ANALYSIS

Scott, D. W. (1992) *Multivariate Density Estimation: Theory, Practice, and Visualization.* New York: Wiley.

Breiman, L. (2001), *Random Forests*, Machine Learning 45(1), 5-32.

RESULTS

Across the three sites and seven years of the project we collected a total of 25,441 bee specimens. Of these specimens, 25,017 (98.33%) were identified to the species level, representing 118 species from five families and 24 genera. Of the remaining 424 specimens, 376 (1.48%) were identified to morphospecies, representing potentially 35 additional species within the genera: Nomada, Osmia, Sphecodes, and Triepeolus. The remaining 48 specimens (0.19%) were damaged, hindering species-level identification, however, these specimens were confidently identified to genus.

Our species accumulation curves did not reach asymptotes, suggesting that the 118 confirmed species is likely an underestimate of the true richness across the three sites (Figure X). Chao richness estimation revealed the total richness across all sites and years as 139 species (using trap and net collected records). Using only trap records, Chao richness was estimated at 124 species. Chao richness for the individual sites, POS, SCL, and BPF was estimated using trap records only at 115, 92, and 80 species, respectively.

A graph of different sites

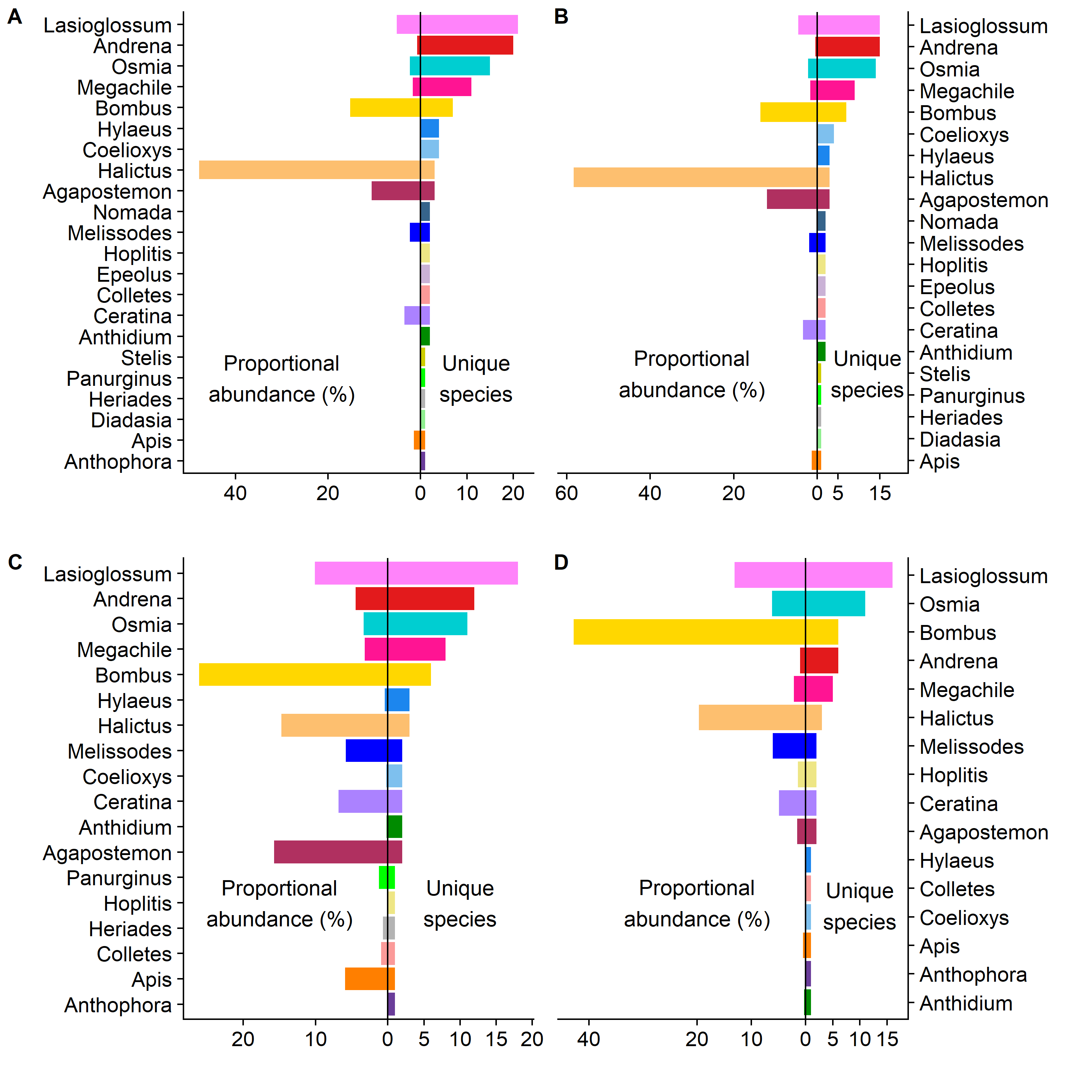
Description automatically generated with medium confidence

