Rapid report

Nonsymbiotic legumes are more invasive, but only if polyploid

Author for correspondence:
Tia L. Harrison
Email: tia.harrison@mail.utoronto.ca

Received: *24 August 2022*
Accepted: *15 October 2022*

**Zoe A. Parshuram[1]*, Tia L. Harrison[1]* (iD), Anna K. Simonsen[2] (iD),
John R. Stinchcombe[1] (iD) and Megan E. Frederickson[1] (iD)**

[1]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON M5S 3B2, Canada; [2]Department of
Biological Sciences, Florida International University, Miami, FL 33199, USA

Summary

• Both mutualism and polyploidy are thought to influence invasion success in plants, but few studies have tested their joint effects. Mutualism can limit range expansion when plants cannot find a compatible partner in a novel habitat, or facilitate range expansion when mutualism increases a plant's niche breadth. Polyploids are also expected to have greater niche breadth because of greater self-compatibility and phenotypic plasticity, increasing invasion success.

• For 847 legume species, we compiled data from published sources to estimate ploidy, symbiotic status with rhizobia, specificity on rhizobia, and the number of introduced ranges.

• We found that diploid species have had limited spread around the globe regardless of whether they are symbiotic or how many rhizobia partners they can host. Polyploids, by contrast, have been successfully introduced to many new ranges, but interactions with rhizobia constrain their range expansion. In a hidden state model of trait evolution, we also found evidence of a high rate of re-diploidization in symbiotic legume lineages, suggesting that symbiosis and ploidy may interact at macroevolutionary scales.

• Overall, our results suggest that symbiosis with rhizobia limits range expansion when legumes are polyploid but not diploid.

Introduction

In plants, both polyploidy and mutualism are thought to contribute to ecological and evolutionary success. Polyploidization is an important driver of plant speciation (Soltis & Soltis, 2016), because it confers instant reproductive isolation and is therefore expected to impact niche evolution and range expansion (Weiss-Schneeweiss *et al.*, 2013). Likewise, mutualism has been linked with high rates of lineage diversification and is thought to increase ecological opportunity by expanding niche breadth and giving rise to coevolution with mutualist partners (Gómez & Verdú, 2012; Hembry *et al.*, 2014; Weber & Agrawal, 2014; Zeng & Wiens, 2021). However, some work has shown that engaging in mutualism can slow diversification (Kaur *et al.*, 2019) and that relying on a mutualistic partner can limit range expansion (Simonsen *et al.*, 2017). Previous research has identified both polyploidy and mutualism as important determinants of where plants establish and persist (Sheth *et al.*, 2020). Polyploidy should also impact plant–microbe interactions (Segraves & Anneberg, 2016), making it

important to determine whether or how these factors interact to shape plant geographic ranges. Here, we test whether ploidy and symbiosis jointly impact invasion success in legumes.

There are several reasons why polyploids may be better invaders than diploids. Polyploids generally have greater genetic variation (Otto & Whitton, 2000) and phenotypic plasticity (Mattingly & Hovick, 2021), both of which may allow polyploids to rapidly adapt to novel habitats. Polyploid plants also often have higher rates of self-fertilization, a trait that is associated with greater invasion success (Barringer & Geber, 2008). However, tests of polyploidy's effects on range size have had mixed results. In *Clarkia*, polyploids have larger ranges than diploids (Lowry & Lester, 2006), consistent with the above predictions. By contrast, in the Potentilleae tribe of Rosaceae, polyploids have smaller range sizes than diploids (Brittingham *et al.*, 2018). Polyploid species may outcompete their diploid relatives only in more extreme habitats due to increased stress tolerance. Therefore, if polyploids can persist only in extreme habitats, they could become niche specialists, making it difficult to spread to new environments (Hummer, 2012; López-Jurado *et al.*, 2019). Overall, broad-scale studies have failed to uncover a significant impact of

---

*These authors contributed equally to this work.

ploidy on range expansion (Hijmans *et al.*, 2007; Martin & Husband, 2009; Glennon *et al.*, 2014), suggesting that polyploidy may facilitate range expansion only under specific conditions, for instance when polyploids are formed through hybridization (Arrigo *et al.*, 2016) or if polyploidy leads to greater allocation toward dispersal-related traits (Coughlan *et al.*, 2017). Although legumes vary markedly in ploidy, no study has tested whether polyploid legumes have spread to more ranges than diploid legumes.

Legumes rely on bacterial partners called rhizobia to grow and reproduce in nutrient-poor soil. Rhizobia form nodules on plant roots where rhizobia fix atmospheric nitrogen into a readily available form of nitrogen for plants (van Rhijn & Vanderleyden, 1995). Although legume–rhizobium mutualisms are context-dependent (Simonsen & Stinchcombe, 2014) and rhizobia can vary in their quality as symbionts (Denison & Kiers, 2004; Heath & Tiffin, 2007), in most cases partnering with rhizobia is beneficial for plants (Friesen, 2012). Therefore, we assume that symbiosis with rhizobia is generally mutualistic and advantageous to plants. Although the ability to make nodules is a common and beneficial trait in species across the legume phylogeny, not all legumes nodulate. The legume–rhizobium symbiosis is over 100 Myr old, and it remains unclear if nodulation has evolved multiple times after an old predisposition event (Doyle, 2011; Werner *et al.*, 2014) or if the symbiotic trait has a single evolutionary origin followed by multiple losses across the legume phylogeny (Griesmann *et al.*, 2018).

Although mutualism can facilitate range expansion in some plant species (Afkhami *et al.*, 2014), a previous global analysis of legumes showed that symbiosis with rhizobia limits range expansion (Simonsen *et al.*, 2017). Legumes that depend on rhizobia for nitrogen may be unable to establish in a new range if they cannot find compatible symbionts there. Expanding on the analysis of Simonsen *et al.* (2017), Harrison *et al.* (2018) found that symbiotic legumes that associate with many rhizobia taxa have spread to more new ranges than symbiotic legumes that specialize on just one or a few rhizobia partners, again suggesting that the availability of compatible rhizobia constrains the spread of legumes around the globe. However, neither of these previous studies considered ploidy.

