Journal of Insect Science, (2021) 21(6): 2; 1–8
https://doi.org/10.1093/jisesa/ieab048
Research

ENTOMOLOGICAL SOCIETY OF AMERICA
SHARING INSECT SCIENCE GLOBALLY

OXFORD

Reproductive and Morphological Quality of Commercial Honey Bee (Hymenoptera: Apidae) Drones in the United States

Bradley N. Metz[1,3,] and David R. Tarpy[1,2]

[1]Department of Entomology & Plant Pathology, NC State Apiculture, Campus Box 7613, North Carolina State University, Raleigh, NC 27695-7613, USA, [2]Biology Graduate Program—Ecology & Evolution, NC State Apiculture, Campus Box 7613, North Carolina State University, Raleigh, NC 27695-7613, USA, and [3]Corresponding author, e-mail: bnmetz@ncsu.edu

Subject Editor: Hongmei Li-Byarlay

Received 26 March 2021; Editorial decision 23 June 2021

Abstract

Exploration into reproductive quality in honey bees (*Apis mellifera* Linneaus (Hymenoptera: Apidae) largely focuses on factors that affect queens, with drones primarily being considered insofar as they pass on effects of environmental stressors to the queen and subsequent offspring. In those studies that consider drone quality explicitly, a primary focus has been on the dimorphic nature of drones laid in worker cells (either through rare queen error or worker reproduction) as compared to drones laid by the queen in the slightly larger drone cells. The implication from these studies is that that there exists a bimodality of drone morphological quality that is related to reproductive quality and competitive ability during mating. Our study quantifies the presence of such small drones in commercial populations, finding that rates of 'low-quality' drones are far higher than theoretically predicted under optimum conditions. Observations from commercial colonies also show significant inter-colony variation among the size and fecundity of drones produced, prompting speculation as to the mechanisms inducing such variation and the potential use of drone-quality variation for the colony- or apiary-level exposure to nutrition, agrichemical, or parasitic stressors.

Key words: honey bee drone, queen quality, reproduction, beekeeping, sperm

Commercial honey bee (Hymenoptera: Apidae: *Apis mellifera* Linneaus) queen rearing is a support industry to greater apiculture supplying queens to beekeepers to grow their operations through splits, replace aged queens, enforce brood breaks for disease and parasite management, or to control genetics (Cobey et al. 2012). Annual queen replacement is almost the norm in beekeeping in the US, where thousands of queens are shipped throughout the country (Cobey et al. 2012). The queen is often a major focus of management effort and response when colony problems occur (e.g., slow growth, excess parasite load, disease prevalence, poor productivity, or declining populations) (Steinhauer et al. 2014). This is in part because she is the sole reproductive member of the colony (usually) but also because in many cases she is a *product* produced outside of the control of the beekeeper, and therefore the subject of contention when colonies fail. Therefore, much attention has been paid to assess and understand the quality of commercial queens and the extent to which colony-level phenotypes commonly attributed to queens are in fact their 'fault' (Tarpy et al. 2012, Lee et al. 2019). A particular finding of the research into queen quality is that the major factors leading to her heading a productive colony or those related to her mating (e.g., number of partners, sperm count, and sperm viability) are correlated with colony growth and survival (Collins 2004, Tarpy et al. 2013, Pettis et al. 2016). These traits are themselves correlated to measures of queen size (e.g., body mass, thorax width, and head width), albeit weakly (Delaney et al. 2010). Commercial queen breeders exert control over queen size, through larval selection during grafting, and control of the 'cell-builder' colonies that raise and feed the larval queen. Control over mating quality is generally less direct. The common practice in the US is to conduct open mating; that is, to allow commercially reared queens to fly freely and mate for themselves. As such, while drone colonies are often provided in adjacent apiaries, little direct manipulation over drones produced is exerted by breeders. As each drone partner contributes to the mating quality of the queen and ultimately the likelihood of her success, understanding the quality of drones during mating may be an important missing element in producing high quality queens.

Unlike the solitary queen, colonies may produce thousands of drones at a time (Page and Metcalf 1984), which are likely subject to pre-mating competition owing to this high ratio of males:females in aerial aggregations (Baer 2005, Brutscher et al. 2019). Evidence for this is that larger drones are more likely to mate successfully and to contribute a higher paternity share to the worker population

(Kraus et al. 2003; Schlüns et al. 2003, 2004; Couvillon et al. 2010). Furthermore, the larger body size is associated with higher fecundity, with larger drones tending to have higher sperm counts and sperm viability, suggesting that—with regard to mating quality—bigger is better (Gençer and Kahya 2020). However, the majority of these studies were performed using experimentally reared drones from worker cells or by laying workers. Because the rate of drone production by workers in queen-right conditions is quite low (<0.1%; Visscher 1989) and normal queens rarely lay unfertilized eggs in worker-sized cells (Ratnieks and Keller 1998), drone size variation may be much more restricted in natural or commercial settings than in experimentally induced ones. Size may not be expected to be the primary factor in the variation of the fecundity of commercial drones, however, and for several reasons. First, adult exposure to the neonicotinoids clothianidin and thiomethoxam through pollen patties at doses similar to that found in the field treated pollen (Straub et al. 2016) or fipronil in sugar syrup at doses reported to be similar to that found in treated plants (Kairo et al. 2017) decreased spermatozoa viability from the seminal vesicles and ejaculated spermatozoa count, respectively. Second, colony-level exposure to commercial formulations of coumaphos, applied according to directions, resulted in decreased viability of ejaculated spermatozoa (Burley, 2007). Combined, these studies indicate that exposure to incidental or targeted agrichemical may play a factor in drone reproductive quality. Finally, damage to drone sperm may confer fragility to the surviving spermatozoa, such that low sperm quality may affect the queen rather than simply preventing the drone from capitalizing on a successful mating attempt (Kairo et al. 2016, Straub et al. 2016).

In this study, we aimed to ascertain the variation in both size and fecundity traits of drones produced in commercial apiaries and to compare the variation in these traits to that of experimentally reared large and small drones. Our goal is to elucidate the breadth of drone variation to provide a baseline for observing changes in those traits as upstream predictors of queen mating quality and colony health.

