Mating frequency estimation and its importance for colony abundance analyses in eusocial pollinators: A case study of Bombus impatiens

Sydney A. Bird (✉ sab6228@psu.edu)  
The Pennsylvania State University

Nathaniel S. Pope  
The Pennsylvania State University

Carley M. McGrady  
The Pennsylvania State University

Shelby J. Fleischer  
The Pennsylvania State University

Margarita M. López-Uribe  
The Pennsylvania State University

Research Article

**Keywords:** polyandry, polygamy, monogamy, COLONY software, conservation, pollinators

**Posted Date:** September 21st, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-2076117/v1

**License:** ☕️ ☝️ This work is licensed under a Creative Commons Attribution 4.0 International License.  
[Read Full License]
Abstract

Bumble bees (genus *Bombus*) includes approximately 250 social species, many of which are in decline in North America and Europe. To estimate colony abundance of bumble bees in natural and agricultural habitats, sib-ship relationships are often reconstructed from genetic data with the assumption that colonies have one monogamous queen. However, some species such as the common Eastern North American bumble bee (*Bombus impatiens*) can display low levels of polyandry that can bias estimates of colony abundance based on sib-ship reconstructions. In order to accurately quantify rates of polyandry in this species, we empirically estimated mating frequencies of queens using a novel statistical model and genotypes from 730 bees. To genotype individuals, we used a highly polymorphic microsatellite set for colonies established from 20 wild caught and 10 commercial queens. We found multiple fathers in 15% of wild colonies and 30% of commercial colonies. This resulted in average effective mating frequencies of 1.07 for wild and 1.15 for commercial colonies. Paternity was also skewed, with the 2nd or 3rd father contributing less than 30% of the offspring. These findings agree with previous reports of polyandry for *B. impatiens*. Using a large empirical dataset, we demonstrate that assuming monogamy for colony abundance estimation in species that violate this assumption may result in a vast overestimation of the number of colonies. Our results emphasize the importance of studying mating frequencies in social species of conservation concern for the appropriate implementation of genetic approaches for colony abundance estimation.

Introduction

In the eusocial Hymenoptera, the number of successful mates per female, known as mating frequency, helps inform the designation of a species’ mating system as monogamous or polygamous. Mating system is an important demographic parameter necessary for the estimation of colony abundance from genetic data (Jones and Wang 2010). For social insects, population size estimates derived from numbers of individuals counted in the wild misrepresents key population-level processes, because the colony is the reproductive unit. Thus, accurate estimates of colony numbers are crucial for species of conservation concerns, as they are used for the identification of drivers of species declines (e.g., correlating population status with land use history) (Maebe et al. 2015), for the monitoring and eradication of invasive species (Toquenaga and Kokuvo 2010), and the development of plans for conservation actions and priorities (Hansen and Jensen 2005, Lepais et al. 2010, McGrady et al. 2021). Many social pollinator species are in decline, which underscores the importance of accurate estimations of colony abundance for these species (Powney et al. 2019).

For colony abundance estimations, sib-ship relationships between wild-collected individuals are assigned based on multilocus genotypes. Under Mendelian inheritance, full sibling groups are genetically distinct. For monogamous haplodiploid species, one full sibling group represents one colony, and workers share on average 75% of alleles among themselves and 50% of the alleles with the queen (Hamilton 1964). Therefore, family relationships can be easily assigned based on genetic data (Lepais et al. 2010, Toquenaga and Kokuvo 2010). However, in a polygamous colony there may be multiple full-sibling
groups, in this case, the average relatedness among members of a colony decreases and more genetic markers are necessary in order to improve the chances of distinguishing between half siblings and full siblings (Wang 2004, Wang and Santure 2009). For example, if one father produces two offspring, they will share on average 100% of the paternal alleles and 50% of the maternal alleles. However if two unrelated fathers each produce one offspring, the half-siblings will only share maternal alleles and not share any paternal alleles. Ultimately, by understanding the levels of polygamy typical of a species, we can better estimate true colony numbers within a population based upon numbers of full-sibling groups are present in a population sample.

