WASPnest: a worldwide assessment of social Polistine nesting behavior

Sara E. Miller[1], Sarah E. Bluher[1], Emily Bell[2], Alessandro Cini[3], Rafael Carvalho da Silva[4], André Rodrigues de Souza[4], Kristine M. Gandia[5], Jennifer Jandt[6], Kevin Loope[7], Amanda Prato[4], Jonathan N. Pruitt[8], David Rankin[7], Erin Rankin[7], Robin J. Southon[9], Floria M. K. Uy[5], Susan Weiner[10], Colin M. Wright[11], Holly Downing[12], Raghavendra Gadagkar[13], M. Cristina Lorenzi[14,15], Lidiya Rusina[16], Seirian Sumner[3], Elizabeth A. Tibbetts[17], Amy Toth[18], Michael J. Sheehan[1,19]

[1]Department of Neurobiology and Behavior, Cornell University, Ithaca, NY, USA

[2]School of Biological Sciences, University of Bristol, Bristol, UK

[3]Centre for Biodiversity and Environmental Research, Department of Genetics, Evolution and Environment, University College London, London, UK

[4]Departamento de Biologia, Faculdade de Filosofia Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, São Paulo, Brazil

[5]Department of Biology, University of Miami, Coral Gables, FL, USA

[6]Department of Zoology, University of Otago, Dunedin, New Zealand

[7]Department of Entomology, University of California, Riverside, CA, USA

[8]Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, ON, Canada

[9]School of Biological Sciences, University of Bristol, Bristol, UK

[10]Department of Biology, Roosevelt University, Chicago, IL, USA

[11]Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, CA, USA

[12]Black Hills State University, Spearfish, SD, USA

[13]Centre for Ecological Sciences, Indian Institute of Science, Bangalore, India

[14]Laboratoire d'Ethologie Expérimentale et Comparée, Université Paris 13, Villetaneuse, France

[15]Department Life Science and Systems Biology, University of Turin, Torino, Italy

[16]Kherson State University, Kherson, Ukraine

[17]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA

[18]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, USA

[19] Corresponding author: msheehan@cornell.edu

METADATA

CLASS I. DATA SET DESCRIPTORS

A. Data set identity:

This dataset describes the frequency and extent of cooperative nest founding for 87 paper wasp species.

B. Data set identification code:

Data file: WASPnest.foundressdata.csv

Metadata files: WASPnest.sources.txt, WASPnest.species.txt

C. Data set description:

1. Originator: Michael Sheehan
   Department of Neurobiology and Behavior, Cornell University, Ithaca, NY, USA

2. Abstract:

Cooperative breeding decreases the direct reproductive output of subordinate individuals, but cooperation can be evolutionarily favored when there are challenges or constraints to breeding independently. Environmental factors, including temperature, precipitation, latitude, high seasonality and environmental harshness have been hypothesized to correlate with the presence of cooperative breeding. However, to test the relationship between cooperation and ecological constraints requires comparative data on the frequency and variation of cooperative breeding across differing environments, ideally replicated across multiple species. Paper wasps are primitively social species, forming colonies composed of reproductively active dominants and foraging subordinates. Adult female wasps, referred to as foundresses, initiate new colonies. Nests can be formed by a single solitary foundress (non-cooperative) or by multiple foundress associations (cooperative). Cooperative behavior varies within and among species, making paper wasps species well suited to disentangling ecological correlates of variation in cooperative behavior. This dataset reports the frequency and extent of cooperative nest founding for 87 paper wasp species. Data was assembled from more than 170 published sources, previously

unpublished field observations, and photographs contributed by citizen scientists to online natural history repositories. The dataset includes 25,872 nest observations and reports the cooperative behavioral decisions for 45,297 foundresses. Species names were updated to reflect modern taxonomic revisions. The type of substrate on which the nest was built is also included, when available. A smaller population-level version of this dataset found that the presence or absence of cooperative nesting in paper wasps was correlated with temperature stability and environmental harshness, but these variables did not predict the extent of cooperation within species. This expanded dataset contains details about individual nests and further increases the power to address the relationship between the environment and the presence and extent of cooperative breeding. Beyond the ecological drivers of cooperation, these high-resolution data will be useful for future studies examining the evolutionary consequences of variation in social behavior. This dataset may be used for research or educational purposes provided that this data paper is cited.