**Species accumulation curves** for all sites with blue vane trap and net sweeping (black), all sites with only blue vane traps (maroon), Port of Seattle (yellow), Seattle City Lights (blue), and Boeing Plane Field (green). Collection effort (number of sampling sites) was derived from permutation resampling of the subsites within each site for each year.

Species which were represented by a single specimen (“singletons”) have a disproportionate influence on some analyses, such as Chao richness and may have special significance regarding collecting technique, conservation, and biogeography (). Twenty of our recovered species (16.95% of total richness) were represented by singletons, which are listed in Table S1. In a similar vein, super-abundant species can reflect importantly on community structure and we therefore list in Table S2 the top 10 species by specimen representation. Breadth of distribution can be important in community considerations, so we also list in Table S3 those species that were collected in all three of our sites. Collecting technique and the importance of supplementing traps with net collecting is non-negligible. For example, we found 11 species collected only by net, representing 9.32% of total richness (Table S4). As males are often underrepresented in collections for various reasons we also list in Table S5, species for which we collected no males by either traps or netting. We collected a total of 575 records of *Apis mellifera*, representing 2.26 % (575/25441). Of the 575 *A. mellifera* specimens, 213 (37.04 %) were collected by net with the remainder in traps, mostly Blue Vane Traps.

This study did not focus on documenting in detail the differences in trap function regarding bowl vs. Blue Vane or between various bowl colors. Anecdotally, however, we can report that Blue Vane traps excelled in collecting large bodied species and a greater diversity than bowl traps, yet the two trap types were complementary in function. What is this? I don’t have a distinction between traps in the data.

Of the 24 genera found, overall, *Lasioglossum* was the most speciose (21 species), while *Halictus* the most abundant, representing 47.01% of all records (Figure Xa). However, two sites, SCL and BPF did not follow the overall abundance pattern. In these sites, *Bombus* was by far the most abundant genus, with 26.10% of records in SCL, and 42.80% in BPF (Figure Xc,d).



**Proportional abundance and unique species.** A) All sites collectively, B) POS, C) SCL, and D) BPF. Data are from trap collected records and exclude all net caught records. Data are pooled across years.

Alternatively:

A group of colorful bars with black text

Description automatically generated with medium confidence

**Proportional abundance and unique species for all genera.** A) trap caught records pooled across all sites and years, B) trap records from POS, C) trap records from SCL, D) trap records from BPF, E) trap and net caught records pooled across all sites and years, and F) all net caught records pooled across sites and years.

Or:

A graph of different colors and numbers

Description automatically generated

**Proportional abundance and unique species** for A) trap caught records and B) net caught records. Data combine records across all sites and years.

Or:

A graph of different species

Description automatically generated

**Proportional abundance and unique species** for A) all net and trap caught records and B) net caught records only. Data combine records across all sites and years.