In addition to variation in mating frequencies, another complication that reduces the ability to distinguish between half and full siblings is the inevitability of genotyping error. Microsatellite markers—short simple repeat regions present throughout the genome—are among the most common markers used for sib-ship reconstruction because of their highly polymorphic nature (Ashley et al. 2008, Conflitti et al. 2022, Geib et al. 2015, Hama-Ali et al. 2015, Li et al. 2013, McGrady et al. 2021). However, these markers are prone to two types of errors: allelic dropout, in which one allele fails to amplify giving the appearance of a homozygous genotype, and mistyping errors, in which alleles are incorrectly called or mistyped during genotyping (Wang 2004). These types of errors can mimic different alleles from different paternal lines and thus erroneously create full- or half-sibling groups. Therefore, estimating genotyping errors when reconstructing sib-ship relationships is of critical importance for accurate colony abundance estimations.

Bumble bees (genus *Bombus*) are a group of social pollinators (excluding a small proportion of kleptoparasitic species) of conservation concern that provide critical pollination services to a variety of crops (Artz and Nault 2011; Goulson et al. 2008, McGrady et al. 2021). While most bumble bee species are monogamous, some species deviate from this assumption (Fig. 1). Still, mating frequencies have been empirically studied for only about 10% of the 220 social species of bumble bees, and in many cases these estimates are derived from a small number of colonies and genetic markers (Table S1). In particular, within the subgenus *Pyrobombus* polygamy has been reported in several species (Fig. 1). This subgenus includes many important pollinators such as *B. impatiens* and *B. bimaculatus* (Cameron et al. 2007). Despite this, most studies estimating colony density in these species assume monogamous mating (Conflitti et al. 2022; Shmid-Hempel and Shmid-Hempel 2000; Takahashi et al. 2008).

Here, we focus on a novel methodological approach to accurately estimate mating frequency in *B. impatiens*, which is one of the most common and important wild pollinators in Eastern North America (Strange and Tripodi, 2019; McGrady et al. 2021). Because of its effectiveness as a pollinator and amenability to management, *B. impatiens* is commercially produced and used to provide pollination services in agricultural fields and greenhouses (Cnaani 2002; Suni et al. 2017). Population monitoring of this species suggests that it is relatively stable in its range and abundance, compared to some other North American bumble bee species (Cameron et al. 2011). One study (Payne et al. 2003) evaluated mating frequencies in 11 colonies of wild *B. impatiens* and found 3 of those 11 queens mated with 3 males while the rest of the colonies were monogamous, showing a low-levels of polygamy compared to
other Bombus species (Fig. 1). However, this information was based on a few loci and number colonies and average mating frequency could not be calculated due to unknown contribution of males.

In this study, we used microsatellite genotypes of queens and workers from both wild and commercial B. impatiens colonies reared in the laboratory to estimate the mating frequency for wild and commercial B. impatiens queens. Additionally, we used a large dataset of genotypes from wild collected workers to assess the impact of assuming ‘monogamy’ for colony abundance reconstructions species that deviate from the monogamy assumption. We obtained genotypes from queen and worker bees from 20 wild and 10 commercial colonies using an optimized set of 11 microsatellite loci (Table S2). We adapted the model used by COLONY (Jones and Wang 2010; Wang 2004) to create a Bayesian inference algorithm that jointly estimates paternity and genotyping error for each colony to increase the accuracy of kinship estimations. We then evaluated how assumptions of monogamy or polygamy affect estimates of colony numbers based on inferred sibling groups, using a previously published dataset of worker genotypes from 30 agricultural sites across Pennsylvania (USA) (McGrady et al. 2021). Our study provides three major findings that have important implications for estimating colony abundance in social bee pollinators: (1) a more robust estimates of for mating frequencies in wild and commercial B. impatiens; (2) a novel statistical approach that may be implemented for future studies to easily determine mating frequencies of other eusocial species; and (3) a demonstration that monogamous or polygamous mating frequency assumptions have enormous impact on the results of colony abundance estimation from inferred sibships. We use our results to make recommendations to improve the accuracy of sib-ship reconstruction in social hymenoptera.