Materials and Methods

Queens and Colonies

Drone-Laying Queens

Drone-laying queens were procured from a commercial apiary where they were identified as such by the beekeeper during post-mating screening. They were shipped overnight to the Lake Wheeler Honey Bee Research Facility, Raleigh NC, where they were temporarily banked in individual cages placed within a strong, queenright colony for 1–6 wk. In April–May 2018, four queens were introduced into individual five-frame, standard deep, nucleus hives each standardized with one frame of honey, one frame of pollen and young brood, one frame of capped brood, and one empty frame to provide open comb space so that the queen could lay eggs. Six frame-faces of adult bees (i.e., three frames front and back) were included in the initial setup to establish equivalent adult bee populations among nuclei (~6,000) with the populations maintained by adding newly emerged bees from unrelated sources.

Normal-Laying Queens

Similar to colonies not being able to rear high- and low-quality queens at the same time (Tarpy and Mayer 2009), we were unable to successfully raise or foster worker- and drone-cell reared drones simultaneously in the same colonies. We, therefore, selected five separate colonies headed by naturally mated queens to raise drones from drone cells. Drone rearing and fostering occurred in the same manner as worker-cell reared drones, albeit in separate colonies. Colonies headed by normal-laying queens were fed supplemental pollen from patties but were restricted from collected forage pollen for an unrelated study.

Drone Rearing and Collection

Drone collection occurred from April to June 2018. All six experimental queens were confirmed to be normal or drone-laying by observation of capped brood. Each queen was confined to a single frame face using a push-in cage constructed of the queen excluder and hardware cloth, sized to completely cover a single frame face (45 × 20 cm). Drone-laying queens were forced to lay on a worker-cell frame (mean cell diameter 5. 83 ± 0.03 mm) producing worker cell-reared drones (WC) while the normal queens laid on a drone-cell frame (mean cell diameter 6. 55 ± 0.03 mm) producing drone cell-reared drones (DC). The experimental frames were removed from each colony when the capped brood were approximately 24 h from emergence (based on eye color and pigmentation of a few pupae that were uncapped and sacrificed to stage the cohort), introduced into individual metal and screen cages (49 × 25 × 12 cm) with ~100 worker bees collected from the brood nest of the source colony (along with a 1:1 sugar syrup feeder to aid in emergence), and placed into an incubator set at 34°C and ~50% RH.

Daily between 8:00 and 9:00 AM, emerged drones were collected into rearing boxes constructed of a wooden frame (127 mm × 127 mm × 25.4 mm) with #8 hardware cloth (~3.2 mm mesh) for sidewalls to prevent escape of the smaller drones. Each cage was stocked with drones from a single source/emergence date. Cages were placed into unrelated foster colonies, which were populous (>30,000 adult workers), headed by a naturally mated queen, and utilized for no other purpose. A 5 cm insert was placed on the top-box and under the lid creating space to place the cages flat atop the frames. This enabled the resident workers access to the drones without disrupting colony dynamics. Each foster colony supported 4–7 cages of up to 100 drones each. Cages from a particular queen source were distributed roughly equally among the foster colonies to negate adult environment as a confounding variable; because the experiment proceeded over several weeks, drones of the two size classes (WC and DC) were not co-fostered. Up to 20 WC drones from each source were collected from the foster colonies when they were aged 6, 12, 19, and 30 d of age. These ages were indicated as biologically relevant in prior studies, with 6 d being the point at which 50% of sperm migrated to the seminal vesicles, 12 being the mean age of mating flights as per, 19 d being the age of maximum sperm count, 30 being the mean age of death (Rueppell et al. 2005, Metz and Tarpy 2019). The study was replicated with four drone-laying queens.

Drones reared in drone cells (=DC drones) were reared to 12 d of age exclusively, owing to changing apiary conditions (e.g., rising temperatures, onset of dearth, and the end of the natural breeding season) resulting in the cessation of successful rearing and fostering of drones. Drones were collected from their foster colonies between 8:00 – 9:00 AM and immediately transported to the lab in rearing cages free of workers. Water was rubbed onto the wire mesh of each cage for hydration, and the cages were placed in a bench-top incubator set to 34°C for processing as below.

Variation of Drones From Commercial Apiaries

From April-May, 2018, April-Mach, 2019, and April 2020, beekeepers from commercial apiaries were solicited for the sampling

of drone size and fecundity. Respondents from 19 operations collected drones from colonies of their choosing and shipped them overnight to NC State where they were processed as below. Sampling and shipping methods differed somewhat throughout the sampling period. However, the most widely applied methods were as follows. In 2018, beekeepers sampled drones by blocking the entrance to the nest with ~6.4 mm mesh hardware cloth during the afternoon (approx. 2–4 pm depending on location and day) and collected returning drones with forceps. From 2019 onward, beekeepers instead opened the colonies and selected an outer frame with a large number of drones. Drones were lightly pressed on the abdomen and drones that responded by buzzing their wings and actively running throughout the frame were selected as being on the verge of flight initiation. As drones of the age to initiate flight are likely to have sperm migrated to their seminal vesicles, flight behavior is taken as reasonable rapid proxy for maturity in this study (Rueppell et al. 2005, Metz and Tarpy 2019).

In either case, approximately 30 drones were collected in this manner and placed into shipping cages made from plastic dishes provided with candy made of 1:1 Karo brand corn syrup and powdered sugar, water provided through a damp dental wick, a TempQueen lure (Mann Lake: Hackensack, MN), and ~100 workers collected from the brood nest. Participants shipped all samples overnight to NC State Apiculture for processing as below.

Drone Dissection and Sperm Testing

Drones were dissected and processed following procedures outlined in Metz and Tarpy (2019). Briefly, each drone was lightly anesthetized with CO_2 delivered via a low flow stream, and the intersegmental membrane between abdominal segments 5–6 was pierced to minimize the chance of ejaculation during dissection. The drone was then weighed and pinned for photography of their head and abdomen. The seminal vesicles and mucus glands were then removed and photographed, with the seminal vesicles cut free for subsequent sperm analysis. The head, abdomen, and wings were then cut away and the thorax and legs weighed to the nearest 0.1 mg. All drone remains were then stored in a 1. 5 ml sterile, enzyme-free microcentrifuge tube and preserved at −80°C for future analysis.