Ploidy is predicted to have important effects on the interaction between legumes and rhizobia. Autotetraploid plants obtain more fixed nitrogen from larger nodules than diploids (Forrester & Ashman, 2020). Polyploid plants can also be more generalized on rhizobia, obtaining greater benefits from a wider diversity of rhizobia partners than diploids (Forrester *et al.*, 2020). Legumes vary substantially in ploidy and polyploidy influences plant interactions with other species (Segraves & Anneberg, 2016), yet we currently lack broad-scale studies of how ploidy and symbiosis with rhizobia jointly influence invasion success in legumes. In this study, we (1) ask whether ploidy and symbiosis interact to affect range expansion in legumes and (2) estimate transition rates in symbiotic status and ploidy across the legume phylogeny to better understand how the symbiosis with rhizobia evolved in plants (e.g. whether it depends on evolutionary transitions in ploidy).

Materials and Methods

We used published data on symbiotic status from Werner *et al.* (2014) and geographic range data originally compiled from the International Legume Database and Information Service (ILDIS) by Simonsen *et al.* (2017) in our analysis. We obtained ploidy information for 847 species from these larger datasets. Overall, our global trait dataset of legume species comprised 232 genera and several subfamilies of legumes (Table S1).

Ploidy

We estimated ploidy (i.e. the number of copies of each chromosome in a cell) for 847 species of legumes using methods adapted from Brittingham *et al.* (2018). Specifically, we extracted total chromosome count values for each species from the Chromosome Count Database (CCDB; Rice *et al.*, 2015), the Index to Plant Chromosome Numbers (IPCN; Goldblatt & Johnson, 1979), or Rice *et al.* (2019). If a species was not found in one of these databases, we searched Web of Science for a genus-level average for the total chromosome count number (Table S2). When multiple sources reported different chromosome counts for the same species, we used the median value in our analyses. To calculate ploidy for each species, we divided the chromosome counts by the genus-level base chromosome number as reported in Fedorov (1969) and Bairiganjan & Patnaik (1989). If the base chromosome number for a genus was missing from the Fedorov (1969) table, chromosome counts were divided by the base chromosome number reported for the legume subfamily: Mimosoideae ($x = 13$; Santos *et al.*, 2012), Caesalpinioideae ($x = 7$; Resende *et al.*, 2013), or an average value for Papilionoideae ($x = 10.5$; Lackey, 1980) to find the ploidy level of the species. If there was no base chromosome number reported for the subfamily, we used the base chromosome number for the previously classified subfamily for that species (e.g. subfamily Detarioideae previously belonged to Caesalpinioideae) (LPWG, 2017). For our main analysis, we categorized all species with ploidy values ≤ 2 as diploids and all species with ploidy values > 2 as polyploids since there were few taxa with high ploidy. Although subfamily base chromosome number and genus base chromosome number values were significantly correlated in Kendall's rank test ($\tau = 0.411$, $P < 0.0001$), we expect the ploidy values calculated from subfamily base chromosome numbers to be imperfect. However, the dataset with ploidy calculated using only the genus-level base chromosome values had fewer species overall ($n = 671$) and only two nonsymbiotic polyploid species (Table S3). It is thus difficult to disentangle the influence of ploidy, lack of nodules, and other shared traits between these two species on the number of introduced ranges. Therefore, for our analyses, we used a combination of genus-level and subfamily-level estimated ploidy levels ($n = 847$). Specifically, if ploidy calculated from the genus-level base chromosome number was available, we used that value, and if it was not available, we used ploidy calculated using the subfamily-level base chromosome number. Since there was a small number of polyploid species calculated from subfamily base chromosome values, we searched the literature to confirm that these species are indeed polyploids (Table S4). Of these 62 species, 28

were confirmed to be polyploids, 22 were changed to diploid in our dataset, and 12 were left as polyploid because we could not find any information on ploidy level. We then chose 62 species at random from the list of diploids in the dataset and looked in the literature to confirm their ploidy level. Of these 62 species, two were changed to polyploid, 14 did not have ploidy levels reported in the literature, and the rest were confirmed to be diploid (Table S5), suggesting that our method incorrectly classified diploids in only *c.* 3% of species. We also confirmed that all eight polyploid generalists in our dataset were polyploid since the sample size was small in this category (Table S6). After corrections, our dataset consisted mostly of diploid species ($n = 570$) and fewer polyploid species ($n = 277$; Table S3).

Mutualists

We determined the symbiotic status of the legumes in our study from Werner *et al.* (2014). If a species is known to form nodules with rhizobia, it was categorized as a symbiotic species. We used the dataset assembled by Harrison *et al.* (2018) to determine whether each symbiotic legume species is a specialist or a generalist with regard to its interactions with rhizobia. Harrison *et al.* (2018) used data from Andrews & Andrews (2017) to determine the number of unique rhizobia genera that interact with a diverse group of legume species. Andrews & Andrews (2017) identified seven unique rhizobia genera based on 16 S rRNA, 16 S-23 S DNA, and nif/nod genes and then performed a literature search for legume species partnered with bacteria belonging to these genera. We classified a legume species as a specialist if it interacts with only one genus of rhizobia. Any legume species that interacts with more than one genus of rhizobia is a generalist.

