UC Davis
UC Davis Electronic Theses and Dissertations

Title

Ecology, genetics, and genomics of biological nitrogen fixation in the wild: Cicer spp. symbiosis with Mesorhizobium spp. as a source for useful alleles in agriculture.

Permalink

https://escholarship.org/uc/item/8951w79b

Author

Perilla Henao, Laura Margarita

Publication Date

2022

Supplemental Material

https://escholarship.org/uc/item/8951w79b#supplemental

Peer reviewed|Thesis/dissertation

Ecology, genetics, and genomics of biological nitrogen fixation in the wild:
Cicer *spp.* symbiosis with *Mesorhizobium* spp.
as a source for useful alleles in agriculture.

By

LAURA MARGARITA PERILLA HENAO
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Plant Pathology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____
Douglas Randal Cook, Chair

_____
Gitta Coaker

_____
Johan Leveau

Committee in Charge

2022

i

"*Gracias a la vida*"

Estoy profundamente agradecida con cada una de las personas que me acompañaron y apoyaron en el camino de estudiar la ciencia de la vida. En este camino han partido varios seres queridos y estar a la distancia ha sido el desafío mas grande para aceptar el destino. Hoy agradezco a mi familia y a la vida por la oportunidad de seguir aprendiendo cada día.

TABLE OF CONTENTS

Laura Margarita Perilla Henao

DISSERTATION ABSTRACT

Nitrogen fixation efficiency is a desired trait in agriculture. Here we study nitrogen fixation traits in chickpea (*Cicer arietium*), which has special relevance given the crop's predominance as an important source of protein and nutrition in the developing world. This study seeks to understand the relevance of genetic and transcriptional diversity of both the crop (*C. arietinum*), its two wild relatives *C. reticulatum* and *C. echinospermum*, and their cognate symbiotic microbes *Mesorhizobium mediterraneum* and *M. ciceri*. The experiments test the hypothesis that local adaptation is an ecological mechanism operating in wild systems to optimize biological nitrogen fixation, by quantifying biomass gain and nodulation phenotypes in a network of naturally-occurring host and bacterial genotype combinations. Plant biomass gain was greater for native (homologous) compared to non-native (heterologous) plant-bacteria combinations, leading to the conclusion that nitrogen fixation per se is more effective in co-evolved plant-microbe pairs, providing strong evidence of local adaptation in natural systems. *C. echinospermun* has effective symbiosis with its native, homologous symbiont *M. ciceri* and it is incompatible (nod-) with the heterologous symbiont *M. mediterraneum*. Conversely, *C. reticulatum* has more effective symbiosis with its native, homologous symbiont *M. mediterraneum* compared to heterologous *M. ciceri*. Interestingly, *C. reticulatum* has similar nodulation phenotypes with both bacterial species, and thus increased biomass gain in the homologous interaction is interpreted as greater efficiency of nitrogen fixation. The analysis of recombinant inbred lines derived from *C. echinospermum* X *C. arietinum* inoculated with either *M. mediterraneum* or *M. ciceri* document genetic segregation of symbiotic specificity, providing the basis for molecular genetic studies. Transcriptional profiling analysis of compatible and incompatible symbiotic pairs revealed responses to nitrogen (symbiotic and inorganic) that are both common and different among plant species. In particular, enhanced transcription of genes related to carbon metabolism and plastid-related functions is characteristic of *M. ciceri* nodules, while interaction with *M. mediterraneum* is associated with upregulation a diversity of processes, including nitrogen related gene

iv

expression, but largely exclusive of carbon metabolism. Whole genome analyses were used to characterize bacterial strains associated with chickpea cultivation in Pakistan and strains from the wild's center of origin in South-Eastern Turkey. The analysis focused, in particular, on the provenance of genes involved in Type III secretion and Nod factor synthesis, informing us about the evolution of bacterial symbiosis in chickpea and bacterial loci relevant in the host-microbe interaction. The understanding gained from these studies informs fundamental questions about *Cicer-Mesorhizobium* co-evolution, and also has two broad practical implications: (1) for breeding of modern cultivars with improved biological nitrogen fixation, and (2) for developing improved microbial inoculants.

CHAPTER I

Introduction

Chickpea (*Cicer arietinum*) is one of the most consumed legumes, with its seeds providing an important source of protein nitrogen in the human diet. Globally the production of chickpea is characterized by large yield gaps that distinguish production in developed from developing countries. For example, in India and Pakistan, which constitute 78% of global chickpea acreage and where the crop is crucial to food security, yields are only 27-65% of that in developed countries (FAO STAT, 2016). Differences in yield outcomes across regions reflect differences in both agronomic practices and genetic gain from breeding programs that have focused on agronomic traits, such as yield, phenology and disease resistance. In recent decades, legume production has increased primarily due to the expansion of cultivated area (Foyer et al., 2016), often into marginal lands where legumes' chief advantage – biological nitrogen fixation (BNF) - is less effective. By comparison, increases in cereal production are largely due to crop genetic improvement and the use of fertilizers (Foyer et al., 2016). The extent to which genetic improvements to BNF might improve legume productivity is unresolved, although theoretically there are several opportunities: improve the effectiveness of microbial strains, increase the responsiveness of host genotypes, and modify agronomic conditions to better support symbiosis (e.g., reduce the use of N fertilizers, improve phosphate availability, and reduce abiotic stress) (Bayer et al., 2016, Mendoza-Suárez et al., 2021, Adams et al., 2018).

Soybean (*Glycine max*), for example, which accounts for 70% of global legume production, is often cultivated with agronomic practices that reduce the benefits of BNF. Nitrogen is typically applied to the crop at planting, which increases seedling establishment, but likely reduces the strength of selection for BNF traits. Similarly, many soybean fields contain highly competitive *Bradyrhizobium* strains that are poor nitrogen fixers and that reduce the efficacy of farmer-provided inoculum (Herridge et al., 2008, Adams et al., 2018). In South America, an estimated 60% of farmers use one or more inoculation doses of *Bradyrhizobium* per year, while in the U. S.

1

only 15% of soybean farmers use inoculants (Herridge et al., 2008). The use of inoculants by soybean farmers in Brazil and Argentina reduces their dependence on N fertilizer while increasing commercial yields (Herridge et al., 2008). Interestingly, in soybean there is evidence that modern breeding may have reduced the effectiveness of symbiosis (Kiers et al., 2007).

Legume BNF integrates plant photosynthesis with microbial nitrogen fixation to create organic nitrogen that sustains both plant and soil nutrition (Jackson et al., 2008). Thus leguminous crops offer a carbon-neutral alternative to nitrogen fertilizers that are otherwise (most commonly) gained from the fossil fuel-intensive Haber-Bosch process (Gu et al., 2013). Efficient symbiotic nitrogen fixation is key to sustainable crop production, not just for the legume but also for crops that follow in rotation using strategies of conservation agriculture (Bayer et al., 2016). However, rates of nitrogen fixation among legume crops can be highly variable (Adams et al., 2018). Surprisingly, the factors that influence the efficiency of nitrogen fixation in legume crops are poorly understood and represent a critical research need. In the majority of legume crops, breeding efforts to improve nitrogen fixation have been either absent or poorly conceived.

Symbiotic efficiency can be quantified by determining the ratio of nitrogen isotopes in plant tissue – the so-called "% Ndfa" method (nitrogen derived from the atmosphere) (Herridge et al., 2008). The natural abundance of N isotopes in the atmosphere is constant across the globe at 99.6337% $\delta^{14}N$ and 0.3663% $\delta^{15}N$. The enzyme responsible for BNF, nitrogenase, maintains this stoichiometry exactly. Conversely, soil sources of reduced nitrogen can have altered ratios due to biological and physical factors or as a consequence of experimental isotope dilution. When the $\delta^{14}N$: $\delta^{15}N$ of biologically available soil N is known, one can calculate lifetime nitrogen fixation histories in a crop. According to estimates of $\delta^{15}N$, nitrogen fixation rates are variable, the cultivated species with lowest efficiency is common bean with ~36% Ndfa; soybean and groundnut have intermediate values of ~58% Ndfa; chickpea, lentil, pea, cowpea, mug bean, pigeon pea, fababean, and lupins have values in the range of 65% Ndfa (Herridge et al., 2008). Nitrogen fixation efficiency can vary as dramatically as 10-fold within species, depending on growth conditions (Den Herder & Parniske, 2009).

Nodulation and Nitrogen fixation in legumes: regulation by host and symbiont.

The majority of legume species (family Fabaceae) associate with facultative legume bacterial endosymbionts that belong to the Alpha- or Beta- proteobacteria and that fix nitrogen in exchange for photosynthate (Mus et al., 2016, Remigi et al., 2016, MacLean et al., 2007). Symbiotic nitrogen fixation occurs in a specialized root organ, the "nodule", with crop legumes contributing an estimated ~40 million tons N/Year (Herridge et al., 2008). The genes and functional pathways controlling nitrogen fixation are increasingly well understood, especially signaling during early symbiotic development (Young et al., 2011, Wang et al., 2010, Oldroyd et al., 2011, Mergaert et al., 2020). Broadly, different stages in the establishment of a functional N-fixing nodule include early recognition at the root hair, morphological changes that entrap bacteria within a root hair curl, infection of the host cell by the bacterium via infection threads, cell division in the root cortex to form a nodule primordium, nodule growth and differentiation, co-differentiation of plant and bacterial cells to form the nitrogen fixing "symbiosome" organelle, nitrogen fixation and related metabolism, maintenance and later senescence of the nodule organ (Udvardi & Poole, 2013, Jardinaud et al., 2022, Lambert et al., 2020, Wang et al., 2018).

Host specificity is a hallmark of legume nitrogen fixation. Symbiotic rhizobia secrete lipochitooligosaccharides signals (Nod-Factors) that are perceived by host receptors and transduced to trigger gene expression and developmental outcomes. Signaling pathways are functionally conserved among legumes. Initial interactions occur at the root hair tip, and many of the molecular players involved in these early recognition steps have been identified using the tools of forward/reverse genetics and transcriptional profiling) (Larrainzar et al., 2015, Jardinaud & Boivin, 2016, Den Herder & Parniske, 2009, Perret et al., 2000, Kawaharada et al., 2015, Ane et al., 2002). Specificity begins with exudates, principally flavonoids, which are secreted from roots and recognized by rhizobial *nodD* gene LysR-type transcriptional regulators. NodD proteins activate the bacterium's "*nod*" (nodulation) regulon, whose gene products synthesize chito-

oligosaccharide signals called Nod-Factors (NF) (Perret et al., 2000). Different legume species produce characteristic flavonoids and each rhizobial strain produces slightly different NF structures (Perret et al., 2000).

Plant recognition of NF is carried out by membrane-localized receptors, eliciting plant signal transduction and response pathways and the rapid, staged expression of thousands of host genes (Larrainzar et al., 2015). The LysM family of receptor-like kinases (RLKs) are membrane localized receptors (Arrighi et al., 2006, Haney et al., 2011) with extracellular chitin-like binding domains that bind NF and determine specificity (Bozsoki et al., 2020). LysM RLKs with established phenotypes include NFP1, LYK3, LYK4 in *M. truncatula* and NFR1, NFR5, and EPR3 in *Lotus japonicus* (Kawaharada et al., 2015). NFP1, NFR1, and NFR5 are associated with early NF signaling (Madsen et al., 2003, Radutoiu et al., 2003). Conversely LYK3 and LYK4, which also signal NF, are implicated in bacterial entry along infection threads (Smit et al., 2007). Interestingly, EPR3 is a LysM RLK that recognizes bacterial exopolysaccharides (EPS) during bacterial colonization (Haney et al., 2011, Kawaharada et al., 2015, Smit et al., 2007).

The plant responds to NF by activating a signaling cascade mediated by calcium spiking (Capoen et al., 2011). The downstream pathway includes several additional receptors and kinases, among which DMI2 is a leucine-rich RLK required for calcium spiking and DMI3 (CCaMK) is a calcium-activated protein kinase that perceives calcium transients. Calcium spiking is localized to the nuclear compartment and requires the action of nuclear cation exchanger DMI1 to gate the process. On activation by calcium oscillations, DMI3 activates transcription factors at the top of a transcriptional cascade, among which are the TFs NSP1, NSP2, and ERN1. Interestingly, at least seven genes from the NF signal transduction pathway are required for arbuscular mycorrhizal (AM) symbiosis (initially reported by Duc et al., 1989), a widespread and ancient form of plant root symbiosis with fungi (Parniske, 2008). Similarities among bacterial and fungal root endosymbiosis indicate that pre-existing plant AM genes were key during the evolution of legume root nodule symbiosis (Staehelin et al., 2011, Markmann & Parniske, 2009, Martin et al., 2017).

The quality of symbiotic interactions can also be regulated by plant hormones, as in the case of negative regulation by ethylene (Larrainzar et al., 2015, Penmetsa & Cook, 1997) or regulation of cell division via cytokinin (Tirichine et al., 2007, Reid et al., 2017). Additional local and systemic regulation occurs in response to the carbon: nitrogen (C:N) ratio, hormonal crosstalk and N metabolites (Coruzzi & Bush, 2001, Alvarez et al., 2012, Tsay et al., 2011). Specialization of transporters in plant organs and subcellular compartments as well as N status sensing involves local and systemic signals, post-transcriptional regulation and small RNAs (Vidal et al., 2015, Yanagisawa, 2014, Alvarez et al., 2012, O'Brien et al., 2016). Fine regulation involves small peptides, for example CLAVATA3/Endosperm surrounding region-related (CLE) peptides. These are systemic signals in pathways inhibitory of nodulation, while c-terminally encoded peptides (CEP) are positive regulators of nodulation (Murray et al., 2016, Reid et al., 2011). Reduced forms of nitrogen, especially ammonium and nitrate, are a key convergence point, with regulatory pathways sensing plant N status and feeding back onto nodulation (Murray et al., 2016). Thus, high levels of reduced nitrogen in the soil can inhibit nodulation and the application of exogenous nitrogen can induce a rapid decline in nodule metabolism leading to nodule senescence (Cabeza et al., 2014, Gallusci et al., 1991).

The efficiency of symbiosis is reflected in the level of assimilated N, which is integrated by the plant and easily measured in terms of plant growth and grain yield (Den Herder & Parniske, 2009, Udvardi & Poole, 2013), with the latter being the most agronomically relevant measure. Effectiveness of nodules is correlated with the degree and maintenance of bacterial development. N-fixation occurs in organelles known as "symbiosomes" composed of specialized plant membranes surrounding the differentiated bacterium known as a "bacteroid" (Maróti & Kondorosi, 2014, Mergaert et al., 2006). Ineffective nodules are often small and lack the characteristic pink coloration of leghemoglobins, the haem-containing proteins abundant in nodules that binds oxygen (Jiang et al., 2021). Cytologically, ineffective nodules contain incompletely developed or senescent bacteroids. Plant cysteine-rich-repeat peptides (NCR's)

guide terminal bacteroid differentiation (Van de Velde et al., 2010) and are also implicated in bacterial cell death leading to ineffective nodules. *Medicago* mutants of the *defective in nitrogen fixation* (*dnf*) genes lack differentiated bacteroids (Wang et al., 2010). DNF1 is a signal peptidase necessary for the maturation of NCR's. Many legume symbionts secrete Type III effector proteins (Okazaki et al., 2013). Symbiotic Type III effectors are more conserved than those of pathogenic bacteria (Kimbrel et al., 2013) and can modulate host responses that either promote (Miwa & Okazaki, 2017) or restrict (Yang et al., 2010) symbiotic development, combined with corresponding host recognition systems (Zipfel & Oldroyd, 2017).

Nitrogen fixation is inextricably linked to nitrogen signaling, metabolism and transport in the plant host. Nitrogen uptake and assimilation, and nitrogen remobilization are increasingly well described in model plants such as *M. truncatula* and *Arabidopsis thaliana*. Improving crop N use efficiency necessarily involves understanding the mechanisms by which plant roots sense and respond to the heterogeneous N compounds in the soil. In the soil the most prevalent form is the nitrate ion (NO_3), but also commonly ammonium (NH_4), amino acids, and urea (Xu et al., 2012). Plants uptake, assimilate and remobilize these ions using specialized proteins that acquire, transport and convert exogenous nitrogen into biomolecules. For example, root cells contain ion membrane transporters, with distinct families of proteins involved in uptake of nitrate (NTR), ammonia (AMT), and urea (DUR) (O'Brien et al., 2016, Leran et al., 2014, Li et al., 2014). Ammonia, which is the immediate product of nitrogen fixation and also an intermediate in the assimilation of urea and nitrate, is toxic above certain species-specific thresholds (Esteban et al., 2016). Thus the production of ammonia from nitrogen fixation and other sources is tightly linked to its assimilation and transport as amino acids or ureides via various transport systems (Xu et al., 2012). Nitrate and amino acids are transported from roots to shoots. In sink organs, N is stored in the form of photosynthetic enzymes that are degraded and remobilized during increased nitrogen demand during senescence, grain filling, or N starvation (Gallardo et al., 2008, Fischer et al., 2013, Vidal et al., 2010, Coruzzi & Bush, 2001). Plant nitrogen metabolism is dynamic in space and time and adjusts according to phenology and physiological demand. However, the

genetic networks for regulation of N metabolism and N starvation are not yet fully understood (Fischer et al., 2013).

Crop domestication

Domesticated species are known for their reduced genetic variability compared to their wild progenitors, usually as the product of bottlenecks imposed by intense selection for a few traits. Conversely, the growth of domesticated species in highly managed agricultural environments can lead to relaxed selection, by reducing dependence on traits that are compensated by agricultural practices. Taken together, these processes create reduced fitness of domesticated species in their natural settings (Ross-Ibarra et al., 2007, Olsen & Wendel, 2013). Domestication traits include both large effect and small effect loci (Zhou et al., 2015, Olsen & Wendel, 2013). Some of the traits selected during domestication are the product of independent phenotypic convergence, such as larger size of edible organs, loss of seed dispersal mechanisms, erect plant architecture, modified growth habit, reduced seed dormancy, clear seed coat, and insensitivity to photoperiod (Ross-Ibarra et al., 2007).

Wild legume species and their symbionts co-evolved in natural environments, without the advent of human inputs. Subsequently, both domestication and crop improvement altered their (agro)ecological contexts, including reduction of host (crop) genetic diversity (von Wettberg et al., 2018), expansion to new geographic areas with different soil chemistries and microbiota (Alford, 2020), increasingly intensive management of soil nutrients and water, and encounters with new symbiotic partners as the crop spread globally (Greenlon et al., 2019).

STUDY SYSTEM

Chickpea (Cicer arietinum* L.) *is important as a crop and as a biological tool to study the domestication of nodulating legumes. Crop domestication represents humankind's first biotechnology, beginning ~12,000 years ago and occurring in parallel in different regions of the

world, leading to genetically distinct modern crop species (Meyer & Purugganan, 2013). An important domestication center is the Fertile Crescent, where wild relatives of legumes and cereals co-exist, including chickpea's wild relatives. Chickpea (*Cicer arietinum*) was domesticated along with other Near-Eastern Neolithic founder crops around 10,000 years ago (Zohary, 1999). The primary domestication site of chickpea is Southeast Anatolia (10 ka), with secondary diversification in India and Ethiopia (~2-6 ka) and recent introduction to the developed world (~100 ka). *C. reticulatum* is the primary wild gene pool for cultivated *C. arietinum*. Both species are self-pollinated, diploid (2n = 16), with genome sizes of ~750 Mbp (von Wettberg et al., 2018).

The Cook lab at UC Davis led an effort to broaden the genetic base of chickpea across production systems and regions, to deliver evidence-based solutions for chickpea growers. The first step in these efforts was a rational and detailed collection of plant and microbial germplasm, expanding the available genetic diversity of host crop wild relatives (*C. reticulatum* and *C. echinospermum*) (von Wettberg et al., 2018) and rhizobial symbionts (*Mesorhizobium* spp.) (Greenlon et al., 2019). Germplasm collection was accompanied by bioclimatic, geographic, population genetic and genomic analyses, permitting assessment of factors shaping modern genetic diversity.

Chickpea's wild relatives' collection comprises a diverse genetic, geographic, edaphic and climatic sampling of *Cicer* accessions, named after the local towns or roads near which they occurred (von Wettberg et al., 2018). The collection spans a broad range of habitats with different soil types, such as calcisols, leptosols, and luvisols. It also has an altitudinal gradient encompassing distinct micro-environments; for example the Cudi site has the highest cooling degrees days (approx. ~4000) in April-May, while Egil, Cermik and Sarikaya sites have the lowest cooling degree days (approx. ~3000) for the same season (von Wettberg et al., 2018). Importantly, certain traits have distributions that suggest local adaptation, which demonstrates the role of local factors shaping plant genetic diversity. For example, seed coat color (quantified as reflectance) has high genetic heritability, is variable between plant populations, and frequently matches spectral characteristics of local soils, suggesting a role for crypsis (camouflage) in

herbivore avoidance. Other traits such as transpiration rates, phenology and pest resistance, display a broad range of phenotypes suggesting potential variation in relevant agronomic traits (Moenga et al., 2020, von Wettberg et al., 2018).

C. reticulatum's and *C. echinospermum's* distinct genetic populations were characterized by a combination of technologies for whole genome (PacBio HiFi and Illumina), and reduced representation (RadSeq) sequencing, which permitted the identification of single nucleotide polymorphisms (SNP's) for population genetic and evolutionary analyses. Wild *C. reticulatum* were most diverse, with subpopulations having a large degree of genetic differentiation as shown by measure of genetic differentiation and variance in allele frequency with Wright's F-statistics (F_{ST}) values of 0.77 (von Wettberg et al., 2018). In parallel to characterization of plant genetic diversity, Greenlon et al (2019) characterized the genomes of a global collection of chickpea symbionts to understand the origins of microbial symbiont diversity. Analysis of ~1,400 *Mesorhizobium* genomes revealed massive species diversity among chickpea symbionts. At the center of origin, *C. reticulatum* (the immediate wild progenitor) is nodulated by only three related *Mesorhizobium* species, while the cultivated *C. arietinum* is nodulated by at least 28 diverse species of *Mesorhizobium.* These new symbiont species were recruited for chickpea symbiosis by horizontal gene transfer of a genomic island comprising ~10% of the genome (Greenlon et al., 2019). The symbiosis island carries genes for Nod factor biosynthesis, Type III secretion (core machinery and effector gene set), and bacterial genes for modulation of host ethylene production. Any one of these systems could contribute to differences in symbiotic effectiveness.