Materials And Methods

Specimen Collection and Rearing

Wild queen B. impatiens were collected between April and May of 2018 from three locations: University Park, Newport, and Landisville (Pennsylvania, USA). Twenty colonies from wild-caught queens and resulting offspring were then reared following standard protocols (Treanore et al. 2021). Colonies were held in a walk-in incubator under constant darkness, at 28–30 ºC, 60% RH, and supplied ad libitum with a 60% sucrose solution and honey bee collected pollen. Colony maintenance and offspring collection were performed under red light. For the assessment of mating frequencies in managed colonies, ten commercial colonies were purchased from Koppert Biological Systems (Howell Michigan, USA). To quantify mating frequencies in wild-caught colonies, we collected 10 workers after the first worker emergence, and 10 additional workers approximately 3–4 weeks after. The queen was removed and frozen along with the rest of the workers after collections of the 2nd cohort. Commercial colonies were purchased from Koppert and left in the field until late summer when the entire colony was removed from the field and frozen. The original queen and 20 sampled workers were removed for microsatellite genotyping analysis.
Microsatellite Genotyping, Paternity Reconstruction, and Power Analysis

After pinning, the right mesothoracic leg from each bee was removed and relocated to individual wells in a 96-well PCR plate for DNA extraction. A total of 150 µl Chelex® 100 (5%, in milli-q H2O) and 5 µl of 10 mg/mL Proteinase K was added to each well plate containing a leg. Each sample was then heated in a Mastercycler® pro thermocycler for 60 minutes at 55°C, 15 minutes at 99°C, 1 minute at 37°C, and 15 minutes at 99°C. Extracted DNA samples were stored for up to 3 weeks in a 4°C refrigerator. We amplified 11 microsatellite loci with different concentrations optimized for multiplex amplification Table S2. Polymerase chain reactions were performed using a Mastercycler® pro thermocycler with the following cycle settings: 3 min 30 s initial denaturation at 95°C, followed by 30 cycles of 30 s at 95 ºC, 1 m 15 s at an annealing temperature of 55 ºC and 45 s at 72 ºC, then a final 15 minute extension at 72 ºC with a hold temperature at 15°C. Following amplification, PCR products were diluted 1:10 with molecular water, then sized on an ABI3730XL® DNA Analyzer (Applied Biosystems) with GeneScan LIZ 500 internal size standard (Applied Biosystems) at the Penn State Genomics Core Facility. Fragment analysis results were imported into Geneious Prime v.2022.0.1 (Biomatters Ltd) for genotyping and alleles were scored manually. To jointly estimate paternity and error rates while including known queen genotypes, we developed a Bayesian approach to paternity reconstruction based around the genotyping error model in Wang (2004). Briefly, the model assumes that observed genotypes are perturbed from true genotypes by two classes of errors: dropouts (1 where a heterozygous genotype appears homozygous) and mistyping (2 where a given allele appears to be another allele). The likelihood is calculated by summing (integrating) over possible maternal, paternal, and offspring genotypes conditional on a paternity assignment. We used a nonparametric prior (a Dirichlet process) on full sibling groups, and developed a Gibbs sampler that targets the joint posterior distribution of error rates and paternity assignments. The model and Gibbs sampler are detailed in Supplementary Methods 3, and are implemented in an R package (https://github.com/nspope/paternity_DP). To assess the accuracy of the method under realistic conditions, we applied it to 100 simulations for each combination of number of fathers (1 to 6) and error rates (0.01, 0.05, and 0.1 for both classes of errors). Each simulation used true parental genotypes that were sampled according to Hardy-Weinberg equilibrium given the observed maternal allele frequencies from wild colonies; and produced observed genotypes for 20 offspring and a mother. In each simulation, one father was assigned the majority of offspring and each additional father was allocated only one offspring, as this represents the most challenging conditions for distinguishing genotyping errors from distinct paternal genotypes.

We then applied this novel paternity reconstruction method to the empirical genotypes from each colony, after removing monomorphic loci and a small number of “floater” bees. These are individuals who originated from a different mother colony, which can result from workers who escape then return to a different colony during colony maintenance procedures. The identification of floaters was accomplished by modifying the model described above to include a second, unobserved queen. After paternity reconstruction, we calculated effective mating frequency – defined as \( \frac{1}{\sum p_i^2} \) where \( p_i \) is the proportion of
each offspring that belongs to father \( i \) (Starr 1984) – by averaging over 1,000 Markov chain Monte Carlo (MCMC) samples for each colony. We also calculated posterior probabilities for possible numbers of fathers for each colony, which gives a natural measure of uncertainty regarding monogamy and polygamy. All following statistical tests were computed in R v.4.1.1 (R Core Team 2020).