Seminal vesicles were ruptured in 1000 micro-l buffer D (Collins and Donoghue 1999, Makarevich et al. 2010), mixed with forceps, and pipetted into an amber glass vial containing 10 micro-l (Invitrogen Live/Dead sperm staining kit #L7011; 1 mM in DMSO) diluted 1:500 into Dimethylsulfoxide (99.8%) and 10 micro-l propidium iodide solution (2.4 mM in water). Vials were lightly vortexed and allowed to rest at room temperature a minimum of 5 min before reading with a Nexcelom Cellometer Vision Sperm Counter. The number of live and dead sperm cells were counted on three different areas on the slide using the Cellometer Vision Software version 2.1.2.1 2018 Nexcelom Bioscience LLC and the resultant average taken as the final live and dead sperm counts.

Images were processed using ImageJ version 1.51m9 (Schneider et al. 2012). Each image was calibrated to an image of a 1.0 mm glass ruler taken alongside the image using the same microscope and camera settings. The following measurements were taken: width of the head at the widest point perpendicular to the longitudinal body axis; distance between the distal tips of each tegula; mean length of the seminal vesicles along the central axis; mean length of the mucus glands along the central axis. We defined fecundity measures as total sperm count, sperm viability, mean seminal vesicle length, and mean mucus gland length. We defined size measures as body mass, thorax mass, inter-tegula distance, and head width.

Statistical Analyses

All statistical analyses were performed in R version 4.0. (R Core Team 2019); all packages used are cited in Table 1. Descriptions of the individual analytical methods are presented alongside the results. A full code and analytical dataset for these analyses and figures may be provided upon request. All means are reported as ±SEM, unless otherwise noted with a range in parenthetical where informative. We used an $\alpha = 0.05$ to signify statistical significance with corrections for multiple comparisons via a Benjamini–Hochberg false discovery rate of 0.05 (Benjamini and Hochberg 1995).

Results

Variation Among Individual Size and Fecundity Characters By Age and Rearing Cell Size

We reared and fostered WC drones from four different colonies at ages 6, 12, 19, and 30 d and DC drones from 2 different colonies at age 12 d (Table 2). Because WC drones are in essence pseudomutants in this experiment and it is of interest to compare each individual reproductive measure to DC drones, we first performed individual ANOVA comparisons between age-matched (12 d) WC and DC drones for their descriptive statistics (Table 3). All individual measures were significantly higher in DC drones with the exception of sperm count. We next analyzed the effects of Colony and Age of WC drones only, using a full-factorial ANOVA model on each individual reproductive measure, modelling Age as a categorical variable as it was discontinuous and variable (unlike the DC drones). Drone source colony was a significant factor for each of the following independent variables: body mass, thorax mass, thorax width, seminal vesicle length, and mucus gland length (Table 4). Age, however, was only significant for mucus gland length ($F_{3,218} = 19.69$; $P < 0.0001$). For the individual measures of sperm viability, sperm count, and head width, there were significant interactions between Colony and Age (all $F > 2.83$; $P < 0.002$), and as such we do not report the main effects. However, since these individual phenotypes are singular dimensions of overall reproductive quality, a more comprehensive simultaneous assessment is more appropriate.

Principal Component Analyses

Spearman's correlations (Sokal and Rolf 1995) of body size and fecundity measures of age-matched WC and DC drones (12 d) were, similarly to previous results (Schlüns et al. 2003, Gençer and Kahya 2011, Metz and Tarpy 2019), significantly positively correlated (Supp Fig. SI 1 [online only]). We then generated the first principal component (Dunteman 1989) for each correlation cluster to serve as a unified measure of *Size* and *Fecundity* (Supp Table S1 [online only]). Expectedly, Size and Fecundity were themselves positively correlated (Spearman's Rho = 0.52, $P < 0.0001$).

Table 1. R-statistical packages used during these analyses

Package	Publication
broom	Robinson, D., A. Hayes, and S. Couch 2020
corrplot	Wei, T. and V. Simko 2017
factoextra	Kassambara, A. and F Mundt. 2020
gridExtra	Auguie, B. 2017
Hmisc	Harrell, FE Jr., C. Dupont, et al. 2020
RColorBrewer	Neuwirth, E. 2014
readxl	Wickham, H. and J. Bryan 2019
tidyverse	Wickham, H. et al. 2019

Variation of Size and Fecundity Principal Components By Age and Rearing Cell Size

We then analyzed the principal components Size and Fecundity for age and colony-level variation among the WC drones, to observe whether the trends observed in our analyses of individual characters was preserved. In a full-factorial model, Fecundity varied significantly by both drone source colony ($F_{3,217}$ = 12.74, P < 0.0001) and Age ($F_{3,217}$ = 14.52, P < 0.0001) with no significant interaction ($F_{7,217}$ = 1.79; P = 0.09) among WC drones. Size of WC drones varied significantly only by Drone source colony ($F_{3,237}$ = 37.19; P < 0.0001) with no significant variation due to Age ($F_{3,237}$ = 0.81; P = 0.64) or an interaction ($F_{7,237}$ =1.70; P = 0.11; Fig. 1a). Comparing Fecundity as a function of colony and size with Age as a covariate revealed a significant drone source colony by size interaction ($F_{3,220}$ =4.73; P < 0.005), and as such we analyzed those colonies separately; Fig. 1b). This effect is mostly driven by Colony C, which scored highly in the Fecundity irrespective of age ($F_{2,52}$ = 0.67; P = 0.52) but increasing levels for colonies A ($F_{3,54}$ = 7.41; P < 0.001), B ($F_{3,69}$ = 5.43; P < 0.005), and D ($F_{2,42}$ =6.04; P < 0.005). Results are illustrated in Fig. 1b.