Legume ranges

We obtained data on the number of introduced ranges for each legume species from Simonsen *et al.* (2017). Using geographic data from ILDIS, Simonsen *et al.* (2017) classified native and invaded ranges for each species as discrete geographical polygons that roughly align with geopolitical boundaries (Fig. S1). If a species was present in a non-native polygon and was touching neighboring non-native polygons, that area was counted as a single successful invasion event, or one introduced range. We assume that this reflects the successful introduction and establishment of a species in a new region followed by spread to contiguous areas. Therefore, the number of introduced ranges in the dataset is the sum of noncontiguous non-native polygons. It is nonetheless possible for a species to spread to multiple polygons from their native range and then later go extinct from some polygons, producing a patchy range with a high number of successful introductions. We included several covariates from the Simonsen *et al.* (2017) dataset in our models including the number of human uses for each legume species, area of the species' native range, the midpoint latitude of a species' native range, and whether or not the species was annual or perennial. On ILDIS, each species contains a report of the multiple ways humans use the plant in different industries (e.g. agriculture, medicine, and chemical). Simonsen *et al.* (2017) categorized these uses and summed the total number of unique uses for each legume species. We expect that humans facilitate the introduction of

legume species that are important in human practices. We included the area of the native range (summed total area of all native polygons) as a covariate since we expect that species with large native ranges may be more likely to establish in new areas because of broader niche breadth or greater propagule pressure. Previous analysis found that temperate species and annuals are more likely to establish in non-native regions (Simonsen *et al.*, 2017), so we also included these as covariates in our models, using the midpoint latitude (i.e. the midpoint of the native polygons for each species) to measure whether a legume is temperate, subtropical, or tropical.

Statistical analyses

All analyses were performed in R (R Core Team, 2020). We used the *gls* function in the package NLME (Pinheiro *et al.*, 2020) to fit phylogenetic least-squares models (PGLS). We pruned the Zanne *et al.* (2014) angiosperm phylogeny to retain only the species in our trait dataset. First, we modeled the number of introduced ranges as a function of the main and interactive effects of symbiotic status (symbiotic/nonsymbiotic) and ploidy (diploid/polyploid or ploidy level). However, because the trait dataset is highly unbalanced with fewer nonsymbiotic species (44 nonsymbiotic diploids and 24 nonsymbiotic polyploids) compared with symbiotic species (526 symbiotic diploids and 253 symbiotic polyploids), we also analyzed the impact of ploidy on range expansion separately for symbiotic and nonsymbiotic legumes. For the symbiotic species in the dataset, we also modeled the main and interactive effects of specialization (specialist/generalist) and ploidy (diploid/polyploid) on introduction success. This analysis had fewer taxa than the symbiosis presence/absence dataset (Table S3); specifically, it comprised 36 diploid generalists, 68 diploid specialists, eight polyploid generalists, and 29 polyploid specialists. We not only treated symbiotic status, specialization, and ploidy as binary, categorical variables but also performed a set of analyses in which ploidy was modeled as a continuous variable. All analyses included the following covariates: number of human uses, scaled value of total native area, life cycle duration (annual, perennial), and absolute latitude of origin. Before fitting a PGLS model, we performed model selection on the saturated model (all covariates and interaction between ploidy and symbiotic status) using the *dredge* function in the package MUMIN (Burnham & Anderson, 2002). The largest difference in Akaike information criterion (AIC) scores between models with and without interaction terms and covariates was below 2; therefore, we included all covariates in all our analyses. We allowed Pagel's lambda parameter to vary and optimize in the PGLS models. The internal branch lengths of the phylogeny are multiplied by this lambda parameter to account for phylogenetic signal in the data. A lambda value of 0 indicates that variation in the response variable is independent of phylogeny, and a lambda value of 1 indicates a Brownian motion model of evolution. We log-transformed (base 10) the number of introduced ranges (plus one to avoid zero values) to improve normality and homoscedasticity. When the PGLS models estimated negative lambda values (suggesting phylogenetic overdispersion in the data) or values very close to zero (suggesting very weak phylogenetic structure in the data), we also performed tests on the data using the *glm* function without accounting for phylogeny. Since the number of introduced ranges

was not normally distributed and overdispersed, we fit the data to a quasi-Poisson distribution. We tested for significance by performing a type III (on models with interaction terms) or a type II (on models without interaction terms) ANOVA using the *Anova* function in the CAR package (Fox & Weisberg, 2019).

State transitions

We modeled the evolution of ploidy and symbiosis across the Zanne *et al.* (2014) phylogeny with the package CORHMM (Beaulieu *et al.*, 2021) to estimate transitions between the following states: nonsymbiotic diploid, nonsymbiotic polyploid, symbiotic diploid, and symbiotic polyploid. We fit a simple CORHMM model with no hidden states (i.e. a single rate category) and also a model with hidden states (i.e. two rate categories) on the genus/subfamily-level ploidy data. The two rate categories represent additional hidden or unmeasured states that may affect evolutionary transitions in ploidy or symbiotic status. Because the hidden state model always outperformed the simple model, we plotted transition rates estimated from the hidden state models on the phylogeny using the *plotSimmap* function in PHYTOOLS (Revell, 2012).

Results

Symbiotic status and ploidy

Nonsymbiotic polyploids have successfully invaded more new ranges than symbiotic polyploids (Fig. 1a). By contrast, diploid species have successfully established in few new ranges regardless of symbiotic status (Fig. 1a). Symbiotic status and ploidy interacted
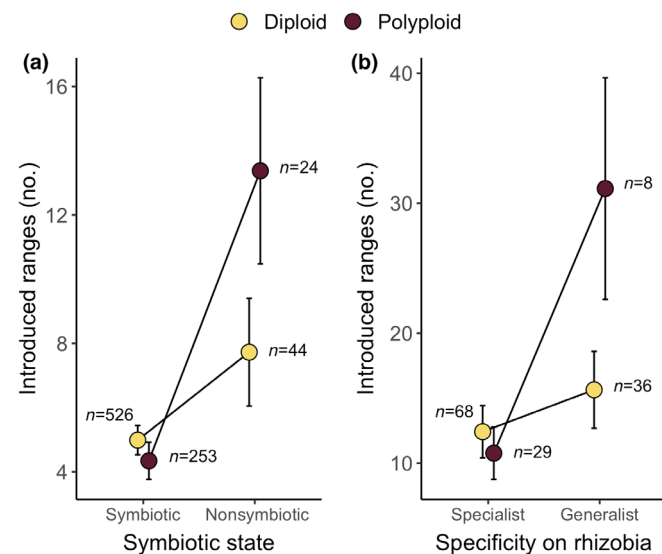

