**Meta-analysis of introduced vs native legumes: A story of persistence**

Kathryn J Bloodworth1.2, Smriti Pehim Limbu3, Kimberly J Komatsu1,2

1Biology Department, University of North Carolina at Greensboro, Greensboro, NC 27403 USA

2Smithsonian Environmental Research Center, Edgewater, MD 21037 USA

3Smriti’s affiliation (JHU? Dartmouth? Both?)

**Abstract**

**Keywords**

**Introduction**

1. Species Invasion (Kathryn)
   1. Theory
   2. Consequences
2. Mutualism and invasion (Smriti)
   1. Theories
   2. Under studied phenomenon

Legumes present a particularly interesting case of plant-bacterial mutualisms. Legumes have a symbiotic relationship with nitrogen (N)-fixing soil bacteria called rhizobia, which infect the legumes’ roots through the formation of nodules (Greenwood and Pankhurst 1977, Bottomley 1992). Rhizobia are not transmitted through seeds; instead, viable rhizobia are released into the soil at nodule senescence and new nodules are formed by independent infections (Greenwood and Pankhurst 1977, Bottomley 1992). Therefore, although an introduced legume may succeed in a novel range by nodulating extensively and obtaining atmospheric N to outcompete native plants (Rodríguez-Echeverría et al. 2008), this strategy depends on finding and associating with compatible rhizobia in the novel range.

Some legume-rhizobial interactions are fairly specific, with a highly conserved association between the leguminous species and specific phylotypes of rhizobia (Woomer et al. 1988, Richardson et al. 2000, Bever et al. 2013). In contrast, other species of legumes may be generalists in terms of the rhizobial phylotypes with which they can form mutualistic associations (Thrall et al. 2000, Richardson et al. 2000, Barrett et al. 2012). The specificity of a leguminous species for particular bacterial phylotypes may have large consequences for a legume’s success in its introduced range. If an introduced legume preferentially associates with specific rhizobial phylotypes, then to succeed it will need to co-invade with those phylotypes (Rodríguez-Echeverría et al. 2008, 2012). On the other hand, if an introduced legume is a rhizobium generalist, then it may associate with native rhizobia (Parker 2001) and therefore have one less limitation impacting its ability to persist in its novel range. As such, it might be expected that legume species that succeed in an introduced range are more likely to be generalists for their rhizobial partners.

1. Ways people study invasion (setting up hypotheses) (Smriti)
   1. Comparing native range to non native range
   2. Comparing non native to natives in non native range
2. What We Did (Kathryn)
   1. Questions
      1. Do species that are globally native have a different strain richness than those that are introduced somewhere in the world? (regardless of where the samples were collected, this is the global plant status)
      2. Are non-native legumes utilizing the same number rhizobial strains as they do in their native range? ((a)across studies that specifically addressed this question and (b) those that did not)
         1. What is the percent overlap of strains between non-natives in their native range compared to non-native range
      3. Do species that are non-native in a given area associate with the same number of rhizobial symbionts as those that are native in the same area?
         1. What is the percent overlap?
   2. Hypotheses

**Methods**

*Journal Article Identification*

Papers were identified through a two-stage process, first a Web of Science search and second a follow-up Google Scholar search. In both cases, multiple searches were performed using different search terms to identify journal articles. These included all combinations of terms related to introduced plant status (*i.e.*, “introduced”, “invasive”, “exotic”, “non-native”, and “alien”) paired with terms related to various genera of rhizobia (*i.e.*, rhizobi\*, bradyrhizobi\*, mesorhizobi\*, azorhizobi\*, sinorhizobi\*, allorhizobi\*, ensifer, cupriavidus, and burkholderia). All papers returned from Web of Science and the first 100 pages of results returned from Google Scholar for each search term combination was examined. For each search, paper titles were examined to determine suitability for this meta-analysis. For papers that appeared relevant, abstracts were examined to confirm suitability. Papers were considered suitable for use in this meta-analysis if they (1) examined a non-crop legume species and (2) assessed rhizobial richness under natural conditions (*e.g.*, not in a greenhouse or plantation) utilizing molecular techniques (*e.g.*, sequencing, AFLP, RFLP, ERIC, BOX, or rep-PCR). This search was conducted first in 2014 and again in 2022. In total, 330 papers were obtained from this search process (Supplemental Table 1).

*Data Extraction* (Kathryn)

Include name cleaning steps here?