Similar to the results for individual characters, same-aged drones reared from drone cells were both larger and more fecund than those reared in worker cells (Table 3). However, colony variation remains important. A main effects ANOVA model including both Rearing cell size and Colony showed that while Rearing cell size was the only significant factor for Fecundity ($F_{1,190}$ = 76.1; P < 0.0001), with no significant effect of Colony ($F_{4,190}$ = 2.10; P = 0.10), this was not true for Size, where both Rearing cell size (Table 3; $F_{1,207}$ = 387.0; P < 0.0001) and Colony ($F_{4,208}$ = 19.2; P < 0.0001) were significant. Experimental colony variation in size and fecundity are visualized in Fig. 2.

Variation of Drones From Commercial Apiaries

We sampled drones from a total of 19 operations each with 2. 11 ± 0.30 (1–4) colonies and a mean of 9. 57 ± 0.98 (1–27) drones each.

Table 2. Sampling of WC and DC drones for each age and colony

		Colony			
Worker cell (WC)	Age	A	B	C	D
	6	12	11	21	20
	12	21	33	22	20
	19	20	20	9	20
	30	10	13	-	-
		Colony			
Drone cell (DC)	Age	E	F		
	12	60	60		

To compare these with our experimental drones, we generated separate Size and Fecundity principal components using the coefficients generated in the prior section. We first compared the variation of individual commercially reared drones with our population of experimental drones. We filtered the experimental drone dataset to only flight-aged drones (≥12 d) to provide the closest match with commercial drones, for which age was not known. We then performed k-means cluster analysis (Hartigan and Wong 1979) on the drones predicting two clusters, referred to here as high- and low-quality for larger, more fecund drones and smaller, less fecund drones, respectively. Euclidean distance was used to form the clusters. We chose two clusters from the *a priori* assumption that clustering would be driven primarily by our experimental populations (WC and DC), although this was confirmed visually by silhouette (Rousseeuw 1987). Experimentally raised drones clustered such that the majority of drones in the high-quality cluster were reared in drone cells and those in the low-quality cluster largely reared from worker cells (Table 5). Commercial drones were classified mostly as high-quality, with about 7% classified as low-quality (Table 5).

We then combined all drones for operational level analysis and calculated colony means for size and fecundity to represent colony and operational variation visually. Colonies from different operations varied significantly in both Size ($F_{18,400}$ = 7.163, P < 0.0001) and Fecundity ($F_{18,371}$ =9.319; P < 0.0001), with the general trend of a positive association among Size and Fecundity being upheld (Fig. 3).

Discussion

Unlike with queens, where each individual is critical for the life and productivity of the colony, thousands of drones are produced and are largely disposable, representing strong analogy to the gametes they attempt to distribute. As such, drones are highly variable, ranging 145.3–289.5 mg in body mass and 0.18–29.0 * 10^6 spermatozoa among commercial colonies in this study.

In our experimental population, about 2% of DC drones classified with the WC drones as 'low quality', suggesting the possibility of other rearing environment factors contributing to reproductive potential besides the dimensions and volume of a given cell. This is similarly true for ~17% of WC drones classified along with the DC drones as 'high quality'. However, we observed some colonies effectively reshaping worker cells, elongating and doming them in such a fashion that overall cell length was much longer than typical. This increased the overall cell volume, suggesting that controlling for cell diameter alone is not necessarily sufficient for generating the size differences among the two experimental drone rearing methods.

Table 3. Comparison of size and fecundity traits among drones reared in worker or drone cells

	Worker cell (WC)	Drone cell (DC)	Stat		*P*-value
N	96	120			
Body mass (mg)	143.7 ± 2.9	196.8 ± 1.8	$F_{1,214}$	256	<0.0001
Thorax mass (mg)	64.5 ± 1.4	91.2 ± 0.7	$F_{1,214}$	333	<0.0001
Head width (mm)	4.07 ± 0.03	4.41 ± 0.01	$F_{1,211}$	134	<0.0001
Thorax width (mm)	5.12 ± 0.04	5.73 ± 0.02	$F_{1,211}$	284	<0.0001
Sperm viability (%)	71.0% ± 1.5%	79.2% ± 1.2%	$F_{1,214}$	19.1	<0.0001
Sperm count (*10^6)	10.50 ± 0.68	10.70 ± 0.41	$F_{1,214}$	0.895	0.345
Seminal vesicle length (mm)	3.21 ± 0.04	3.69 ± 0.04	$F_{1,213}$	82.6	<0.0001
Mucus gland length (mm)	3.91 ± 0.05	4.51 ± 0.04	$F_{1,194}$	86.8	<0.0001
Size	−1.60 ± 0.16	1.31 ± 0.08	$F_{1,211}$	308	<0.0001
Fecundity	−0.84 ± 0.13	0.69 ± 0.10	$F_{1,194}$	74.9	<0.0001

Table 4. Full-factorial ANOVA models for the effects of colony and age on worker cell-reared drone size and fecundity measures

Character	Colony		*P*-value	Age		*P*-value	Interaction		*P*-value
Body mass	$F_{3,238}$	24.01	<0.0001	$F_{3,238}$	1.19	0.316	$F_{7,238}$	1.17	0.320
Thorax mass	$F_{3,238}$	26.89	<0.0001	$F_{3,238}$	1.35	0.259	$F_{7,238}$	0.99	0.438
Sperm viability							$F_{7,238}$	5.21	<0.0001
Sperm count							$F_{7,238}$	2.83	0.008
Head width							$F_{7,237}$	3.43	0.002
Thorax width	$F_{3,237}$	26.27	<0.0001	$F_{3,237}$	0.98	0.405	$F_{7,237}$	1.81	0.086
Seminal vesicle length	$F_{3,234}$	6.03	<0.001	$F_{3,234}$	2.52	0.059	$F_{7,234}$	0.52	0.820
Mucus gland length	$F_{3,218}$	5.65	0.0010	$F_{3,218}$	19.69	<0.0001	$F_{7,218}$	1.48	0.177