These previous collections and the corresponding characterization of host and microbe genetic diversity provide the foundation for the present study. Experimental tools and resources include (1) characterized accessions of wild chickpeas (*Cicer reticulatum* and *Cicer echinospermum*), (2) their respective cognate microbial symbionts (*Mesorhizobium mediterraneum* and *Mesorhizobium ciceri*), (3) domesticated chickpea whose germplasm separated from the wild progenitor *C. reticulatum* ~10 ka and the sister species *C. echinospermum* ~100 ka, and (4) advanced recombinant inbred populations of wild x cultivated

parentage. Chapter II evaluates the interaction between a network of naturally-occurring host and bacterial genotypes. Phenotyping of wild and cultivated plant accessions demonstrates that homologous, native combinations of host and microbial genotypes have greater efficiencies than heterologous, non-native combinations. Subsequent analysis of corresponding wild x cultivated recombinant inbred lines document genetic segregation of symbiotic specificity, providing the basis for molecular genetic studies. Chapter III examines transcriptional signatures of *Cicer* species in response to homologous (co-evolved) and heterologous bacterial symbionts, as well as to differing inorganic nitrogen treatments. The data reveal that transcriptional profiles differ depending on the source of nitrogen and the nature of the bacterial symbiont. Chapter IV presents data on the evolution of bacterial symbiosis in chickpea, by examining the genomic relationship between strains associated with chickpea cultivation in Pakistan and strains at the wild center of origin in South-Eastern Turkey, focusing in particular on the provenance of genes involved in Type III secretion and Nod factor synthesis.

REFERENCES

Adams MA, Buchmann N, Sprent J, Buckley TN, Turnbull TL, 2018. Crops, Nitrogen, Water: Are Legumes Friend, Foe, or Misunderstood Ally? *Trends Plant Sci* **23**, 539-50.

Alford BA, 2020. *Microbial Ecology of Wild and Domesticated Chickpeas*. Ann Arbor: University of California, DavisPh.D.

Alvarez JM, Vidal EA, Gutierrez RA, 2012. Integration of local and systemic signaling pathways for plant N responses. *Curr Opin Plant Biol* **15**, 185-91.

Ane JM, Levy J, Thoquet P, *et al.*, 2002. Genetic and cytogenetic mapping of DMI1, DMI2, and DMI3 genes of Medicago truncatula involved in Nod factor transduction, nodulation, and mycorrhization. *Mol Plant Microbe Interact* **15**, 1108-18.

Arrighi JF, Barre A, Ben Amor B, *et al.*, 2006. The Medicago truncatula lysin [corrected] motif-receptor-like kinase gene family includes NFP and new nodule-expressed genes. *Plant Physiol* **142**, 265-79.

Bayer C, Gomes J, Zanatta JA, Vieira FCB, Dieckow J, 2016. Mitigating greenhouse gas emissions from a subtropical Ultisol by using long-term no-tillage in combination with legume cover crops. *Soil and Tillage Research* **161**, 86-94.

Bozsoki Z, Gysel K, Hansen SB, *et al.*, 2020. Ligand-recognizing motifs in plant LysM receptors are major determinants of specificity. *Science* **369**, 663-70.

Cabeza R, Koester B, Liese R, *et al.*, 2014. An RNA sequencing transcriptome analysis reveals novel insights into molecular aspects of the nitrate impact on the nodule activity of Medicago truncatula. *Plant Physiol* **164**, 400-11.

Capoen W, Sun J, Wysham D, *et al.*, 2011. Nuclear membranes control symbiotic calcium signaling of legumes. *PNAS* **108**, 14348-53.

Coruzzi, Bush, 2001. N and carbon nutrients and metabolits in plants as signal Plant Physiol.-2001-Coruzzi-61-4. *Plant Physiol.*

Den Herder G, Parniske M, 2009. The unbearable naivety of legumes in symbiosis. *Curr Opin Plant Biol* **12**, 491-9.

Esteban R, Ariz I, Cruz C, Moran JF, 2016. Review: Mechanisms of ammonium toxicity and the quest for tolerance. *Plant Sci* **248**, 92-101.

Fischer JJ, Beatty PH, Good AG, Muench DG, 2013. Manipulation of microRNA expression to improve nitrogen use efficiency. *Plant Science* **210**, 70-81.

Foyer CH, Lam H-M, Nguyen HT, *et al.*, 2016. Neglecting legumes has compromised human health and sustainable food production. *Nature Plants* **2**, 16112.

Gallardo K, Thompson R, Burstin J, 2008. Reserve accumulation in legume seeds. *C R Biol* **331**, 755-62.

Gallusci P, Dedieu A, Journet EP, Huguet T, Barker DG, 1991. Synchronous expression of leghaemoglobin genes inMedicago truncatula during nitrogen-fixing root nodule development and response to exogenously supplied nitrate. *Plant Molecular Biology* **17**, 335-49.

Greenlon A, Chang PL, Damtew ZM, *et al.*, 2019. Global-level population genomics reveals differential effects of geography and phylogeny on horizontal gene transfer in soil bacteria. *Proc Natl Acad Sci U S A* **116**, 15200-9.

Gu B, Chang J, Min Y, *et al.*, 2013. The role of industrial nitrogen in the global nitrogen biogeochemical cycle. *Scientific Reports* **3**, 2579.

Haney CH, Riely BK, Tricoli DM, Cook DR, Ehrhardt DW, Long SR, 2011. Symbiotic rhizobia bacteria trigger a change in localization and dynamics of the Medicago truncatula receptor kinase LYK3. *Plant Cell* **23**, 2774-87.

Herridge DF, Peoples MB, Boddey RM, 2008. Global inputs of biological nitrogen fixation in agricultural systems. *Plant and Soil* **311**, 1-18.

Jackson LE, Burger M, Cavagnaro TR, 2008. Roots, nitrogen transformations, and ecosystem services. *Annu Rev Plant Biol* **59**, 341-63.

Jardinaud M-F, Carrere S, Gourion B, Gamas P, 2022. Symbiotic Nodule Development and Efficiency in the Medicago truncatula Mtefd-1 Mutant Is Highly Dependent on Sinorhizobium Strains. *Plant and Cell Physiology*, pcac134.

Jardinaud MF, Boivin S, 2016. A Laser Dissection-RNAseq Analysis Highlights the Activation of Cytokinin Pathways by Nod Factors in the Medicago truncatula Root Epidermis. **171**, 2256-76.

Jiang S, Jardinaud MF, Gao J*, et al.*, 2021. NIN-like protein transcription factors regulate leghemoglobin genes in legume nodules. *Science* **374**, 625-8.

Kawaharada Y, Kelly S, Nielsen MW*, et al.*, 2015. Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature* **523**, 308-12.

Kiers ET, Hutton MG, Denison RF, 2007. Human selection and the relaxation of legume defences against ineffective rhizobia. *Proc. R. Soc. B* **274**, 3119-26.

Kimbrel JA, Thomas WJ, Jiang Y*, et al.*, 2013. Mutualistic Co-evolution of Type III Effector Genes in Sinorhizobium fredii and Bradyrhizobium japonicum. *PLOS Pathogens* **9**, e1003204.

Lambert I, Pervent M, Le Queré A*, et al.*, 2020. Responses of mature symbiotic nodules to the whole-plant systemic nitrogen signaling. *J Exp Bot* **71**, 5039-52.

Larrainzar E, Riely BK, Kim SC*, et al.*, 2015. Deep Sequencing of the Medicago truncatula Root Transcriptome Reveals a Massive and Early Interaction between Nodulation Factor and Ethylene Signals. *Plant Physiol* **169**, 233-65.

Leran S, Varala K, Boyer JC*, et al.*, 2014. A unified nomenclature of NITRATE TRANSPORTER 1/PEPTIDE TRANSPORTER family members in plants. *Trends Plant Sci* **19**, 5-9.

Li Y, Krouk G, Coruzzi GM, Ruffel S, 2014. Finding a nitrogen niche: a systems integration of local and systemic nitrogen signalling in plants. *J Exp Bot* **65**, 5601-10.

Maclean AM, Finan TM, Sadowsky MJ, 2007. Genomes of the symbiotic nitrogen-fixing bacteria of legumes. *Plant Physiol* **144**, 615-22.

Madsen EB, Madsen LH, Radutoiu S*, et al.*, 2003. A receptor kinase gene of the LysM type is involved in legumeperception of rhizobial signals. *Nature* **425**, 637-40.

Markmann K, Parniske M, 2009. Evolution of root endosymbiosis with bacteria: How novel are nodules? *Trends Plant Sci* **14**, 77-86.

Maróti G, Kondorosi É, 2014. Nitrogen-fixing Rhizobium-legume symbiosis: are polyploidy and host peptide-governed symbiont differentiation general principles of endosymbiosis? *Frontiers in Microbiology* **5**.

Martin FM, Uroz S, Barker DG, 2017. Ancestral alliances: Plant mutualistic symbioses with fungi and bacteria. *Science* **356**.

Mendoza-Suárez M, Andersen SU, Poole PS, Sánchez-Cañizares C, 2021. Competition, Nodule Occupancy, and Persistence of Inoculant Strains: Key Factors in the Rhizobium-Legume Symbioses. *Frontiers in Plant Science* **12**.

Mergaert P, Kereszt A, Kondorosi E, 2020. Gene Expression in Nitrogen-Fixing Symbiotic Nodule Cells in Medicago truncatula and Other Nodulating Plants. *Plant Cell* **32**, 42-68.

Mergaert P, Uchiumi T, Alunni B, *et al.*, 2006. Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium-legume symbiosis. *Proc Natl Acad Sci U S A* **103**, 5230-5.

Meyer RS, Purugganan MD, 2013. Evolution of crop species: genetics of domestication and diversification. *Nat Rev Genet* **14**, 840-52.

Miwa H, Okazaki S, 2017. How effectors promote beneficial interactions. *Current Opinion in Plant Biology* **38**, 148-54.

Moenga SM, Gai Y, Carrasquilla-Garcia N, Perilla-Henao LM, Cook DR, 2020. Gene co-expression analysis reveals transcriptome divergence between wild and cultivated chickpea under drought stress. *Plant J* **104**, 1195-214.

Murray JD, Liu CW, Chen Y, Miller AJ, 2016. Nitrogen sensing in legumes. *J Exp Bot*.

Mus F, Crook MB, Garcia K, *et al.*, 2016. Symbiotic Nitrogen Fixation and the Challenges to Its Extension to Nonlegumes. *Appl Environ Microbiol* **82**, 3698-710.

O'brien JA, Vega A, Bouguyon E, *et al.*, 2016. Nitrate Transport, Sensing, and Responses in Plants. *Mol Plant* **9**, 837-56.

Okazaki S, Kaneko T, Sato S, Saeki K, 2013. Hijacking of leguminous nodulation signaling by the rhizobial type III secretion system. *Proceedings of the National Academy of Sciences* **110**, 17131-6.

Oldroyd GE, Murray JD, Poole PS, Downie JA, 2011. The rules of engagement in the legume-rhizobial symbiosis. *Annu Rev Genet* **45**, 119-44.

Olsen KM, Wendel JF, 2013. A bountiful harvest: genomic insights into crop domestication phenotypes. *Annu Rev Plant Biol* **64**, 47-70.

Parniske M, 2008. Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nat Rev Microbiol* **6**, 763-75.

Penmetsa RV, Cook DR, 1997. A Legume Ethylene-Insensitive Mutant Hyperinfected by Its Rhizobial Symbiont. *Science* **275**, 527-30.

Perret X, Staehelin C, Broughton WJ, 2000. Molecular Basis of Symbiotic Promiscuity. *Microbiology and Molecular Biology Reviews* **64**, 180-201.

Radutoiu S, Madsen LH, Madsen EB, *et al.*, 2003. Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* **425**, 585-92.

Reid DE, Ferguson BJ, Hayashi S, Lin Y-H, Gresshoff PM, 2011. Molecular mechanisms controlling legume autoregulation of nodulation. *Ann Bot* **108**, 789-95.

Reid DE, Nadzieja M, Novak O, Heckmann AB, Sandal N, Stougaard J, 2017. Cytokinin biosynthesis promotes cortical cell responses during nodule development. *Plant Physiol*.

Remigi P, Zhu J, Young JPW, Masson-Boivin C, 2016. Symbiosis within Symbiosis: Evolving Nitrogen-Fixing Legume Symbionts. *Trends in Microbiology* **24**, 63-75.

Ross-Ibarra J, Morrell PL, Gaut BS, 2007. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc Natl Acad Sci U S A* **104 Suppl 1**, 8641-8.

Smit P, Limpens E, Geurts R, *et al.*, 2007. Medicago LYK3, an Entry Receptor in Rhizobial Nodulation Factor Signaling. *Plant Physiol* **145**, 183-91.

Staehelin C, Xie ZP, Illana A, Vierheilig H, 2011. Long-distance transport of signals during symbiosis: are nodule formation and mycorrhization autoregulated in a similar way? *Plant Signal Behav* **6**, 372-7.

Tirichine L, Sandal N, Madsen LH, *et al.*, 2007. A gain-of-function mutation in a cytokinin receptor triggers spontaneous root nodule organogenesis. *Science* **315**, 104-7.

Tsay YF, Ho CH, Chen HY, Lin SH, 2011. Integration of nitrogen and potassium signaling. *Annu Rev Plant Biol* **62**, 207-26.

Udvardi M, Poole PS, 2013. Transport and metabolism in legume-rhizobia symbioses. *Annu Rev Plant Biol* **64**, 781-805.

Van De Velde W, Zehirov G, Szatmari A*, et al.*, 2010. Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science* **327**, 1122-6.

Vidal EA, Alvarez JM, Moyano TC, Gutierrez RA, 2015. Transcriptional networks in the nitrate response of Arabidopsis thaliana. *Curr Opin Plant Biol* **27**, 125-32.

Vidal EA, Tamayo KP, Gutierrez RA, 2010. Gene networks for N-sensing, signaling and response in Arabidopsis thaliana. *Wiley Interdiscip Rev Syst Biol Med* **2**, 683-93.

Von Wettberg EJB, Chang PL, Başdemir F*, et al.*, 2018. Ecology and genomics of an important crop wild relative as a prelude to agricultural innovation. *Nature Communications* **9**, 649.

Wang D, Griffitts J, Starker C*, et al.*, 2010. A nodule-specific protein secretory pathway required for nitrogen-fixing symbiosis. *Science* **327**, 1126-9.

Wang Q, Liu J, Li H*, et al.*, 2018. Nodule-Specific Cysteine-Rich Peptides Negatively Regulate Nitrogen-Fixing Symbiosis in a Strain-Specific Manner in Medicago truncatula. *Mol Plant Microbe Interact* **31**, 240-8.

Xu G, Fan X, Miller AJ, 2012. Plant nitrogen assimilation and use efficiency. *Annu Rev Plant Biol* **63**, 153-82.

Yanagisawa S, 2014. Transcription factors involved in controlling the expression of nitrate reductase genes in higher plants. *Plant Sci* **229**, 167-71.

Yang S, Tang F, Gao M, Krishnan HB, Zhu H, 2010. R gene-controlled host specificity in the legume–rhizobia symbiosis. *Proceedings of the National Academy of Sciences* **107**, 18735-40.

Young ND, Debelle F, Oldroyd GED*, et al.*, 2011. The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**, 520-4.

Zhou Z, Jiang Y, Wang Z*, et al.*, 2015. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat Biotech* **33**, 408-14.

Zipfel C, Oldroyd GED, 2017. Plant signalling in symbiosis and immunity. *Nature* **543**, 328-36.

Zohary D, 1999. Monophyletic vs. polyphyletic origin of the crops on which agriculture was founded in the Near East. *Genetic Resources and Crop Evolution* **46**, 133-42.

CHAPTER II

Efficient nitrogen fixation in *Cicer* spp.

ABSTRACT

The wild relatives of *Cicer*, *C. reticulatum* and *C. echinospermum*, have distinct geographic distributions that are associated with distinct soil types and elevations in South-East Anatolia. Microbial symbiont occurrence displays a similar pattern, with *Mesorhizobium mediterraneum* being native to the limestone soils that are preferred by *C. reticulatum* and *M. ciceri* being native to the basaltic soils where *C. echinospermun* grows. These patterns of co-occurrence could reflect local selection for optimal symbiotic partners, or they could be unrelated to symbiotic function and instead derive from demographic processes or local edaphic and environmental factors that influence the distribution of bacterial species independent of host selection.

In this research we tested whether the co-occurrence of *Cicer* spp. wild relatives reflect local adaptation by cross-inoculations in home (native) and away (non-native) genotypes. We also evaluated whether domestication involved functional changes in nitrogen responses, especially responses to symbiotic microbes and inorganic nitrogen.

We found that crop wild relatives have more efficient interactions with sympatric, homologous symbionts compared to allopatric, heterologous symbionts. Moreover, efficiency segregated at the level of plant and bacterial species. The observed changes in rank order of performance between homologous and heterologous symbiont pairs strongly supports an ecological hypothesis of local adaptation, where wild plant and microbial species are co-evolved for efficient nitrogen fixation and non-co-evolved partners have low performance. Interestingly, the strain specificity of cultivated chickpea genotypes appears to segregate within the germplasm, with some chickpea genotypes possessing the strain specificity of the wild progenitor and other chickpea genotypes lacking strain discrimination. Using recombinant inbred lines derived from cultivated *C. arietinum* x wild *C. echinospermum*, we tested the hypothesis that specificity for bacterial species has a simple genetic control. Genetic data are consistent with a

single recessive gene for specificity segregating within cultivated germplasm. The relevance of this study is that efficient nodulation and nitrogen fixation as genetic traits can be tractable to host wild alleles that favor a particular symbiont. One practical application is the identification of genetic markers associated with efficient biological nitrogen fixation (BNF) per symbiont, to be incorporated in chickpea breeding programs.

INTRODUCTION

Legumes are key components of both natural and agricultural systems, due in large part to their ability to associate with "nodule bacteria", creating a symbiotic relationship that produces vital reduced nitrogen.

<u>Nod factor specificity</u>. One of the best understood mechanisms for specificity in legume nodulation is the host's perception of a bacterial lipo-chito-oligosaccharide signal, known as "Nod Factor", and the initiation of nodule organogenesis. Different legume species produce specific (iso)flavonoid exudates from roots that activate the bacterial nodD family of transcriptional regulators, leading to Nod-Factor (NF) production (Perret et al., 2000). Each rhizobial strain produces slightly different NF structures (Broughton et al., 2000) recognized by plant root membrane-localized LysM receptor-like kinases (Haney et al., 2011). The receptor-like kinases are serine/threonine kinases with extracellular binding LysM domains (Bozsoki et al., 2020, Smit et al., 2007). LysM receptors of Nod Factors elicit plant signal transduction and nodulation response pathways that modify the root developmental program and initiate nodule organogenesis (Zipfel & Oldroyd, 2017, Smit et al., 2007).

Signaling downstream of Nod factor perception triggers a dramatic shift in root gene expression (Larrainzar et al., 2015) activating two parallel cellular reprograming responses, one in the root epidermis and a second, simultaneously, in the root cortex (Cook et al., 1995, Yang et al., 1994). Nod Factor signaling in the epidermis is characterized by nuclear-localized Ca^{2+} spiking, root hair curling, infection thread development, and bacterial infection (Oldroyd & Downie, 2008,

Oldroyd et al., 2011). Subsequent responses in the root cortex lead to cell divisions that form the nascent nodule primordium, giving rise to the differentiation of nodule tissues. Once infection threads reach developing nodule cells in the root cortex, the infecting bacteria are released into the cytoplasm as plant membrane-bound "bacteroids" that differentiate into facultative, nitrogen-fixing organelles termed "symbiosomes" (Oke & Long, 1999).

Nodule organogenesis is regulated locally via plant hormones, most notably ethylene, cytokinin and auxin (Penmetsa & Cook, 1997, Penmetsa et al., 2003, Oldroyd & Downie, 2008, Oldroyd et al., 2011). Ethylene is a negative regulator or Nod factor signaling and effectively fine-tunes Nod factor perception in a concentration-dependent manner (Oldroyd et al., 2001). Beyond hormones and Nod factor signaling, numerous other genetic factors can modulate nodulation, for example EPR3 is a LysM receptor-like kinase that distinguishes the molecular structure of bacterial exopolysaccharides (EPS) and can either promote or impede nodule formation (Haney et al., 2011, Kawaharada et al., 2015). Similarly, soybean Rj2/Rfg1 are TIR-NBS-LRR R gene alleles that restricts nodulation by certain strains of rhizobia through recognition of bacterial Type III (T3SS) effectors (Yang et al., 2010).

The impairment of nodule organogenesis (phenotype Nod-) was also a tool for screening mutant collections, revealing the genetic components of the common symbiosis pathway in nodulation and mycorrhization (Ane et al., 2002). The transcription factor *NIN* is among the earliest-acting symbiosis-specific transcriptional regulators, required for essentially all subsequent steps of nodulation (Schauser et al., 1999, Griesmann et al., 2018). Numerous genes act downstream of nod factor recognition and *NIN*, including components of the so-called "common sym pathway" that is also required for mycorrhizal symbiosis. In *Medicago truncatula* these genes are designated as DMI1, DMI2, and DMI3 (does not make infections) and have roles in Ca^{2+} spiking and perception (Middleton et al., 2007). The common sym pathway activates transcription factors, including NSP1 and NSP2 which are GRAS family transcriptional regulators,

leading to further gene expression changes for Early Nodulation Genes (ENOD) (Larrainzar et al., 2015).

