**Metadata**

**for**

# **Phylogenetic community structure and stable isotope analysis of the parasitoid community associated with Eastern spruce budworm (Lepidoptera: Tortricidae)**

**Authors**

Christopher J. Greyson-Gaito1, Sarah J. Dolson2, Glen Forbes3, Rosanna Lamb3, Wayne E. MacKinnon3, Kevin S. McCann1, M. Alex Smith1, Eldon S. Eveleigh3,4

\*Corresponding Author - christopher@greyson-gaito.com

**Affiliations**

1. Department of Integrative Biology, University of Guelph, Guelph, Ontario

2. Department of Biology, University of Ottawa, Ottawa, Ontario, Canada

3. Natural Resources Canada, Canadian Forest Service, Atlantic Forestry Centre, Fredericton, NB, Canada

4. Population Ecology, Faculty of Forestry and Environmental Management, University of New Brunswick, Fredericton, New Brunswick, Canada

**ORCID**

CJGG – 0000-0001-8716-0290

SJD – 0000-0001-9312-2282

KSM - 0000-0001-6031-7913

MAS – 0000-0002-8650-2575

ESE – 0000-0001-5060-8565

### Instructions

1. Download the GitHub/Zenodo repo
2. Open RStudio, and open the project called SpruceBudworm\_Parasitoids\_Hardwood.Rproj
3. Then open the R file called Hardwood\_SBW\_GreysonGaitoetal2021\_RScript.R.

Because you have opened a R project, your working directory has been set to this folder and you can immediately start running the code.

If you do not use RStudio, open the R file Hardwood\_SBW\_GreysonGaitoetal2021\_RScript.R, the normal way you would. Then change your working directory to this folder. You will need to change the paths to the dataset and the figures in the R script.

### Folder and file structure

* data (see below for the structure of each data file)
  + SI\_data\_GreysonGaitoetal2021.csv
  + malaise2016\_reared2015\_barcoded\_metadata\_GreysonGaitoetal2021.csv
  + malaise2016\_reared2015\_barcoded\_tree\_GreysonGaitoetal2021.nwk
  + reared1980s\_barcoded\_metadata\_GreysonGaitoetal2021.csv
  + reared1980s\_barcoded\_tree\_GreysonGaitoetal2021.nwk
* figs – folder to place figures when created
* .gitignore – file containing files and folders that git should ignore
* Hardwood\_SBW\_GreysonGaitoetal2021\_RScript.R – R script with statistical analyses and figure creation
* LICENSE.txt – CC by 4.0 License for this repository
* README.md – Readme file
* SpruceBudworm\_Parasitoid\_Hardwood.Rproj – Opens R project
* metadata\_SpruceBudworm\_Parasitoids\_Hardwood\_GreysonGaitoetal2021.docx – contains information on each datafile

**Data Structure**

**SI\_data\_GreysonGaitoetal2021.csv**

* Column: Identifier
  + Unique identifier for each data value in the format plot-basetree-trophiclevel-month or group-month-year where:
  + For plot-basefoliage-trophiclevel-month ids
    - plot ranges from 01 to 09. And where 01,02,03 are balsam fir dominated plots, 04,05,06 are mixed stand plots, 07,08,09 are hardwood dominated plots
    - basetree is either BF (sampled from balsam fir trees) or HW (sampled from hardwood trees)
    - trophiclevel is either P (for plants), or SBW (for budworm caterpillars), or ALT (for other caterpillar species).
    - month is either JUNE, JULY, or AUG.
  + For group-month-year ids
    - group is either GR01, GR02, GR03, GR04, GR05 (for different functional groups – see below for what these original functional groups were). NOTE we combined GRO1 and GR02 into one group (now called group one in the manuscript) and we did not use GR04. This was because there weren’t enough data points in the original groupings for meaningful analysis. GR03 became group two and GR05 became group three in the manuscript.
    - Month is either JUNE, JULY, AUG. Some may have two months which means the samples from each month were combined.
    - Year is either 82, 83, 86, 87
  + NOTE for some ids there are A, B, or C at the end. These were SI repetitions of the same sample.
* Column: cald13C – δ13Carbon value
* Column: calpercentC - % Carbon in the sample
* Column: cald15N – δ15Nitrogen value
* Column: calpercentN - % Nitrogen in the sample
* Column: CNratio – Carbon:Nitrogen ratio in the sample