Species composition varied significantly across sites (PERMANOVA; *F*2 = 2.11, *P* = 0.005), with BPF hosting a statistically distinct bee community from the POS and SCL sites (Figure XX). Moreover, our random forest model classified each site by species composition with an overall out of bag error rate of 15.38%. However, while the model struggled to distinguish between the POS and SCL sites, the BPF site was never misclassified (0.0% class error). Major changes in abundance primarily delineate the BPF site from the others. Notably, when compared to the POS and SCL sites, the BPF site had drastic reductions in Halictus tripartitus, Agapostemon texanus, Apis mellifera, and Bombus fervidus, and major increases in Bombus melanopygus, Halictus confusus, and Bombus mixtus

A diagram of different types of objects

Description automatically generated

**Variation in community composition across sites.** Bee species are plotted on the first two axes of a three-dimensional non-metric multidimensional scaling ordination of the 69 combinations of station (subsite) and year, across the three sites. Small points are the individual station/year combinations. Large points are the centroids of the three sites. Ellipses are 95% confidence intervals around the site centroids. Bee species shown are the most representative (top 10th percentile of a random forest analysis) of the compositional differences among sites. Text size of the labels is proportional to variable importance score (mean decrease in Gini).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Species** | **BPF** | **POS** | **SCL** | **µ ↓ Accuracy** | **µ ↓ Gini** |
| Halictus tripartitus | 2.1 (38) | 54.3 (11603) | 7.8 (146) | 12.93 | 3.44 |
| Agapostemon texanus | 1.3 (24) | 10.6 (2269) | 14.7 (275) | 10.85 | 2.76 |
| Bombus melanopygus | 5 (89) | 0.1 (23) | 0.1 (2) | 8.66 | 1.8 |
| Osmia albolateralis | 0.1 (2) | 1.0 (204) | 0.5 (9) | 7.94 | 1.64 |
| Apis mellifera | 0.4 (8) | 2.1 (444) | 6.6 (123) | 7.62 | 1.62 |
| Melissodes rivalis | 1.8 (32) | 0.1 (17) | 1.5 (28) | 7.53 | 1.61 |
| Halictus confusus | 5.4 (97) | 0.5 (108) | 0.2 (3) | 7.62 | 1.41 |
| Bombus fervidus | 0.5 (9) | 0.9 (196) | 2.7 (51) | 6.88 | 1.31 |
| Bombus vosnesenskii | 13.9 (248) | 7.3 (1556) | 18.4 (344) | 6.57 | 1.16 |
| Bombus mixtus | 7.6 (135) | 2 (433) | 2.7 (51) | 2.97 | 0.93 |