Organogenesis. Legumes produce either determinate or indeterminate nodules. The differences in the type of nodule are reflected in the tissue origin of the nodule primordium, the nature of bacteroid differentiation, and the persistence of a nodule meristem. Determinate nodules originate from outer cortical tissue, have non-persistent meristems, produce round-shaped nodules, and contain differentiated bacteroids -the nitrogen-fixing form- that have the same morphology as free-living rhizobia (Maróti & Kondorosi, 2014). Examples of determinate nodules are those produced by *L. japonicus*, soybean, Phaseolus, and other tropical legumes. Indeterminate nodules on the contrary, have distinguishable histological organization, forming cylindrical nodules with a gradient of developmental zones, an apical meristem (Zone I), an infection zone (Zone II), then a fixation zone (Zone III), and a senescence zone (Zone VI) (Van de Velde et al., 2006). Indeterminate nodules originate from inner cortical cells and their bacteroids undergo enlargement, endoreduplication, and terminal differentiation (Alunni & Gourion, 2016, Mergaert et al., 2006). The degree of differentiation and bacteroid development is correlated with the nodule function in BNF (Vasse et al., 1990). Indeterminate nodules are characteristic of *M. truncatula* and other inverted repeat lacking clade (IRLC) legumes such as chickpea (Montiel et al., 2016).

Incompatible interactions (Nod-) are a significant factor in legume nodulation. Incompatibility is often manifest in a strain-specific manner, with plant allelic or copy number variation conferring preference for strains of rhizobia. Contrasting phenotypic responses to rhizobia have been documented in multiple plant species from the genera *Glycine*, *Trifolium*, *Pisum*, *Amphicarpaea*, and *Medicago* (Parker, 1999, Béna et al., 2005).

One type of genetic control of incompatibility occurs via recessive alleles and resembles gene-for-gene genetics, including as *Sym2* in pea (Geurts et al., 1997), *Sym37* in pea (Zhukov et

al., 2008), and the allele *Rj1* in soybean (Lee et al., 2011). These genes are lysine motif receptor-like kinases (LysM) and thus may function as receptors for Nod factors, other bacterial ligands that are perceived by LysM-like kinases (e.g., bacterial EPS, Kawaharada et al., 2015).

Dominant host alleles can also cause a lack of functional nodules, often in a strain-specific manner. Several genes/alleles act to restrict nodulation in soybean (*Glycine max*), including Rj4 restricts nodulation with B. *japonicum* USDA62; Rj2 restricts nodulation by *Bradyrhizobium japonicum* USDA122; and the Rfg1 restricts nodulation with *Sinorhizobium fredii* USDA257. Interestingly, Rj2 and Rfg1 are alleles of the same TIR-NBS-LRR R gene (Yang et al., 2010) that recognizes a specific T3SS effector from the rhizobia (Tsukui et al., 2013). Incompatibility occurs after induction of root hair curling, with limited infection-thread growth and rare advancement to formation of nodule primordia (Yang et al., 2010). Another naturally occurring dominant gene, *Mt*-NS1, restricts nodulation of *Medicago* specifically by *S. meliloti* Rm41, the nodule primordium is formed but not colonized by rhizobia (Liu et al., 2014).

Efficiency of symbiosis. The initiation of nodule organogenesis and rhizobial infection does not guarantee proper nodule functioning. Inefficient symbiotic phenotypes (Nod+/Fix-) segregate in natural populations of *M. truncatula*. In combination with specific rhizobial strains (e.g. *S. meliloti* A145) these genetic variants produce poorly developed nodules that are of small size, white in appearance and are not fully functional for N fixation. One such host factor, Mt-Sym6, was originally described as a single gene (Tirichine et al., 2000), but is in fact two linked host genes encoding nodule cysteine rich peptides (Wang et al., 2018).

Multiple *M. truncatula* mutants (*dnf1-1*, *dnf1-2*, *dnf2*, *dnf3*, *dnf5*, *dnf6*) are unable to support nitrogen fixation with *S. meliloti*. Dnf mutants have altered nodule anatomies and incomplete bacteroid development (Starker et al., 2006). DNF1 (DEFECTIVE IN NITROGEN FIXATION 1) is part of the Signal Peptidase Complex (SPC) and its expression is induced upon nodulation. DNF1 a critical for symbiosome development and bacteroid differentiation (Wang et al., 2010), and is

essential for late nodule associated gene expression (Feng et al., 2021). DNF2 is required for bacteroid differentiation, symbiosome persistence, and early nodule senesce (Bourcy et al., 2013). Two Fix- mutants *dnf4* and *dnf7* were found to be genes encoding NCR Peptides (Pan & Wang, 2017).

The family of Nodule cysteine rich (NCR) peptides - analogous to defensins in plants and animals - are expressed characteristically in the nodule and are critical in bacteroid differentiation (Van de Velde et al., 2010). In the mutant *dnf1*, defective in the nodule-specific subunit of the SPC, the signal peptide of NCRs is not cleaved and bacteria accumulate in the endoplasmic reticulum where they remain undifferentiated (Van de Velde et al., 2010). The *M. truncatula* locus *nsf1* encodes an NCR peptide and controls host-strain compatibility at the level of nitrogen fixation, negatively impacting symbiotic performance (Yang et al., 2017).

Fix- phenotypes (Fix-) have been extensively studied in mutant screens, but also occur as natural variants. For example, wild type *Lotus japonicus* efficiently forms nodules with *Rhizobium etli*, but the resulting nodules exhibit limited nitrogen fixation and internal nodule structures show hallmarks of senescence, such as cell membrane disintegration. Nodules formed by *R. etli* also have higher starch content and sparse bacteroid colonization, both of which are hallmarks of ineffective symbiosis (Banba et al., 2001). Natural variation for the efficiency of symbiosis has also been observed in *Medicago truncatula* vs *Sinorhizobium* interactions. For example, although *S. meliloti* has been the primary experimental symbiont with *M. truncatula*, in nature a more efficient symbiotic partner is *S. medicae*. Inoculation with *S. medicae* WMS419 yields higher photosynthetic rates, nitrogenase activity, nodule biomass and carbohydrate metabolism (Larrainzar et al., 2014).

Nodule number control

20

Nodule number production per plant is tightly regulated by the plant host with multiple mechanisms operating in the root system. Super nodulator mutants in several legume species have been key biological tools to discover local and systemic pathways that inhibit or promote nodulation. One level of local regulation occurs with the hormone ethylene that negatively regulates the number and distribution of successful infection events, as studied in the ethylene insensitive mutant Mt*skl* (Penmetsa & Cook, 1997).

A distinct genetic pathway is <u>A</u>utoregulation <u>of</u> <u>N</u>odulation (AON), a systemic feedback repression pathway that controls nodule number. Super nodulator mutants such as *sunn- M. truncatula, har1- L. japonicus, sym29-Pisum sativum*, and *nark-* soybean, are defective in autoregulation of nodulation (Penmetsa et al., 2003, Schnabel et al., 2005, Reid et al., 2011). These genes are orthologs of the CLV1 receptor kinase (CLAVATA1 like leucine-rich repeat) (Schnabel et al., 2005), which perceives specific peptides, known as CLEs (Clavata4/Embryo Signaling Region-Related), and regulates apical meristem differentiation in *Arabidopsis*. CLE peptides also function as root-derived signals in nodulation, including *Medicago* MtCLE12 and MtCLE3. CLE posttranslational modification is necessary for signaling, for example MtCLE12 is arabinosylated by RDN1 (ROOT-DETERMINED NODULATION 1) to enable perception by the CLV1 ortholog MtSUNN (Imin et al., 2018, Schnabel et al., 2011). In the shoot, CLV1-like kinases require other proteins such as of CLV2, KLV and CRN (Reid et al., 2011). The corresponding protein complex elicits a shoot-to-root signal that restricts further nodule development, requiring the function in the root of the F-box protein TML (TOO MUCH LOVE); however the mechanism of negative regulation is not fully elucidated. At least one mobile microRNA -i.e. miR2111- targets and post-transcriptionally regulates TML (Tsikou et al., 2018). Mutation of the AON pathway yields a hyper nodulation phenotype (Tsikou et al., 2018), demonstrating its role as a negative regulator of nodule meristem formation.

A different group of CLE peptides (MtCLE34 and MtCLE35) represents signaling components of the nitrate-dependent regulation of nodulation pathway, responsive to nitrogen status in the root (Djordjevic et al., 2015, Mens et al., 2021). While MtCLE35 and MtCLE34 are induced by both

rhizobia and nitrogen, only MtCLE35 is required for the suppression of nodulation (Lebedeva et al., 2020, Mens et al., 2021). The AON pathway is complex since there are multiple regulatory signals depending on the plant status, host species, the type of CLE peptide activated, and time of action, suggesting multiple systemic regulatory circuits that act at specific times (Kassaw et al., 2015)

A different systemic pathway that promotes infection and nodule number in *M. truncatula* is mediated by C-terminally encoded peptides (CEPs). Under nitrogen deficient conditions, these CEPs peptides promote nodulation via the shoot receptor LRR-RLK CRA gene (COMPACT ROOT ARCHITECTURE2) (Laffont et al., 2019). MtCEP1 enhances nodulation and inhibits lateral root formation (Mohd-Radzman et al., 2016). MtCEP7 is expressed 1 day following rhizobial infection, with expression mediated by cytokinin and by Nod-Factor, and requiring *NIN* transcription factor (Laffont et al., 2020).

In summary, nodule number is a complex trait that integrates multiple signal peptides derived from the root, hormones and dynamic temporal regulatory signals fine-tuning nodulation along plant development.

<u>Ecological complexity of symbiosis.</u>

There are more than 17,000 species of legumes that have the potential to nodulate in natural environments. Current evidence points to a single evolutionary origin for nodulation traits in angiosperms, with multiple losses in particular taxa (Griesmann et al., 2018). Despite the taxonomic breadth of the nodulation trait, ecological studies using naturally occurring symbionts are limited to only a few legumes, including (but not only) the model legumes *Medicago truncatula* (Béna et al., 2005) and *Lotus japonicus* (Bamba, 2019). Even fewer studies provide substantive insights about variation for symbiosis in an agricultural context (Parker, 1999).

Variation in symbiosis specificity and efficiency has been evaluated at a single genus level, testing 35 species of *Medicago* in association with strains of *Sinorhizobium medicae* and *S. meliloti*. Béna et al (2005) report that specificity is a dynamic trait and that the geographic distribution of rhizobia impacts the distribution and establishment of host species. Understanding the relevance of natural variation often involves methods distinct from those used to identify single genes. Ecologists often test genotypes in "home vs. away" environments, "common garden" experiments in which diverse genotypes are tested in the same environment, and "local vs. foreign" and "sympatric vs allopatric" combinations, giving power to detect local adaptation for microbes (Kraemer & Boynton, 2017).

Ecologists develop theory to describe the directionality of the mutualism, and the type of selection pressures operating in the mutualistic interaction under efficient or inefficient symbiosis. Under nitrogen-limiting conditions plants that successfully establish an efficient symbiosis are expected to have growth and reproductive advantages – fitness – suggesting that given a diversity of partners varying in degrees of 'efficiency', host plants will select efficient symbionts, as demonstrated by Batstone et al (2020). As a facultative symbiosis, legume-rhizobia associations are not maintained vertically from host to progeny and the interaction needs to be re-established in each generation. Several researchers have focused on determining host-mediated selection mechanisms for effective nitrogen fixing symbionts, using terms such as *host sanctions* or *partner choice* as reviewed by Frederickson (2013). Because symbiosis is the culmination of many genetic and physiological processes, the efficiency of BNF is often inferred from proxies, such as variation in the number of nodules, the status of nodule development (biomass size), or bacterial density recovered form nodules in co-inoculations (Westhoek et al., 2017, Mendoza-Suárez et al., 2021, Kiers et al., 2007). However interpretation of these proxies is not straight forward, because outcomes can vary in their meaning depending on the host species tested, the bacterial strain, and the metrics used for quantifying fitness in the symbiont (Sachs et al., 2010, Burghardt et al., 2018).

Biological Nitrogen Fixation (BNF) in agricultural systems is estimated to account for 50-70 Tg of fixed nitrogen (Herridge et al., 2008). Paradoxically, most large-scale legume cultivation uses synthetic fertilizers at planting to promote seedling establishment, while use of inoculum technology is more sporadic. For example, traditional farming of legumes relies on indigenous or naturalized soil bacteria, rather than inoculant applications. Commercial inoculants must perform effectively in local environments, for example overcoming competition by native microbes in farmers' soils, and having compatibility for nodulation with local host genotypes. The host then needs to metabolize fixed nitrogen from the nodule and translocate it to the grain efficiently. The multiple layers of complexity in biological nitrogen fixation described above are reflected in a high level of variability for nitrogen fixation in most agricultural legumes. Limitations in developing compatible and efficient inoculants for legumes, despite extensive knowledge from molecular mechanisms of nodulation in model legumes, expose the gaps in applications to develop inoculants that perform consistently across a range of agronomic situations.

In agricultural legumes, symbiotic conditions for both partners may be different in natural systems. Global agricultural systems are typically more homogeneous than natural systems (McDonald & Stukenbrock, 2016). Symbiosis in the natural environment spans a broader range of soil ecology, host and symbiont genetics, and the specific environmental conditions in which the legume-rhizobia co-evolved. Cultivated species are often grown in monoculture and their gene pools tend to have narrower diversity as a product of domestication, selection and breeding (Warschefsky et al., 2014). The potential impacts of genetic shifts among domesticated and cultivated genotypes in legume species may include changes in symbiotic compatibility, influenced, for example, by large geographic shifts in the distribution of both host (Sokolkova et al., 2020) and bacterial genotypes (Greenlon et al., 2018) subsequent to domestication, geographic differences in root-associated microbial communities (Alford, 2020), or a difference in host mechanisms for selection and maintenance of symbiosis (Liu et al., 2020).

Porter and Sachs (2020) conjecture that differences among wild vs. cultivated legumes derive from evolutionary tradeoffs in which artificial selection for agricultural traits increased the frequency of alleles that disrupt symbiosis (Porter & Sachs, 2020). However, the opposite has been observed for soybean genes/alleles that confer partner choice. For example, soybean Rj4 confers strain-specific restrictive nodulation. Although Rj4 segregates in both wild and cultivated populations, (bio)geographic patterns indicate that breeding has increased allele frequencies of Rj4 under environments where enhanced selection would be beneficial for symbiosis (Liu et al., 2020). By contrast, the distribution of alleles Rfg/rfg1 suggest a lack of selection (Liu et al., 2020).

A potential cost of domestication, given population bottlenecks and inbreeding, is that the potential for selection response among symbiosis traits has been diminished due to reduced germplasm diversity. Kiers et al (2007) describe a reduction in the symbiotic effectiveness of elite soybean when inoculants contain mixtures of effective and ineffective symbionts, but that historic cultivars are unaffected by the same strain mixture. An additional scenario is relaxation of selection, where agricultural inputs and cultivation practices reduce the need for efficient symbiotic traits, and therefore alleles that are less favorable to symbiosis can accumulate (Porter & Sachs, 2020).

The work reported here examined the relevance of both plant and bacterial genetic variation to the efficiency of symbiotic nitrogen fixation in the *Cicer – Mesorhizobium* interaction. *Cicer arietinum* – cultivated chickpea – was derived from wild *C. reticulatum* ~10 KYA in the ancient Mesopotamian region of modern Southeastern Turkey. *C. echinospermum* (Ce) is the 100 KYA sister to *C. reticulatum* (Cr)(von Wettberg et al., 2018). The two species are broadly co-incident (Greenlon et al., 2019), but inhabit distinct soil types with distinct microbial communities (Alford, 2020). The wild nitrogen-fixing microbial symbionts of *Cr* and *Ce*, exclusively nodulates with species of *Mesorhizobium*, are also genetically distinct and have geographic distributions that parallel host genetics. The data presented here support two broad conclusions. (1) Plant and

bacterial biogeographic patterns reflect local adaptation for efficient symbiosis in wild systems. (2) Post-domestication genetic processes have reduced the effectiveness of symbiosis in certain crop genotypes.

RESULTS

I. Establishing a quantitative assay for plant responsiveness to nitrogen and nitrogen fixation.

1.1 Plant responses to nitrogen during late vegetative stage plants.

Given that domestication and breeding practices caused a narrowing of genetic diversity in the primary gene pool of modern *Cicer arietinum* relative to wild *C. reticulatum* (von Wettberg et al., 2018), we tested whether domestication also involved functional changes in nitrogen responses, especially responses to symbiotic microbes and inorganic nitrogen. Initially we investigated the response to inorganic nitrogen in the form of NH_4NO_3. Understanding plant responses to exogenous nitrogen also serves as a baseline for subsequent characterization of the response to nitrogen fixing symbionts.

Two cultivated genotypes, including an early-flowering breeding line from India (ICCV96029) and a commercial cultivar from Canada (CDC Consul), and four wild accessions representing distinct genetic populations of *Cicer reticulatum*, were assayed. Nitrogen was added weekly as fertilizer using successive 10-fold increments from 0.001 mM to 100 mM NH_4NO_3. Plant were harvested at 5 weeks post inoculation, corresponding to late vegetative stage growth. Despite the genetic diversity of tested accessions, their response to exogenous nitrogen was similar (Figure 1, Supplemental Figure 1). The highest level of NH_4NO_3 (100 mM) was inhibitory to growth, with plants displaying symptoms of nitrogen toxicity. Hence 100 mM concentration data were excluded from statistical analyses. Stimulation of plant growth was only observed at 10 mM NH_4NO_3 (Supplemental Table 1) and the response was restricted to shoot tissues; thus in all genotypes, root biomass was unresponsive to inorganic nitrogen treatment.

Despite similar patterns in the response to mineral nitrogen, plant genotypes might differ in the strength of their response. To examine this possibility, we expressed growth response as , which corrects for inherent differences in plant size among genotypes (Figure 1). As shown in Table 1 and Supplemental Table1 individual genotypes exhibit differences in their growth response to mineral nitrogen, with the two cultivated accessions being more responsive than the four wild genotypes.

Imaging of roots with the WinRhizo system provided an opportunity to evaluate plant root architecture responses at 5 weeks. At this time point, only cultivar 96029 had transitioned to reproductive phase (flowering), while the remaining 5 genotypes were in the vegetative stage. Similar to the absence of stimulation in root biomass, there were no significant differences among treatments in root length or root diameter (Supplemental Table S1). In summary, there was a main effect of nitrogen treatment across all plant genotypes, with inhibition at 100 mM and subtle differences among genotypes in the degree of stimulation at 10 mM.

1.2 The effect of bacterial species in late vegetative stage plants.

As a prelude to extensive phenotyping, bacterial x plant genotype interactions were tested using a subset of four bacterial species and six plant genotypes. Plant genotypes consisted of two cultivated accessions from Canada (Consul) and India (ICCV96029), respectively, and four *C. reticulatum* accessions obtained from four genetically differentiated wild populations. Bacterial genotypes consisted of two wild strains from Turkey, Kar-203 (*M. ciceri*) and B2O3/Rse19 (*M. mediterraneum*), and two novel agronomic symbiont species from Ethiopia, 43P5 and 43P2 (related to *M. plurifarium*).

Seeds were inoculated at sowing and harvested during vegetative growth after 5 weeks for determination of root and shoot traits. Although we observed differences that could be attributed

to host and bacterial genotype (Figure 1 and Supplemental Table 1), effect sizes were small. Indeed the majority of microbial strains failed to stimulate plant biomass compared to low nitrogen controls (0, 0.001, 0.01, 0.1 to 1 mM NH_4NO_3).

1.3 Examining the effect of plant phenology.

Although plants require reduced nitrogen throughout all phases of plant development, nitrogen requirements are highest during reproduction (Schiltz et al., 2005). Moreover, within and among plant species their rates of development ("phenology") can vary widely, and indeed many traits in legumes exhibit phenological dependence (Schiltz et al., 2005). To understand the effect of plant phenology on nitrogen responsiveness, we assayed the symbiotic performance of cultivated *C. arietinum* (ICCV96029) and wild *C. reticulatum* (Oyali 84) in a time course assay, spanning vegetative and reproductive development.

Plants were inoculated with *M. ciceri* (Kar203) and *M. mediterraneum* (B203) or treated with a series of mineral nitrogen (0.1 to 10 mM NH_4NO_3). Nitrogen responses (mineral or symbiotic) were assessed by comparing biomass gain among treatments within sampling dates, using low mineral nitrogen treatments as a comparator. The two plant genotypes differ in their phenology, with ICCV96029 being an early flowering cultivated accession. For purposes of comparing between plant genotypes we monitored plant phenology and assayed plant biomass during (1) vegetative growth, (2) at the onset of flowering, and (3) at mid-pod set when 50% of reproductive nodes had mature-sized pods (Fig. 2).

For both cultivated *C. arietinum* ICCV96029 and wild *C. reticulatum* Oyali 84, the greatest response to either symbiotic rhizobia or mineral nitrogen was evident in plants with maturing pods (Fig. 2), which is consistent with the high demand for reduced nitrogen during pod filling. This observation likely explains the small effect of *Mesorhizobium* inoculants observed in section 1.2, above, and established the developmental window for future experiments. Although both

plant genotypes formed functional nodules with *M. ciceri* (Kar203) and *M. mediterraneum* (B203), there were differences in symbiotic response that could be attributed to a combination of host and bacterial genotype effects. Thus, both wild and cultivated accessions exhibited greater biomass gain on inoculation with *M. mediterraneum*, the cognate symbiont of *C. reticulatum*, compared to *M. ciceri* (Fig. 2A). At mid-pod set we quantify 45% and 35% more biomass with Mm-B203 compared with Mc-Kar203, in wild and cultivated genotypes respectively (Fig. 2A).