Fig. 1 Mean (±1 SE) number of introduced ranges for diploid (yellow) and polyploid (purple) legumes that (a) form nodules with rhizobia ('symbiotic') or do not form nodules ('nonsymbiotic') and (b) associate with more than one genus of rhizobia ('specialist') or associate with more than one genus of rhizobia ('generalist'). Total number of legume species in each category is listed as *n*. The analysis in (a) included 847 plant species across the legume phylogeny.

significantly to predict legume introduced ranges, although only in a nonphylogenetic model (Table 1). In the PGLS model that accounts for phylogeny, there was a nonsignificant interaction effect between symbiotic status and ploidy on the number of introduced ranges (Table S7). However, the PGLS model also showed a weak phylogenetic signal in the number of introduced ranges ($\lambda = 0.0641$), suggesting that phylogenetic correction is unnecessary. Furthermore, when we analyzed ploidy's effects separately within nonsymbiotic and symbiotic legumes, we found that ploidy significantly increased the number of introduced ranges in nonsymbiotic legumes ($\lambda = 0.3555$) but not in symbiotic legumes (Table S8). When we modeled ploidy as a continuous variable, we found a marginally significant interaction between ploidy level and symbiosis in a nonphylogenetic model (Table S9). We observed a positive relationship between ploidy level and introduced ranges in nonsymbiotic species but not in symbiotic species (Fig. S2); however, there were very few data points representing higher levels of ploidy in this analysis. In our second set of analyses on symbiotic species only, polyploid generalists were introduced to many more ranges than diploid generalists and both diploid and polyploid specialists (Fig. 1b). There was a significant interaction between specialization on rhizobia and ploidy in both our phylogenetically corrected (Table S7) and uncorrected models (Table 1). In both sets of analyses, the number of human uses was a significant and positive predictor of introduction success, which is consistent with past results (Simonsen *et al.*, 2017; Harrison *et al.*, 2018). Annual plants were associated with more introduced ranges, although this result was not always significant. Both midpoint latitude and area of the native range were significant and negative predictors of introduction success, which is consistent with Harrison *et al.* (2018) but in contrast with Simonsen *et al.* (2017).

State transitions

The hidden state model performed better ($-\log_e L = -398.327$, AIC = 832.655, $n = 847$) than the simple model ($-\log_e L = -448.216$, AIC = 912.432, $n = 847$); therefore, we present results only for the hidden state model. The transition rate from class 1 to class 2 was slightly higher (0.0400) than the transition rate from class 2 to class 1 (0.0200). The highest transition rate (3.5698) was from polyploidy to diploidy within nonsymbiotic lineages in the first rate class (Fig. 2). Overall, transitions in ploidy were higher within nonsymbiotic lineages than in symbiotic lineages. Within symbiotic lineages, transitions to diploidy were higher than transitions to polyploidy (Fig. 2).

Discussion

Symbiosis and ploidy interact to affect range expansion

Overall, we found that symbiosis only limits range expansion within polyploid species and that diploid legumes spread to few new ranges regardless of symbiotic status. Therefore, our results generally support the hypothesis that polyploid legumes are niche generalists and better able to colonize new habitats. However, when polyploids are also symbiotic or specialized on rhizobia, they are

Table 1 Estimates and results for effects of ploidy and symbiotic status and their interactions (denoted by ×) on the number of introduced ranges in legumes obtained from glm models fit with a quasi-Poisson distribution.

Factor	Estimate	SE	df	Wald χ^2	P
Symbiosis status, df = 818					
Ploidy (polyploid)	0.579	0.204	1	6.470	0.0110
Symbiosis (symbiotic)	0.262	0.175	1	2.510	0.1134
Total native area	−0.088	0.048	1	5.29	0.0215
Human uses	0.425	0.018	1	558.440	< 0.0001
Absolute latitude	−0.010	0.004	1	9.540	0.0020
Annual	0.368	0.135	1	9.670	0.0019
Ploidy × Symbiosis	**−0.618**	**0.238**	**1**	**6.100**	**0.0135**
Specialist status, df = 137					
Ploidy (polyploid)	0.417	0.185	1	2.530	0.1117
Specificity (specialist)	0.038	0.185	1	0.043	0.8359
Total native area	−0.171	0.050	1	10.218	0.0014
Human uses	0.273	0.030	1	75.270	< 0.0001
Absolute latitude	−0.015	0.004	1	9.141	0.0025
Annual	0.204	0.158	1	1.229	0.2676
Ploidy × Specificity	**−0.709**	**0.258**	**1**	**4.346**	**0.0371**

Ploidy is estimated from combined genus- and subfamily-level base chromosome numbers. Italicized factors are significant at $P < 0.05$, and P values reported here are the results of type III ANOVAs. Bolded factors highlight significant interaction effects on legume introductions. Ploidy was coded as 0 = diploid and 1 = polyploid, while symbiotic status was coded as 0 = nonsymbiotic and 1 = symbiotic in the model. Therefore, the intercept represents nonsymbiotic diploids. Specialists were coded as 1 and generalists as 0 in the model. Here, SE represents standard error, and df represents degrees of freedom.

restricted in their range expansion, suggesting that symbiotic legumes have difficulty finding a compatible rhizobia partner when they are introduced to a novel habitat (Simonsen *et al.*, 2017; Harrison *et al.*, 2018). Nonetheless, in legumes, diploidy seems to be the main factor restricting a plant's range; both nonsymbiotic diploids and generalist diploids successfully established in few new ranges despite not requiring or not specializing on rhizobia. Therefore, the impact of symbiosis and specialization on rhizobia (Simonsen *et al.*, 2017; Harrison *et al.*, 2018) on range expansion appears to apply only to polyploid legume species.

Polyploids might be good colonizers because they have more genetic material in their large genomes giving them greater adaptive potential (Otto & Whitton, 2000). Polyploids also often have greater capacity for phenotypic plasticity and can be better competitors than diploids, since they have fast germination and growth rates (te Beest *et al.*, 2012). However, being unable to associate with compatible rhizobia seems to cancel out the advantages of being polyploid. One explanation is that these symbiotic polyploids are largely specialists in terms of which rhizobia species they can form nodules with, making it more challenging to find rhizobia in new habitats. There were more specialist polyploids (29 species) than generalists (eight species) in our dataset, in contrast to previous results that report generalization on rhizobia in polyploids (Forrester *et al.*, 2020). In our analyses on symbiotic species only, we found that generalist polyploids were introduced to far more ranges than any other category of legume (Fig. 1b), providing further support for the theory that having many potential rhizobia partners is beneficial for range expansion, but only when paired with the benefits of being a polyploid.

