

## Symbioses with nitrogen-fixing bacteria: nodulation and phylogenetic data across legume genera

MICHELLE E. AFKHAMİ,<sup>1,8</sup> D. LUKE MAHLER,<sup>2</sup> JEAN H. BURNS,<sup>3</sup> MARJORIE G. WEBER,<sup>4</sup>  
MARTIN F. WOJCIECHOWSKI,<sup>5</sup> JANET SPRENT,<sup>6</sup> AND SHARON Y. STRAUSS<sup>7</sup>

<sup>1</sup>*Department of Biology, University of Miami, 1301 Memorial Dr, #215, Coral Gables, Florida 33146 USA*

<sup>2</sup>*Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks Street, Toronto, Ontario M5S 3B2 Canada*

<sup>3</sup>*Department of Biology, Case Western Reserve University, 307 DeGrace Hall, Cleveland, Ohio 4410 USA*

<sup>4</sup>*Department of Plant Biology, Michigan State University, East Lansing, Michigan 48824 USA*

<sup>5</sup>*School of Life Sciences, Arizona State University, PO Box 874501, Tempe, Arizona 85287 USA*

<sup>6</sup>*College of Life Sciences, University of Dundee at JHI, Dundee DD2 5DA UK*

<sup>7</sup>*Department of Evolution and Ecology, University of California, Davis, One Shields Avenue, 2320 Storer Hall, Davis, California 95616 USA*

**Abstract.** How species interactions shape global biodiversity and influence diversification is a central – *but also data-hungry* – question in evolutionary ecology. Microbially based mutualisms are widespread and could cause diversification by ameliorating stress and thus allowing organisms to colonize and adapt to otherwise unsuitable habitats. Yet the role of these interactions in generating species diversity has received limited attention, especially across large taxonomic groups. In the massive angiosperm family Leguminosae, plants often associate with root-nodulating bacteria that ameliorate nutrient stress by fixing atmospheric nitrogen. These symbioses are ecologically-important interactions, influencing community assembly, diversity, and succession, contributing ~100–290 million tons of N annually to natural ecosystems, and enhancing growth of agronomically-important forage and crop plants worldwide. In recent work attempting to determine whether mutualism with N-fixing bacteria led to increased diversification across legumes, we were unable to definitively resolve the relationship between diversification and nodulation. We did, however, succeed in compiling a very large searchable, analysis-ready database of nodulation data for 749 legume genera (98% of Leguminosae genera; LPWG 2017), which, along with associated phylogenetic information, will provide a valuable resource for future work addressing this question and others. For each legume genus, we provide information about the species richness, frequency of nodulation, subfamily association, and topological correspondence with an additional data set of 100 phylogenetic trees curated for database compatibility. We found 386 legume genera were confirmed nodulators (i.e., all species examined for nodulation nodulated), 116 were non-nodulating, four were variable (i.e., containing both confirmed nodulators and confirmed non-nodulators), and 243 had not been examined for nodulation in published studies. Interestingly, data exploration revealed that nodulating legume genera are ~3 × more species-rich than non-nodulating genera, but we did not find evidence that this difference in diversity was due to differences in net diversification rate. Our metadata file describes in more detail the structure of these data that provide a foundational resource for future work as more nodulation data become available, and as greater phylogenetic resolution of this ca. 19,500-species family comes into focus. We release this data set under the Creative Commons 4.0 Attribution-ShareAlike License (<https://creativecommons.org/licenses/by-sa/4.0/>). The data may be used, distributed, and reproduced with proper citation of this article.

**Key words:** *bacteria; biodiversity; database; diversification; leguminosae; mutualism; nitrogen fixation; nodulation; phylogeny; rhizobia; species interactions; tree.*

The complete data sets corresponding to abstracts published in the Data Papers section in the journal are published electronically as Supporting Information in the online version of this article at <https://doi.org/onlineibrary.wiley.com/doi/10.1002/ecy.2110/supinfo>