Differences in shoot biomass gain (a measure of symbiotic effectiveness) were not correlated with differences in nodule biomass, but rather with differences in the number and size of nodules. Thus, plants inoculated with the more effective *M. mediterraneum* (B203) had significantly larger and less numerous nodules compared to the less effective *M. ciceri* (Kar-203) (Fig. 2C-E) (t-test p-values = 0.0052 and 0.00011, respectively). Expressing shoot biomass as a function of nodule number yields a measure of nitrogen fixation efficiency. By this measure, *M. mediterraneum* was significantly more efficient than *M. ciceri* on both ICC96029 (93.43 mg shoot/nodule versus 23.32 mg shoot/nodule) and Oyali 84 (91.03 mg shoot/nodule versus 30.16 mg shoot/nodule).

The inverse relationship of nodule size and nodule number was especially pronounced for cultivated ICCV96029 (compare Figure 2C and 2E). *M. ciceri* produced on average 50% more nodules at all developmental stages in the cultivated genotype, much greater than in wild Oyali 84, suggesting contributions from both bacterial and host genetics. Despite microbial and host genotype effects, total nodule biomass was similar between microbial species at all developmental stages (Figure 2D).

In addition to nodulation phenotypes, we tested multiple levels of inorganic nitrogen across plant development. Consistent with previous results, supplementation with 10 mM ammonium nitrate resulted in a significant increase in shoot biomass in both wild and cultivated genotypes at all developmental stages, with the effect being most pronounced during pod filling (compare Figure 1 with Figure 2A). By contrast, neither shoot nor root biomass were measurably impacted by low

29

nitrogen treatment; indeed, the relative distribution of biomass to roots was constant at 0, 0.1- and 1-mM ammonium nitrate. Treatment with 10 mM ammonium nitrate, however, caused a dramatic shift in biomass distribution (Figure 2B), with increased biomass directed primarily to shoots. Interestingly, inoculation with either strain of *Mesorhizobium* was associated with reduced relative root biomass even during vegetative plant growth, when treatment effects on shoot biomass were not yet evident.

II.   Understanding nitrogen fixation specificity in wild and cultivated gene pools of *Cicer*.

1. Wild relatives *C. echinospermun* and *C. reticulatum* prefer their co-evolved microbial symbionts.

Results from initial greenhouse experiments (above) demonstrate that symbiotic performance depends on both plant and bacterial genotype. In particular, the tested accessions of *C. reticulatum* and *C. arietinum* favored *M. mediterraneum*, the co-evolved symbiont of the crop wild progenitor *C. reticulatum* (Fig. 1 and Fig. 2). However, due to the limited sampling within plant and bacterial species it was not possible to know at what taxonomic level symbiosis traits segregate.

To determine if individual plant species display specificity for individual bacterial species, we focused on homologous (co-evolved) and heterologous (non-co-evolved) interactions. Symbiotic performance was determined in cross inoculation experiments involving multiple accessions of three *Cicer* and two *Mesorhizobium* species. Among twelve plant genotypes were eight accessions of the genetically diverse *C. reticulatum*, which is the immediate wild progenitor of chickpea; two accessions of the genetically narrow 100 Kya sister species *C. echinospermum*; and two genotypes of the genetically narrow cultivated species *C. arietinum*. Among six wild *Mesorhizobium* strains were three representatives of each of two bacterial species, *M. mediterraneum* (Mm1-Rse11, Mm2-B203, Mm3-Rse20) and *M. ciceri* (Mc1-Gun202, Mc2-Rse17, Mc3-Kar203), which are the native symbionts of *C. reticulatum* and *C. echinospermum*, respectively. As in the case of wild plant

accessions, the wild bacterial strains were selected to represent both geographically and genetically distinct members of each species (von Wettberg et al., 2018; Greenlon et al., 2019).

In total 72 *Cicer* x *Mesorhizobium* genotype combinations were analyzed (Table 2). Plants were harvested at mid-podset and divided into roots, shoots and nodules. Biomass of root and shoot tissues was determined after drying at 45°C for at least two weeks. The number and size distributions of nodules were quantified using a Videometer 3 imaging system after which nodules were dried and biomass was determined (Fig. 6B).

Comparisons at the highest taxonomic levels reveal significant plant species x bacterial species interactions, with wild species exhibiting greater performance in homologous (co-evolved) compared to heterologous (non-co-evolved) combinations. Substantial treatment effects were evident even based on simple visual inspection (Figure 3). The most apparent differences were increased plant size in high inorganic nitrogen treatments and in wild plant species inoculated with their native bacterial symbionts, i.e., *C. reticulatum* inoculated with *M. mediterraneum* and *C. echinospermum* inoculated with *M. ciceri*. Conversely, treatment with low inorganic nitrogen or with non-native bacterial species produced smaller plants, often with symptoms of nitrogen-deficiency chlorosis.

Symbiotic benefit was quantified as shoot biomass gain, which was the most sensitive measure. Within each genotype shoot biomass was normalized to the 10 mM NH_4NO_3 control, which we refer to as the Nitrogen Fixation Index (NFI). NFI permits direct comparison among genotypes and treatments (Figure 4). Combining data among genotypes from the same plant species revealed that *C. echinospermum* had significantly greater shoot biomass gain when inoculated with homologous *M. ciceri* (NFI = 0.44 +/- 0.05 SE) compared to *M. mediterraneum* (NFI = 0.16 +/- 0.04 SE). Conversely, *C. reticulatum* had greater shoot biomass in response to its homologous symbiont *M. mediterraneum* (NFI = 0.32 +/- 0.02 SE) compared to *M. ciceri* (NFI = 0.19 +/- 0.02 SE). Combined data for two cultivated accessions of *C. arietinum* did not reveal significant

differences in NFI between bacterial species (*M. ciceri* NFI = 0.32 +/- 0.04 SE vs *M. mediterraneum* NFI = 0.36 +/- 0.04 SE) (Fig. 4).

Analyzing NFI of individual genotypes revealed similar patterns of specificity, but also suggested genotype-specific differences within both plant (Fig. 5) and bacterial species (Table S2A). In the case of *C. echinospermun*, Gunas_62 exhibited the greatest differential specificity between bacterial species (t-test, P ≤ 0.0001). Mean NFI values for *C. echinospermum* SDR65 were also substantially different between bacterial species, but the difference was not significant by t-test; instead NFI tested differently when pairwise contrasts were calculated using the estimated marginal means and Tukey HSD methods (Table S2B). Similarly, all eight *C. reticulatum* genotypes, representing the diversity of the crop wild progenitor, exhibited greater NFI in association with homologous *M. mediterraneum* compared to heterologous *M. ciceri* (Fig. 4A), with genotype Egill73 showing the largest differential benefit from *M. mediterraneum* compared to *M. ciceri*. In contrast to wild species, NFI of crop genotypes was not influenced by bacterial species.

The observed changes in rank order of performance between homologous and heterologous symbiont pairs strongly supports an ecological hypothesis of local adaptation, where wild plant and microbial species are co-evolved for efficient nitrogen fixation and non-co-evolved partners have low performance.

2.2 Nodule trait variation in wild relatives.

Multiple nodulation traits were quantified in the wild genotypes, including the number of nodules, nodule biomass, nodule area (Fig. 6, Fig. 7) and morphometric parameters (Table S3). Although both wild plant species nodulated efficiently in response to their homologous bacteria, they had contrasting outcomes for nodule development with heterologous microbes. *C. echinospermum* was largely refractory to nodulation by heterologous *M. mediterraneum*, with many roots entirely lacking nodules. This is typified by genotype Gunas_62 with significantly

more (11.3 vs 3.2, t-test p value = 0.00035) and larger (51.5 mg +/- 5.9 vs 24.5 mg +/- 6.0) nodules on treatment with homologous *M. ciceri* compared to heterologous *M. mediterraneum*. Digital area projection of individual nodules on *C. echinospermum* support this differing specificity (Fig 7C). Indeed *C. echinospermun* genotypes differed in multiple nodule-associated parameters in a bacterial species-dependent manner (Supplemental Table S3).

To account for the potential effect of root size on nodulation traits, nodule number values were normalized to root biomass, permitting direct comparison among genotypes and treatments (Fig. 5B). *C. echinospermun* exhibited the greatest difference in normalized nodule number, with greater numbers of nodules on inoculation with *M. ciceri* compared to *M. mediterraneum*. In *C. arietinum* the normalized number of nodules was comparable for both sources of inoculum. Conversely, *C. reticulatum* genotypes either had a significant negative relationship between nodule number and NFI (t-test p-value p ≤ 0.001; Bari1_92, Egill_73, and Oyali_84) or nodule number was similar among bacterial species irrespective of differences in NFI (Figure 5B; Supplemental Table S3).

Analogous to water use efficiency and nitrogen use efficiency, we expressed nitrogen fixation efficiency as shoot biomass gain per mg of nodule tissue. In the case of *C. reticulatum* there was a consistent trend (Fig. 5C), with homologous, co-evolved interactions having higher nitrogen fixation efficiency than heterologous non-co-evolved interactions. This relationship was significant at p-value ≤ 0.01 for six of the eight *C. reticulatum* genotypes. Similar comparisons were not appropriate for *C. echinospermum* due to poor or absent nodulation with *M. mediterraneum*.

Pairwise comparisons of nodule number produced in *C. arietinum* and *C. reticulatum* genotypes, generally were similar among *M. ciceri* or *M. mediterraneum* (Fig6B), suggesting that efficient biomass conversion was unrelated to absolute nodule number values. Moreover, in *C. reticulatum* Oyali84 produced many small nodules with the heterologous symbiont *M. ciceri*,

while Sirna60 produced many small nodules with strains from homologous *M. mediterraneum* (Fig6D).

Nevertheless, measures of nodulation and biomass gain clearly establish the functional superiority of homologous bacteria on *C. echinospermum* (Fig. 4, Fig. 5 A and C). Among *C. reticulatum* genotypes, in multiple experiments Egill 073 exhibited the greatest discrimination among bacterial species, and nodule number was correlated with shoot biomass in the case inoculation with *M. mediterraneum* (R^2 =0.64), but not with *M. ciceri* (R^2 = 0.09) (Supplemental Fig. 2A).

Incidental to imaging of nodules for morphometric features, multispectral reflectance data were collected for a subset of the twelve plant genotypes at 19 wavelengths (spanning 375 nm - 970 nm) and aspects of visual color (CIE L, A, B, saturation, hue) (Sup. Fig. 3E). Spectral differences reflect biochemical properties, which if understood can be used predictively to infer physiological status. This logic is the basis of remote sensing in agriculture (Khanal et al., 2020). Due to the large number of nodules under analysis it was necessary to first preserve tissue by ethanol dehydration and storage at 4°C and then to rehydrate with water prior to imaging. Data were visualized in a principal component analysis (PCA) plot, with the first two dimensions capturing 53% of variation (Fig. 7C). Discrimination among individual nodules was driven by multiple distinct wavelengths as well as by components of visual color (Fig. S3D). Interestingly, spectral properties resolved nodules by plant genotype, with nodule reflectance sufficient to resolve plant species: *C. reticulatum* > *C. echinospermum* > *C. arietinum* (Fig 7C,D). Even within plant species, individual genotypes could be resolved based on reflectance (Fig. S3B) and across the continuum of genotype rankings, *C. reticulatum* Sirna_60 had top rank and *C. arietinum* 96029 had the lowest rank. Despite differences among plant genotypes, we did not observe a main effect of bacterial species (Fig. S3A), although there was an effect of host genotype x bacterial genotype interaction. Within each plant genotype, nodules formed with *M. mediterraneum* had higher reflectance values than nodules formed with *M. ciceri* (Fig 7D).

Although we lack sufficient information to understand the nature of differences in spectral properties, the strong effect of plant genotype and of plant genotype x bacterial genotype interactions suggests an opportunity for future investigation.

2.3 Root trait variation in wild relatives

At mid-pod set, the root systems of greenhouse grown plants were large and constrained by pot size, and thus systematic analysis of root system architecture was not feasible. Nevertheless, gross differences in root system development were obvious when plants were harvested. In general, homologous interactions had larger, more branched root systems than heterologous interactions. This situation is shown in Figure 3C for the examples of *C. reticulatum* Sirna_60 and *C. echinospermum* Gunas_62, where images are representative of outcomes for these particular plant x bacterial genotype combinations. Interestingly, the visual differences in root architecture shown in Figure 3 were not reflected in average root biomass. Indeed, across all plant x bacterial genotype combinations, total root biomass was not affected by bacterial genotype despite strong effects on shoot biomass (Table. S2A, Table. S2E).

2. Cultivated *C. arietinum* variation in nitrogen fixation efficiency.

In multiple experiments, two cultivated genotypes (CDC Consul and 96029) exhibited limited or no preference for *M. ciceri* versus *M. mediterraneum*. To discern whether this situation is peculiar to these particular genotypes or a more general attribute of the cultivated species, the responses of nine *C. arietinum* cultivars were quantified. The nine cultivars are the product of distinct breeding programs in India, Ethiopia, Canada and the United States. For comparison, two *C. echinospermun* genotypes and three *C. reticulatum* genotypes, were included in the analysis. The

inoculants tested were the six native strains previously used: *M. ciceri* (Mc1-Gun202, Mc2-Rse17, Mc3-Kar203) and *M. mediterraneum* (Mm1-Rse11, Mm2-B203, Mm3-Rse20) (Table 3).

Consistent with previous experiments, wild *Cicer* species gained significantly greater benefit from homologous compared to heterologous *Mesorhizobium* (with the exception of Ce S2Drd_65) (Fig. 8). As a group the domesticated *C. arietinum* accessions exhibited a preference for *M. mediterraneum* (Fig. 8A), however when genotypes were considered individually cultivated accessions divided into two categories (Figure 8B). As observed previously CDC_Consul and 96029 failed to discriminate between bacterial species, as did CDC_Leader and CDC_Orion. Conversely, cultivars Billy Bean, Habru, Hattrick, Kyabra and Minjar exhibited a significant preference for *M. mediterraneum* over *M. ciceri*. Although specificity was statistically significant, the effect sizes based on Cohens'D estimates were classified "small" or "very small", compared to "medium" effect sizes observed in wild *C. reticulatum* Egill_73 and Sirna_60 (Sup. Table 4E).

III. Genetic analysis of nitrogen fixation specificity in recombinant inbred populations of *C. echinospermun* x *C. arietinum*.

To test the hypothesis that host specificity for *Mesorhizobium* species is controlled by simple genetics, we selected F4 interspecific recombinant inbred lines (RILs) derived from crosses between ICCV 96029 or CDC Consul and *C. echinospermum* Gunas_62 or Gunas_63. Genotypes from the Gunasan population were selected as the wild donor because Gunasan_62 has low rates of nodulation with *M. mediterraneum* compared to *M. ciceri* and thus we anticipated that segregating traits could be scored unambiguously. Use of multiple cultivated and Gunasan parents was necessary to maximize use of available germplasm. Importantly, genomic data demonstrates that the Gunasan population is genetically narrow and thus all Gunasan genotypes are highly similar (von Wettberg et al., 2018).

This initial experiment involved analysis of 71 RIL accessions with 10-fold replication. Tested accessions were F4 families, expected to segregate at ~25% of loci, which is the average heterozygosity of F3 parents. Each RIL received treatments consisting of two bacterial species (*M. mediterraneum* and *M. ciceri*). Plants were harvested at mid-pod set, when nitrogen demand is high. Subsequent to harvest, plants were separated into roots, shoots and nodules. Nodules were further analyzed by quantitative imaging to determine the status of nodule development and to generate nodule size distributions. All tissues were then dried and total biomass was determined.

Specificity was calculated using the estimated marginal means statistic, which is a least squares method. As shown in Figure 7, imposing a p-value threshold of ≤ 0.05, thirteen RILs exhibited a preference *M. mediterraneum* over *M. ciceri* (Table 4). No F4 families exhibited a significant preference for *M. ciceri* over *M. mediterraneum* and in only two cases was there intra-family variation for low nodulation with *M. mediterraneum* (Fig. 10A). Thus, contrary to expectations, the trait of the wild Gunasan parents (i.e., preference for *M. ciceri*) was not penetrant as a simple trait in F4 progeny. The wild parent trait (Nod- with Mm) was observed in individuals with poor nodulation and plant development (Fig. 10A), but not captured with statistical tests. Notwithstanding caveats about continued segregation in F4 families and the heterogeneous parentage of the population, the ratio of 13:58 (specificity:no-specificity) is not significantly different from the expected segregation of a single recessive gene determining the trait (i.e., 18:53, chi-square = 1.861, two-tailed p-value = 0.1726). In summary, the genetic control of symbiotic specificity and compatibility among *C. echinospermun* and *Mesorhizobium* is a complex trait.

DISCUSSION

Research in this chapter focused on understanding the evolution of symbiotic performance in the wild progenitors of the chickpea crop and the ways in which human selection may have reshaped symbiotic effectiveness during domestication and breeding. Initial

experiments quantified plant performance in co-evolved and non-co-evolved *Cicer-Mesorhizobium* pairs to test the ecological hypothesis of local adaptation. Subsequent experiments quantified performance and specificity in a panel of globally disperse cultivated accessions. Given observed differences in strain specificities among species, a panel of recombinant inbred lines were evaluated to assess genetic control of symbiont specificity between wild and cultivated species.

Wild *Cicer* species exhibit distinct geographic and soil type distributions, which is mirrored by the distribution of the cognate *Mesorhizobium* symbiont. In a common garden greenhouse experiment involving both homologous and heterologous symbiont species, we observed a strong preference of wild plant species for their sympatric microbial partners. This specificity segregated at the species level, being consistent among multiple plant and microbial accessions from each of the tested species. These data provide strong evidence of local adaptation.

Interestingly, wild species' specificity has two distinct patterns. *Cicer reticulatum* exhibits compatibility for nodule development, both in terms of nodule biomass and nodule number, with both homologous (*M. mediterraneum*) and heterologous (*M. ciceri*) symbiont species. However, shoot biomass gain is significantly greater in the interaction with the homologous *M. mediterraneum* compared to heterologous *M. ciceri*, demonstrating that symbiosis is more effective among co-evolved partners. Greater efficiency of nodules colonized by homologous bacteria took two forms, which suggests variation within *C. reticulatum*: cases in which homologous bacteria provided greater benefit with fewer nodules and cases where homologous bacteria provided greater benefit despite comparable rates of nodule development. Although mechanisms are uncertain, the simplest interpretation is that nitrogen acquisition is more efficient in developed nodules for native, compared to non-native, plant-bacteria combinations, and thus that nitrogen fixation per se is more effective. In the interaction *M. truncatula- S. medicae* WSM 419, nitrogen fixation efficiency was associated with higher carbon catabolic activity (Larrainzar et al., 2014). More generally, in both *M. truncatula* and *L. japonicus*, nodules

38

with inefficient Nod+/Fix- phenotypes have high starch content, which is also feature of the pre-fixation zone of mature nodules (Vasse et al., 1990). Another possible explanation for differences in symbiotic outcome, despite similar nodulation phenotypes, is the early onset of nodule senescence (Schumpp et al., 2008). Within the nodule, the status of bacteroid differentiation can impact the efficiency of nitrogen fixation (Vasse et al., 1990). More recent studies have determined that, in many cases, such variation in bacterial differentiation derives from the action of plant NCR peptides that can control development and maintenance of the microsymbiont (Alunni & Gourion, 2016, Yang et al., 2017).

In contrast, specificity in *C. echinospermum* accessions was characterized by low or absent nodulation with heterologous bacteria, a situation that was especially pronounced in the case of accessions from the Gunasan population. Reduced interaction is reminiscent of classical legume-rhizobia specificity, where early reciprocal signaling between host and microbe is either incompatible (Oldroyd et al., 2011, Wang et al., 2012) or nodule development is arrested at an early stage (Tsukui et al., 2013).

In cases of strain specificity among cultivated genotypes, two distinct scenarios were observed. Approximately half of tested plant accessions exhibited preference for *M. mediterraneum* over *M. ciceri*, similar to *C. reticulatum* from which the crop derived. Conversely, the other roughly half of tested accessions did not distinguish between *M. mediterraneum* and *M. ciceri*. Despite these categorical groupings, we did observe variation in the response within individual genotypes; for example, ICCV 96069 was the most tested genotype and it displayed absent specificity in 3 of 4 independent experiments. The most parsimonious explanation is that host preference for *M. mediterraneum* (relative *to M. ciceri*) is ancestral to the *C. arietinum* gene pool and inherited during domestication from *C. reticulatum*. Under this scenario, absent specificity among cultivated accessions arose post-domestication and continues to segregate in the cultivated germplasm. Interestingly, domesticated genotypes that lack strain specificity also have lower nitrogen fixation efficiency (Fig. 8B), suggesting that the two traits might have a common basis. Altered nitrogen fixation in cultivated accessions could derive from selection relaxation in agricultural soils with relatively high nitrogen content and/or from genetic

tradeoffs. Among cultivated accessions with absent specificity, CDC Consul, CDC Leader and CDC Orion all originate from the same Canadian breeding program, raising the possibility of a common genetic basis. However, ICCV 96029, which also has absent specificity, has origins in a distinct Indian breeding program, suggesting either an older common origin of the absent specificity trait or that the trait arose multiple times. In any case, phenotypic variation for nitrogen fixation within the cultivated germplasm suggests opportunities for chickpea improvement by breeding for nitrogen fixation traits. It is noteworthy that the majority of commercial inoculants involve *M. ciceri*, the less effective symbiont of cultivated *C. arietinum*. Thus, there are likely also opportunities for agronomic improvement by development of superior inocula that contain the more effective *M. mediterraneum*.