Some caveats are in order, however. There were somewhat conflicting results for the interaction between symbiosis and ploidy in our PGLS analysis (nonsignificant) and nonphylogenetic GLM (significant). Because of the phylogenetic correction, the PGLS has reduced power. Given that the phylogenetic signal was low, this correction is unlikely necessary (Revell, 2010). We also used a combination of ploidy estimates calculated from genus- and subfamily-level base chromosome values. Although this allowed us to compile a larger dataset, it also introduced some errors in ploidy estimates since subfamily base values are less reliable. The species for which we have more accurate ploidy estimates (i.e. genus-level data) are likely well-studied species that have spread widely across the globe. We corrected the data for some taxa initially categorized as polyploids, and a smaller number initially categorized as diploids, by searching the literature to help deal with this issue. However, in order to compile large datasets, we accept there will be some inaccuracies in the data, especially when some species have reports of multiple ploidy levels (Hijmans *et al.*, 2007). The problems with ploidy estimation that we encountered are not unique to our study (Brittingham *et al.*, 2018). Overall, we were able to compile a large broad-scale dataset from online sources to estimate an overall pattern in range expansion in legumes and minimize some errors in the data by categorizing species into a
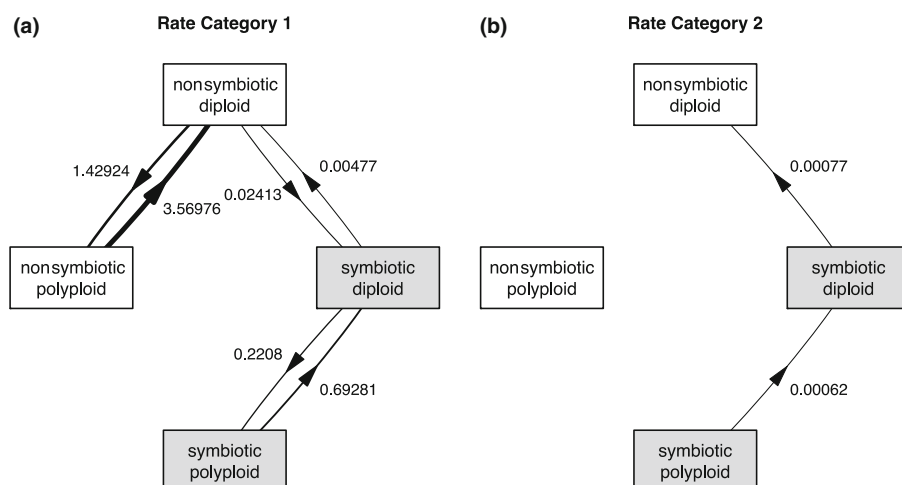

(a) Rate Category 1

(b) Rate Category 2

Fig. 2 corHMM results for hidden state model on the evolution of ploidy and symbiosis for (a) Rate Category 1 (R1) and (b) Rate Category 2 (R2). Ploidy is estimated from combined genus- and subfamily-level base chromosome numbers. Arrows in the plot show the direction of the transition rates between the four states of legumes: 'nonsymbiotic, diploid', 'nonsymbiotic, polyploid', 'symbiotic, diploid', and 'symbiotic, polyploid'. Symbiotic states are in gray, and nonsymbiotic states are in white. The thickness of the lines corresponds to the strength of the transition rate. Extremely low transition rates < 0.00001 were not plotted in the figure. The transition rate from R2 to R1 was 0.02, and the transition rate from R1 to R2 was 0.04.
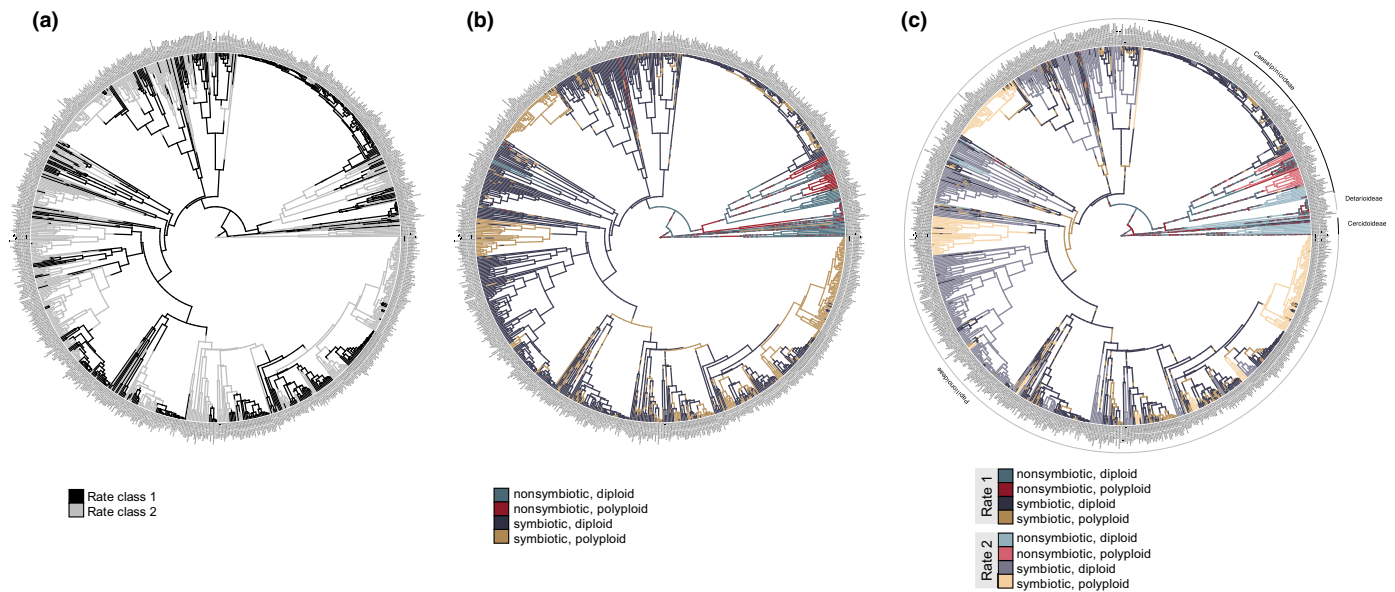
Fig. 3 Transition rates between four states ('nonsymbiotic diploid', 'nonsymbiotic polyploid', 'symbiotic diploid', and 'nonsymbiotic polyploid') in two rate classes (R1 and R2) estimated from the corHMM (hidden-state) model. Ploidy is estimated from combined genus and subfamily-level base chromosome numbers. In (a), only the two rate classes are depicted on the phylogeny (R1 = black, R2 = gray). Panel (b) shows the four states (rate class information not shown). Panel (c) shows all the information on one phylogeny, the four categories in both rate classes with lighter colors representing the four states in R2. Major subfamilies are labeled on the outside edges of the tree in panel (c).