D. Key words: *Belonogaster, cooperation, eusociality; foundress; Mischocyttarus; Parapolybia, Polistes; paper wasp; Ropalidia; social insects*

CLASS II. RESEARCH ORIGIN DESCRIPTORS

A. Overall project description:

1. Identity: A dataset of the frequency of cooperative nest founding for 87 species of paper wasps.

2. Originator: Michael Sheehan, Cornell University, Department of Neurobiology and Behavior, 215 Tower Rd, Ithaca, NY, 14853, USA.

3. Period of study: Includes nest observations made between 1931-2017. The database was compiled between 2015-2018.

4. Objectives: Our aim was to quantify inter- and intra-specific variation in cooperative nest founding in the paper wasps. Secondarily, observations were scattered throughout a wide diversity of sources, therefore we sought to increase the accessibility of this data by compiling data into an easily searchable database.

5. Abstract: See Class I.C.2

6. Sources of funding: NIH-DP2-GM128202 to MJS

B. Specific project description:

1. Site description: Observations were made at paper wasp nesting sites.

2. Data collection methodology: We collected observations of the number of foundresses at pre-emergence nests. Observations of foundresses at nests were collected from three types of data sources. Specific methods for each type of data are briefly described below and further elaborated in the supplemental methods in Sheehan *et al.* 2015. However, we note that the number of foundresses observed at a nest can fluctuate over time for two reasons. In the short term, foundresses may not be counted if they are off the nest during the observation period. At the seasonal scale, during the development of the colony, foundress number can change as individuals join the nest or die. As a result, individual observations are considered an estimate and lower bound of the number of foundresses.

a. Published data: A literature survey was conducted using Google Scholar with the search terms “Polistes”, “Ropalidia”, “Mischocyttarus”, “Belonogaster”, and “Parapolybia”. The text and figures of papers matching these search terms were scanned for reports of foundress numbers. Only wasps specifically reported as foundresses or female wasps from nests described as pre-emergence were included in the dataset. Swarm founding species were excluded from the dataset. Observations from nest boxes were included, but we excluded observations of cooperative founding behavior following experimental manipulation. Published data formed the bulk of our observations (83.3%; N=21,542 nests).

b. *Unpublished observations of nests*: Trained scientists made observations of the number of foundresses on nests in the field. Reports were generally made during the early morning or late evening when all foundresses are on the nest (15.4%; N=3,985 nests).

c. *Data from online natural history databases*: We searched Bugguide (http://bugguide.net), the Atlas of Australia (http://www.ala.org.au), and iNaturalist (http://www.inaturalists.org) for images using the search terms “Polistes”, “Ropalidia”, “Mischocyttarus”, “Belonogaster”, and “Parapolybia”. To be included in the dataset, a photograph must show wasps on a nest and include a clear visual of all nest cells to ensure that the nest is in the pre-emergence phase. The number of foundresses was estimated by counting the number of individuals in each photograph and by reading any associated comments for reports of additional wasps that were observed but not photographed (1.3%; N=345 nests). Photos taken during the early morning or late evening, or accompanied by notes detailing additional observations, are given special designation in the dataset.

CLASS III. DATA SET STATUS AND ACCESSIBILITY

A. Status:

1. Latest update: April 2018
2. Latest archive date: April 2018
3. Metadata status: Metadata is complete
4. Data verification: The second author, SEB, verified all records for accuracy.