With the intent of understanding host genetic control of symbiotic specificity, recombinant inbred lines derived from *C. echinospermum* Gunasan X *C. arietinum* 96029 or CDC Consul were phenotyped for shoot biomass gain following inoculation with either *M. mediterraneum* or *M. ciceri*. In designing this genetic experiment, we hypothesized that the incompatibility of *C. echinospermum* for *M. mediterraneum* would segregate as a single major locus, as reported for other Nod- phenotypes in legumes (Geurts et al., 1997, Lee et al., 2011, Yang et al., 2010). Surprisingly, the phenotype of the wild parent (recalcitrance to nodulation with *M. mediterraneum*) was not penetrant in the assayed progeny. It is possible that specificity for bacterial species in *C. echinospermum* is polygenic and therefore not frequent in small segregating populations (e.g., two recessive genes would occur as co-homozygotes at a frequency of 1:64, etc.). Instead, preference for *M. mediterraneum* segregated at high frequency. This result was unexpected because preference for *M. mediterraneum* is not characteristic of either of the tested cultivated parents (96029 and Consul) nor of the wild Gunasan parents. Instead, preference for *M. mediterraneum* is fixed in the crop's wild progenitor (*C. reticulatum*) and segregates in approximately half of cultivated accessions. One scenario is that ICCV 96029 and CDC Consul contain a dominant allele that suppresses preference for *M. mediterraneum*. If wild *C. echinospermum* donates a recessive, non-suppressing version of the causal locus, then strain

preference within recombinant inbred progeny would be unmasked at a rate of 1:3, close to the observed frequency. Importantly, these data suggest that genetic analysis of *Cicer* specificity for *Mesorhizobium* is a tractable problem, both for trait characterization and for breeding. Towards this end, we have developed an Illumina iSelect genotype array containing 5K SNP that distinguish wild from cultivated accessions. This marker density is sufficient to tag all recombinant intervals and to complete an initial small-scale genetic analysis.

Although one can collect a variety of data about symbiosis, including the quantity and developmental features of nodule organs, the attribute most indicative of symbiotic performance was the biomass of shoots on pod-filling plants. A potentially confounding factor is that different plant genotypes have inherent differences in root and shoot biomass, independent of nitrogen fixation. However because root biomass was relatively insensitive to nitrogen status, shoot biomass could be normalized to root biomass, facilitating comparison among plant genotypes with inherent differences in plant size.

In addition to the main effect of nitrogen fixation on shoot biomass, we observed numerous other factors that varied depending on bacterial and host genotypes. For example, root architecture was often strikingly different within a single plant genotype depending on the bacterial species (Figure 3C). Architectural differences were generally not reflected in total root biomass, suggesting that changes in root density, especially changes in the relative proportion of fine versus thick roots, may underlie the different architectures. Ethylene is a plant hormone that can increase the production of thick versus fine roots and is also a negative regulator of nodulation (Penmetsa & Cook, 1997, Larrainzar et al., 2015). It is plausible that increased ethylene production during inefficient symbiosis shifts root development towards shorter, thicker roots. Indeed ethylene-related induction of the "thick short root" (*tsr*) phenotype is a hindrance to *in vitro* nodulation assays in *Medicago* (Barker et al., 2006), leading some researchers (e.g., Starker et al. (2006)) to conduct nodulation assays in the presence of ethylene biosynthesis inhibitors.

Root system architecture can also respond to nutritional status, including during symbiosis, as reviewed by (Concha & Doerner, 2020). In general, under conditions of low nitrogen status plants invest relatively more carbon into root systems, while under nitrogen sufficient conditions carbon is preferentially invested into shoot systems. However, the nature and extent of this phenomenon appears to vary according to host species, genotype and soil conditions. In the current study, plants inoculated with better-performing, homologous symbionts invested relatively less carbon into their root systems (Figure 5D). Although this trend was consistent among the wild plant genotypes, differences were significant only in three interactions (e.g., Gunas_62, CudiB_22C and Egil_73). This change in biomass allocation is entirely due to increased allocation to shoots, as root systems are largely unaffected by symbiont species (Table S6). The same situation was observed in high mineral nitrogen control treatments, which were significantly more stimulatory to shoot growth than was symbiosis, potentially reflecting an interaction between regulation of root growth and systems for uptake and sensing of soil nitrate and ammonium (Kiba & Krapp, 2016). Interestingly, inoculation with either *Mesorhizobium* species was associated with reduced relative root biomass even during vegetative plant growth when effects on shoot biomass were not evident (Figure 2B). Compatible rhizobia are known to impact the synthesis and transport of mobile CLE and/or CEP peptides (Lebedeva et al., 2020) even during early stages of symbiotic development, which provides a potential mechanism to modulate root: shoot ratios during early plant development in a manner independent of nitrogen. In any case, it is likely that these analyses are confounded by the physical limitations of pot size used in greenhouse experiments; thus, a more complete study of the relationship between nitrogen status and root architecture will require a revised experimental design.

The number of nodules formed varied considerably between plants even within a single genotype and treatment (Figure 6B). Nevertheless, in several cases there was clear statistical evidence of plant genotype x bacterial genotype interactions. As mentioned previously, *C. echinospermum* is largely refractory to nodulation by *M. mediterraneum*, and therefore the number of nodules was significantly greater in the homologous interaction with *M. ciceri* (Figure

6B) and nodule number and shoot biomass were positively correlated. The is different from the situation with *C. reticulatum*, where nodule number was either significantly increased in the heterologous (*M. ciceri*) interaction, or not different between homologous and heterologous *Cicer-Mesorhizobium* pairs. In other legumes (soybean, *Medicago* and *Lotus*), nodule number is under control of a systemic Clavata 1-like signaling system that is responsive to nitrogen status (Carroll et al., 1985, Searle et al., 2003), which could potentially explain the increased number of nodules in less efficient interactions, for example Bari1_92, Egil_73 and Oyali_84 on inoculation with the less efficient *M. ciceri* symbiont (Figure 4B).

The effect of plant genotype x bacterial genotypes combination on nodule number was especially pronounced in the experiment testing the effect of plant phenology, in which cultivated *C. arietinum* ICCV 96029 was compared with wild *C. reticulatum* Oyali_84. Although total nodule biomass was similar, nodule size and nodule number varied significantly. Cultivated ICCV 96029 had many fewer and larger nodules in the interaction with *M. mediterraneum* than with *M. ciceri*. Moreover, although these differences were ultimately correlated with differences in symbiotic benefit (shoot biomass gain, Figure 2A-E), strain-dependent differences in nodule size and number were also evident during vegetative growth (27 and 34 days), well in advance of measurable strain-dependent differences in shoot biomass gain (compare genotype 96029 in Figure 2A with Figures 2C and E).

An additional plant genotype-dependent difference in nodulation was observed in the case of multi-spectral reflectance data, collected incidental to imaging for nodule size and morphology. Sampled wavelengths extended beyond the human-visible spectrum. The technology has widespread application to seed quality, where imaging provides the basis to remove diseased and low quality seed (Vrešak et al., 2016). Moreover, reflectance data can indicate differences in underlying (bio)chemistry. Plotting of nodule reflectance data in a principal components analysis reveals clear separation of plant genotypes, and an effect of plant-bacterial genotype interaction, but not of bacterial genotype alone. The origin of spectral differences is uncertain, but likely derives from biochemical properties of nodules that differ repeatably between plant genotypes.

For example, nodules with incomplete differentiation will have lower levels of leghemoglobin, while nodules with premature senescence would have both reduced leghemoglobin and accumulation of bilirubin (Pislariu et al., 2019).

CONCLUSION

The research presented in this chapter demonstrate that biological variation in wild *Cicer-Mesorhizobium* symbiosis supports two broad conclusions. (1) That plant and bacterial biogeographic patterns reflect local adaptation for efficient symbiosis in wild systems. (2) That post-domestication genetic processes have reduced the effectiveness of symbiosis in certain crop genotypes.

The genetic control of symbiotic specificity and compatibility among *C. echinospermun* and *Mesorhizobium* is a complex trait. More broadly, other crop wild relative traits for symbiosis efficiency are associated with host genetic control, and can be harnessed for transferring compatible alleles to modern cultivated genotypes via interspecific crosses, and targeted breeding for nitrogen fixation.

MATERIALS AND METHODS

Biological material

Wild *Mesorhizobium* spp. strains were isolated and cultured from nodules of wild *Cicer* spp. in across different geographical location in Turkey by Greenlon et al (2019). Wild strains belong to two different *Mesorhizobium* species, the host were naturally growing in different geographical locations in Turkey (von Wettberg et al., 2018). For most of the experiments wild strains were used *M. ciceri* strains were Kar-203 (Karabache soil), Rse17 (Siverek- Diyarbakir roadside soil), and Gun202 (Gunnisan soil). *M. mediterraneum* strains were B2O3/Rse19 (Oyali soil), Rse11 (Kalkan Soil), and Rse20 (Sarikaya soil). Two agricultural strains (43P5 and 43P2) were

tested in one (von Wettberg et al., 2018)experiment, are two novel symbiont species, related to *M. plurifarium* and isolated from agricultural chickpea in Ethiopia (Greenlon et al., 2019).

Wild *Cicer* spp. accessions were collected by von Wettberg et al. (2018) from natural field sites in Turkey. We tested a subset of 14 genotypes that represent the known diversity of the wild's primary (*C. reticulatum*) and secondary (*C. echinospermum*) gene pools. Together these wild accessions represent 13 physical locations and 10 distinct genetic populations. The genotypes were chosen to represent diverse geographical locations from wild accessions and cultivated genotypes representative from different breeding programs. *C. echinospermun* genotypes were Gunas_62, Gunas_61 and S2Drd_65. *C. reticulatum* genotypes were Bari1_92, Bari3_106D, Besev_79, CudiA_152, CudiB_22C, Egill_73, Kalkan_64, Kesen_75, Oyali_84, Savur_63, and Sirna_60. Cultivated *C. arietinum* genotypes were Billy Bean (US Pacific Northwest), CDC_Consul (Canada), CDC_Leader (Canada), CDC_Orion (Canada), Habru (Ethiopia), Hattrick (Australia), ICCV 96029 (India), Kyabra (Australia) and Minjar (Ethiopia). Seed material was available in the germplasm collection at the Cook Lab, University of California- Davis.

Greenhouse growth conditions.

All experiments were performed in greenhouse 705 of the UC Davis Core Greenhouse Facilities. Average daytime temperatures were 20C – 22 C and night time temperatures were 16C - 18C. To provide a nutrient-deficient growth substrate we used 80% sand mixed with 20% UC Mix. One cubic yard of UC Mix contains a 1:1 mixture of sand and sphagnum peat moss (0.5 yd^3 each), supplemented with potassium nitrate (0.25 lb), potassium sulfate (0.25 lb), triple super phosphate (2.5 lb), oyster shell lime (2.5 lb), dolomite lime (7.5 lb). To reduce unwanted microorganisms, the substrate was steam sterilized twice. *Rhizobium*-inoculated plants were fertilized with nitrogen free nutrients: KH_2PO_4 5.5 mM, K_2SO_4 0.52 mM, $MgSO_4.7H_2O$, 0.25 mM, $CaCl_2.2H2O$ 1 mM, $FeSO_4.7H_2O$ 50 µM, Na_2EDTA 50 µM, H_3BO_3 30 µM, $MnSO_4.H_2O$ 10 µM, $ZnSO_4.7H_2O$ 0.7 µM, $CuSO_4.5H_2O$ 0.2 µM, $Na_2MoO_4.2H_2O$ 1 µM, and $CoCl_2.6H_20$ 0.04 µM. Control, inorganic nitrogen treatment was with NH_4NO_3.

The inorganic nitrogen dose response experiment (Fig. 1) was conducted under 14 hours of light, with individual plants sowed separately into 2.5 inches diameter Deepot cells (Stuewe & Sons). To minimize the pot-size effect on root development (Poorter et al., 2012), pot size was increased to a minimum of 1 gallon in subsequent experiments. With the exception of the inorganic nitrogen dose response experiment, experimental designs targeted 10 biological replicates for each genotype and treatment tested.

For the three independent experiments testing the performance of native *Mesorhizobium* in *Cicer* genotypes, (Fig. 2- Fig.8), 1 gal pots were used (approx. volume ~600 g), sowing two seeds per pot and thinning within the first week of germination. The first was phenology experiment, sowing was simultaneous for the two species and full replicates were harvested at desired developmental stages.

For genetic analysis of F4 recombinant inbred lines (Fig. 9), five seeds of a single RIL accession were co-planted into a single 3-gal pots. This strategy conserved space and made this exploratory experiment feasible in terms of logistics. Plants were harvested and phenotyped individually.

At the time of harvest, plant organs were separated into shoots and roots. Shoots were bagged and oven dried. Roots were carefully washed with water to retain nodules. For phenotyping that required fresh tissue, such as root imaging, plant samples were labeled and briefly stored moist at 4 °C. For large (n ~1000 plants) experiments, roots and separated nodules were stored in ethanol 70%, with rehydration in water at the time of analysis. After image processing, roots and nodules were oven dried. Dry weight biomass was recorded using a Mettler Toledo micro- or ultra-balance.

Bacterial culture conditions

Glycerol stocks were revived on solid YMA (Yeast Mannitol Agar) containing D-glucose 3g/l, Mannitol 2g/L, Yeast extract 1g/L, K_2HPO_4 0.5 g/L, $MgSO_4.7H_2O$ 0.2g/l, NaCl 0.1 g/L, $CaSO_4.2H_2O$ 0.05 g/L, NH4Cl 0.1 g/L and Agar 15 g/L, pH 6.8. For bacterial growth, agar plates

were maintained at 28°C. For liquid cultures YM medium was prepared with the same composition but without agar. Starter liquid cultures consisted of 1-5 ml volume, which were transferred to liquid media in 250 ml Erlenmeyer flasks with constant rotation, for 3-5 days. Typically, *M. ciceri* strains grew faster than *M. mediterraneum* strains.

Seed treatment and inoculation

Seeds were surface-sterilized by 1 min treatment with 4% Bleach, followed by extensive washing with sterile water, 1 min treatment with 70% ethanol, and three successive rinses with sterile water. After sterilization, seeds were nicked with nail cutters to break the seed coat under sterile conditions.

Inoculum consisted of freshly grown bacteria in YM medium, adjusted to an optical density of 0.1 OD with ½ strength YM medium. Seeds were imbibed in liquid inoculum in sterile 50 ml Falcon tubes for 3-4 h. Treated seed were immediately sown into pre-moistened substrate in pots. After sowing, 1ml of fresh inoculum (0.1 OD) was added to the surface of the soil substrate using sterile pipettes.

Digital Imaging

Roots were harvested and carefully spread horizontally on transparent trays for digital imaging. Image acquisition for root architecture analysis was conducted the WinRhizo (Regent Instruments Inc.) package equipped with an Epson scanner.

Nodules were imaged for morphology and reflectance using a Videometer3 (Analytik Ltd) multi-spectral imaging system, with data acquired and processed according to manufacturer's specifications.

Statistical analysis

To test the effect of nitrogen and *Mesorhizobium* spp. treatments on each *Cicer* spp. genotype, linear models for each quantitative trait (Shoot dry weight, Root Dry weight, Shoot-to-Root- ratio, Nodule number, etc) were computed with the interaction terms Strain and Genotype.

Derived calculations

The fitted linear models were the input for calculating the least-square means (or estimated marginal means) using the R package *emmeans*. The estimated means were contrasted using the pairwise method. Significant differences were denoted using compact letter display for Tukey-adjusted p-values. Treatments sharing the same letter are not significantly different at the 95% confidence interval, although having the same letter does not imply that the means are equal. To calculate the differences among *M. mediterraneum* and *M. ciceri*, paired t-test analysis was performed, and standardized effect sizes were calculated using Cohen's D (observed difference divided by population SD). All analyses were performed using the computational background of R studio Version 1.1.456.

Genetic analysis

F4 generation recombinant inbred line populations were derived from *C. echinospermun* Gunas_62 crossed with *C. arietinum* CDC Consul and *C. echinospermun* Gunas_63 crossed with *C. arietinum* ICCV96029. Thirty-five and 36 F4 families were used from each cross, respectively. Crosses involving CDC Consul and ICCV 96029 were harvested at 63 and 58 days, respectively, to accommodate for differences in time to maturity.

For each of the F4 families, 20 seeds were selected and divided in two experimental groups: 10 seeds for each F4 families were inoculated with *M. ciceri* Rse17 and 10 with *M. mediterraneum* Rse19. Five seeds from the same treatment and genotype were planted in two independents 3-gallon pots, giving 10 individual plants per treatment divided into two independent blocks. The inoculation volume was 5 ml of bacterial suspension OD 0.1 per pot. Plants were fertilized as described above.

REFERENCES

Alford BA, 2020. *Microbial Ecology of Wild and Domesticated Chickpeas*. Ann Arbor: University of California, DavisPh.D.

Alunni B, Gourion B, 2016. Terminal bacteroid differentiation in the legume-rhizobium symbiosis: nodule-specific cysteine-rich peptides and beyond. *New Phytol* **211**, 411-7.

Ane JM, Levy J, Thoquet P*, et al.*, 2002. Genetic and cytogenetic mapping of DMI1, DMI2, and DMI3 genes of Medicago truncatula involved in Nod factor transduction, nodulation, and mycorrhization. *Mol Plant Microbe Interact* **15**, 1108-18.

Banba M, Siddique A-BM, Kouchi H, Izui K, Hata S, 2001. Lotus japonicus Forms Early Senescent Root Nodules with Rhizobium etli. *Molecular Plant-Microbe Interactions®* **14**, 173-80.

Béna G, Lyet A, Huguet T, Olivieri I, 2005. Medicago-Sinorhizobium symbiotic specificity evolution and the geographic expansion of Medicago. *J Evol Biol* **18**, 1547-58.

Bourcy M, Brocard L, Pislariu CI*, et al.*, 2013. Medicago truncatula DNF2 is a PI-PLC-XD-containing protein required for bacteroid persistence and prevention of nodule early senescence and defense-like reactions. *New Phytologist* **197**, 1250-61.

Bozsoki Z, Gysel K, Hansen SB*, et al.*, 2020. Ligand-recognizing motifs in plant LysM receptors are major determinants of specificity. *Science* **369**, 663-70.

Broughton WJ, Jabbouri S, Perret X, 2000. Keys to Symbiotic Harmony. *Journal of Bacteriology* **182**, 5641-52.

Burghardt LT, Epstein B, Guhlin J*, et al.*, 2018. Select and resequence reveals relative fitness of bacteria in symbiotic and free-living environments. *Proceedings of the National Academy of Sciences* **115**, 2425-30.

Carroll BJ, Mcneil DL, Gresshoff PM, 1985. Isolation and properties of soybean [Glycine max (L.) Merr.] mutants that nodulate in the presence of high nitrate concentrations. *Proc Natl Acad Sci U S A* **82**, 4162-6.

Concha C, Doerner P, 2020. The impact of the rhizobia-legume symbiosis on host root system architecture. *J Exp Bot* **71**, 3902-21.

Cook D, Dreyer D, Bonnet D, Howell M, Nony E, Vandenbosch K, 1995. Transient induction of a peroxidase gene in Medicago truncatula precedes infection by Rhizobium meliloti. *The Plant Cell* **7**, 43-55.

Djordjevic MA, Mohd-Radzman NA, Imin N, 2015. Small-peptide signals that control root nodule number, development, and symbiosis. *J Exp Bot* **66**, 5171-81.

Feng J, Lee T, Schiessl K, Oldroyd GED, 2021. Processing of NODULE INCEPTION controls the transition to nitrogen fixation in root nodules. *Science* **374**, 629-32.

Frederickson ME, 2013. Rethinking mutualism stability: cheaters and the evolution of sanctions. *Q Rev Biol* **88**, 269-95.

Geurts R, Heidstra R, Hadri AE*, et al.*, 1997. Sym2 of Pea Is Involved in a Nodulation Factor-Perception Mechanism That Controls the Infection Process in the Epidermis. *Plant Physiol* **115**, 351-9.

Greenlon A, Chang PL, Damtew ZM*, et al.*, 2019. Global-level population genomics reveals differential effects of geography and phylogeny on horizontal gene transfer in soil bacteria. *Proc Natl Acad Sci U S A* **116**, 15200-9.

Griesmann M, Chang Y, Liu X*, et al.*, 2018. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* **361**, eaat1743.

Haney CH, Riely BK, Tricoli DM, Cook DR, Ehrhardt DW, Long SR, 2011. Symbiotic rhizobia bacteria trigger a change in localization and dynamics of the Medicago truncatula receptor kinase LYK3. *Plant Cell* **23**, 2774-87.

Herridge DF, Peoples MB, Boddey RM, 2008. Global inputs of biological nitrogen fixation in agricultural systems. *Plant and Soil* **311**, 1-18.

Imin N, Patel N, Corcilius L, Payne RJ, Djordjevic MA, 2018. CLE peptide tri-arabinosylation and peptide domain sequence composition are essential for SUNN-dependent autoregulation of nodulation in Medicago truncatula. *New Phytol* **218**, 73-80.

Kassaw T, Bridges W, Jr., Frugoli J, 2015. Multiple Autoregulation of Nodulation (AON) Signals Identified through Split Root Analysis of Medicago truncatula sunn and rdn1 Mutants. *Plants (Basel, Switzerland)* **4**, 209-24.

Kawaharada Y, Kelly S, Nielsen MW*, et al.*, 2015. Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature* **523**, 308-12.

Khanal S, Kc K, Fulton JP, Shearer S, Ozkan E, 2020. Remote Sensing in Agriculture—Accomplishments, Limitations, and Opportunities. *Remote Sensing* **12**.

Kiba T, Krapp A, 2016. Plant Nitrogen Acquisition Under Low Availability: Regulation of Uptake and Root Architecture. *Plant and Cell Physiology* **57**, 707-14.

Kiers ET, Hutton MG, Denison RF, 2007. Human selection and the relaxation of legume defences against ineffective rhizobia. *Proc. R. Soc. B* **274**, 3119-26.

Kraemer SA, Boynton PJ, 2017. Evidence for microbial local adaptation in nature. *Molecular Ecology* **26**, 1860-76.

Laffont C, Huault E, Gautrat P*, et al.*, 2019. Independent Regulation of Symbiotic Nodulation by the SUNN Negative and CRA2 Positive Systemic Pathways. *Plant Physiol* **180**, 559-70.