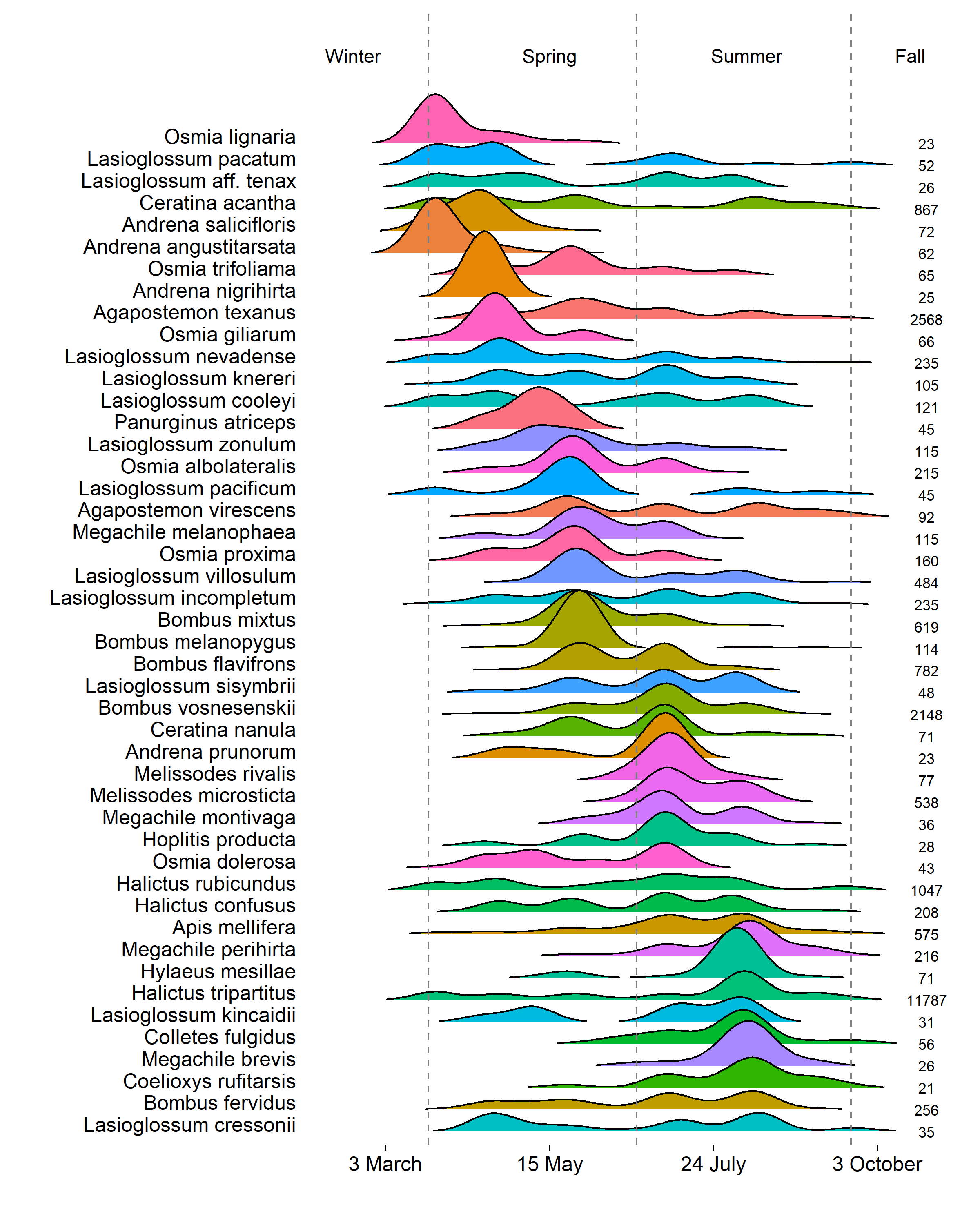
**Table X. The top 10th percentile of species ranked by variable importance score (mean decrease in Gini) from a random forest classification of site by species composition.** Proportional abundance of each species is listed for each site, BPF, POS, and SCL. In parentheses are the counts of each species. Mean decrease in accuracy is the decrease in model accuracy by removing each species. Similarly, mean decrease in Gini is the total decrease in node impurities from splitting each node on each species, averaged over all trees. In both cases, a larger value indicates greater variable importance in the model’s ability to correctly classify sites by the parameters, in this case, a matrix of species composition.

Our sampling, spanning seven years and three sites allowed the estimation of bee abundance across time. At the genus level, 20 of the 24 total genera had at least 20 records, and 46 of the 118 species had at least 20 records. Kernel density estimation revealed several patterns; 1) at the genera level, multimodal distributions may reflect multivoltinism, multiple species, sociality, or a combination of these variables. 2) At the species level, well-defined multimodality was rare, with most species reaching a single peak abundance, then tapering off throughout the season. Some species appear to follow more uniform distributions, flying throughout the season with limited variation in abundance. Lastly, our comparison of parasitic genera and their presumptive hosts…(not sure where we are going with this part, are we saying there is overlap? That host species decline after parasite peaks? There isn’t a clear pattern across all of the combinations)…

