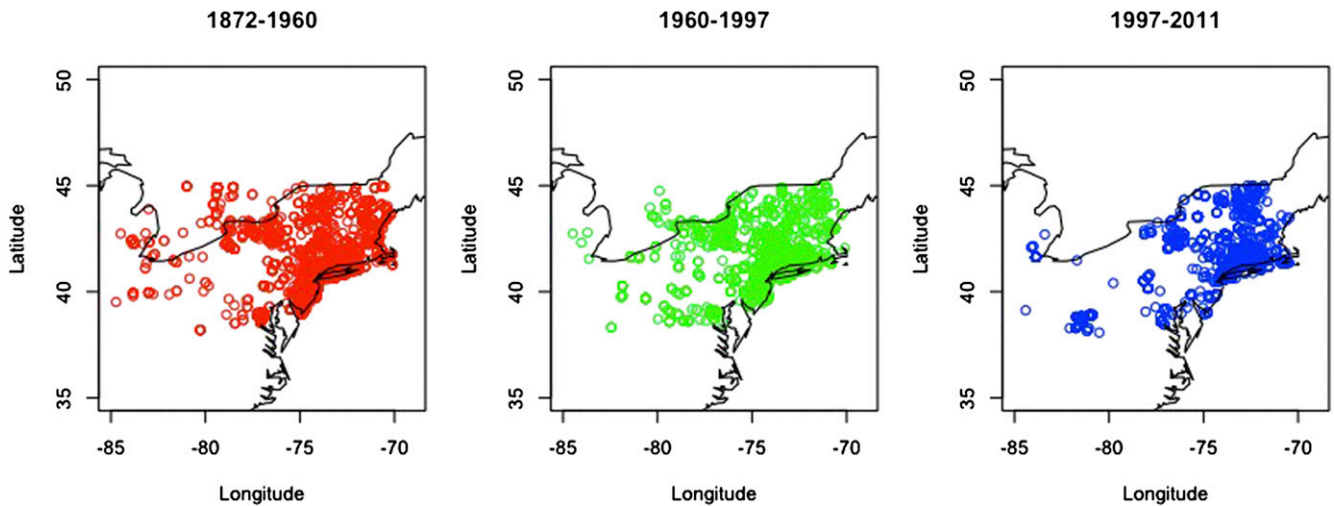

Fig. S3. Map of the study area with the collection events for three time periods of similar sample size (same time periods as used in the sensitivity analysis for $n = 3$).
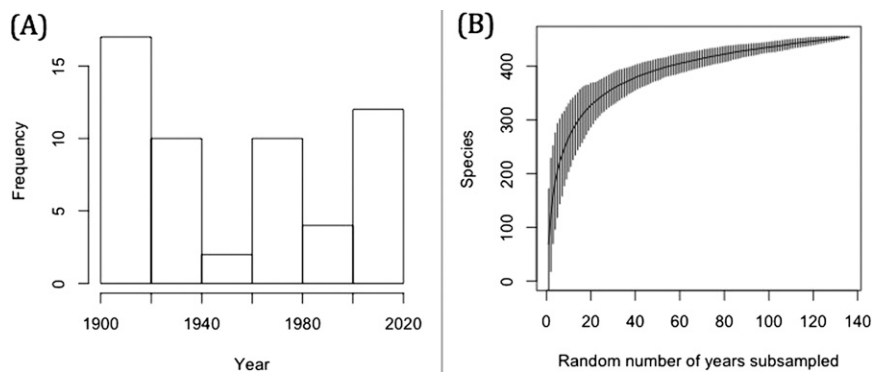

Fig. S4. (*A*) Histogram of the number of singletons present in the entire dataset as a function of the collection year. (*B*) Species accumulation curve showing richness when randomly selecting number of sampling years (black bars = SE).
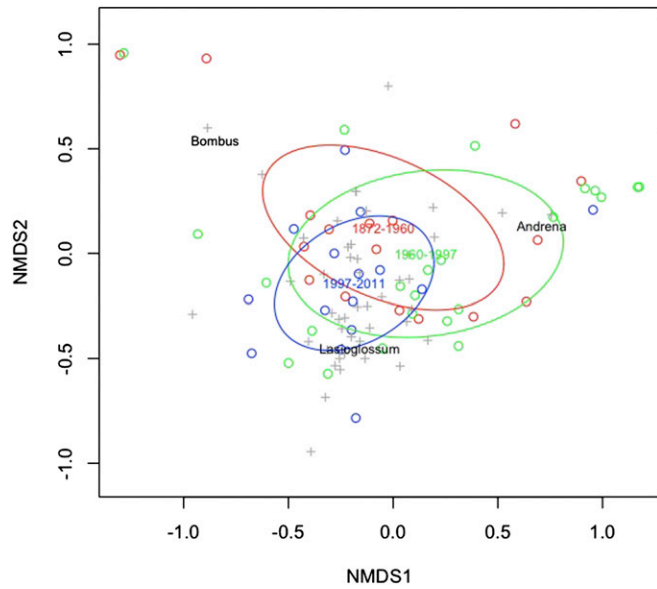
Fig. S5. Nonparametric multidimensional scaling plot (NMDS) of the 55 top collectors as a function of similarity among their collections shows that most collectors are generalists. Bee genera are represented by crosses and collectors by circles. Colors denote the activity period based on the three time periods (same time periods as used in the sensitivity analysis for *n* = 3 and in Fig. S3). Three of the more speciose genera (*Bombus*, *Andrena*, and *Lasioglossum*) are labeled. Nonlabeled peripheral genera are in all cases rare. The ellipse shows the centroid and dispersion of the three time periods.
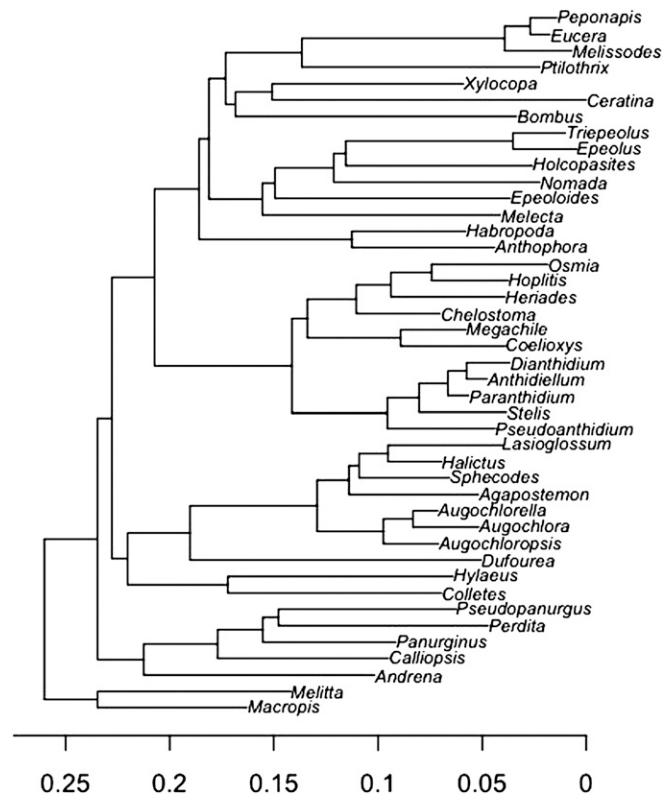

Fig. S6. Maximum-likelihood estimate of phylogenetic relationships pruned to include only the subset of bee genera represented in our study. Outgroups not shown.

Other Supporting Information Files

Table S1 (DOCX)
Table S2 (DOCX)