binary variable of diploids and polyploids. We also expect our results are conservative; when we double-checked our ploidy estimates for a subset of taxa, far more polyploids turned out to be diploids than the other way around (22/62 diploids incorrectly classified as polyploids vs 2/62 diploids incorrectly classified as polyploids). Thus, some polyploids in our dataset might be diploid species with small ranges, and the removal of these diploids from the polyploid category should strengthen our results. We do not know whether most of the polyploid species in our dataset were formed through autopolyploidization (genome duplication) or allopolyploidization (hybridization), although recent work has estimated higher rates of autopolyploidization in plants (Barker *et al.*, 2015). We expect allopolyploids to exhibit novel phenotypes and contain more genetic variation than autopolyploids, potentially making it easier for allopolyploids to establish in new habitats than autopolyploids. Alternatively, allopolyploidization may be an outcome of species invading new ranges, because range expansion promotes contact between species and increases the likelihood of hybridization (Parisod *et al.*, 2009). Future studies should investigate the interaction between mechanisms of polyploid formation and symbiosis in legumes.

High rate of re-diploidization in legumes

Overall, the best-fitting model of trait evolution was a hidden state model, likely because other unmeasured traits contribute to rate heterogeneity across such a large and old clade as the legumes (Beaulieu & O'Meara, 2016). This model found that evolutionary transitions in ploidy were generally higher within nonsymbiotic than symbiotic lineages. Within symbiotic lineages, we observed a high rate of re-diploidization. Re-diploidization is a common process that occurs in many plant lineages (Tamayo-Ordóñez *et al.*, 2016) and

seems to occur more frequently in symbiotic than nonsymbiotic legume lineages. Polyploid species may undergo re-diploidization if the extra genetic material in the genome causes dosage imbalance, errors in mitosis or meiosis, disruption to gene regulation, or epigenetic instability (Comai, 2005). Larger genomes also tend to be more costly to maintain. After polyploid legumes evolved genes important for interactions with rhizobia, they may have experienced extensive gene loss (and thus genome reduction) during re-diploidization, which would reduce these costs.

The ancestral state in the legume tree was nonsymbiotic and diploid. Our analysis suggests that polyploidy directly evolved from diploidy, leading to the evolution of symbiosis followed by re-diploidization in symbiotic species (Fig. 3). There is evidence in the family Brassicaceae that whole-genome duplication has facilitated the evolution of diverse chemical defense compounds against herbivores (Edger *et al.*, 2015). If polyploidy did indeed evolve before symbiosis with rhizobia, the extra genetic material in the polyploid genome could provide more opportunity for mutations and the evolution of symbiosis to occur. However, previous in-depth analysis of transcriptome data in a few key legume species suggests that the evolution of polyploidy and symbiosis in legumes are unrelated (Cannon *et al.*, 2010). Overall, our analysis still suggests that polyploidy evolved first and is a potential predisposition event to the evolution of symbiosis. However, this is an important question that warrants further exploration with a larger dataset and better ploidy estimates.

Conclusion

Our results show the importance of considering the effects of both ploidy and symbiosis on invasion success; here, we showed that ploidy can influence the impact mutualism has on the spread of invasive species across the globe. Previous work has shown that

symbiosis (Simonsen *et al.*, 2017) and specialization on a small number of rhizobia partners (Harrison *et al.*, 2018) introduces a barrier to range expansion in legumes. Our study supports these results but shows that this barrier affects only polyploid legumes, which otherwise have an advantage over diploids in terms of successful introductions to new ranges. Looking ahead, the joint influence of ploidy and mutualism on invasion success should be evaluated in other mutualisms, such as plant–pollinator interactions; polyploids tend to be self-fertilizing, which is thought to facilitate range expansion (Barringer & Geber, 2008), but some work has suggested that invasive plants can take advantage of local pollinators (Graves & Shapiro, 2003). Our results also suggest that evolutionary transitions in ploidy and symbiosis might be linked at macroevolutionary scales; polyploidy appears to have arisen before symbiosis in legumes, perhaps setting the stage for the evolution of the legume–rhizobium mutualism. Nonetheless, once symbiotic, legumes frequently revert to their ancestral diploid state, likely explaining why so many extant symbiotic legumes are diploids.

Acknowledgements

Author contributions

TLH and MEF designed the study. AKS, ZAP and TLH collected data for the project. TLH, ZAP, AKS, JRS and MEF analyzed the data, interpreted the results, and wrote the manuscript. ZAP and TLH contributed equally to this work.

ORCID

Megan E. Frederickson https://orcid.org/0000-0002-9058-7137

Tia L. Harrison https://orcid.org/0000-0002-0282-1094

Anna K. Simonsen https://orcid.org/0000-0002-5091-261X

John R. Stinchcombe https://orcid.org/0000-0003-3349-2964

Data availability

All data and code for reproducing our analyses are available on the GitHub repository name: symbiotic-polyploid-legumes (https://github.com/harri318/symbiotic-polyploid-legumes).