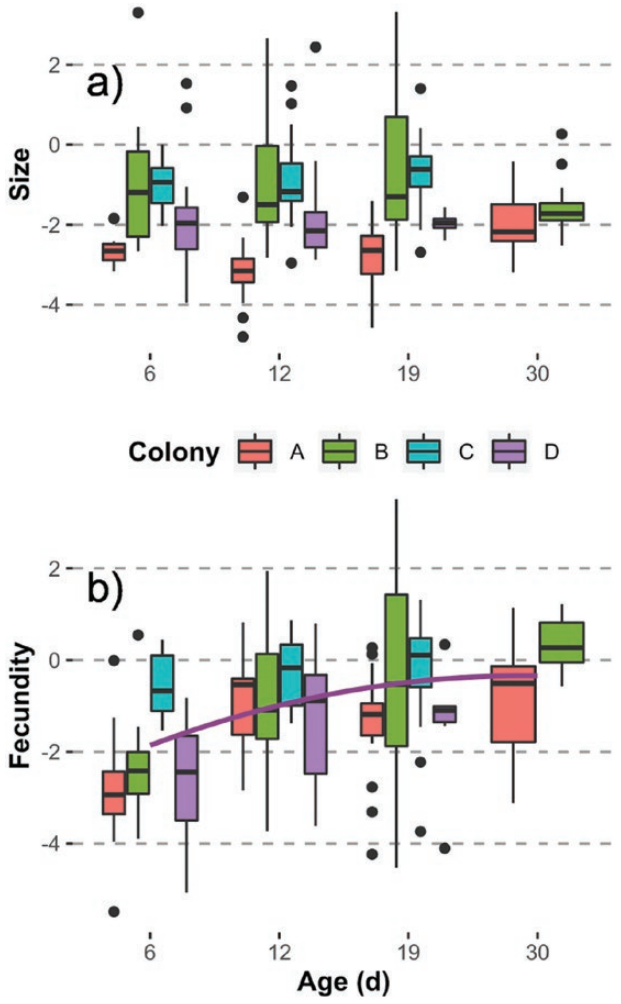

Fig. 1. Variation in size and fecundity among worker cell-reared drones Worker cell-reared drones from four colonies were assessed at five ages for size measures (a) and fecundity measures (b) defined as in the methods. Standard box plots for each group are provided that illustrate the mean, four quartiles, and outliers. Variation in these measures shows a significant size difference among drones from different colony sources that remains largely constant at various ages. Age and Colony both elicited a difference in Fecundity (b). Trend line represents the overall relationship between Age and Fecundity.

It is unclear why some colonies perform this but others do not. Cell size clearly matters, although there appears to be other rearing environment factors at play when determining the size and fecundity of drones.

Commercially reared drones classified primarily with our experimental DC drones (Table 5). The proportion of commercial drones classified as WC in this study (~6.5%) was in line with but somewhat less than the 9.1% reported previously (Berg 1991). If we base our predictions off of a normal laying colony, with a predicted 0.12% of worker-produced drones (Visscher 1989) and no queen laying error (Ratnieks and Keller 1998), this proportion appears to be outsized. One possible cause is population-level variation. While all drones sampled were from mixed lineage new world blends of purportedly European stock populations, temperate and tropically adapted bees vary significantly in size (Rinderer et al. 1985), and drones sampled from apiaries with different populations than that of our experimental population may have different trait ranges despite being ostensibly from similar lineages. Second, when sampling free-flying drones from commercial colonies, it is possible that the drones sampled were not the progeny of the queen from that colony. Drones may drift, sometimes significantly, so that they return to colonies from which they did not originate—up to 150 m away (Currie and Jay 1991), a distance far greater than the ~2 m typically recommended for placing hives in an apiary (Sammataro and Avitabile 2011). It is possible that drones from laying-worker colonies may have been inadvertently sampled, inflating the proportion of small drones, although there is no reason to suggest that this was commonplace. Finally, there is the possibility that the rates of queen laying errors need to be reevaluated.

In the experimental population, inter-colony variation in drone reproductive potential was significant even when accounting for cell size (Fig. 1a), and we observed significant inter-colony variation in the commercial population as well (Fig. 3). In essence, some colonies seem to produce larger, more fecund drones than others, despite being of the same population and every effort made to minimize the possibility of egg-laying error by the queen by caging her onto frames with selected cell size. As has been established elsewhere, correlations of size and fecundity characteristics in honey bee drones are positive and our results confirm this (Schlüns et al. 2003, Gençer and Kahya 2011, Brutscher et al. 2019, Metz and Tarpy 2019). Even within a subset of abnormally small, experimentally derived individuals, larger drones are more fecund. Generally, drone reproductive ontogeny proceeds as expected, with reproductive development increasing throughout the first week of life to a maximum (Fig. 1b). Stürup et al. (2013) showed a variable decline in sperm viability among older drones, a result we did not observe in our time course with WC drones. As discussed above, we observed low numbers of drones surviving to older ages (data not documented). We therefore cannot discount a survivor effect, where those drones (WC or DC) that are more fecund are also more likely to survive. While there are established trade-offs between reproduction and longevity in many model systems (reviewed in Blacher et al. 2017), honey bee females seem to exhibit a lack of such relationships (Rueppell et al. 2004)