B. Accessibility

1. Storage location and medium: The full dataset and metadata files were posted in a github repository (https://github.com/waspfoundressdata/Foundress-Data; DOI: https://doi.org/10.5281/zenodo.1236498). Backup copies of the data are saved on computers at Cornell University. Data from published observations are available in the original publications. The original observer has saved data from unpublished observations.

2. Contact person: Michael J. Sheehan, Department of Neurobiology and Behavior, Cornell University, 215 Tower Rd, Ithaca, NY, 14853, USA. Email: msheehan@cornell.edu.

3. Copyright restrictions: This dataset may be used for research or educational purposes provided that this data paper is cited.

4. Proprietary restrictions: None.

5. Disclaimer: This dataset is being actively updated with new observations from all three data sources. Researchers with unpublished observations of cooperative foundress behavior are encouraged to contact the authors. Periodic revisions to the dataset will be uploaded to the github repository (https://github.com/waspfoundressdata/Foundress-Data; DOI: https://doi.org/10.5281/zenodo.1236498).

CLASS IV: DATA STRUCTURAL DESCRIPTORS

A. Data set files:

Identity: WASPnest.foundressdata.csv
Size: 23820 rows, 22 columns, including header row.
Format and storage mode: Text file, semi-colon delimited, uncompressed
Header information: Header names.
Alphanumeric attributes: Mixed
Missing data codes: Missing values for dates are indicated by 999 or NA.

Identity: WASPnest.sources.txt
Size: 173 rows, 2 columns, including header row
Format and storage mode: Text file, tab delimited, uncompressed
Header information: Header names.
Alphanumeric attributes: Mixed

Identity: WASPnest.species.txt
Size: 93 rows, no header row
Format and storage mode: Text file, tab delimited, uncompressed

Header information: no header
Alphanumeric attributes: Mixed

B. Variable Information

Identification Variables

ID: A unique identification number for each observation
Genus: Genus of observed species
Species: Name of observed species. Includes subspecies information when given. Subspecies were combined when reporting the number of species in database.
Binomial: Full species binomial. Records using revised species names have been updated to modern taxonomic designations.
Day: The day the observation was made. Observations spanning multiple days are indicated with an underscore (e.g. May 1-5: "1_5"), and observations made on non-consecutive days are reported with a semi-colon (e.g. May 1 and 5: "1; 5"). The data in this example should be interpreted as nests examined between May 1st and May 5th ("1_5") or observed on either May 1st or May 5th ("1; 5"). Missing data are reported as 999.
Month: The month the observation was made. Observations spanning multiple months are indicated with an underscore (e.g. May through July: "5_7"), and observations made on non-consecutive months are reported with a semi-colon (e.g. May and July: "5; 7"). Interpretation of month ranges are the same as for days. Missing data are reported as 999.
Year: The year the observation was made. Observations spanning multiple years are indicated with an underscore (e.g. 2012 through 2017: "2012_2017"), and observations made on non-consecutive years are reported with a semi-colon (e.g. 2012 and 2017: "2012; 2017"). Interpretation of year ranges are the same as for days. Observation for which a year was not indicated were reported as occurring before the date of the publication (e.g. a paper published in 2012 would have the year <2012).
Source: Abbreviated reference for the source of the data. The complete citations for all sources are given in the waspfoundressdata.sources.txt metadata file.
Source_Type: Sources were classified by the data collection methodology in section B.2. Paper refers to a peer reviewed publication. Book and thesis are published data that was not peer

reviewed. Unpublished refers to unpublished field observations. Image is data collected from photographs.

Nest Foundress Variables

The wide range of sources for our data required a reporting methodology that was flexible to various types of input. Table 1 illustrates how four different data reporting methods were recorded in the dataset. In this example, a single foundress nest and one nest with two foundresses were reported for *P. fuscatus* (ID 1-2); a combined observation of three nests with five foundresses was reported for *P. annularis* (ID 3); *P. dominula* observations (ID 4-6) were taken from a table reporting one nest with three foundresses and two nests with 5 or more foundresses; and for *P. carnifex* (ID 9), two single foundress nests and 4 nests that were cooperatively founded by an aggregate of 21 females, with the author reporting that multiple foundress nests had a range of 4-6 foundresses. The number of foundresses and nests for each species with at least 20 observed nests is shown in Figure 1.

