Running head:

Bombus spp. show different responses to agricultural intensification...I hope?

Jenna B. Melanson[1*], Tyler T. Kelly[2], Natalia Clermont[2], Jonathan B. Koch[3], Claire Kremen[1,2]

1. Biodiversity Research Center
   Department of Zoology
   University of British Columbia
   2212 Main Mall
   Vancouver, BC, Canada, V6T 1Z4
2. Institute for Resources, Environment, and Sustainability
   University of British Columbia
   2202 Main Mall
   Vancouver, BC, Canada, V6T 1Z4
3. USDA

⋆ Corresponding author: jenna.melanson@ubc.ca

- Type of article:
- Abstract word count:
- Word count:
- Number of figures and tables:
- Number of references:
- Author contributions:

Abstract

abstract text

Keywords: keyword1, keyword2, keyword3

1 Introduction

2 Methods

2.1 Study system

Our study took place in the Lower Fraser Valley in southwestern British Columbia, Canada, in an agricultural system dominated by mixed vegetable, hay, and perennial berry production. From 1984-2018 the Lower Fraser Valley underwent a 13% decrease in forest patch area, mainly resulting from conversion to urban or agricultural land use (**paulTrackingChangesSoil2020a**). .... *decide what else needs to be said in this section later on!*

Field surveys were conducted across six replicate landscapes distributed throughout the Lower Fraser Valley. Each landscape encompassed roughly 3 sq km of farmland interspersed with rural/suburban residence. Landscapes were initially chosen to span a gradient of configurational and composition diversity metrics, including Shannon's diversity, edge density, and the ratio of annual to perennial crop cultivation.

Thirty sampling transects (50 meters x 2 meters) were established in each landscape, spaced as evenly as possible based on land-access and the availability of foraging resources on which to observe bees. We did not survey in active crop fields except for high-bush blueberry, which offers floral resources during its spring bloom and can sometimes host other flowers such as clovers (*Trifolium spp.*) and flatweed (*Hypochaeris radicata*) later in

the season. A total of *INSERT TOTAL TRANSECT NUMBER HERE* were surveyed over the course of two years (2022-2023).

2.2 *Bombus* collections and floral surveys

Each landscape was surveyed during 10 sampling rounds in year one (May-August 2022) and 17 sampling rounds in year two (March-August 2023). During each round of sampling, surveys were conducted on *INSERT MEAN PLUS OR MINUS SD OF SAMPLING EFFORT* (mean $\pm$ SE) transects.

Bombus surveys at each transect entailed 5 minutes of active search time (totaling 140 hours in 2022 and *INSERT TOTAL SAMPLING EFFORT 2023* in 2023), during which the stopwatch was paused whenever a foraging bumble bee was sighted. Specimens were captured by netting, placed into sterile 15 mL tubes, and immediately placed on ice before transfer to a -80°C freezer at the end of the day. Surveys were conducted on days when the temperature was above 12°C (10°C for queen surveys) and wind speeds below 2.5 m/s. In 2022, all *Bombus* species were collected; in 2023 only the focal species (*B. mixtus* and *B. impatiens* were collected.

To assess floral quality, all flowering plants within the transect area were identified to species or genus level. Abundance estimates were taken for each species on the log-scale (i.e., 0 = 1-10 inflorescences, 1 = 11-100 inflorescences, 2 = 101-1000 inflorescences, 3 = 1001-10,000 inflorescences, 4 = 10,000+ inflorescences). Floral survey data was later filtered to exclude

species which bumble bees were never observed visiting (based on over 3,400 visitation events in 2022, and 3,500 visitation events in 2023). This filtering step was included to reduce the noise introduced by a variety of herbaceous weeds with flowers too small to attract or support bumble bee foragers, but which were frequently observed on the transects in high abundance.

2.3 Landscape characterization

Land cover maps were developed for each study site based on manual classification of Google Earth satellite imagery (2021) and site visits. Briefly, land cover was classified into 16 categories: annual row crops, blueberry, cranberry, other perennials, polyculture, hay meadows, pasture, fallow, grassy field margins, hedgerows (tree-dominated), hedgerows (blackberry-dominated), forest, wetlands, urban/suburban, roads, and water). These land cover types were chosen based on their hypothesized provisioning of nesting/floral resources and differences in disturbance regimes (see Table **??** for details). Land cover was mapped at 2-meter resolution in QGIS (**QGIS_software**).

INSERT EXPLANATION OF LANDSCAPE METRICS AND HOW WE CALCULATED THEM

2.4 PCR Amplification and Fragment Analysis

DNA was extracted from the mid-leg basitarsus and tarsus (distal tarsus only for queens) using the HOTSHOT protocol (**<empty citation>**)

Table 1: Description of land cover classifications, including hypothesized resource provisioning for bumble bees. Based on personal observations and expert opinion.