Original Functional Groupings (the functional groupings in the manuscript can be found in the supporting information of the manuscript)

**Group 1**

*Apantales fumiferanae*

*Charmon extensor*

*Glypta fumiferanae*

**Group 2**

*Lypha fumipennis (Lypha setifacies)*

*Smidtia fumiferanae (Winthemia fumiferanae)*

**Group 3**

*Actia interrupta*

*Agria affinis (Psuedosarcophaga affinis)*

*Compsilura concinnata*

*Eumea caesar*

*Hemisturmia parva (Hemistermia tortricis)*

*Nilea erecta (Pseudoperichaeta erecta)*

*Sarcophaga aldrichi*

*Tachinomyia nigricans*

**Group 4**

*Apechthis ontario (Ephialtes ontario)*

*Dirophanes hariolus (Phaeogenes hariolus)*

*Itoplectis conquisitor*

**Group 5**

*Ceromasia auricaudata (Ceromasia aurifrons)*

*Madremyia saundersii*

*Meteorus trachynotus*

*Nemorilla psyte*

*Phryxe pecosensis*

**malaise2016\_reared2015\_barcoded\_metadata\_GreysonGaitoetal2021.csv**

NOTE each row is an individual parasitoid that has been sequenced

* Column: ProcessID
  + BOLD created ID depending on the BOLD project
* Column: SampleID
  + original sample ID for sequencing
* Column: CollectionNotes
  + contains information on what type of forest the plot was in, the plot number, and whether malaise caught or reared parasitoids.
* Column: HWGrad
  + code for what type of forest the plot was in where BFBF are balsam fir dominated plots, BFMX are mixed stand plots, and HWBF are hardwood dominated plots.
* Column: Plot
  + where Plots 1,2,3 are balsam fir dominated plots, Plots 4,5,6 are mixed stand plots, and Plots 7,8,9 are hardwood dominated plots.
* Column: Method
  + Either Malaise (malaise caught parasitoids) or Reared (emerged from sampled caterpillars reared on standard diet)
* Column: HWGrad\_Description – description of type of forest stand plots are in
* Column: CONTIG
* Column: Project Code
  + BOLD identifier for different projects
  + ASSBW is for malaise caught parasitoids from 2016 and ASNBA is for reared parasitoids from 2015
* Column: BIN
  + BOLD defined DNA barcoded BIN
* Column: COI-5P Seq. Length
* Column: # ambiguities

**malaise2016\_reared2015\_barcoded\_tree\_GreysonGaitoetal2021.nwk**

* Newick tree format data file of a single-representative maximum likelihood tree in MEGA6 based on estimation of the best substitution models in MEGA6 (Nei and Kumar 2000; Tamura et al. 2013) for parasitoids that were caught in malaise traps in 2016 and parasitoids that emerged from reared caterpillars sampled in 2015.

**reared1980s\_barcoded\_metadata\_GreysonGaitoetal2021.csv**

NOTE each row is an individual parasitoid that has been sequenced

* Column: ProcessID
  + BOLD created ID depending on the BOLD project
* Column: SampleID
  + original sample id for sequencing
* Column: CollectionNotes
  + contains information of which plot the parasitoid was collected from and what was the composition of the trees in that plot.
* Column: Project Code
  + BOLD identifier for different projects
* Column: BIN
  + BOLD defined DNA barcoded BIN
* Column: OTU
* Column: Collection Date

**reared1980s\_barcoded\_tree\_GreysonGaitoetal2021.nwk**

* Newick tree format data file of a single-representative maximum likelihood tree in MEGA6 based on estimation of the best substitution models in MEGA6 (Nei and Kumar 2000; Tamura et al. 2013) for a subset of the parasitoids that emerged from reared caterpillars sampled in the 1980s.