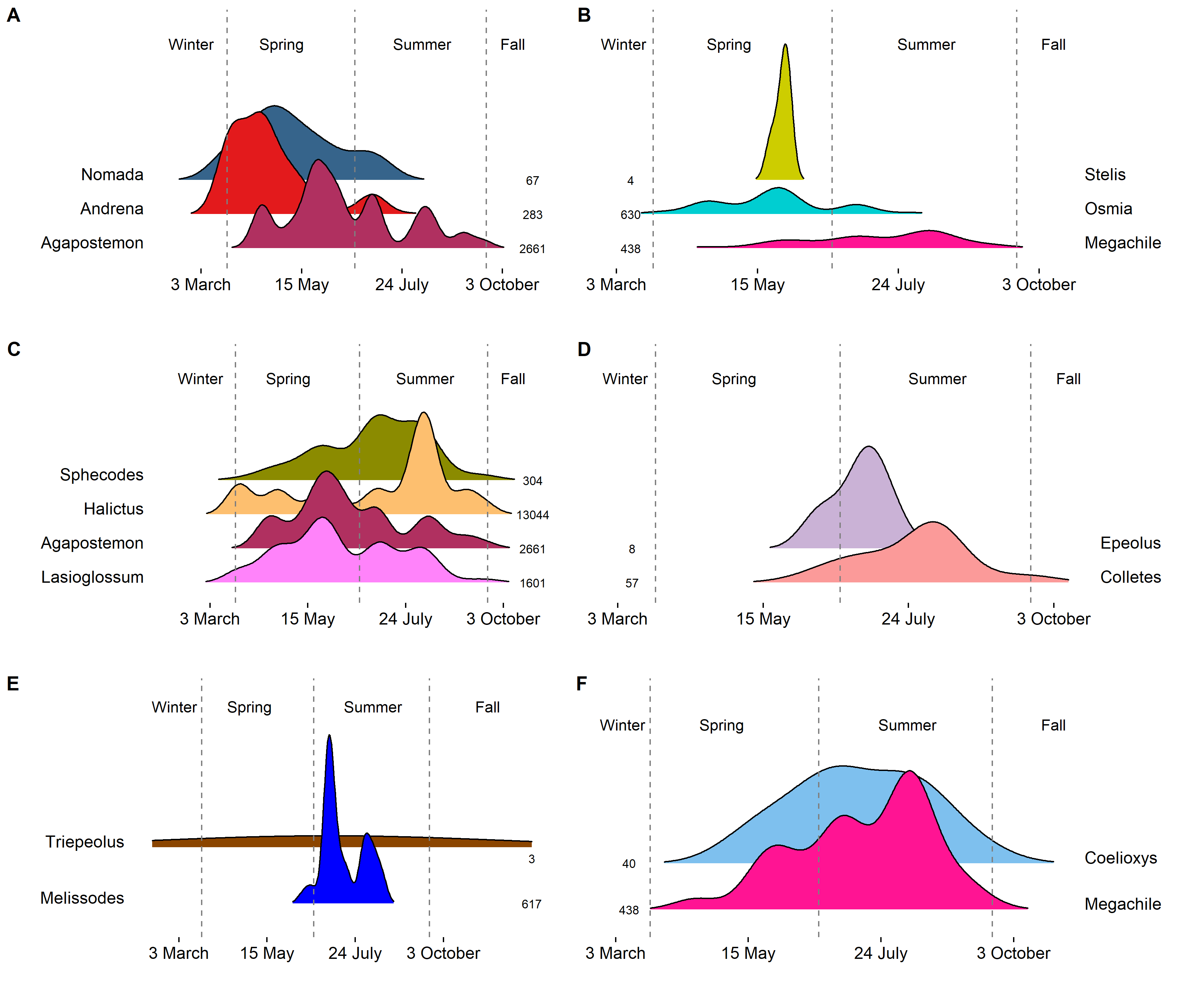
A graph of different colors

Description automatically generated

**Genus-level seasonal distributions.**Density is estimated at the genus level with a smoothing parameter determined from biased cross validation. Values displayed on the right are the total number of records for each genus. Only genera with sample sizes ≥ 20 are shown. Vertical dashed lines correspond to 21 March, 21 June, and 21 September.