Landcover Class	Description	Vegetation Level	Flowers	Nesting	Disturbance Rank
annual	monoculture annual crops, tilled yearly	bare	sometimes (resource pulse)	no	5
polyculture	mixed annual crops, tilled yearly	bare	sometimes	no	5
hay	multiple cuts per year, tilled every 1-7 yrs	grass	sometimes (clover)	some species	4
fallow	fields taken out of production temporarily	grass	yes	some species	2
pasture	grazed hay meadow	grass	yes	some species	4
blueberry	described in text	woody	yes (resource pulse)	some species	3
cranberry	perennial, flooded yearly	herbaceous	yes (resource pulse)	no	5
other perennial	orchards, tree farms, etc.	woody	sometimes	yes	3
hedgerow (blackberry)	field margins dominated by *Rubus armeniacus* and *R. laciniatus*	woody	yes (resource pulse)	yes	2
forest	forest fragments (primarily native species)	woody	yes	yes	1
hedgerow (woody)	planted or remnant hedgerows dominated by trees; planted hedgerows contain flowering species selected for pollinators	woody	yes	yes	2
road/ industrial	paved/impermeable surfaces	bare	no	no	5
grassy margins	unmanaged field margins without woody vegetation	grass	yes	some species	3
suburban/ urban	residential properties, including gardens	woody	yes	yes	4
water	lakes, rivers, irrigation ditches, ocean	bare	no	no	5

CITE TRUIT. We utilized the following microsatellite loci from the existing literature: BT10, BTERN01, BL13, BL15, B126, BTMS0057, BTMS0059, BTMS0062, BTMS0083 (both species), BTMS0066, BTMS0072, BTMS0086, BTMS0104, BTMS0126, BTMS0136 (*B. mixtus* only), BT28, BT30, B10, B96, B124, BTMS0073, and BTMS0081 (*B. impatiens* only) CITE STOLLE, REBERFUNK, ESTOUP–CHECK WHICH ESTOUP PAPERS!. One primer for each locus was individually dye-labelled using 6FAM, NED, PET, or VIC, and loci were amplified in two multiplex reactions per individual. Each multiplex reaction contained 4μL of template DNA, 1μL of 10X primer mix, and 5μL of Qiagen 2X Multiplex PCR Master Mix (Qiagen, Hilden, Germany). See **??** for plexes and primer concentrations. Diluted PCR products were submitted for automated fragment sizing at the UBC Sequencing and Bioinformatics Consortium (*B. mixtus*) and the Utah State Center for Integrated Biosystems (*B. impatiens*). 1μL of each sample was added to 8.85μL of Hi-Di formamide and 0.15μL of LIZ500 size standard; *B. mixtus* fragments were analyzed on an Applied Biosystems 3730XL 96-capillary DNA Analyzer and *B. impatiens* fragments were analyzed on an Applied Biosystems 3730 DNA Analyzer (Applied Biosystems, Foster City, CA, USA).

Scoring error rates were assessed by re-genotyping a panel of 96 individuals per species, and loci with observed error rates $\geq$3% were discarded from further analyses (BL15 in both species). BTMS0072 was also discarded for *B. mixtus* due to poor amplification and difficulty assigning

peak-calls.

2.5 Colony Assignments

We used an iterative approach to assign workers and queens to their natal colonies. First, using the pedigree reconstruction software COLONY 2.0 (CITE JONES AND WANG), we assigned full siblingships based on all available microsatellite data, assuming male and female monogamy and no inbreeding. A single run was carried out for each species and year, using the software's full-likelihood approach and no siblingship size scaling or priors. Sibling pairs were maintained when $P_{fullsibdyad} = 1$. A single individual from each putative colony (including non-circular colonies, described below) was maintained for downstream analyses of locus quality.

While COLONY can account for inbreeding at the population level, locus-specific estimates of F_{is} should generally be similar to one another for a given population, reflecting their shared evolutionary history. Loci with F_{is} estimates that deviate significantly from the species mean (across loci) may suffer from null alleles or other types of scoring errors that can bias siblingship assignment. We therefore tested individual locus deviations from population mean F_{is} as a criterion for marker inclusion/exclusion.