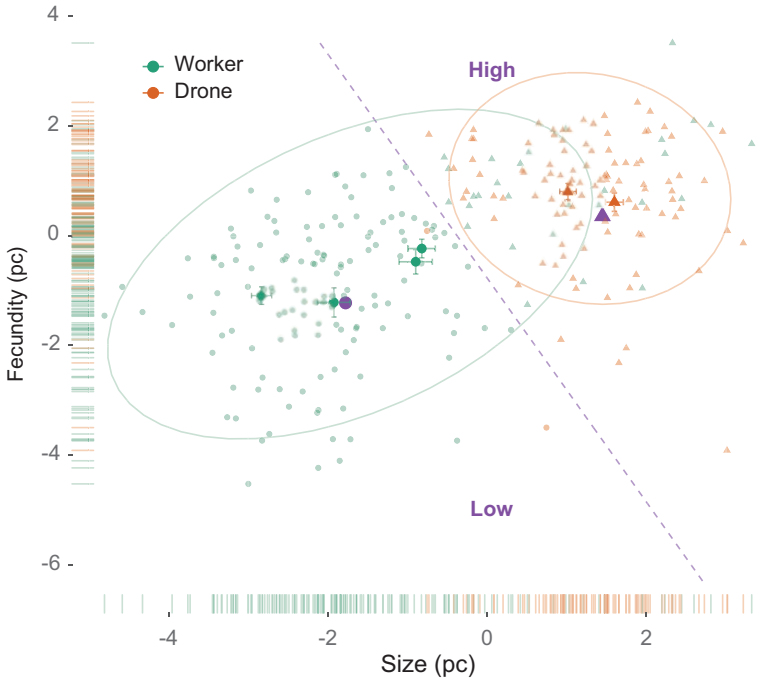

Fig. 2. Classification of high- and low-quality clusters based on comparison to experimental drones reared in worker and drone cells Clusters of worker cell-(green) and drone cell-(red) reared drones. Ellipses show the 95% confidence interval of each class. Colored dashes along the axes show the distribution of DC and WC drones for Fecundity and Size. The shape of each point indicates whether drone was classified into either a high- or low-quality cluster. Larger, colored points represent experimental colony means and standard errors. The purple points represent cluster centroids and the dashed, purple line indicates the cluster border, calculated as the line at which the Euclidean distance from each cluster centroid was equivalent.

Table 5. Classification of drones based on size and fecundity into high and low quality clusters

Drone	Cluster		Total	Proportion (Low/Total)
	High	Low		
Worker-cell	29	141	170	82.9%
Drone-cell	103	2	105	1.9%
Commercial	358	25	383	6.5%
Total	489	169	658	25.5%

and exploring whether fecundity and size are consistent measures of other life history parameters leading to increased longevity is worth future exploration. Body mass of WC drones did not decrease with age, countering previously reported results (Metz and Tarpy 2019) in DC drones. It is possible that these drones, already at a minimum weight, do not have energy or tissue reserves to lose during development. Alternatively, Goins and Schneider (2013) suggested that smaller adult drones may be targets of increased worker attention; therefore, these drones might counterintuitively be more well cared for than their larger brethren and more likely to retain their mass as they age. We observed significant age-by-colony interactions among measures presumed to be static (e.g., head and thorax width). This may well be caused by sampling bias or differential survival among larger and smaller drones, as previously supposed (Czekońska et al. 2019). Trends observed in the principal components followed those of the individual measures, indicating that these may be a useful proxy for full phenotypic variation. Colonies differed in both the size and fecundity of the drones produced, with size not entirely obviating colony differences in fecundity. This suggests that there is additional variation in fecundity not solely explainable by variation in size. Possible environmental exposures to temperature fluctuations (Czekońska et al. 2013, Rousseau et al. 2020), agrichemicals (Burley 2007, Straub et al. 2016), or parasites (Duay et al. 2002) may have occurred in these drones, all of which have been shown to elicit negative effects on reproductive traits, but quantifying all of the possible permutations is beyond the scope of the present study and only adds to the observed variation.

Variation among drones may persist because any given individual drone does not contribute significantly to a colony's fitness. However, looking at drone variation among commercial colonies may yet provide an intriguing insight into colony phenotypes and perhaps health. In this, drones may serve as a useful bioindicator of colony stressors. On the one hand, the presence of drones already serves as a method for identifying laying problems with the queen, as once she is no longer capable of laying fertilized eggs she will only produce drones. On the other hand, a small proportion of drones present in the colony is *also* a measure that the colony has the resources to produce them, as drone production is socially regulated by both the queen and workers (Boes 2010). The quality and variability of drones produced in a colony may serve as indicators of queen laying problems, declining suppression of worker reproduction, or environmental stressors that may be affecting the colony but not yet immediately apparent in other classes. Honey bee males follow a well-established trend of being the 'fragile' sex (Retschnig et al. 2014, Friedli et al. 2020). Drone quality may serve as a bioindicator for colony or apiary level health, which may provide an opportunity for targeted management to occur before potential problems being identified in the subsequent generation of queens.

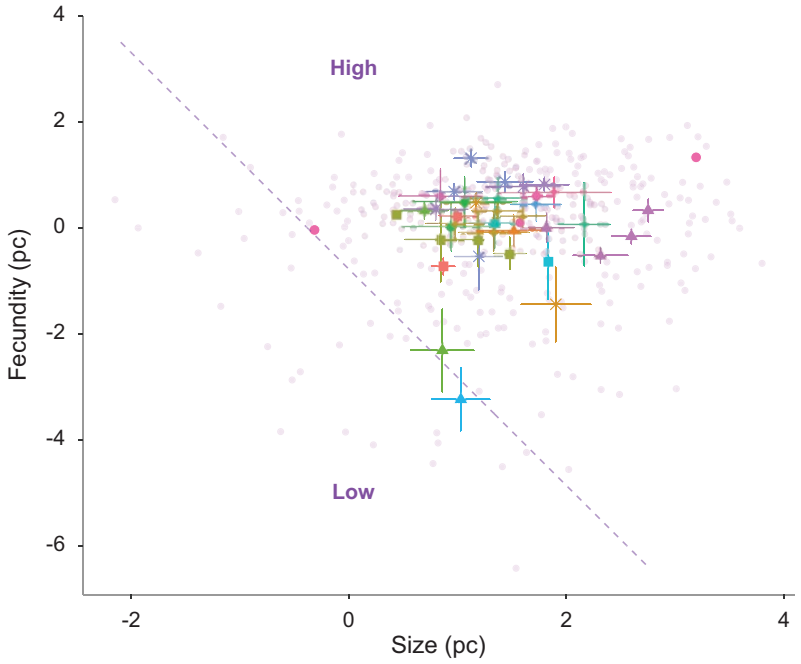

Fig. 3. Size and fecundity variation of colonies and drones sampled from different apiary operations and their classification into high- and low-quality clusters Means and standard errors for each commercial colony. Each operation is represented by a unique combination of color and shape. The dashed, purple line represents the cluster border between high- (upper right) and low- (lower left) quality drones, calculated as the line of Euclidean equidistance from the two cluster centroids. Individual commercial drones in the background are shown in purple.

Supplementary Data

Supplementary data are available at *Journal of Insect Science* online.