ID	Binomial	Nests	Foundresses	+	Grouped_nestinfo
1	Polistes_fuscatus	1	1		
2	Polistes_fuscatus	1	2		
3	Polistes_annularis	3	5		
4	Polistes_dominula	1	3		
5	Polistes_dominula	1	5	+	
6	Polistes_dominula	1	5	+	
7	Polistes_carnifex	1	1		
8	Polistes_carnifex	1	1		
9	Polistes_carnifex	4	21		MF; 4-6

Table 1: Example foundress data illustrating data from three different reporting methods. See main text for details.

Nests: The number of nests reported for each observation.

Foundresses: The total number of foundresses observed on the nest(s).

Plus (+): This symbol indicates that the number of foundresses was reported in a bin (e.g. 5 or more) and the exact number of foundresses is unknown. We report the minimum number for that bin in the *Foundresses* column.

Grouped_nestinfo: When nest observations were grouped together, this column adds any additional available information on foundress number.

MF All nests contained multiple foundresses

MF; a-b	All nests reported had multiple foundress associations ranging from a to b foundresses. (e.g. 4-6 foundresses, MF; 4-6).
All	Grouped nest information contains both single and multiple foundress nest data

Nest_method: The number of foundresses was determined using one of seven possible methods.

1. Directly reported. This can include in the text of a paper, in a table, or a direct observation in the field.
2. Estimated from the text of paper.
3. Derived from a figure using WebPlot Digitizer (https://automeris.io/WebPlotDigitizer/).
4. Genetic estimation. For more detail, see the supplemental information for Sheehan *et al.* 2015.
5. Morphological estimation. For more detail, see the supplemental information for Sheehan *et al.* 2015.
6. Photograph taken during foraging hours, or time not reported.
   6n: notes include additional information
7. Photograph taken after sundown or before 8am when temperature was $\leq 18^{o}$C. These photographs are more likely to capture all foundresses.
   7n: notes include additional information

Nest_substrate: The type of substrate on which the nest was built. Unknown substrates are indicated with "NA". Substrates are classified as follows:

1. Nest box
2. Human structure
3. Plant
4. Rock

Location Information

Locality: A general description of the nest location

Province/State: The state or province where the nest was observed.

Country: The country where the nest was observed

Latitude: Latitude of observation (decimal degrees).

Longitude: Longitude of observation (decimal degrees).

Coordinate_method: GPS coordinates were either directly reported in the original source or were estimated using available locality information.

Coordinate_accuracy: The scale of GPS coordinates varied widely across the dataset depending on how location information was reported. We classified the coordinate information into five categories based on the accuracy at which the nest location could be determined.

1. ≤ 500 m
2. 500 m – 2000 m
3. 2 km – 20 km
4. 20 km – 100 km
5. > 100 km

CLASS V: SUPPLEMENTAL DESCRIPTORS

A. Publication: Sheehan, M. J., Botero, C. A., Hendry, T. A., Sedio, B. E., Jandt, J. M., Weiner, S., Toth A.L. & Tibbetts, E. A. (2015). Different axes of environmental variation explain the presence vs. extent of cooperative nest founding associations in *Polistes* paper wasps. *Ecology letters*, *18*(10), 1057-1067.

Acknowledgments:

We would like to acknowledge the many citizen scientists who submitted data and images to bugguide, iNaturalist, and the Atlas of Australia.