Single locus F_{is} estimates were computed for all site, year, species groups following CITE WEIR in the package *genepop* [CITE GENEPOP]. The distributions of F_{is} estimates for each species at each site can be found in Ap-

pendix 1, **??**. We then fit linear models to F_{is} estimates with locus and site as fixed predictors. We used sum-to-zero coding so that model intercepts represented mean F_{is} across all loci, and calculated the estimated marginal mean of each locus using the *emmeans* package (CITE EMMEANS PACKAGE)(Appendix 1, **??**A-B). We iteratively removed loci with F_{is} significantly different from the species mean F_{is}, starting with the locus with the greatest deviation, and re-running the model after each removal (i.e., because removing a locus with a high or low inbreeding coefficient will change the species mean estimate and therefore all comparisons to the mean). We did not apply an adjustment for multiple-hypothesis testing, but instead utilized a relatively stringent p-value (α = 0.01) for removal of loci. This process resulted in the removal of loci BTERN01, BTMS0104, and BTMS0059 from downstream analyses for *B. mixtus* and removal of BTMS0073 and BT28 for *B. impatiens*. F_{is} estimates for the remaining loci (n = 10 for *B. mixtus*, n = 13 for *B. impatiens*) can be found in Appendix 1, **??**C-D.

Next, we checked locus pairs for linkage disequilibrium (LD). Marker linkage can lead to non-independent assortment, a condition which violates the assumptions of most parentage reconstruction software and can result in overconfidence in estimated sibling pairs. LD was calculated for each locus pair in each population (site, year, species groups) in *genepop*. We applied a Bonferroni correction for multiple hypothesis testing, and flagged locus pairs which showed significant deviations in $>$ 2 popula-

tions. There was not strong support for linkage disequilibrium between locus pairs in *B. mixtus*. Four pairs showed signs of LD, but each occurred at only a single site in a single year. For *B. impatiens*, there was strong evidence for linkage disequilibrium between BTMS0057 and BL13 (7 out of 12 populations showed significant LD following Bonferroni correction). To determine which locus to maintain, we calculated the polymorphic information content (PIC) using the package *PopGenUtils* CITE TOURVAS POPGEN UTILS. We found that BTMS0057 had higher PIC in both years (PIC = 0.78) compared to BL13 (PIC = 0.47 in 2022 and 0.45 in 2023). For this reason, we chose to maintain BTMS0057 and remove BL13 from further analyses for *B. impatiens*.

After locus quality screening, we calculated global and pairwise F_{st} following CITE NEIL 1987 in the package *hierfstat* CITE GOUDET HIERFSTAT. For both species and years, estimates of global F_{st} were less than 0.005. Pairwise F_{st} ranged from -0.002 to 0.01, indicating little or no genetic differentiation between surveyed sub-populations. Finally, we computed the marginal mean F_{is} for each site to determine whether there was evidence for inbreeding following locus removal, and whether it varied between sites (Appendix 1, **??**). Because we found evidence for only minor inbreeding (e.g., $F_{is} < 0.05$) we ran all final colony assignments using the no-inbreeding model. Given the very low estimates of global and pairwise F_{st}, we chose to combine sites for siblingship assignments to maximize the accuracy of allele frequency estimation.

To determine the most effective method for inferring true sibling pairs while minimizing falsely inferred pairs,, we tested the performance of the COLONY software on simulated datasets using allelic frequencies derived from our true data (see Appendix 1, Section "Testing COLONY on simulated data"). Final assignments were performed based on a single COLONY run for each species-year combination, where early-season queens from 2023 (defined below) were grouped with workers from 2022. We used exclusion tables to prevent the assignment of siblingships between individuals originating from different sites, based on our intuition that such relationships would be highly unlikely due to species biology and the fact that potential foraging area increases as the square of foraging radius (see Appendix 1, Section "Observing colonymates at multiple sites" for support). Each run assumed male and female monogamy and no inbreeding, and utilized an informative prior on siblingship size (n = 1, e.g., approximately the harmonic mean of expected siblingship sizes) but no siblingship size scaling. All runs were performed using the Linux command line version of the software, with run length set to "long", likelihood set to "full-likelihood" and precision set to "high precision." Sibling pairs were maintained with $P_{fullsibdyad} = 1$. To account for non-circular families (e.g., individual A related to individual B, B related to C, but A *not* related to C) we maintained "missing-links" (e.g., A related to C in the example above) when $P_{AC} > 0.95$. If the missing link was rejected at the 95% confidence threshold, we maintained the largest possible clique (all

individuals related), using a random draw in cases where the two remaining cliques were of equal size. In total this resulted in the loss of XXX highly probable sibling pairs in *B. mixtus* and XXX pairs in *B. impatiens*. We acknowledge the violations on the assumption of male and female monogamy may result in a heightened number of false-negatives while using the monogamous specification in COLONY—however, our simulations showed clearly that our dataset was not sufficient to distinguish half-sib relationships, and allowing for polyandry resulted in a very high rate of falsely-inferred sibling pairs.

3 Results

4 Discussion

Supporting Information