Include filtering steps to remove row crops, unknown plant status, extinct species, plants only identified to genera

*Calculating Overlap* (Smriti)

*Statistical Analysis* (everyone does their own)

Average strain richness associated with 350 legume species was calculated across all studies and genetic methods (*i.e.*, legume species as replicates). The relationship between strain richness and global plant status was determined using an ANOVA, with strain richness as the dependent variable and global plant status as the independent variable. Data were subset to legume individuals where 3 or more nodules were molecularly assessed to determine rhizobial strain identity, based on an assessment of strain richness vs nodule number for all data points (Supplemental Figure 1). At 3 or more nodules, strain richness was not significantly correlated with number of nodules sampled. Strain richness was ln transformed to improve normality of data.

Question 2 part A (Smriti)

To compare rhizobial richness for legume species found across multiple studies, data were subset to include only those where the same genetic method was used to assess rhizobial richness for a legume species in both its native and introduced range (N=16 legume species, many of which were assessed using more than one genetic method). Average strain richness associated with these 16 legume species was calculated across all studies and genetic methods (*i.e.*, legume species as replicates). The relationship between strain richness and local plant status was determined using a mixed effects model, with strain richness as the dependent variable, local plant status as the independent variable, and legume species identity as a random effect. As above, data were subset to legume individuals where 3 or more nodules were molecularly assessed to determine rhizobial strain identity and strain richness was ln transformed to improve normality of data.

~~For the subset of legume species within the Home-Away studies whose strains had associated Genbank Accession Numbers for the 16S gene region (8 species:~~ *~~Caragana aborescens~~*~~,~~ *~~Cytisus scoparius~~*~~,~~ *~~Galega officinalis~~*~~,~~ *~~Lathyrus pratensis~~*~~,~~ *~~Leucaena leucocephala~~*~~,~~ *~~Lotus corniculatus~~*~~,~~ *~~Lo. tenuis~~*~~, and~~ *~~Mimosa pudica~~*~~), strains were genetically compared to determine percent overlap in strain identity between native and introduced ranges.~~

Question 3 (Kathryn)

**Results** (everyone does their own)

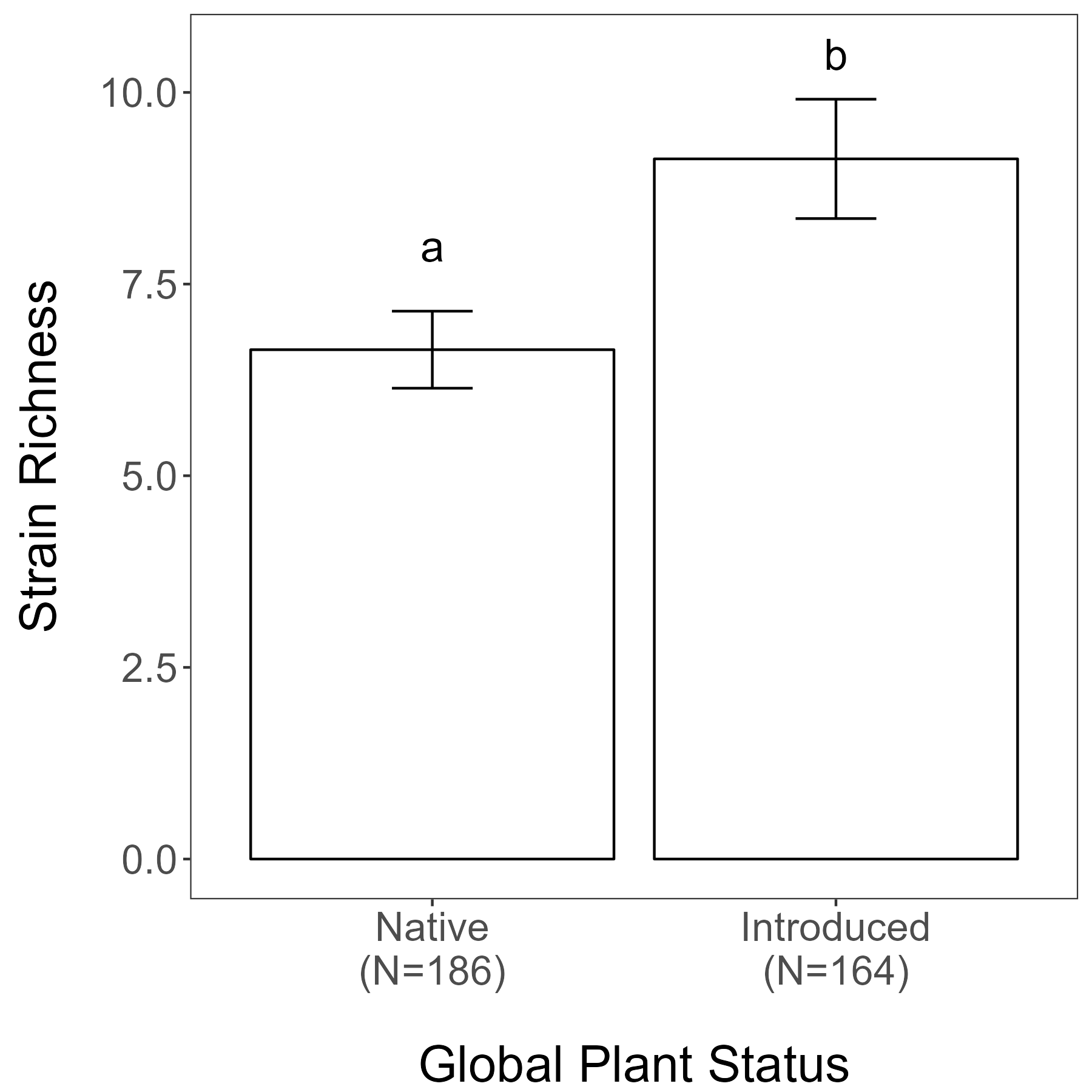
Average strain richness associated with legume species that are globally native was significantly lower than those that have been successfully introduced somewhere in the world (Fig. 1; F1,348=7.77, p=0.006; means ± standard error: Globally Native 6.64 ± 0.50 strains; Introduced Somewhere 9.13 ± 0.78 strains).