Acknowledgments

We wish to thank Jennifer Keller, Erin McDermott, and Hannah Levenson for their assistance in beekeeping and processing hundreds of drones. We also thank the Tarpy Lab Think Tank for constructive feedback and discussion. This study was funded by a generous grant from California State Beekeeper's Association and US DEVCOM Army Research Laboratory grant W911NF1920306.

Author Contributions

BNM: incepted and designed the experiments, analyzed the data, and wrote the original draft. **DRT**: supervised design and conduct, managed all financial concerns, including grant submission, and finalized the writing of the manuscript.

References Cited

Auguie, B. 2017. gridExtra: miscellaneous functions for 'grid' graphics. R package version 2.3. https://CRAN.R-project.org/package=gridExtra

Baer, B. 2005. Sexual selection in Apis bees. Apidologie. 36: 187–200.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate : a practical and powerful approach to multiple resting. J. R. Stat. Soc. Ser. B. 57: 289–300.

Berg, S. 1991. Investigation on the rates of large and small drones at a drone congregation area. Apidologie. 22: 437–438.

Blacher, P., T. J. Huggins, and A. F. G. Bourke. 2017. Evolution of ageing, costs of reproduction and the fecundity–longevity trade-off in eusocial insects. Proc. R. Soc. B Biol. Sci. 284: 20170380. doi:10.1098/rspb.2017.0380

Boes, K. E. 2010. Honeybee colony drone production and maintenance in accordance with environmental factors: an interplay of queen and worker decisions. Insectes Soc. 57: 1–9.

Brutscher, L. M., B. Baer, and E. L. Niño. 2019. Putative drone copulation factors regulating honey bee (*Apis mellifera*) queen reproduction and health: a review. Insects. 10:8.

Burley, L. M. 2007. The effects of miticides on the reproductive physiology of honey bee (*Apis mellifera* L.) queens and drones. M.S. Thesis, Virginia Tech, Blacksburg, VA.

Cobey, S. W., W. S. Sheppard, and D. R. Tarpy. 2012. Status of breeding practices and genetic diversity in domestic U. S. honey bees. pp 39–53, in D. Sammataro & J. A. Yoder (eds.), Honey bee colony health: challenges and sustainable solutions. CRC Press, Boca Raton, FL.

Collins, A. M. 2004. Functional longevity of honey bee, *Apis mellifera*, queens inseminated with low viability semen. J. Apic. Res. 43: 167–171.

Collins, A. M., and A. M. Donoghue. 1999. Viability assessment of honey bee, *Apis mellifera*, sperm using dual fluorescent staining. Theriogenology. 51: 1513–1523.

Couvillon, M. J., W. O. H. Hughes, J. A. Perez-Sato, S. J. Martin, G. G. F. Roy, and F. L. W. Ratnieks. 2010. Sexual selection in honey bees: colony variation and the importance of size in male mating success. Behav. Ecol. 21: 520–525.

Currie, R. W., and S. C. Jay 1991. Drifting behaviour of drone honey bees (*Apis mellifera* L.). J Apic. Res. 30: 61–68.

Czekońska, K., B. Chuda-Mickiewicz, and P. Chorbiński. 2013. The effect of brood incubation temperature on the reproductive value of honey bee (*Apis mellifera*) drones. J. Apic. Res. 52: 96–105.

Czekońska, K., H. Szentgyörgyi, and A. Tofilski. 2019. Body mass but not wing size or symmetry correlates with life span of honey bee drones. Bull. Entomol. Res. 109: 383–389. doi:10.1017/S0007485318000664

Delaney, D. A., J. J. Keller, J. R. Caren, and D. R. Tarpy. 2010. The physical, insemination, and reproductive quality of honey bee queens (*Apis mellifera* L.). Apidologie. 42: 1–13.

Duay, P., D. De Jong, and W. Engels. 2002. Decreased flight performance and sperm production in drones of the honey bee (*Apis mellifera*) slightly infested by Varroa destructor mites during pupal. Genet. Mol. Res. 1: 227–232.

Dunteman, G. H. 1989. Principal components analysis. SAGE 69. Sage Publications, Inc. Newbury Park, CA.

Friedli, A., G. R. Williams, S. Bruckner, P. Neumann, and L. Straub. 2020. The weakest link: Haploid honey bees are more susceptible to neonicotinoid insecticides. Chemosphere. 242: 125145.

Gençer, H. V., and Y. Kahya. 2011. Are sperm traits of drones (*Apis mellifera* L.) from laying worker colonies noteworthy? J. Apic. Res. 50: 130–137.

Gençer, H. V., and Y. Kahya. 2020. Sperm competition in honey bees (*Apis mellifera* L.): the role of body size dimorphism in drones. Apidologie. 51: 1–17.

Goins, A., and S. S. Schneider. 2013. Drone 'quality' and caste interactions in the honey bee, *Apis mellifera* L. Insectes Soc. 60: 453–461.

Harrell, F. E. Jr., C. Dupont, et al. 2020. Hmisc: Harrell Miscellaneous. R package version 4.4-1. https://CRAN.R-project.org/package=Hmisc

Hartigan, J. A., and M. A. Wong. 1979. A K-means clustering algorithm. J. R. Stat. Soc. Ser. C Appl. Stat. 28: 100–108.

Kairo, G., B. Provost, S. Tchamitchian, F. B. Abdelkader, M. Bonnet, M. Cousin, J. Sénéchal, P. Benet, A. Kretzschmar, L. P. Belzunces, et al. 2016. Drone exposure to the systemic insecticide Fipronil indirectly impairs queen reproductive potential. Sci. Rep. 6: 31094.

Kairo, G., Y. Poquet, H. Haji, S. Tchamitchian, M. Cousin, M. Bonnet, M. Pelissier, A. Kretzschmar, L. P. Belzunces, and J. L. Brunet. 2017. Assessment of the toxic effect of pesticides on honey bee drone fertility using laboratory and semifield approaches: a case study of fipronil. Environ. Toxicol. Chem. 36: 2345–2351.