References

Afkhami ME, McIntyre PJ, Strauss SY. 2014. Mutualist-mediated effects on species' range limits across large geographic scales. *Ecology Letters* 17: 1265–1273.

Andrews M, Andrews ME. 2017. Specificity in legume–rhizobia symbioses. *International Journal of Molecular Sciences* 18: 705.

Arrigo N, de La Harpe M, Litsios G, Zozomová-Lihová J, Španiel S, Marhold K, Barker MS, Alvarez N. 2016. Is hybridization driving the evolution of climatic niche in *Alyssum montanum*. *American Journal of Botany* 103: 1348–1357.

Bairiganjan GC, Patnaik SN. 1989. Chromosomal evolution in Fabaceae. *Cytologia* 54: 51–64.

Barker MS, Arrigo N, Baniaga AE, Li Z, Levin DA. 2015. On the relative abundance of autopolyploids and allopolyploids. *New Phytologist* 210: 391–398.

Barringer BC, Geber MA. 2008. Mating system and ploidy influence levels of inbreeding depression in Clarkia (onagraceae). *Evolution* 62: 1040–1051.

Beaulieu J, O'Meara B, Oliver J, Boyko J. 2021. *corHMM: hidden markov models of character evolution*. R package v.2.6. [WWW document] URL https://CRAN.R-project.org/package=corHMM [accessed 1 October 2021].

Beaulieu JM, O'Meara BC. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Systematic Biology* 65: 583–601.

te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubešová M, Pyšek P. 2012. The more the better? The role of polyploidy in facilitating plant invasions. *Annals of Botany* 109: 19–45.

Brittingham HA, Koski MH, Ashman T-L. 2018. Higher ploidy is associated with reduced range breadth in the Potentilleae tribe. *American Journal of Botany* 105: 700–710.

Burnham KP, Anderson DP. 2002. *Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn*. New York, NY, USA: Springer-Verlag.

Cannon SB, Ilut D, Farmer AD, Maki SL, May GD, Singer SR, Doyle JJ. 2010. Polyploidy did not predate the evolution of nodulation in all legumes. *PLoS ONE* 5: e11630.

Comai L. 2005. The advantages and disadvantages of being polyploid. *Nature Reviews Genetics* 6: 836–846.

Coughlan JM, Han S, Stefanović S, Dickinson TA. 2017. Widespread generalist clones are associated with range and niche expansion in allopolyploids of Pacific Northwest Hawthorns (*Crataegus* L.). *Molecular Ecology* 26: 5484–5499.

Denison FR, Kiers TE. 2004. Lifestyle alternatives for rhizobia: mutualism, parasitism, and forgoing symbiosis. *FEMS Microbiology Letters* 237: 187–193.

Doyle JJ. 2011. Phylogenetic perspectives on the origins of nodulation. *Molecular Plant–Microbe Interactions* 24: 1289–1295.

Edger PP, Heidel-Fischer HM, Bekaert M, Rota J, Glöckner G, Platts AE, Heckel DG, Der JP, Wafula EK, Tang M et al. 2015. The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences, USA* 112: 8362–8366.

Fedorov AA. 1969. *Chromosomal numbers of flowering plants*. Leningrad, Russia: Academy of Sciences of the USSR, the Komarov VL Botanical Institute, Nauka.

Forrester NJ, Ashman T-L. 2020. Autopolyploidy alters nodule-level interactions in the legume–rhizobium mutualism. *American Journal of Botany* 107: 179–185.

Forrester NJ, Rebolleda-Gómez M, Sachs JL, Ashman T. 2020. Polyploid plants obtain greater fitness benefits from a nutrient acquisition mutualism. *New Phytologist* 227: 944–954.

Fox J, Weisberg S. 2019. *An R companion to applied regression, 3rd edn*. Thousand Oaks, CA, USA: Sage. [WWW document] URL https://socialsciences.mcmaster.ca/jfox/Books/Companion/ [accessed 4 June 2020].

Friesen ML. 2012. Widespread fitness alignment in the legume–rhizobium symbiosis. *New Phytologist* 194: 1096–1111.

Glennon JL, Ritchie ME, Segraves KA. 2014. Evidence for shared broad-scale climatic niches of diploid and polyploid plants. *Ecology Letters* 17: 574–582.

Goldblatt P, Johnson DE. 1979. *Index to plant chromosome numbers*. St. Louis, MO, USA: Missouri Botanical Garden.

Gómez JM, Verdú M. 2012. Mutualism with plants drives primate diversification. *Systematic Biology* 61: 567–577.

Graves SD, Shapiro AM. 2003. Exotics as host plants of the California butterfly fauna. *Biological Conservation* 110: 413–433.

Griesmann M, Chang Y, Liu X, Song Y, Haberer G, Crook MB, Billault-Penneteau B, Lauressergues D, Keller J, Imanishi L et al. 2018. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* 361: eaat1743.

Harrison TL, Simonsen AK, Stinchcombe JR, Frederickson ME. 2018. More partners, more ranges: generalist legumes spread more easily around the globe. *Biology Letters* 14: 20180616.

Heath KD, Tiffin P. 2007. Context dependence in the coevolution of plant and rhizobial mutualists. *Proceedings of the Royal Society B: Biological Sciences* 274: 1905–1912.

Hembry DH, Yoder JB, Goodman KR. 2014. Coevolution and the diversification of life. *The American Naturalist* **184**: 425–438.

Hijmans RJ, Gavrilenko T, Stephenson S, Bamberg J, Salas A, Spooner DM. 2007. Geographical and environmental range expansion through polyploidy in wild potatoes (Solanum section Petota). *Global Ecology and Biogeography* **16**: 485–495.

Hummer KE. 2012. A new species of fragaria (Rosaceae) from Oregon. *Journal of the Botanical Research Institute of Texas* **6**: 9–15.