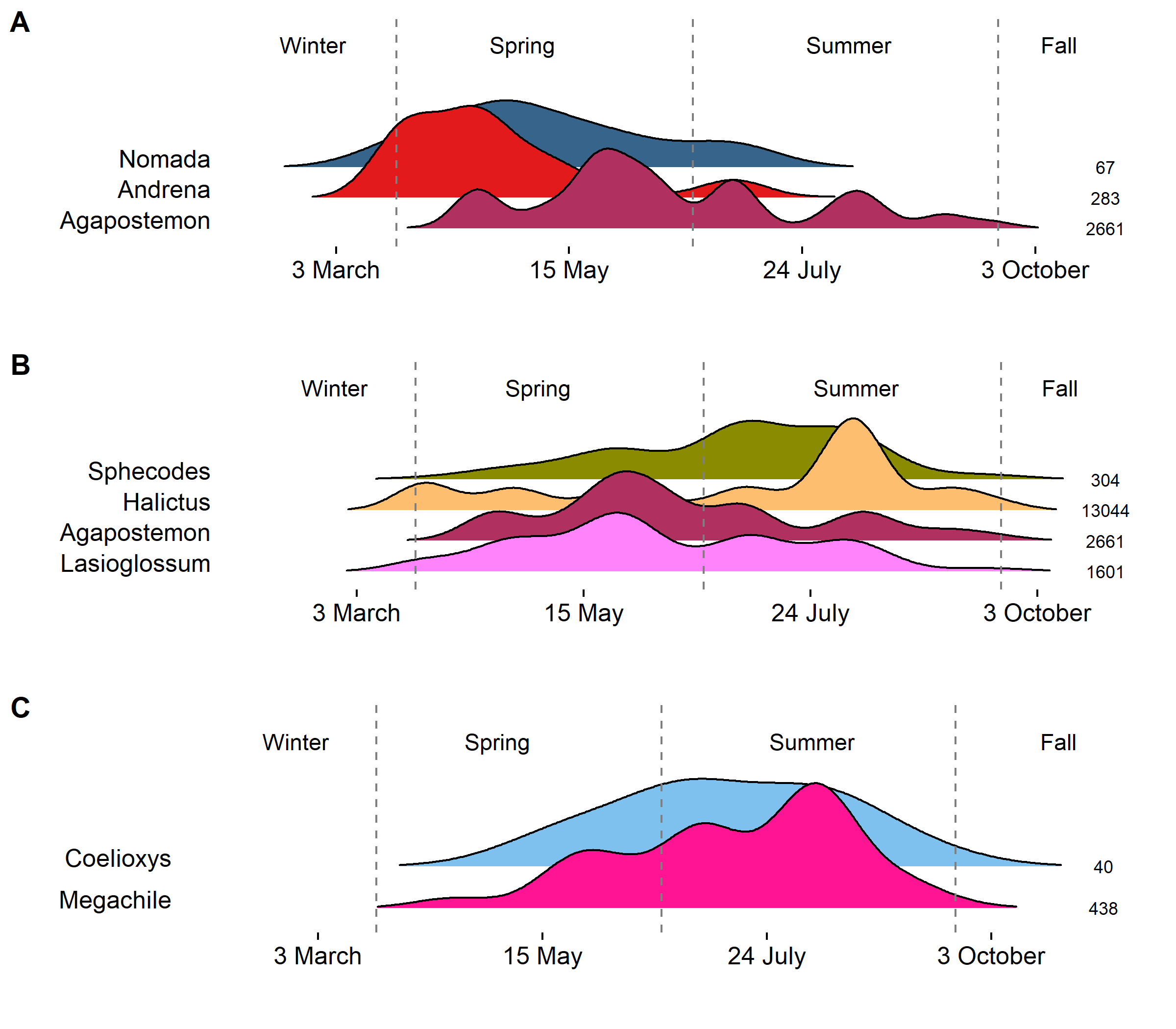


**Species-level seasonal distributions.**Density is estimated uniformly across all species with Silverman’s method (Gaussian smoothing parameter). Values displayed on the right are the total number of records for each species. Only species for which sample sizes were ≥ 20 are shown. Vertical dashed lines correspond to 21 March, 21 June, and 21 September.