Question 2 part A (Smriti)

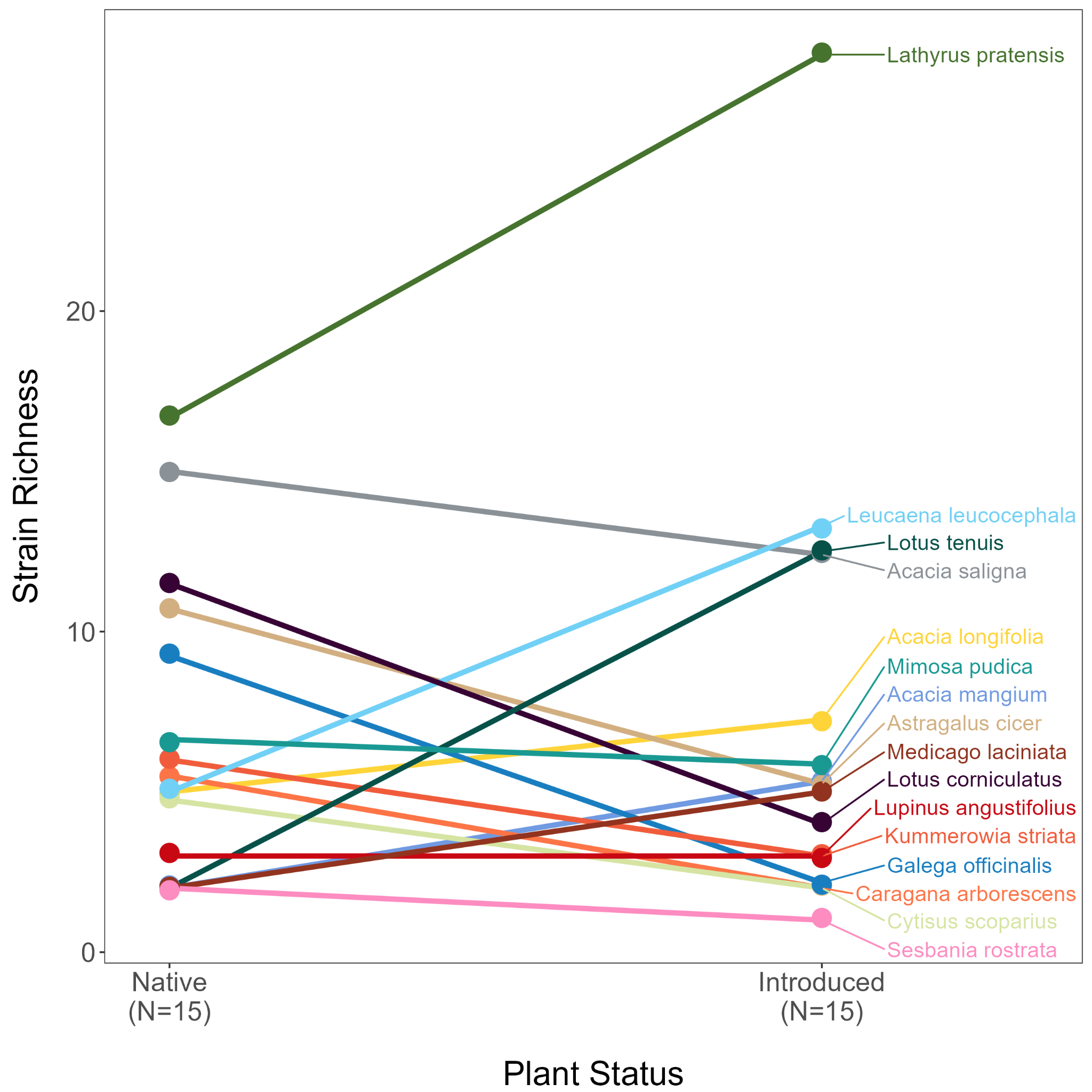
When looking across studies, average strain richness associated with legume species did not significantly differ by local plant status (Fig. 2b; F1,15=0.10, p=0.75), although the response differed greatly across legume species.