**Colony Abundance Estimation From Sampled Worker Genotypes**

In order to determine the influence of polygamous or monogamous assumptions when recreating sibships from genetic data, we reanalyzed a previously published data set of 6,306 *B. impatiens* workers collected from 30 sites in Pennsylvania (USA) (McGrady et al. 2021). For each site, the genotypes of the workers were imported into COLONY v.2.0.6.5 (Jones and Wang 2010; Wang 2004), and analyzed twice: once with the assumption of female monogamy, and once with the assumption of female polygamy, as the software only provides a binary parameter for mating frequency. For monogamy, the number of full-sibling groups was recorded as the number of detected colonies and the number of individuals comprising each colony was recorded. For polygamy, the estimated number of mothers was recorded as the number of detected colonies in order to include half-siblings. Additionally, the number of workers comprising each colony was also recorded. Mating frequencies were estimated from the output of COLONY for polygamy to compare it with our empirical estimation of mating frequencies in *B. impatiens*. A Wilcoxon rank-sum test was used in order to determine if mating frequency differed between monogamous and polygamous colonies. Total estimated colony abundance was estimated in R with the package CAPWIRE (Miller et al. 2005) for both monogamous and polygamous colonies from each site. A Welch’s paired \( t \)-test was performed in R v.4.1.1 (R Core Team 2020) to determine if the number of detected and estimated colonies differed when calculated under assumptions of monogamy and polygamy.

**Results**

**Amplification Success**

We genotyped a total of 730 bees from 30 different colonies. Workers were discarded from analysis if genotypes were missing for more than 50% of loci. We removed a total of 74 bees or an average of 2.47 bees per colony from the analysis due to missing genotypes. Loci with greater than 50% missing genotypes were also removed from each analysis on a colony by colony basis. Of the 11 loci analyzed, locus BT28 was entirely monomorphic and removed from all subsequent analyses. Of the remaining loci, 9 were highly polymorphic with \( \geq 4 \) alleles per locus (Table S2). Each colony was evaluated with no fewer than 6 polymorphic loci, following the minimum set by Lepais et al. (2010). Due to data quality and insufficient number of workers, colony 13 was only evaluated with 3 polymorphic loci and was removed from subsequent analyses.

**Method Validation through Power Analysis**
In order to demonstrate the accuracy of our paternity reconstruction method, we performed a power analysis using colonies simulated using the observed maternal allele frequencies at the 11 microsatellite loci (Fig. 2). These simulations illustrate that the estimated number of fathers is accurate regardless of error rate when there are 1–3 fathers (Fig. 2A), and accurate for arbitrary numbers of fathers when error rates are low (1 = 0.01 and 2 = 0.01). When error rates were high (1 and 2 = 0.1), the model did not distinguish high numbers of fathers (>3) from genotyping errors. Further, error rate estimates are accurate regardless of the number of fathers (Fig. 2B). Power to distinguish monogamy versus polygamy was high regardless of both the true number of fathers or the error rate.

**Paternity Reconstruction**

Of the 20 wild caught colonies, we found 17 monogamous colonies, 2 colonies with 2 fathers, and 1 colony with 3 fathers (Fig. 3A). Of the 10 commercial colonies, 7 were monogamous, 2 showed 2 fathers, and 1 showed 3 fathers within the sibling set. Contributions of the 2nd or 3rd father to the offspring pool were less than 30% of the progeny (Fig. 3B). Average effective mating frequency was $1.075 \pm 0.18$ for wild colonies and $1.154 \pm 0.25$ in commercial colonies. A Wilcoxon rank sum test showed that wild and commercial colony effective mating were not significantly different ($W = 102$, $p = 0.38$). In 14/30 colonies, we identified individuals that were not related to the original queen suggesting the presence of floaters. In two polygamous colonies, the offspring of the secondary fathers were all from the 2nd cohort of workers. However in the other polygamous colonies the minority fathers offspring were from both the 1st and 2nd cohort of workers. Estimated genotyping error rates across colonies for dropouts (1) ranged from 0.0025 to 0.0282 (median 0.0094) and for mistyping (2) ranged from 0.0027 to 0.0476 (2 median 0.0092).