Kaur KM, Malé P-JG, Spence E, Gomez C, Frederickson ME. 2019. Using text-mined trait data to test for cooperate-and-radiate co-evolution between ants and plants. *PLoS Computational Biology* **15**: e1007323.

Lackey JA. 1980. Chromosome numbers in the Phaseoleae (Fabaceae: Faboideae) and their relation to taxonomy. *American Journal of Botany* **67**: 595–602.

López-Jurado J, Mateos-Naranjo E, Balao F. 2019. Niche divergence and limits to expansion in the high polyploid *Dianthus broteri* complex. *New Phytologist* **222**: 1076–1087.

Lowry E, Lester SE. 2006. The biogeography of plant reproduction: potential determinants of species' range sizes. *Journal of Biogeography* **33**: 1975–1982.

Martin SL, Husband BC. 2009. Influence of phylogeny and ploidy on species ranges of North American angiosperms. *Journal of Ecology* **97**: 913–922.

Mattingly KZ, Hovick SM. 2021. Autopolyploids of *Arabidopsis thaliana* are more phenotypically plastic than their diploid progenitors. *Annals of Botany*. doi: 10. 1093/aob/mcab081.

Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annual Review of Genetics* **34**: 401–437.

Parisod C, Holderegger R, Brochmann C. 2009. Evolutionary consequences of autopolyploidy. *New Phytologist* **186**: 5–17.

Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. 2020. *NLME: linear and nonlinear mixed effects models.* R package v.3.1-150. [WWW documnt] URL https://CRAN.R-project.org/package=nlme [accessed 4 June 2020].

R Core Team. 2020. *R: a language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. [WWW document] URL https://www.R-project.org/ [accessed 4 June 2020].

Resende KFM, Davide LC, Torres GA. 2013. Chromosome number and meiosis in populations of *Senna* species (Caesalpinioideae – Fabaceae) from Southeast Brazil. *Caryologia* **66**: 1–5.

Revell L. 2010. Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution* **1**: 319–329.

Revell LJ. 2012. PHYTOOLS: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**: 217–223.

van Rhijn P, Vanderleyden J. 1995. The rhizobium-plant symbiosis. *Microbiological Reviews* **59**: 124–142.

Rice A, Glick L, Abadi S, Einhorn M, Kopelman NM, Salman-Minkov A, Mayzel J, Chay O, Mayrose I. 2015. The chromosome counts database (CCDB) – a community resource of plant chromosome numbers. *New Phytologist* **206**: 19–26.

Rice A, Šmarda P, Novosolov M, Drori M, Glick L, Sabath N, Meiri S, Belmaker J, Mayrose I. 2019. The global biogeography of polyploid plants. *Nature Ecology & Evolution* **3**: 265–273.

Santos ECXR, Carvalho R, Almeida EM, Felix LP. 2012. Chromosome number variation and evolution in neotropical Leguminoseae (Mimosoideae) from northeastern Brazil. *Genetics and Molecular Research* **11**: 2451–2475.

Segraves KA, Anneberg TJ. 2016. Species interactions and plant polyploidy. *American Journal of Botany* **103**: 1326–1335.

Sheth SN, Morueta-Holme N, Angert AL. 2020. Determinants of geographic range size in plants. *New Phytologist* **226**: 650–665.

Simonsen AK, Dinnage R, Barrett LG, Prober SM, Thrall PH. 2017. Symbiosis limits establishment of legumes outside their native range at a global scale. *Nature Communications* **8**: 14790.

Simonsen AK, Stinchcombe RS. 2014. Herbivory eliminates fitness costs of mutualism exploiters. *New Phytologist* **202**: 651–661.

Soltis PS, Soltis DE. 2016. Ancient WGD events as drivers of key innovations in angiosperms. *Current Opinion in Plant Biology* **30**: 159–165.

Tamayo-Ordóñez MC, Espinosa-Barrera LA, Tamayo-Ordóñez YJ, Ayil-Gutiérrez B, Sánchez-Teyer LF. 2016. Advances and perspectives in the generation of polyploid plant species. *Euphytica* **209**: 1–22.

The Legume Phylogeny Working Group (LPWG). 2017. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. *Taxon* **66**: 44–77.

Weber MG, Agrawal AA. 2014. Defense mutualisms enhance plant diversification. *Proceedings of the National Academy of Sciences, USA* **111**: 16442–16447.

Weiss-Schneeweiss H, Emadzade K, Jang T-S, Schneeweiss GM. 2013. Evolutionary consequences, constraints and potential of polyploidy in plants. *Cytogenetic and Genome Research* **140**: 137–150.

Werner GDA, Cornwell WK, Sprent JI, Kattge J, Kiers ET. 2014. A single evolutionary innovation drives the deep evolution of symbiotic N$_2$-fixation in angiosperms. *Nature Communications* **5**: 4087.

Zanne AE, Tank DC, Cornwell WK, Eastman JM, Smith SA, FitzJohn RG, McGlinn DJ, O'Meara BC, Moles AT, Reich PB *et al.* 2014. Three keys to the radiation of angiosperms into freezing environments. *Nature* **506**: 89–92.

Zeng Y, Wiens JJ. 2021. Species interactions have predictable impacts on diversification. *Ecology Letters* **24**: 239–248.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Global geographic polygons for calculation of introduced ranges.

Fig. S2 Impact of ploidy level on introduced ranges.

Table S1 Sample sizes for the different subfamilies of legumes in the global dataset.

Table S2 Base chromosome counts for different legume genera and subfamilies obtained from the literature.

Table S3 Summary of the number of plant species in each category of data.

Table S4 Polyploid species identified in the global legume dataset.

Table S5 Diploid species identified in the global legume dataset.

Table S6 Species identified as polyploid and generalists in the global legume dataset.

Table S7 PGLS estimates and results for effects of ploidy and symbiotic status and their interactions on the number of introduced ranges in legumes.

Table S8 Results of PGLS and GLM models testing the effect of ploidy on introduction success in two separate categories of the data: symbiotic and nonsymbiotic species.

Table S9 PGLS and GLM estimates and results for the effect of ploidy level and symbiotic status on the number of introduced ranges in legumes.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.