Question 3 (Kathryn)

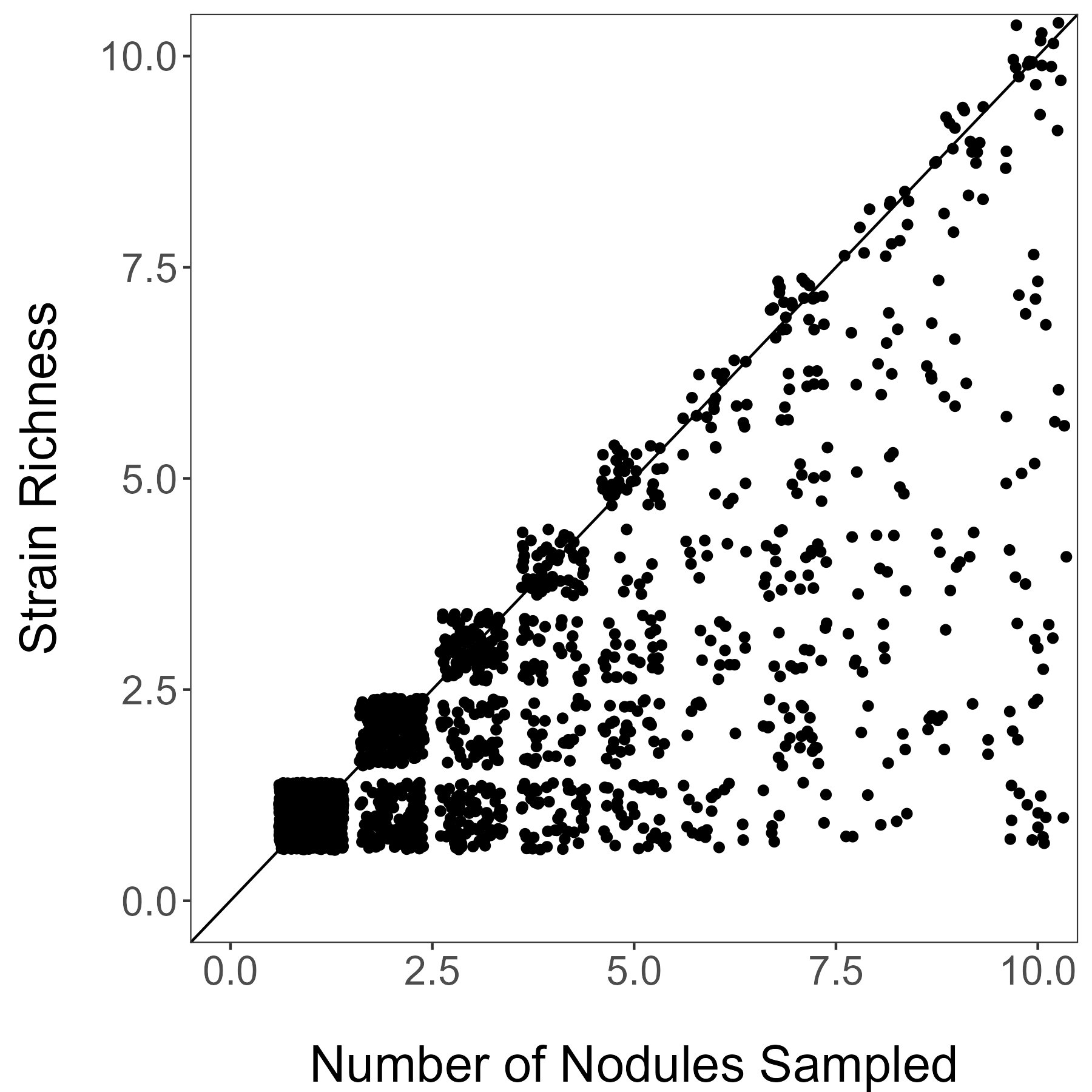
**Discussion**



**Figure 1.**



**Figure 2b.**



**Supplemental Figure 1.** Number of nodules sampled from a legume individual compared to the strain richness across all of those nodules. Maximum strain richness is limited to the maximum number of nodules sampled. With three or more nodules sampled, the strain richness appears to be evenly distributed across all possible values (from 1 to number of nodules sampled).