**Impact of Mating Frequency Assumptions on Colony Abundance Estimation**

From the 6,306 bees collected from 30 sites, we found that the average number of detected colonies per site was greater using the assumption of monogamy (177 colonies) than using the assumption of polygamy (79 colonies) ($df = 35.87$, $p < 0.001$ (Fig. 4C). The estimated effective mating frequency based on COLONY output for polygamy was 2.31. Similarly, the total estimated number of colonies derived from CAPWIRE was greater when monogamy was the chosen COLONY parameter (861 colonies per site) than when polygamy was the chosen COLONY parameter (102 colonies per site) ($df = 29.057$, $p$-value $< 0.001$) (Fig. 4D). The inflated numbers of total estimated monogamous colonies compared to total polygamous colonies were driven by an excess of singleton colonies (colonies represented by only one detected worker) (Fig. 4A and B). When assuming monogamy, singleton colonies were by far the most common, compared to polygamy where singleton and doubleton colonies are more equally represented (Fig. 4A and B).

**Discussion**
Here, we estimated effective mating frequency of wild and commercial *B. impatiens* colonies using a novel statistical method that jointly estimates paternity and error rates, and a robust number of microsatellite markers. We found that the effective mating frequencies were similar for commercial and wild colonies (1.15 and 1.07, respectively). We verified our novel statistical method through a power analysis that demonstrates its accuracy in predicting both number of effective mates per queen and dropout and genotyping error. We also demonstrated that colony abundance estimations from genetic data are very sensitive to mating system assumptions. The number of detected and estimated colonies in the population was twice as high when using the assumption of monogamy than polygamy highlighting the need to incorporate mating frequency information as a parameter for colony abundance estimation.

Due to the highly biased contributions of each father in polygamous colonies, *B. impatiens* still has mating frequencies very close to monogamy despite violations of this assumption in about 30% of the colonies. Future software developments should allow for the incorporation of accurate mating frequency information to improve accuracy of sib-ship reconstruction.

The accurate estimation of mating frequency in eusocial pollinators has important implications for their conservation, ecology and sociobiology. However, for bumble bees, we found that only 27 of the 250 *Bombus* species have been evaluated for mating frequency and 7 of those species show deviations from monogamy, particularly within the subgenus *Pyrobombus* (Fig. 1). Due to findings of complete genetic monogamy in *Bombus* species, many studies have assumed monogamy when reconstructing sib-ships from genetic data in *Bombus* (Conflitti et al. 2022; Jha and Kremen 2013; McGrady et al. 2021; Schmid-Hemple 2000; Takahashi et al. 2008; Timberlake et al. 2021). Our results demonstrate that assuming monogamy for colony abundance estimation in *Bombus* may lead to an overestimation of the number of colonies. This overestimation of colony abundance could cause flawed decisions for the protection of species that may be of conservation concern (Carvell 2002; Goulson et al. 2008; Memmott et al. 2010).

Similarly, estimates of the number of bumble bee colonies providing pollination services in agroecosystems may be overestimated leading to erroneous conclusions about the abundance and resilience of wild bumble bees that inhabit agroecosystems (McGrady et al. 2021). However, we also show that the estimated mating frequency in the polygamous COLONY reconstruction was significantly higher than the actual mating frequency of *B. impatiens*. Our data supports that using monogamous assumptions is closer to the actual mating frequencies in *B. impatiens* than the polygamous assumption. Future iterations of sib-ship reconstruction software (such as COLONY) should consider the incorporation of a mating frequency parameter in which the actual mating frequency of the species can be used in the estimations.

Accounting for genotyping errors is crucial for accurate sib-ship reconstruction for methodologies to distinguish between half-siblings and erroneous genotypes. Through a power analysis, we showed that our method is robust and has greater than a 90% chance of accurately predicting monogamy when true error rates are less than 0.1 (Fig. 2A). When error rates were low (~ 0.01), there was also a greater than 90% chance of accurately determining the number of fathers for queens that mated between 2 and 4 times. However, this accuracy to detect multiple fathers was limited above 5 matings, especially as error rates increase. Our power analysis also showed that we can have confidence in the accuracy of our
empirically estimated median error rates (Fig. 2B). Given that our estimated error rates were less than 0.01, we also have confidence in the accuracy of our sib-ship reconstruction and paternity estimates. Our method also allowed us to accurately detect and remove identified bees that were likely unrelated to the original queen. These floaters likely escaped from their original colony and rejoined a different colony during the process of colony care and sampling workers in the walk-in growth incubator. Finding individuals that are unrelated to the queen — also known as “drifters” — is common in the wild and has been observed in various social species including bumble bees (MacKenzie et al. 2021; Zannette et al. 2014). Unaccounted for, these floaters would erroneously increase mating frequency estimations within colonies.