Laffont C, Ivanovici A, Gautrat P, Brault M, Djordjevic MA, Frugier F, 2020. The NIN transcription factor coordinates CEP and CLE signaling peptides that regulate nodulation antagonistically. *Nature Communications* **11**, 3167.

Larrainzar E, Gil-Quintana E, Seminario A, Arrese-Igor C, González EM, 2014. Nodule carbohydrate catabolism is enhanced in the Medicago truncatula A17-Sinorhizobium medicae WSM419 symbiosis. *Frontiers in Microbiology* **5**.

Larrainzar E, Riely BK, Kim SC*, et al.*, 2015. Deep Sequencing of the Medicago truncatula Root Transcriptome Reveals a Massive and Early Interaction between Nodulation Factor and Ethylene Signals. *Plant Physiol* **169**, 233-65.

Lebedeva M, Azarakhsh M, Yashenkova Y, Lutova L, 2020. Nitrate-Induced CLE Peptide Systemically Inhibits Nodulation in Medicago truncatula. *Plants (Basel, Switzerland)* **9**.

Lee WK, Jeong N, Indrasumunar A, Gresshoff PM, Jeong SC, 2011. Glycine max non-nodulation locus rj1: a recombinogenic region encompassing a SNP in a lysine motif receptor-like kinase (GmNFR1α). *Theor Appl Genet* **122**, 875-84.

Liu J, Yang S, Zheng Q, Zhu H, 2014. Identification of a dominant gene in Medicago truncatula that restricts nodulation by Sinorhizobium meliloti strain Rm41. *BMC Plant Biol* **14**, 167.

Liu J, Yu X, Qin Q, Dinkins RD, Zhu H, 2020. The Impacts of Domestication and Breeding on Nitrogen Fixation Symbiosis in Legumes. *Frontiers in Genetics* **11**.

Maróti G, Kondorosi É, 2014. Nitrogen-fixing Rhizobium-legume symbiosis: are polyploidy and host peptide-governed symbiont differentiation general principles of endosymbiosis? *Frontiers in Microbiology* **5**.

Mcdonald BA, Stukenbrock EH, 2016. Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**, 20160026.

Mendoza-Suárez M, Andersen SU, Poole PS, Sánchez-Cañizares C, 2021. Competition, Nodule Occupancy, and Persistence of Inoculant Strains: Key Factors in the Rhizobium-Legume Symbioses. *Frontiers in Plant Science* **12**.

Mens C, Hastwell AH, Su H, Gresshoff PM, Mathesius U, Ferguson BJ, 2021. Characterisation of Medicago truncatula CLE34 and CLE35 in nitrate and rhizobia regulation of nodulation. *New Phytol* **229**, 2525-34.

Mergaert P, Uchiumi T, Alunni B*, et al.*, 2006. Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium-legume symbiosis. *Proc Natl Acad Sci U S A* **103**, 5230-5.

Middleton PH, Jakab J, Penmetsa RV*, et al.*, 2007. An ERF transcription factor in Medicago truncatula that is essential for Nod factor signal transduction. *Plant Cell* **19**, 1221-34.

Mohd-Radzman NA, Laffont C, Ivanovici A*, et al.*, 2016. Different Pathways Act Downstream of the CEP Peptide Receptor CRA2 to Regulate Lateral Root and Nodule Development. *Plant Physiol* **171**, 2536-48.

Montiel J, Szucs A, Boboescu IZ, Gherman VD, Kondorosi E, Kereszt A, 2016. Terminal Bacteroid Differentiation Is Associated With Variable Morphological Changes in Legume Species Belonging to the Inverted Repeat-Lacking Clade. *Mol Plant Microbe Interact* **29**, 210-9.

Oke V, Long SR, 1999. Bacteroid formation in the Rhizobium-legume symbiosis. *Curr Opin Microbiol* **2**, 641-6.

Oldroyd GE, Downie JA, 2008. Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu Rev Plant Biol* **59**, 519-46.

Oldroyd GE, Engstrom EM, Long SR, 2001. Ethylene inhibits the Nod factor signal transduction pathway of Medicago truncatula. *Plant Cell* **13**, 1835-49.

Oldroyd GE, Murray JD, Poole PS, Downie JA, 2011. The rules of engagement in the legume-rhizobial symbiosis. *Annu Rev Genet* **45**, 119-44.

Pan H, Wang D, 2017. Nodule cysteine-rich peptides maintain a working balance during nitrogen-fixing symbiosis. *Nature Plants* **3**, 17048.

Parker MA, 1999. Mutualism in Metapopulations of Legumes and Rhizobia. *Am Nat* **153**, S48-s60.

Penmetsa RV, Cook DR, 1997. A Legume Ethylene-Insensitive Mutant Hyperinfected by Its Rhizobial Symbiont. *Science* **275**, 527-30.

Penmetsa RV, Frugoli JA, Smith LS, Long SR, Cook DR, 2003. Dual genetic pathways controlling nodule number in Medicago truncatula. *Plant Physiol* **131**, 998-1008.

Perret X, Staehelin C, Broughton WJ, 2000. Molecular Basis of Symbiotic Promiscuity. *Microbiology and Molecular Biology Reviews* **64**, 180-201.

Pislariu CI, Sinharoy S, Torres-Jerez I, Nakashima J, Blancaflor EB, Udvardi MK, 2019. The Nodule-Specific PLAT Domain Protein NPD1 Is Required for Nitrogen-Fixing Symbiosis *Plant Physiol* **180**, 1480-97.

Poorter H, Niklas KJ, Reich PB, Oleksyn J, Poot P, Mommer L, 2012. Biomass allocation to leaves, stems and roots: meta-analyses of interspecific variation and environmental control. *New Phytologist* **193**, 30-50.

Porter SS, Sachs JL, 2020. Agriculture and the Disruption of Plant-Microbial Symbiosis. *Trends Ecol Evol* **35**, 426-39.

Reid DE, Ferguson BJ, Hayashi S, Lin Y-H, Gresshoff PM, 2011. Molecular mechanisms controlling legume autoregulation of nodulation. *Ann Bot* **108**, 789-95.

Sachs JL, Russell JE, Lii YE, Black KC, Lopez G, Patil AS, 2010. Host control over infection and proliferation of a cheater symbiont. *Journal of Evolutionary Biology* **23**, 1919-27.

Schauser L, Roussis A, Stiller J, Stougaard J, 1999. A plant regulator controlling development of symbiotic root nodules. *Nature* **402**, 191-5.

Schiltz S, Munier-Jolain N, Jeudy C, Burstin J, Salon C, 2005. Dynamics of Exogenous Nitrogen Partitioning and Nitrogen Remobilization from Vegetative Organs in Pea Revealed by (15)N in Vivo Labeling throughout Seed Filling. *Plant Physiol* **137**, 1463-73.

Schnabel E, Journet EP, De Carvalho-Niebel F, Duc G, Frugoli J, 2005. The Medicago truncatula SUNN gene encodes a CLV1-like leucine-rich repeat receptor kinase that regulates nodule number and root length. *Plant Mol Biol* **58**, 809-22.

Schnabel EL, Kassaw TK, Smith LS*, et al.*, 2011. The ROOT DETERMINED NODULATION1 gene regulates nodule number in roots of Medicago truncatula and defines a highly conserved, uncharacterized plant gene family. *Plant Physiol* **157**, 328-40.

Schumpp O, Crèvecoeur M, Broughton WJ, Deakin WJ, 2008. Delayed maturation of nodules reduces symbiotic effectiveness of the Lotus japonicus–Rhizobium sp. NGR234 interaction. *J Exp Bot* **60**, 581-90.

Searle IR, Men AE, Laniya TS*, et al.*, 2003. Long-Distance Signaling in Nodulation Directed by a CLAVATA1-Like Receptor Kinase. *Science* **299**, 109-12.

Smit P, Limpens E, Geurts R*, et al.*, 2007. Medicago LYK3, an Entry Receptor in Rhizobial Nodulation Factor Signaling. *Plant Physiol* **145**, 183-91.

Sokolkova A, Bulyntsev SV, Chang PL*, et al.*, 2020. Genomic Analysis of Vavilov's Historic Chickpea Landraces Reveals Footprints of Environmental and Human Selection. *Int J Mol Sci* **21**.

Starker CG, Parra-Colmenares AL, Smith L, Mitra RM, Long SR, 2006. Nitrogen Fixation Mutants of Medicago truncatula Fail to Support Plant and Bacterial Symbiotic Gene Expression   *Plant Physiol* **140**, 671-80.

Tirichine L, De Billy F, Huguet T, 2000. Mtsym6, a gene conditioning Sinorhizobium strain-specific nitrogen fixation in Medicago truncatula. *Plant Physiol* **123**, 845-51.

Tsikou D, Yan Z, Holt DB*, et al.*, 2018. Systemic control of legume susceptibility to rhizobial infection by a mobile microRNA. *Science* **362**, 233-6.

Tsukui T, Eda S, Kaneko T, *et al.*, 2013. The Type III Secretion System of Bradyrhizobium japonicum USDA122 Mediates Symbiotic Incompatibility with Rj2 Soybean Plants. *Applied and Environmental Microbiology* **79**, 1048-51.

Van De Velde W, Pérez-Guerra JC, De Keyser A, *et al.*, 2006. Aging in legume symbiosis: a molecular view on nodule senescence in Medicago truncatula. *Plant Physiol* **141**.

Van De Velde W, Zehirov G, Szatmari A, *et al.*, 2010. Plant Peptides Govern Terminal Differentiation of Bacteria in Symbiosis. *Science* **327**, 1122-6.

Vasse J, De Billy F, Camut S, Truchet G, 1990. Correlation between ultrastructural differentiation of bacteroids and nitrogen fixation in alfalfa nodules. *J Bacteriol* **172**, 4295-306.

Von Wettberg EJB, Chang PL, Başdemir F, *et al.*, 2018. Ecology and genomics of an important crop wild relative as a prelude to agricultural innovation. *Nature Communications* **9**, 649.

Vrešak M, Halkjaer Olesen M, Gislum R, Bavec F, Ravn Jørgensen J, 2016. The Use of Image-Spectroscopy Technology as a Diagnostic Method for Seed Health Testing and Variety Identification. *PLoS One* **11**, e0152011.

Wang D, Griffitts J, Starker C, *et al.*, 2010. A nodule-specific protein secretory pathway required for nitrogen-fixing symbiosis. *Science* **327**, 1126-9.

Wang D, Yang S, Tang F, Zhu H, 2012. Symbiosis specificity in the legume: rhizobial mutualism. *Cell Microbiol* **14**, 334-42.

Wang Q, Liu J, Li H, *et al.*, 2018. Nodule-Specific Cysteine-Rich Peptides Negatively Regulate Nitrogen-Fixing Symbiosis in a Strain-Specific Manner in Medicago truncatula. *Mol Plant Microbe Interact* **31**, 240-8.

Warschefsky E, Penmetsa RV, Cook DR, Von Wettberg EJ, 2014. Back to the wilds: tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. *Am J Bot* **101**, 1791-800.

Westhoek A, Field E, Rehling F, *et al.*, 2017. Policing the legume-Rhizobium symbiosis: a critical test of partner choice. *Scientific Reports* **7**, 1419.

Yang S, Tang F, Gao M, Krishnan HB, Zhu H, 2010. R gene-controlled host specificity in the legume–rhizobia symbiosis. *Proceedings of the National Academy of Sciences* **107**, 18735-40.

Yang S, Wang Q, Fedorova E, *et al.*, 2017. Microsymbiont discrimination mediated by a host-secreted peptide in Medicago truncatula. *Proc Natl Acad Sci U S A* **114**, 6848-53.

Yang WC, De Blank C, Meskiene I, *et al.*, 1994. Rhizobium nod factors reactivate the cell cycle during infection and nodule primordium formation, but the cycle is only completed in primordium formation. *The Plant Cell* **6**, 1415-26.

Zhukov V, Radutoiu S, Madsen LH, *et al.*, 2008. The pea Sym37 receptor kinase gene controls infection-thread initiation and nodule development. *Mol Plant Microbe Interact* **21**, 1600-8.

Zipfel C, Oldroyd GED, 2017. Plant signalling in symbiosis and immunity. *Nature* **543**, 328-36.

At 10 mM NH$_4$NO$_3$		Shoot/Root ratio	
		Mean	± SE
Cultivated genotypes	Ca_96029	5.3862	0.1393
	Ca_Consul	3.3780	0.1393
Wild Genotypes	Cr_CudiA_152	2.5300	0.1558
	Cr_Kalkan_64	2.6642	0.1393
	Cr_Oyali_84	2.4080	0.1393
	Cr_Savur_63	2.9360	0.1393

Table 1. **Comparative Shoot/root ratio response to inorganic nitrogen (10mM), Experiment 1.** Cultivated genotypes showed a higher Shoot/Root ratio than wild *Cicer reticulatum* genotypes.

Number of plants	*M. ciceri*			*M. mediterraneum*			
Host Genotype	Mc1-Gun202	Mc2-Rse17	Mc3-Kar203	Mm1-Rse11	Mm2-B203	Mm3-Rse20	N-10
Ca_96029	8	7	8	9	7	9	9
Ca_Consul	4	3	6	2	3	4	5
Ce_Gunas62	7	6	6	8	8	9	6
Ce_S2Drd_65	3	3	2	4	2	NA	5
Cr_Bari_092	7	8	9	6	9	8	8
Cr_Bari3_106D	8	7	9	8	9	9	6
Cr_CudiB_22C	9	8	9	9	9	9	9
Cr_Egill73	6	5	4	5	6	8	5
Cr_Kesen75	9	6	6	8	7	4	4
Cr_Oyali84	9	7	6	7	4	8	7
Cr_Savur63	9	9	9	8	9	9	9
Cr_Sirna60	7	7	8	9	10	6	5

Table 2. **Biological replicates per host genotypes and *Mesorhizobium*/Nitrogen treatment with emphasis in wild genotypes tested in experiment 3.**

Number of plants	*M. ciceri*	*M. mediterraneum*	N-Zero	N-10

Host Genotype	Mc1-Gun202	Mc2-Rse17	Mc3-Kar203	Mm1-Rse11	Mm2-B203	Mm3-Rse20		
Ca_Billy Bean	8	8	8	9	8	8	8	9
Ca_CDC_Consul	9	9	9	9	7	12	9	9
Ca_CDC_Leader	9	9	9	9	9	9	9	4
Ca_CDC_Orion	9	7	9	9	4	8	8	8
Ca_Habru	9	9	9	7	9	9	7	8
Ca_Hattrick	9	10	9	9	9	9	9	9
Ca_ICCV96029	9	9	9	9	9	9	8	9
Ca_Kyabra	9	9	9	9	9	9	9	9
Ca_Minjar	8	8	9	9	9	6	4	9
Ce_Gunas_62	8	8	8	8	3	9	8	9
Ce_S2Drd_065	3	6	6	7	7	8	6	3
Cr_Besev_79	8	9	9	9	9	9	9	9
Cr_Egill_73	5	9	5	9	10	8	5	7
Cr_Sirna_60	9	9	9	9	9	9	7	3

Table 3. **Biological replicates per host genotypes and *Mesorhizobium*/Nitrogen treatment for experiment 4, with focus in cultivated genotypes.**

Population F4_96xGun63

F4.Plant.ID	Mc-Rse17 Shoot DW (g)			Mm-Rse19 Shoot DW (g)			T-test	
	N	Mean	se	N	mean	se	p value	Sig.
326a	10	0.947	0.100	10	1.3589	0.158	0.0437	*
371a	9	0.742	0.120	9	1.28019	0.177	0.0243	*
375a	10	0.992	0.129	10	1.45507	0.170	0.0445	*
10315b	9	0.501	0.034	10	0.89454	0.120	0.0097	**
386b	10	0.746	0.092	9	1.37498	0.128	0.0012	**
400a	10	0.711	0.054	10	1.50219	0.215	0.0051	**
376a	10	0.687	0.108	9	1.28127	0.087	0.0005	***

Population F4_conXGun62

F4.Plant.ID	Mc-Rse17 Shoot DW (g)			Mm-Rse19 Shoot DW (g)			T-test	
	N	Mean	se	N	mean	se	p value	Sig.
5860b	10	0.894	0.197	9	1.54404	0.203	0.0346	*
5892a	9	0.656	0.097	9	1.19237	0.187	0.0257	*
5827a	9	0.877	0.101	10	1.367	0.119	0.0059	**
5925b	10	0.968	0.178	8	1.80974	0.129	0.0015	**
5938b	10	0.942	0.053	9	1.23214	0.080	0.0088	**
5880a	10	0.601	0.081	11	1.6085	0.196	0.0004	***

Table 4. Hybrid F4 families with significant differences in nitrogen fixation.
 among *M. ciceri* (Rse17) and *M. mediterraneum* (Rse19). Significance t-test ("*" $p<0.05$, "**"
$p<0.001$, "***" $p \leq 0.001$).
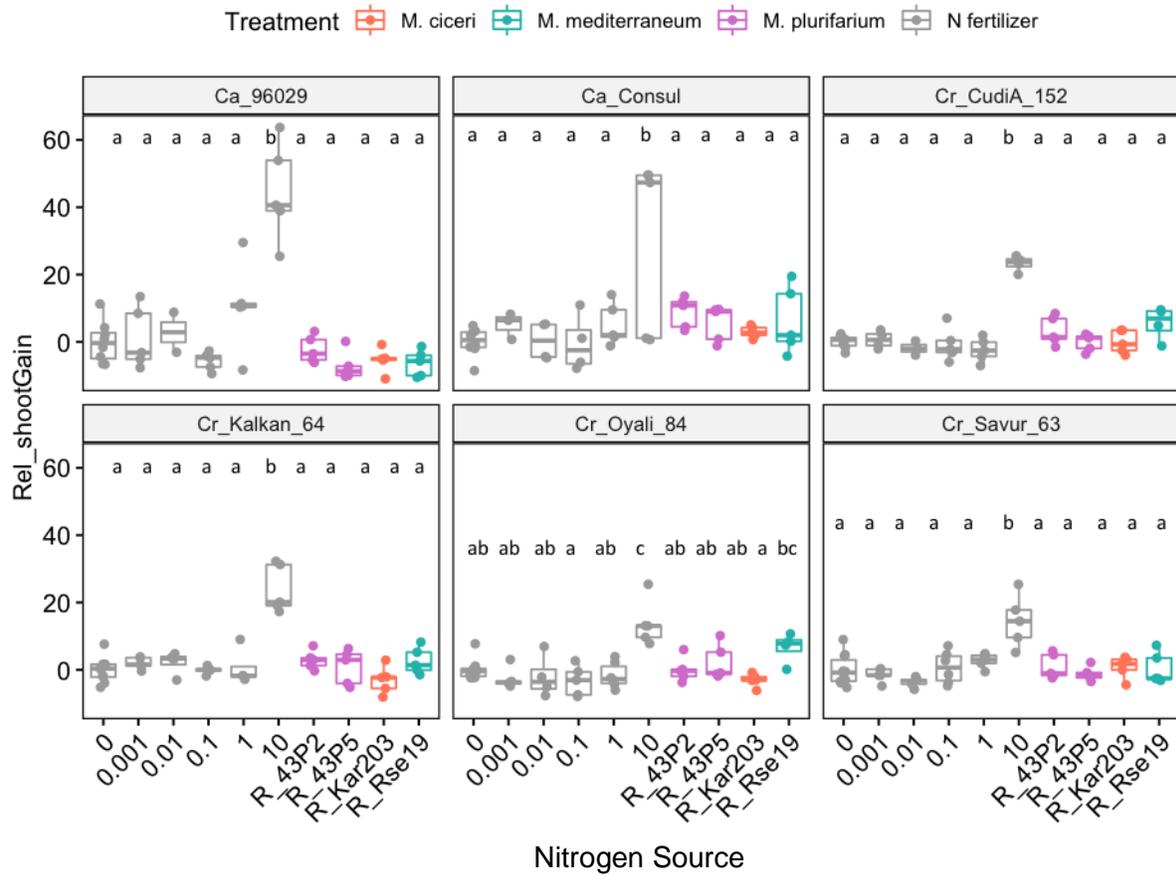
Figure 1. **Relative shoot gain (Index) response to inorganic nitrogen and rhizobia. We tested a range of nitrogen concentrations in a dose-response manner, and four *Mesorhizobium* spp. strains from different geographic origins (Experiment 1). Each panel shows the response of each host genotype to all the treatments. Boxplots with different letter(s) on top are significantly different at p < 0.05 (Tukey HSD) probability level. Gray boxplots show different concentrations of NH_4NO_3. Biological replications (n=5) per genotype/treatment condition. Relative shoot gain was calculated as percentage of gained weight above treatment of zero nitrogen per genotype.**
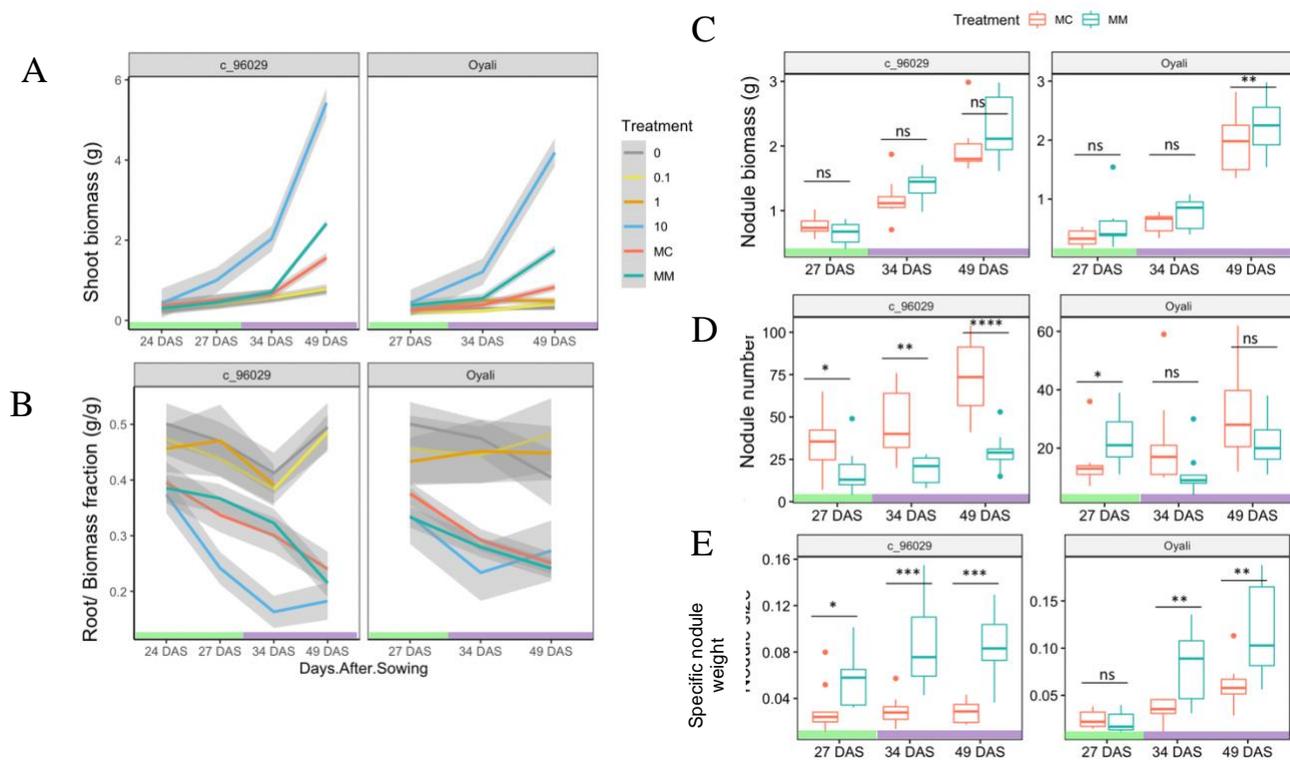
Figure 2. Symbiotic performance of native rhizobia with wild and cultivated chickpea along plant developmental gradient. Developmental stages are notes with color above the x axis, green=vegetative stage, purple= reproductive stage (Experiment 2).