Kassambara, A. and F. Mundt. 2020. factoextra: extract and visualize the results of multivariate data analyses. R package version 1.0.7. https://CRAN.R-project.org/package=factoextra

Kraus, F. B., P. Neumann, H. Scharpenberg, J. van Praagh, and R. F. Moritz. 2003. Male fitness of honeybee colonies (Apis mellifera L.). J. Evol. Biol. 16: 914–920.

Lee, K. V., M. Goblirsch, E. McDermott, D. R. Tarpy, and M. Spivak. 2019. Is the brood pattern within a honey bee colony a reliable indicator of queen quality? Insects. 10: 1–17.

Makarevich, A. V., E. Kubovicova, A. V. Sirotkin, and J. Pivko. 2010. Demonstration of the effect of epidermal growth factor on ram sperm parameters using two fluorescent assays. Vet. Med. (Praha). 2010: 581–589.

Metz, B. N., and D. R. Tarpy. 2019. Reproductive senescence in drones of the honey bee (*Apis mellifera*). Insects, Vol. 10, Page 11. 10: 11.

Neuwirth, E. 2014. RColorBrewer: ColorBrewer Palettes. R package version 1.1–2. https://CRAN.R-project.org/package=RColorBrewer

Page, R. E., and R. A. Metcalf. 1984. A population investment sex ratio for the honey bee (*Apis mellifera*). Am. Nat. 124: 680–702.

Pettis, J. S., N. Rice, K. Joselow, D. vanEngelsdorp, and V. Chaimanee. 2016. Colony failure linked to low sperm viability in honey bee (*Apis mellifera*) queens and an exploration of potential causative factors. PLoS One. 11: e0147220.

R Core Team. 2019. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ratnieks, F. L. W., and L. Keller. 1998. Queen control of egg fertilization in the honey bee. Behav. Ecol. Sociobiol. 44: 57–61.

Retschnig, G., G. R. Williams, M. M. Mehmann, O. Yañez, J. R. de Miranda, and P. Neumann. 2014. Sex-specific differences in pathogen susceptibility in honey bees (*Apis mellifera*). PLoS One. 9: e85261.

Rinderer, T. E., A. M. Collins, D. Pesante, R. Daniel, V. Lancaster, and J. Baxter. 1985. A comparison of Africanized and European drones: weights, mucus gland and seminal vesicle weights, and counts of spermatozoa. Apidologie. 16: 407–412.

Robinson, D., A. Hayes, and S. Couch. 2020. broom: convert statistical objects into tidy tibbles. R package version 0.7.0. https://CRAN.R-project.org/package=broom

Rousseeuw, P. J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comp. Appl. Math. 20: 53–65

Rousseau, A., É. Houle, and P. Giovenazzo. 2020. Effect of shipping boxes, attendant bees, and temperature on honey bee queen sperm quality (*Apis mellifera*). Apidologie. 51: 724–735.

Rueppell, O., G. V. Amdam, R. E. Page, Jr, and J. R. Carey. 2004. From genes to societies. Sci. Aging Knowledge Environ. 2004: pe5.

Rueppell, O., M. K. Fondrk, and R. E. Page, Jr. 2005. Biodemographic analysis of male honey bee mortality. Aging Cell. 4: 13–19.

Sammataro, D., and A. Avitable. 2011. The beekeeper's handbook, 4th edn. Comstock Publishing Associates, Ithaca, NY.

Schlüns, H., G. Koeniger, N. Koeniger, and R. F. A. Moritz. 2004. Sperm utilization pattern in the honeybee (*Apis mellifera*). Behav. Ecol. Sociobiol. 56: 458–463.

Schlüns, H., E. A. Schlüns, J. van Praagh, and R. F. A. Moritz. 2003. Sperm numbers in drone honeybees (*Apis mellifera*) depend on body size. Apidologie. 34: 577–584.

Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods. 9: 671–675.

Sokal, R. R., and F. J. Rolf. 1995. Biometry: the principles and practice of statistics in biological research, 11th ed. W. H. Freeman & Co., New York.

Steinhauer, N. A., K. Rennich, M. E. Wilson, D. M. Caron, E. J. Lengerich, J. S. Pettis, R. Rose, J. A. Skinner, D. R. Tarpy, J. T. Wilkes, et al. 2014. A national survey of managed honey bee 2012–2013 annual colony losses in the USA: results from the Bee Informed Partnership. J. Apic. Res. 53: 1–18.

Straub, L., L. Villamar-Bouza, S. Bruckner, P. Chantawannakul, L. Gauthier, K. Khongphinitbunjong, G. Retschnig, A. Troxler, B. Vidondo, P. Neumann, et al. 2016. Neonicotinoid insecticides can serve as inadvertent insect contraceptives. Proc. R. Soc. London B Biol. Sci. 283: 20160506.

Stürup, M., B. Baer-Imhoof, D. R. Nash, J. J. Boomsma, and B. Baer. 2013. When every sperm counts: factors affecting male fertility in the honeybee Apis mellifera. Behav. Ecol. 24: 1192–1198.

Tarpy, D. R., J. J. Keller, J. R. Caren, and D. A. Delaney. 2012. Assessing the mating 'health' of commercial honey bee queens. J. Econ. Entomol. 105: 20–25.

Tarpy, D. R., and M. K. Mayer. 2009. The effects of size and reproductive quality on the outcomes of duels between honey bee queens (*Apis mellifera* L.). Ethol. Ecol. Evol. 21: 147–153.

Tarpy, D. R., D. Vanengelsdorp, and J. S. Pettis. 2013. Genetic diversity affects colony survivorship in commercial honey bee colonies. Naturwissenschaften. 100: 723–728.

Visscher, K. P. 1989. A quantitative study of worker reproduction in honeybee colonies. Behav. Ecol. Sociobiol. 25: 247–254.

Wei, T. and V. Simko 2017. R package 'corrplot': visualization of a correlation matrix (version 0.84). https://github.com/taiyun/corrplot

Wickham, H. et al. 2019. Welcome to the tidyverse. J. Open Source Softw. 4(43): 1686. doi:10.21105/joss.01686

Wickham, H. and J. Bryan 2019. readxl: read excel files. R package version 1.3.1. https://CRAN.R-project.org/package=readxl