In general, it is assumed that species in the genus *Bombus* are mostly monogamous (unlike honey bees, genus *Apis*) due to the importance of maintaining kin selection in primitively eusocial colonies (Boomsma 2009). Even though monogamy increases the genetic relatedness among individuals in a colony and facilitates kin selection through increased inclusive fitness (Arnvist and Nilsson 2000; Boomsma 2009; Fromhage and Kokko 2011), monogamy also results in lower genetic diversity of offspring. Lower genetic diversity at the colony level increases susceptibility to pathogens and decreases colony survival (Baer and Schmid-Hempel 1999, 2001, 2003). This cost of reductions in genetic diversity may outweigh the benefits to kin selection in certain eusocial species, resulting in the evolution of polyandry or multiple female mating (Boomsma 2009). Thus, some species have secondarily evolved polyandry as observed in honey bees and some stingless bees (Neumann et al. 1999; Paxton et al. 1999; Tarpy et al. 2015). Additionally, increasing colony genetic diversity may have other benefits to the colony through reductions of parasitism (Baer and Schmid-Hempel 1999, 2001, 2003, Shykoff and Schmid-Hempel 1991, Tarpy 2003), sex ratio conflict (Ratnieks and Boomsma 1995), diploid male production (Tarpy and Page 2002), and increases in the efficiency of worker performance (Fuchs and Moritz 1999, Jones et al. 2004) and tolerance to variable climatic conditions (Crozier and Page 1985). Therefore, the observed pattern of polyandry in some species of *Bombus* may be the result of unique selective pressures or life history traits that make increased colony genetic diversity more advantageous (Baer 2005).

In this study, we confirm that *B. impatiens* exhibits a strong paternity skew meaning that one father contributed to a much higher majority of the offspring than other mates. We found that 70% or more of the offspring in the colony shared one father (Fig. 3B). One strategy used by primitively eusocial species to reduce the costs of reduced intracolony relatedness while increasing genetic diversity is to develop progenies with high paternity skews (Jaffé et al. 2012). Hypotheses proposed to explain paternity skew in eusocial bees include cryptic female choice, which implies that queen bees are able to influence which sperm is used after copulation, and sperm competition (Eberhard 1996; Baer 2005). Sperm storage in bees, especially in non-*Apis* species, is relatively understudied. Franck (1999, 2002) found that due to the deterioration in sperm clumping over time, *A. mellifera* queens have more variation in the contribution to offspring of each father directly after mating, and as time goes on the contributions become more equal as sperm mixes. However, studies of paternity in highly polygamous *A. mellifera* queens have found that queens use the sperm of all mates in representative proportions, with no bias towards the first or last
Due to these findings, the alternative hypothesis of "genotype scrambling" has been suggested for *Apis*, in which after discarding any unwanted ejaculate, queens may purposefully mix sperm in order to increase genetic diversity of the offspring (Boomsma and Ratnieks 1996). Our findings corroborate other studies which have shown there is a large bias of paternity towards a particular male in bumble bees (Holmes 1974; Huth-Schwarz et al. 2011; Wilkes 1966). Our data also showed that in 2 out of the 3 polygamous colonies that were sampled temporally, offspring of the secondary father were only present in the second generation worker cohort. Due to the small sample size of polygamous colonies, we are unable to determine whether the paternity skews occur by cohorts or seasonally. However, future studies that evaluate mating frequency in *Bombus* should also sample workers that emerge at multiple timepoints in order to determine how paternity or paternity skew varies throughout the queen and colony’s lifecycle.