The cultivated genotype was *C. arietinum*- ICCV96029 and the wild genotype was *C. reticulatum*-Oyali_84. The wild *C. reticulatum* started reproductive phase 27 days after sowing (DAS), while cultivated *C. arietinum* started reproductive phase 34 DAS. The rhizobia strains were *M. ciceri*-Kar203 (MC) and *M. mediterraneum*- B203 (MM). The inorganic nitrogen treatment had incremental dose of ammonium nitrate (NH_4NO_3) in a range from 0.1 mM to 10 mM. Biological replication per condition (n=10).

A. Aboveground dry weight. **B.** Root/ Biomass fraction **C.** Nodule fresh weight (g). **D.** Nodule number **E.** Nodule biomass/Nodule number. Shaded area in A and B is standard error (SE), in B-D, outliers are shown as points outside the boxplot. Significance t-test ("*" p<0.05, "**" p<0.001, "***" p ≤ 0.001, "****" P ≤ 0.0001).
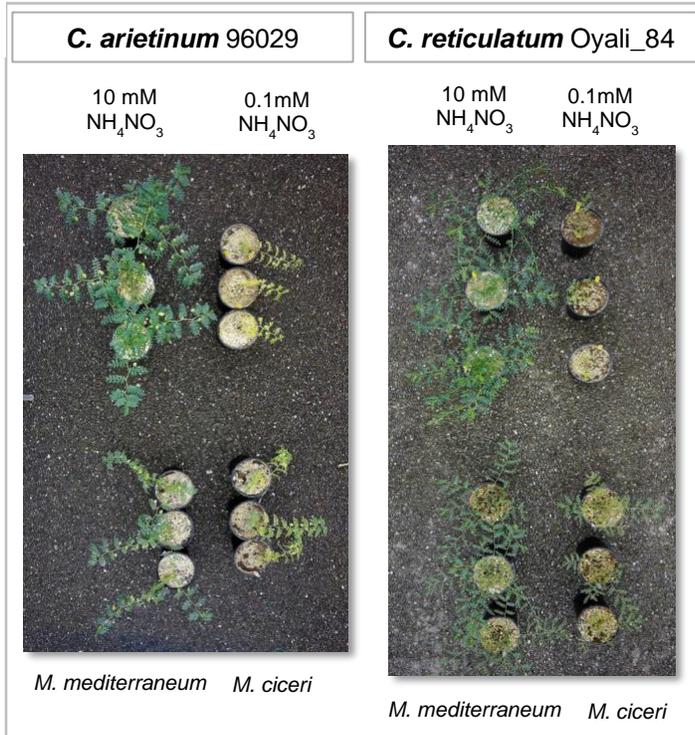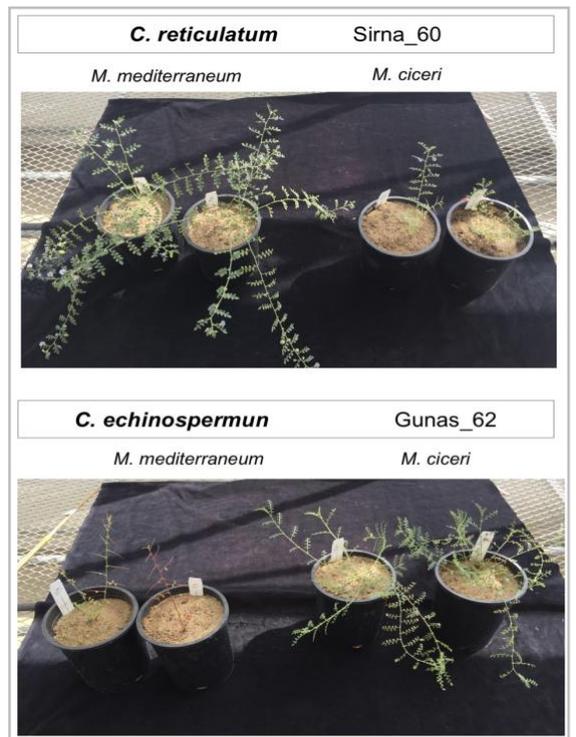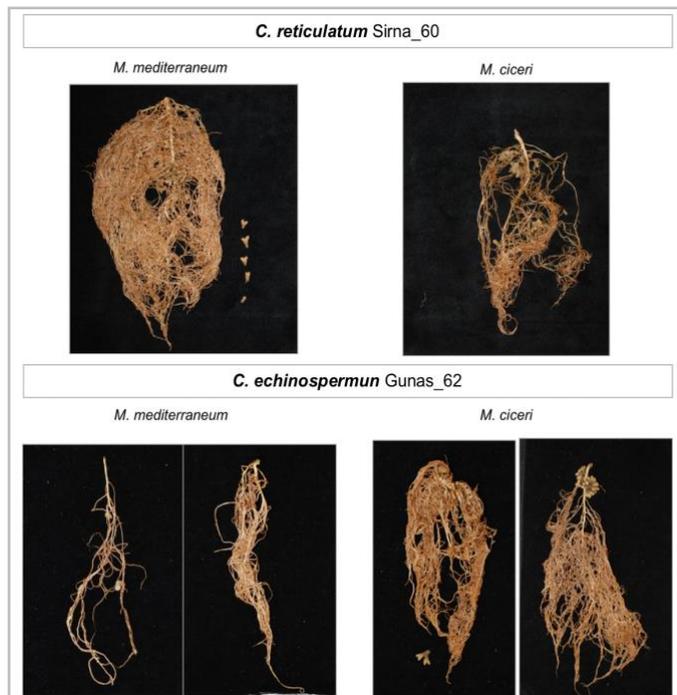
Figure 3. Plant growth aboveground and belowground under different nitrogen sources. **A.** Aboveground view for plants in Experiment 2 (Fig 2) at 49 DAS harvest. **B.** Contrasting shoot growth response to inoculation in wild Cicer genotypes **C.** Gross root response to inoculation in wild Cicer genotypes.
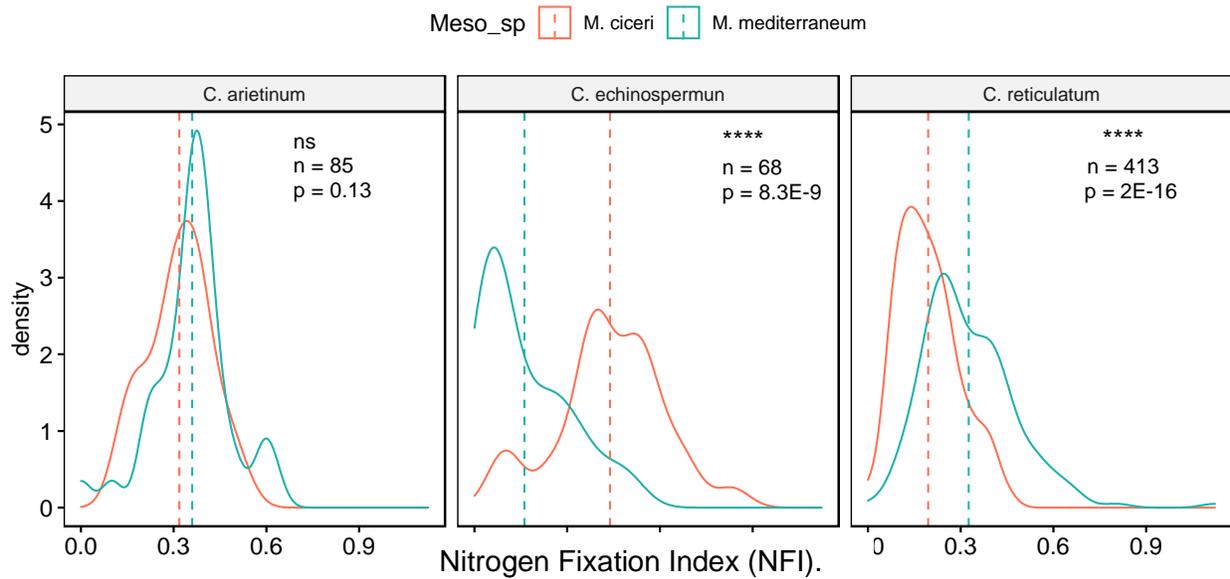
Figure 4. Wild *C. reticulatum* and *C. echinospermun* fix differentially N depending on *Mesorhizobium* species. Density plots represent aggregated genotypes per species, index for Nitrogen fixation efficiency (normalized shoot dry weight value to high nitrogen treatment) for the three *Cicer* species. The normalized index value of 1 represents the maximum shoot biomass under 10 mM of inorganic nitrogen. It allows to combine multiple genotypes within each species, *Cicer arietinum* (2 genotypes), *C. echinospermun* (2 genotypes), and *C. reticulatum* (8 genotypes). (Experiment 3). Inset show sample size per host species and significance t-test (ns= Not significative, "****" P ≤ 0.0001)

Figure 5. Biomass phenotypes for twelve host *Cicer* genotypes (Experiment 3). For each *Mesorhizobium* species, the data of the three strains of Mesorhizobium was aggregated. **A.** Normalized shoot dry weight value to high nitrogen treatment. **B.** Nodule number (n)/ root weight (mg). **C.** Shoot (g)/ Nodule dry weight(mg) ratio. Significance t-test ("*" p<0.05, "**" p<0.001, "***" p ≤ 0.001, "****" P ≤ 0.0001). **D.** Root (g)/biomass (g) ratio. Tukey HSD test for multiple comparisons found differences in the mean values, different letter(s) above boxplot correspond to the same group with p-value <0.05, 95% C.I.
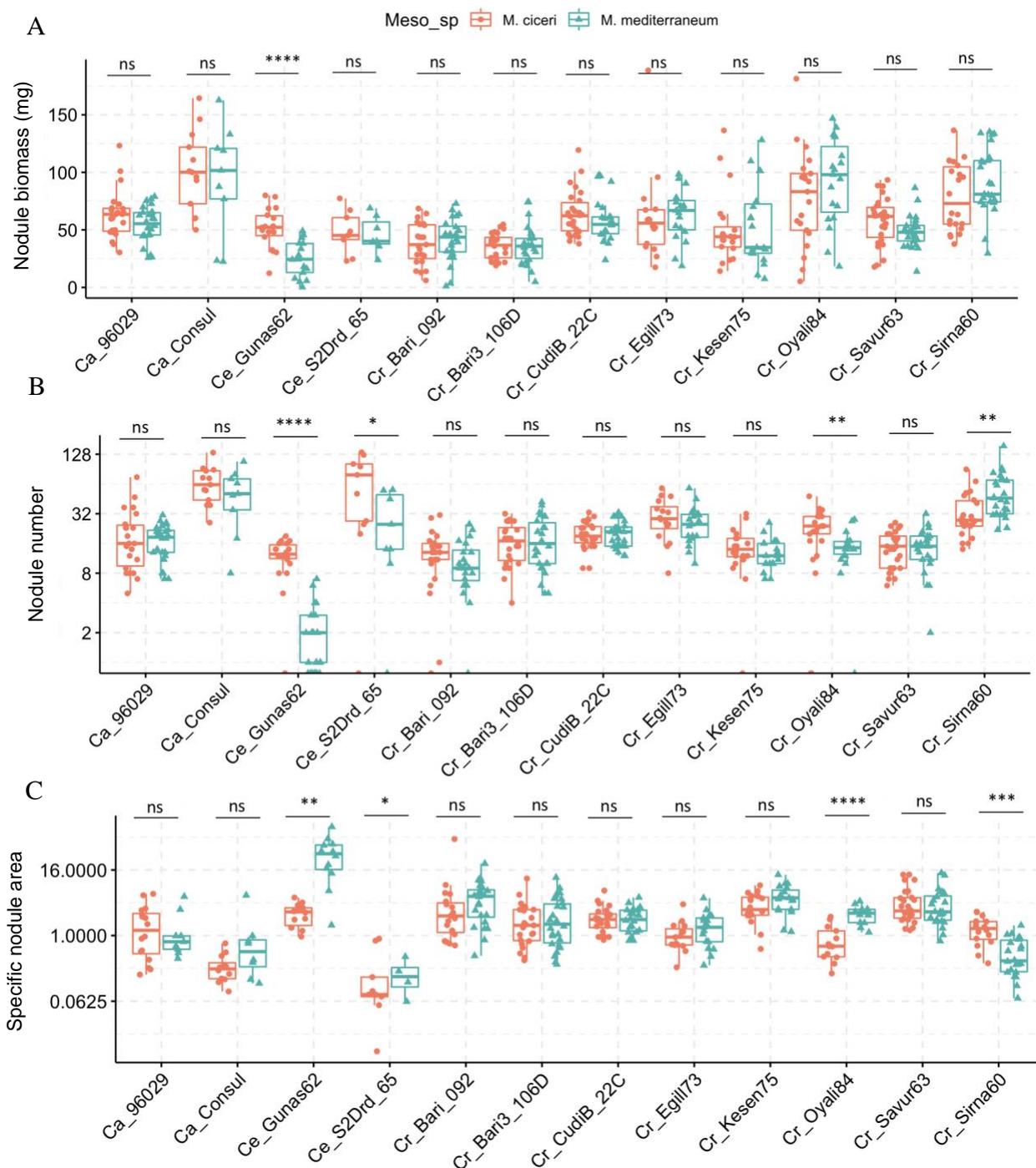
Figure 6. Nodulation Phenotypes.
Classical nodulation parameters (Experiment 3). **A.** Nodule biomass (g). **B.** Nodule Number. **C.** Specific nodule area (mm²) (Average nodule area/Nodule number). Significance t-test ("*" p<0.05, "**" p<0.001, "***" p ≤ 0.001, "****" P ≤ 0.0001).

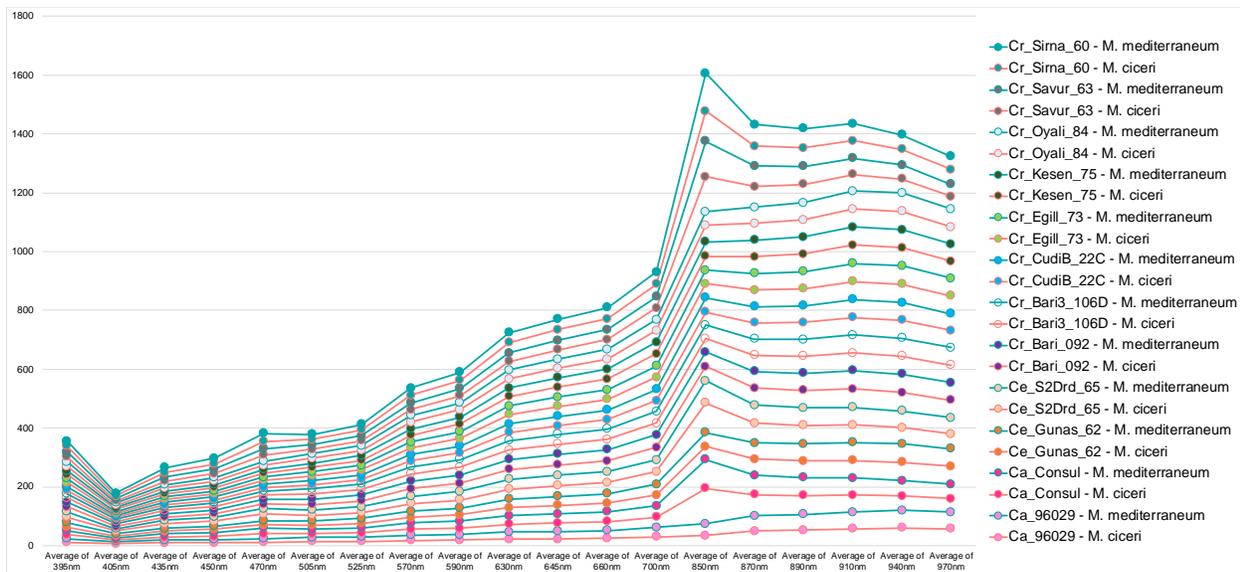A.

1. Multispectral Image acquisition

2. Transformation and segmentation of image objects

3. Computation of quantitative data per nodule
   A. Morphometric parameters
   B. Reflectance value in 19 wavelengths (375 nm -970 nm)

B.

Strain
Mc1-Gun202   Mc3-Kar203   Mm2-B203
Mc2-Rse17   Mm1-Rse11   Mm3-Rse20

C.

PCA - Biplot

Host Cicer, Genotypes
Ce_S2Drd_65
Ce_Gunas_62
Cr_Sirna_60
Cr_Savur_63
Cr_Oyali_84
Cr_Kesen_75
Cr_Egill_73
Cr_CudiB_22C
Cr_Bari3_106D
Cr_Bari_092
Ca_Consul
Ca_96029

contrib
0.05
0.10

D.

Figure 7. Digital nodule analysis.

A. Infographic of nodule digital analysis (Experiment 3). Digital images of nodules were captured with Videometer Lab 3 (Analitik), the reflectance of the fixed tissue in ethanol was captured in 19 wavelengths. After image acquisition, segmentation parameters were applied to extract individual nodule parameters.

B. Distribution of individual nodule area.

C. Principal component analysis (PCA) for 9480 nodules (Fig 6.), multiespectral parameters and morphometric parameters. Along the PCA x axis, Dimension1 represented 35.5% of variation, despite some degree of overlap, there is differentiation among the *Cicer* species. The PCA biplot include the two more relevant parameters: 470nm (spectral region blue), and 630 nm (spectral region red).