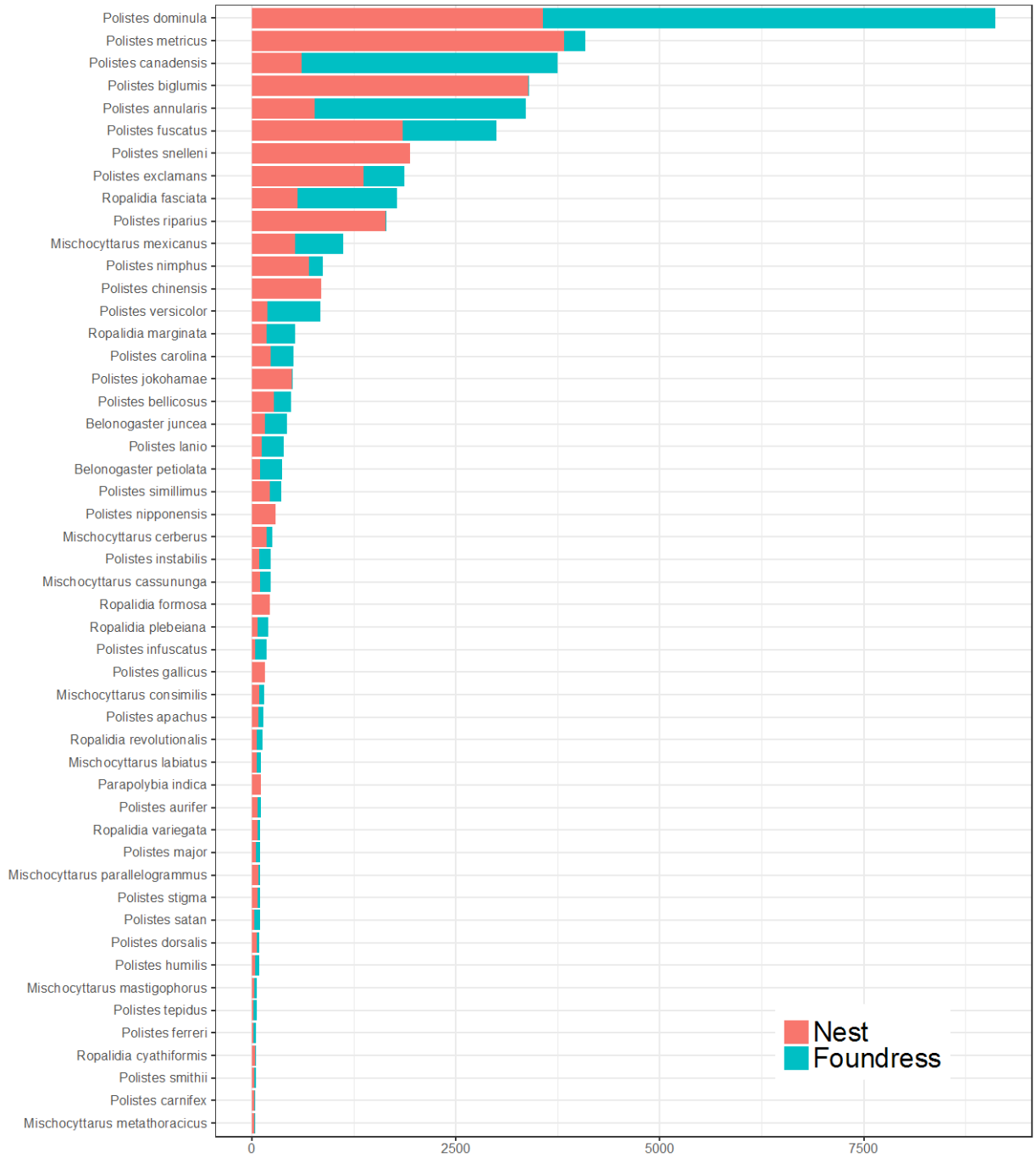

Figure 1: The average number of foundresses varies widely across paper wasp species. Here we show the number of nests and the number of foundresses in our dataset for all species with 20 or more observed nests.