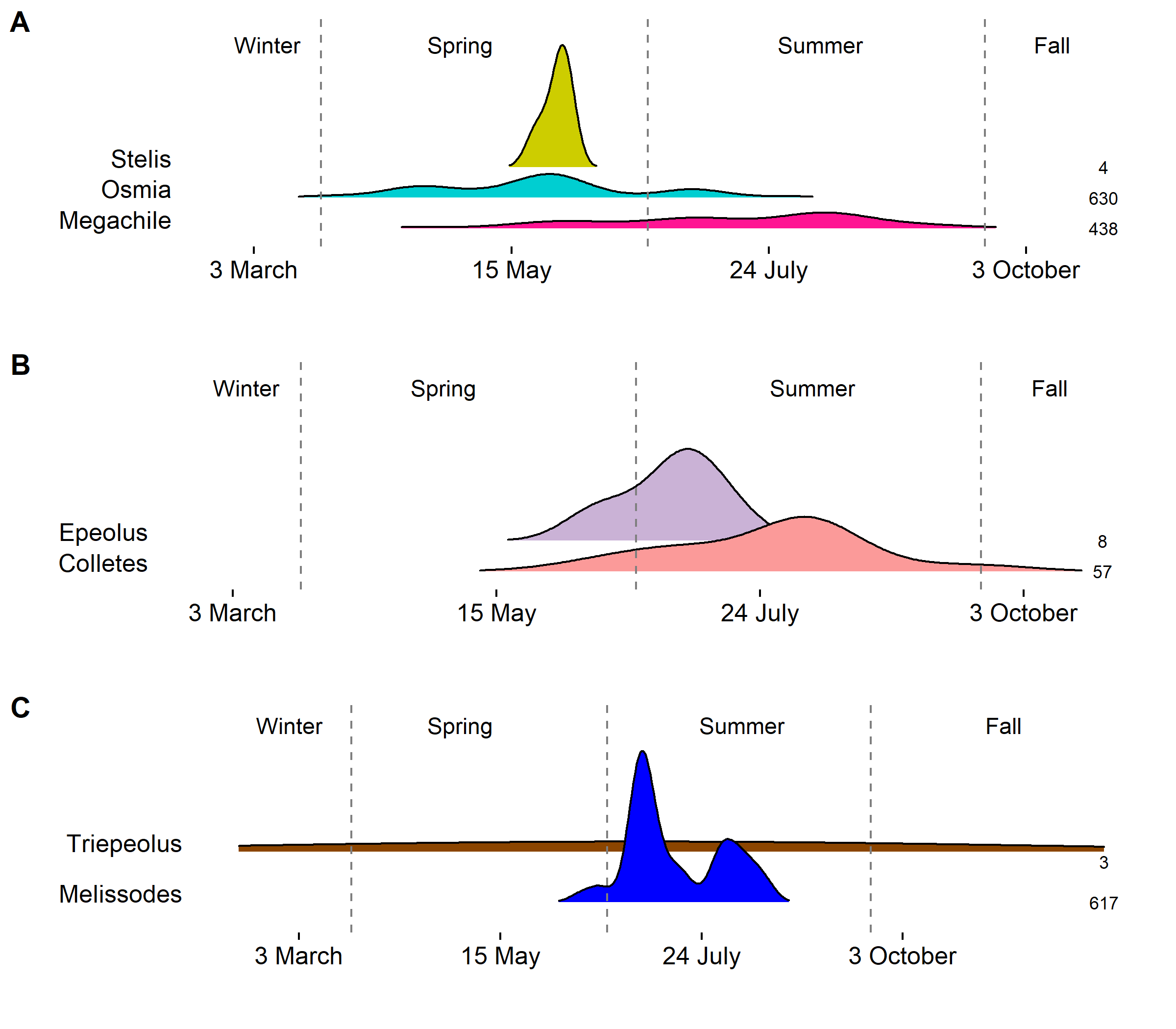
**Genus-level seasonal distributions for the parasites:** A) Nomada, B) Stelis, C) Sphecodes, D) Epeolus, E) Triepeolus, and F) Coelioxys. Beneath each parasite genera are the presumed host genera. Density is estimated at the genus level with a smoothing parameter determined from biased cross validation. Values displayed on the right (plots A, C, E) and left (plots B, D, F) are the total number of records for each genus. Vertical dashed lines correspond to 21 March, 21 June, and 21 September.

Or:



**Genus-level seasonal distributions** for the parasites: A) Nomada, B) Sphecodes, and C) Coelioxys. Beneath each parasite genera are the presumed host genera. Density is estimated at the genus level with a smoothing parameter determined from biased cross validation. Values displayed on the right are the total number of records for each genus. Vertical dashed lines correspond to 21 March, 21 June, and 21 September.

And supplement:



**Genus-level seasonal distributions for the parasites:** A) Stelis, B) Epeolus, and C) Triepeolus. Beneath each parasite genera are the presumed host genera. Density is estimated at the genus level with a smoothing parameter determined from biased cross validation. Values displayed on the right are the total number of records for each genus. Vertical dashed lines correspond to 21 March, 21 June, and 21 September. These parasites have low sample sizes limiting distributional estimation.