**GENERAL INFORMATION**

**1. Title of Dataset**: “Monarch (Danaus plexippus) larval records in the western U.S”

**2. Author Information**

 A. Person responsible for collecting the data

 Name: Alma Nalleli Carvajal Acosta

 Institution: University of California, Irvine

 Address: 321 Steinhaus Hall, Irvine, CA 92697 USA

 Email: ancarvaj@uci.edu

 B. Principal Investigator

 Name: Kailen Mooney

 Institution: University of California, Irvine

 Address: 321 Steinhaus Hall, Irvine, CA 92697 USA

 Email: mooneyk@uci.edu

**3. Date of data collection (single date, range, approximate date)**

May 2017-February 2019

**4. Geographic location of data collection**

Western United States: California, Nevada, Oregon, Colorado, New Mexico, Utah, Arizona, Idaho, Washington.

**5. Keywords used to describe the data topic:** Danaus plexippus, western monarch, species occurrences, species distributions models, citizen science

**6. Language information:** English

**7. Information about funding sources that supported the collection of the data:**

UC-MEXUS/ CONACyT

National Science Foundation, Award: DEB-1354734

National Science Foundation, Award: DEB-1457029

**DATA & FILE OVERVIEW**

**1. For each filename, a short description of what data it contains**

“Monarch (Danaus plexippus) larval records in the western U.S” contains larval occurrence information for the western monarch within their breeding range collected from multiple open-source databases.

**2. Format of the file if not obvious from the file name:**

comma separated values, file extension “.csv”

**3. Additional related data collected that was not included in the current data package:**

No

**4. Are there multiple versions of the dataset?** yes/no

 A. If yes, name of file(s) that was updated:

 i. Why was the file updated?

 ii. When was the file updated?

**SHARING/ACCESS INFORMATION**

**1. Licenses/restrictions placed on the data:**

N/A

**2. Links to publications that cite or use the data:**

Carvajal-Acosta, A. N. and Mooney, K.A*. 2021.* Effects of geographic variation in hostplant resources for a specialist herbivore's contemporary and future distribution. Ecosphere.​

<https://doi.org/10.1002/ecs2.3822>

Rx non-revised version:

Carvajal-Acosta, A. N. and Mooney, K.A*.* 2020*.* Roles of hostplant availability and quality for the distribution and climate change response of a dietary specialist

doi: <https://doi.org/10.1101/2020.12.03.410225>

**3. Was data derived from another source?** yes/no

 A. If yes, list source(s): data was compiled from the following open-source databases:

* GBIF
* BISON
* iNaturalist
* Vertnet
* iDigBio
* MLMP (Monarch Larvae Monitoring Program)

**4. Recommended citation for this dataset:**

Carvajal Acosta, Nalleli (2021, Monarch (Danaus plexippus) larval records in the western U.S Dryad, Dataset, https://doi.org/10.7280/D1SQ5K

**METHODOLOGICAL INFORMATION**

**1. Description of methods used for collection/generation of data:**

We retrieved monarch larval records for the United States using R Studio (R Studio Team 2015) from multiple open-source databases using the R packages SPOCC, Ecoengine, rbison (Chamberlain et al. 2014, Karthik 2014, Chamberlain 2019) and by accessing species occurrences directly from GBIF and iNaturalist databases (“GBIF Occurrence Download” 2019, “Naturalist [online]. Website” 2019). We also obtain monarch larval records directly from the Monarch Larvae Monitoring Program (MLMP).

**2. Methods for processing the data:**

Data cleaning and preparation was performed in R Studio following protocols described in Hijman (2019). To focus on the western monarch population, we selected monarch larval records from states corresponding to this region: California, Nevada, Colorado, Washington, New Mexico, Arizona, Utah, Oregon, and Idaho. First, we removed incorrect coordinates (i.e., over oceans) and inaccurate coordinates with >1000 meters uncertainty. Then we removed duplicated records.

**3. Instrument- or software-specific information needed to interpret the data:**

Species occurrences can be visualized using any geographic information software (ArcMap, QGIS) or using the package rgdal in R studio.

**4. Describe any quality-assurance procedures performed on the data:**

To ensure accuracy, when permitted, we used filters that only retrieved records confirmed by experts and/or records classified as of research quality.

Inaccurate coordinates with over 1000 meters GPS uncertainty were discarded.

Incorrect coordinates located over oceans or in other countries were also discarded.

We also control for potential sampling biases by using a spatial filtering approach described in Kramer-Schadt et al. (2013).

We visited the image url when available to confirm the identity of the species as well as the recorded life stage were correct.

**7. People involved with sample collection, processing, analysis and/or submission:**

Dr. Collen Nell, Dr. Will Petry, Dr. Kailen Mooney

**DATA-SPECIFIC INFORMATION FOR:** “Monarch (Danaus plexippus) larval records in the western U.S”

**1. Number of variables:**

22

**2. Number of cases/rows:**

7,916 Asclepias occurrences

**3. Variable List:**

Scientific name: Scientific name of monarch larval

latitude: latitude coordinate in WGS1994

longitude: longitude coordinate in WGS1994

Uncertainty: GPS uncertainty

Verbatum\_location: described location by the person recording the occurrence

Date: date when sighting occurred

Year: year when the sighting occurred

State: state where the sighting occurred

City: city where the species sighting occurred

occurrenceRemarks: any special remarks made by the person recording the occurrence

BasisOfRecord: how was the occurrence documented, PRESERVED from herbarium records, HUMAN\_OBSERVATION for field sightings

institution: database or herbarium where the species occurrence was originally recorded (MLMP, GBIF, iNaturalist, RSA, GMDRC, UCR, SD, SEINET, CHSC, JOTR, CAS-BOT-BC, BISON, NA, KU, MO, SPOCC, etc)

speciesKey: taxon identification for each opensource database

Occurrence\_url: link to occurrence

**4. Missing data codes:**

N/A

**5. Specialized formats or other abbreviations used:**

**References**

Chamberlain, S. 2019. rbison: Interface to the “USGS” “BISON” API.

Chamberlain, S., K. Ram, and T. Hart. 2014. spocc: R interface to many species occurrence data

sources.

Karthik, R. 2014. ecoengine: Programmatic interface to the API serving UC Berkeley’s Natural History Data.

GBIF.org. 2019. <https://doi.org/10.15468/dl.sjg2d2>.

Hijman, R. 2019. Spatial Data in R. https://rspatial.org/spatial/Spatialdata.pdf

Naturalist [online]. Website. 2019, January. https://www.inaturalist.org.