In summary, we confirm low levels of polyandry and strong paternity skews in *B. impatiens* using (1) a large number of wild and commercial colonies, (2) a large set of microsatellite markers, and (3) a novel and more accurate statistical methodology. The combination of our results and the implementation of a novel methodology highlights the power of jointly estimating paternity and genotyping error when reconstructing sib-ship relationships. By demonstrating that the assumptions of either monogamy or polygamy have huge impact on colony estimations, our results call attention to the need of incorporating mating frequency into statistical approaches that estimate colony abundance from genetic data rather than relying on a binary mating system parameter. Our results also highlight interesting areas for future research including the mechanisms of sperm competition and drifting behaviors in *Bombus*.

**Declarations**

**Acknowledgements**

We would like to thank the Hines and Amsalem Labs at Penn State and in particular Briana Wham and Erin Treanore for assisting with collection and rearing of the *B. impatiens* colonies used in this study. We would also like to thank the Youngsteadt Lab at NC State for providing the commercial *B. impatiens* colonies used in this study. Finally we would like to acknowledge the López-Uribe lab for comments on earlier versions of the manuscript.

**Funding**

SB was supported by the Pennsylvania State University Graduate Black Fellowship. NP was supported by NSF DEB-2046474. CMM was supported by USDA NIFA under award number 2012-51181-20105, Developing Sustainable Pollination Strategies for U.S. Specialty Crops. MML-U was funded through the NSF DEB-2046474 and the USDA NIFA Appropriations under Projects PEN04716 Accession No. 102052.

**Competing Interests**

The authors declare that they have no competing interests.
Author Contributions

SAB: Performed lab work, data analysis, and drafted manuscript. CMM: Conceived project idea and experimental design, collected worker bees, obtained reagents and primer sets. NSP: Developed code and statistical approach for paternity analysis, drafted manuscript, provided supervision for data analysis. SJF: Obtained funding, provided supervision, and edited manuscript. MML-U: Provided supervision for molecular work and data analysis, drafted and revised manuscript, provided funding. Manuscript was approved by all authors before submission.

Data Availability

The datasets generated and analyzed during the current study are available through a Dryad repository [Link].

References


**Figures**
Figure 1

Reported estimates of mating frequency in the literature for species in the genus *Bombus*. A phylogeny showing relationships between each species is shown to the left of the y axis. Phylogeny was adapted from Cameron et al. (2007). The x axis shows the effective mating frequency as calculated by Starr (1984). Mating frequency of 1 represents a monogamous species. *Bombus* species are organized phylogenetically on the y axis. Sample sizes (number of colonies evaluated) are shown next to bars. Additional information and citations can be found in Table S1.
Figure 2

Accuracy of paternity reconstruction method for estimating number of fathers and genotyping error rates on simulated data. (A) Average posterior probabilities (over 100 simulated colonies) for the estimated number of fathers (y-axis) given the true number of fathers (x-axis). Note each “column” sums to 1. Additional fathers beyond the first are only assigned one offspring (out of 20 total offspring/simulation) to emulate high paternity skew. The method is accurate for low error rates, but at high error rates cannot distinguish many fathers (>3-4) from errors. (B) Estimated genotyping error rates across simulations. For each class of errors a global rate was used across loci. Boxplots show the distribution of the posterior mean across simulations. Regardless of the true number of fathers, error rates are estimated accurately.

Figure 3
Results of paternity estimation for 20 wild and 10 commercial *B. impatiens* colonies. (A) Paternity estimates for each colony showing the posterior probabilities for the possible number of different fathers on a gradient of blue based on the accuracy of each paternity scenario per colony. The best estimate of the true number of fathers represented in each colony is marked with a black dot. Average effective mating frequency for wild colonies was 1.07 and for commercial colonies was 1.15. (B) Progeny contribution of each father in polygamous colonies showing that one father contributes a majority proportion of offspring in polygamous colonies. The x axis shows each polygamous colony. The proportion of offspring is shown on the y axis and bars are shaded based on which father contributed each proportion of offspring.

Figure 4
Testing the effect of the monogamy assumption on estimates from COLONY and Capwire analyses of field collected workers from 30 sites. (A - B) Histograms show the frequency of the number of individuals per colony detected in COLONY under the assumption of monogamy (A) and polygamy (B). (C) Boxplots showing the number of detected colonies by COLONY using assumptions of monogamy or polygamy (D) Boxplot showing the colony abundance per site estimated by CAPWIRE using the two outputs from COLONY.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary1.pdf
- Supplementary2.pdf
- Supplementary3.pdf