D. Average spectral profile per host genotype, Y values are multicolor mean reflectance upon incidence of each wavelength in X axis. Each genotype- *Mesorhizobium* category with unique spectra profile.
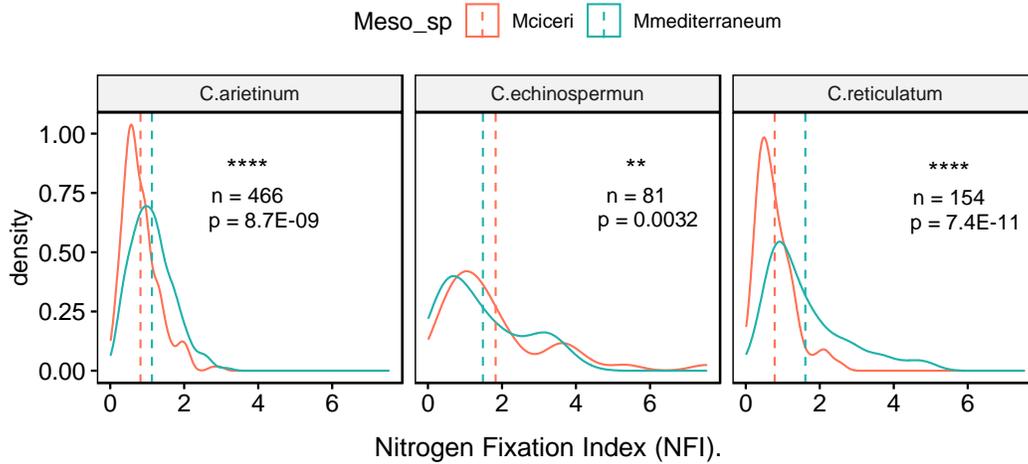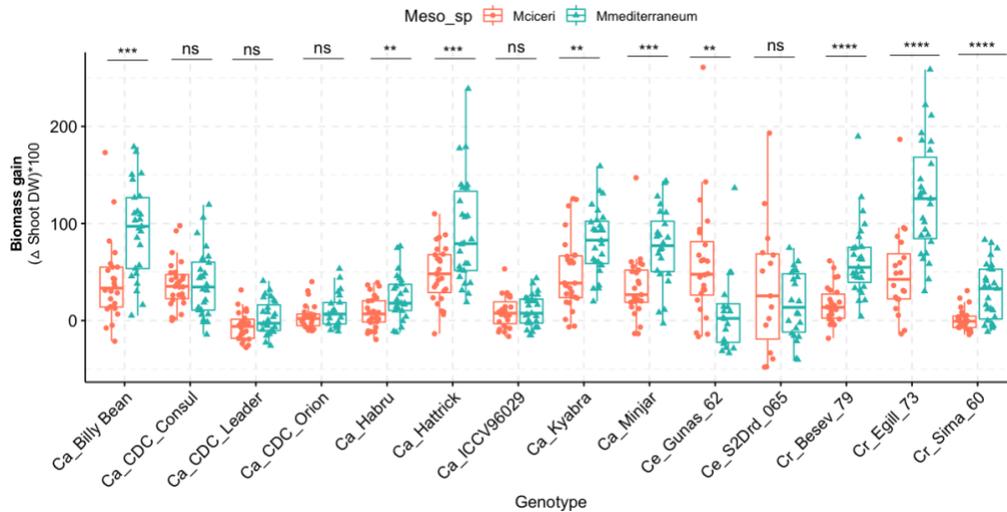
Figure 8. Cultivated genotypes of *C. arietinum* varies in their response to native
Mesorhizobium.
A. Density plots for Nitrogen fixation efficiency in cultivated diversity panel (Experiment 4).
Cicer arietinum (9 genotypes), *C. echinospermun* (2 genotypes), and *C. reticulatum* (3
genotypes). Inset show sample size per host species and significance t-test ("**" p<0.001,, "****"
P ≤ 0.0001)
B. Biomass gain per genotype, cultivars Billy Bean, Habru, Hattrick, Kyabra and Minjar responds
with more biomass increase with native strains of *Mesorhizobium mediterraneum* than with *M.
ciceri* native strains. Significance t.test ("**" p<0.001, "***" p ≤ 0.001, "****" P ≤ 0.0001).
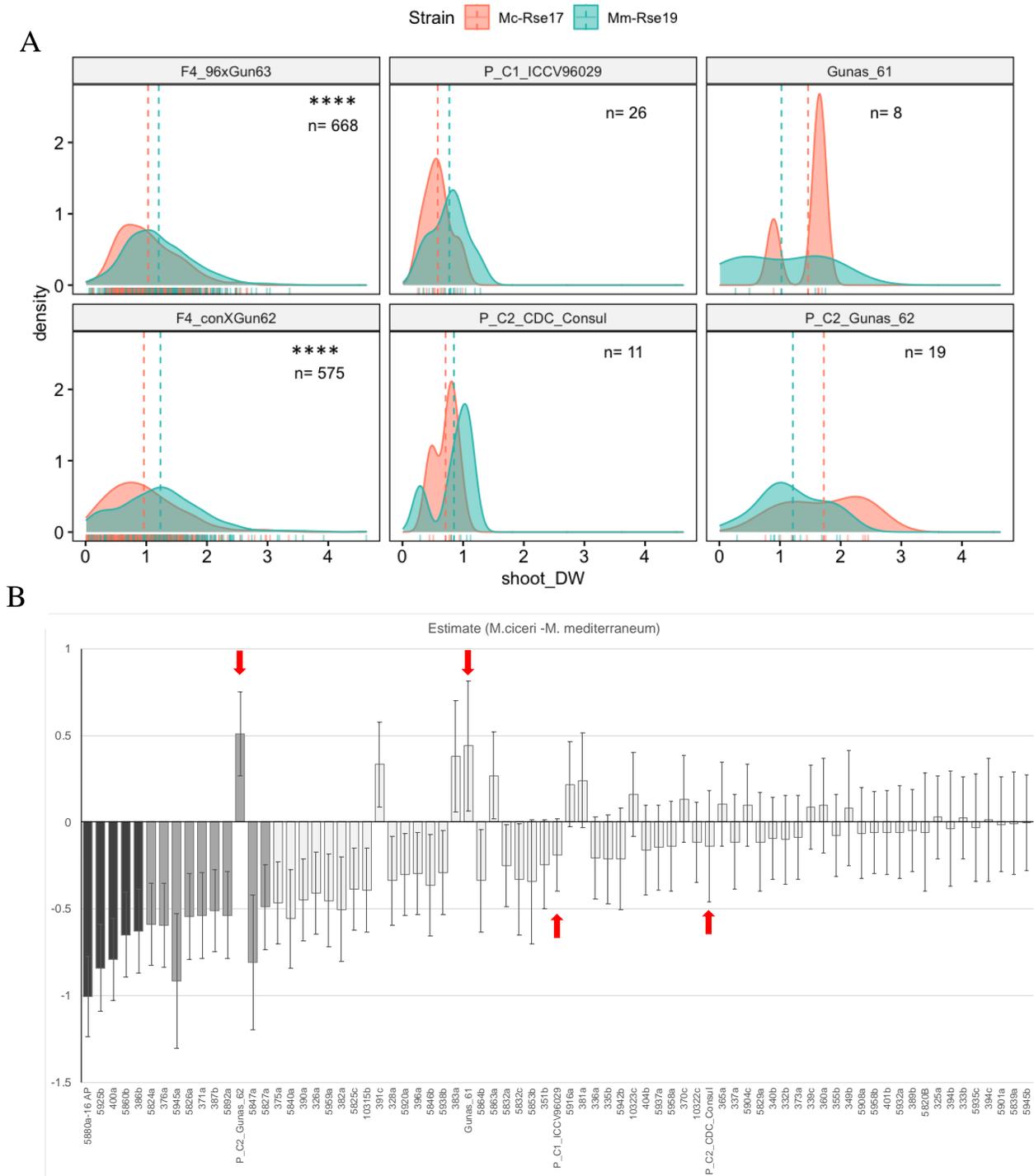
Figure 9. *C. arietinum* x *C. echinospermun* hybrid population (F4) and parental genotypes. **A.** Density plots of shoot dry weight in F4 biparental population (Experiment 5). Significance Wilcoxon test ("****" P ≤ 0.0001). **B.** F4 RILs ranked based on estimated marginal means statistic for pairwise differences in shoot biomass. Parental genotypes pointed with arrows. Dark Grey bars: p-value ≤ 0.01, Medium gray: p-value ≤ 0.05, light gray: p-value of > 0.05.
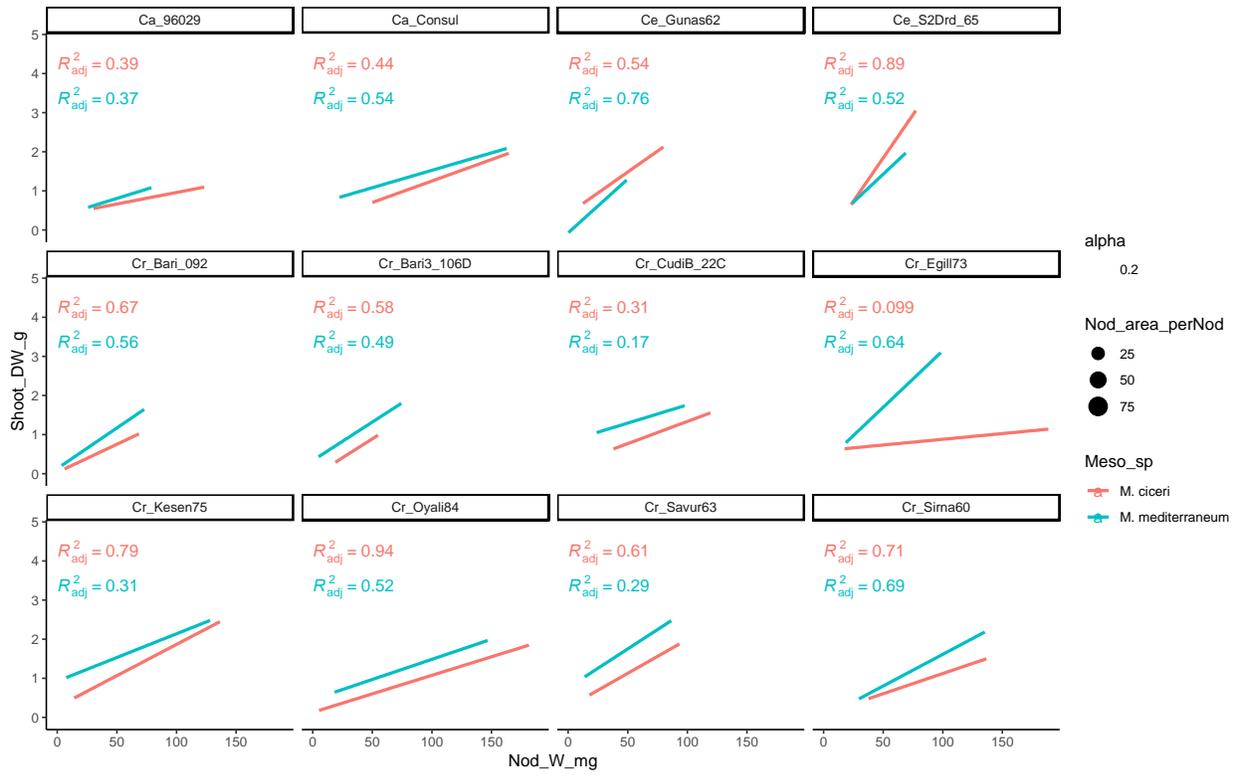
A

ICCV96029 xGunas63
F4 -333b   *M. mediterraneum*

F4- *C. arietinum* x *C. echinospermun*

Normal root-nodules   Abnormal root-nodules

B.

Normal flowers   Abnormal flower

Figure 10. Plant growth of *C. arietinum* x *C. echinospermun* hybrid population (F4)
A. Example of one plant from a F4 family that had poor growth and few nodules, as compared with siblings inoculated with *M. mediterraneum*. **B.** Some Hybrids F4 has morphological abnormalities
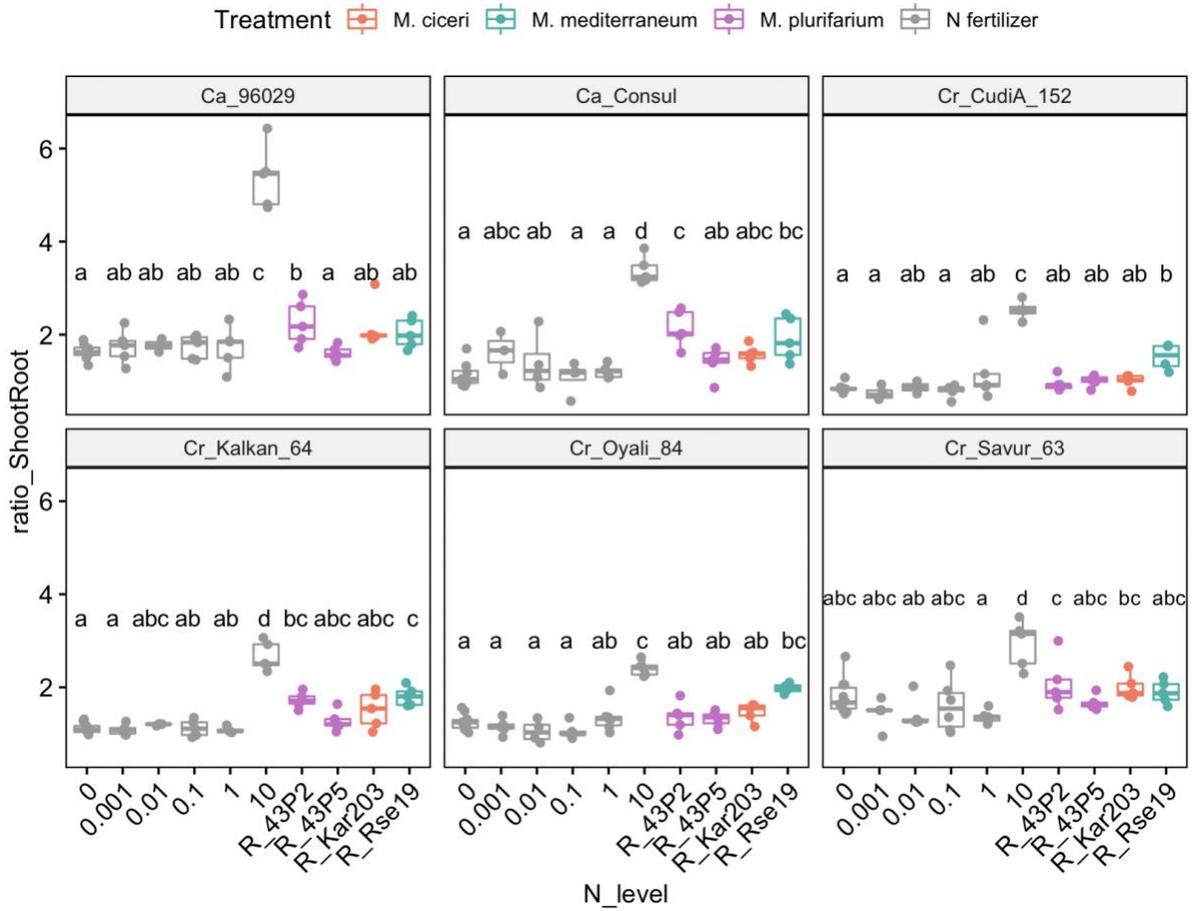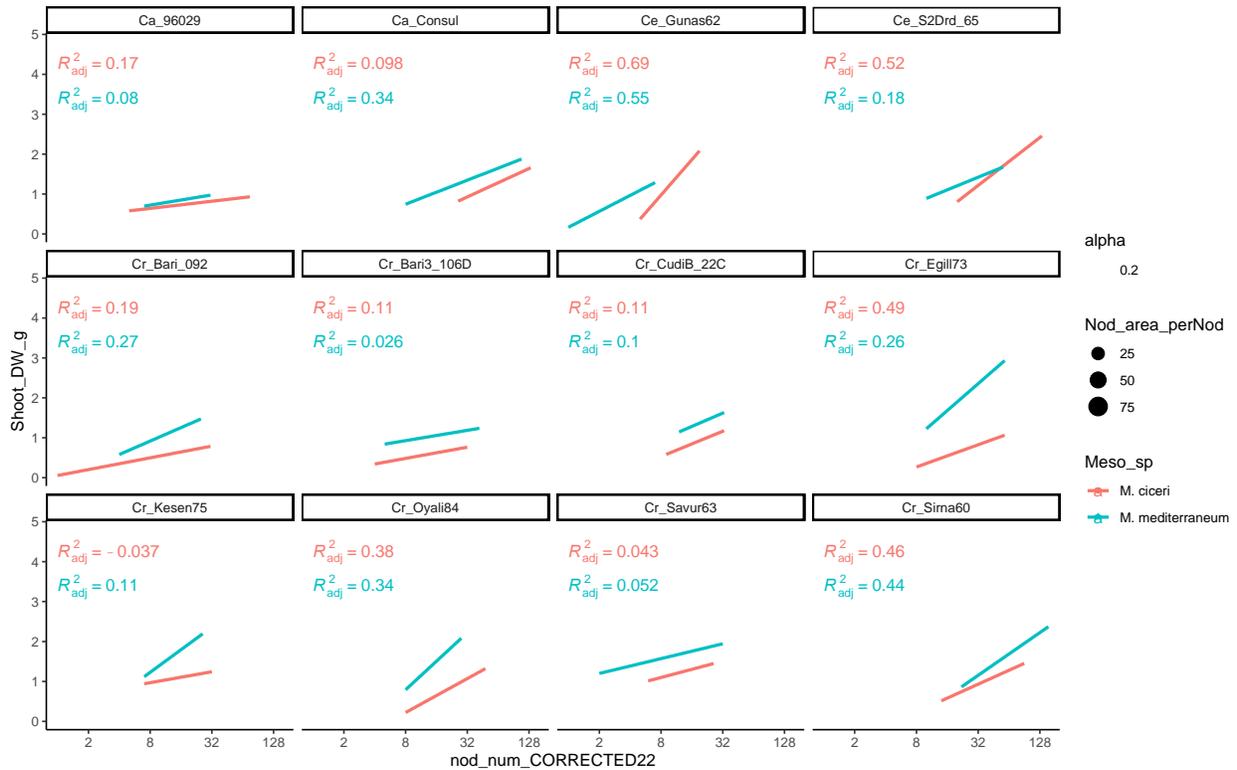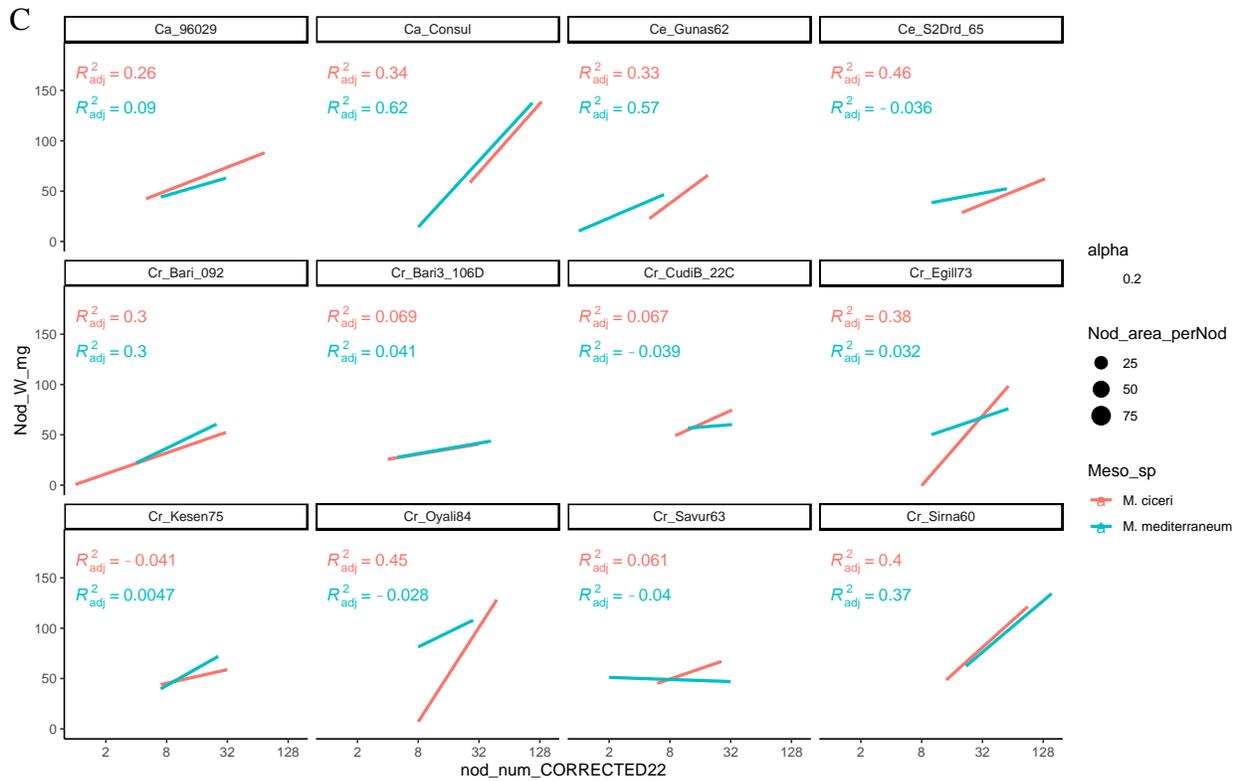
Supplemental Figure 1.
Shoot/root ratio in response to inorganic nitrogen and rhizobia.

A

| Ca_96029	Ca_Consul	Ce_Gunas62	Ce_S2Drd_65
$R^2_{adj} = 0.39$ $R^2_{adj} = 0.37$	$R^2_{adj} = 0.44$ $R^2_{adj} = 0.54$	$R^2_{adj} = 0.54$ $R^2_{adj} = 0.76$	$R^2_{adj} = 0.89$ $R^2_{adj} = 0.52$

Cr_Bari_092	Cr_Bari3_106D	Cr_CudiB_22C	Cr_Egill73
$R^2_{adj} = 0.67$ $R^2_{adj} = 0.56$	$R^2_{adj} = 0.58$ $R^2_{adj} = 0.49$	$R^2_{adj} = 0.31$ $R^2_{adj} = 0.17$	$R^2_{adj} = 0.099$ $R^2_{adj} = 0.64$

Cr_Kesen75	Cr_Oyali84	Cr_Savur63	Cr_Sirna60
$R^2_{adj} = 0.79$ $R^2_{adj} = 0.31$	$R^2_{adj} = 0.94$ $R^2_{adj} = 0.52$	$R^2_{adj} = 0.61$ $R^2_{adj} = 0.29$	$R^2_{adj} = 0.71$ $R^2_{adj} = 0.69$

B

Ca_96029	Ca_Consul	Ce_Gunas62	Ce_S2Drd_65
$R^2_{adj} = 0.17$ $R^2_{adj} = 0.08$	$R^2_{adj} = 0.098$ $R^2_{adj} = 0.34$	$R^2_{adj} = 0.69$ $R^2_{adj} = 0.55$	$R^2_{adj} = 0.52$ $R^2_{adj} = 0.18$

Cr_Bari_092	Cr_Bari3_106D	Cr_CudiB_22C	Cr_Egill73
$R^2_{adj} = 0.19$ $R^2_{adj} = 0.27$	$R^2_{adj} = 0.11$ $R^2_{adj} = 0.026$	$R^2_{adj} = 0.11$ $R^2_{adj} = 0.1$	$R^2_{adj} = 0.49$ $R^2_{adj} = 0.26$

Cr_Kesen75	Cr_Oyali84	Cr_Savur63	Cr_Sirna60
$R^2_{adj} = -0.037$ $R^2_{adj} = 0.11$	$R^2_{adj} = 0.38$ $R^2_{adj} = 0.34$	$R^2_{adj} = 0.043$ $R^2_{adj} = 0.052$	$R^2_{adj} = 0.46$ $R^2_{adj} = 0.44$

Supplemental Figure 2.
A. Shoot dry weight as function of nodule weight. **B.** Shoot dry weight as a function of nodule number. **C.** Nodule weight as a function of nodule number
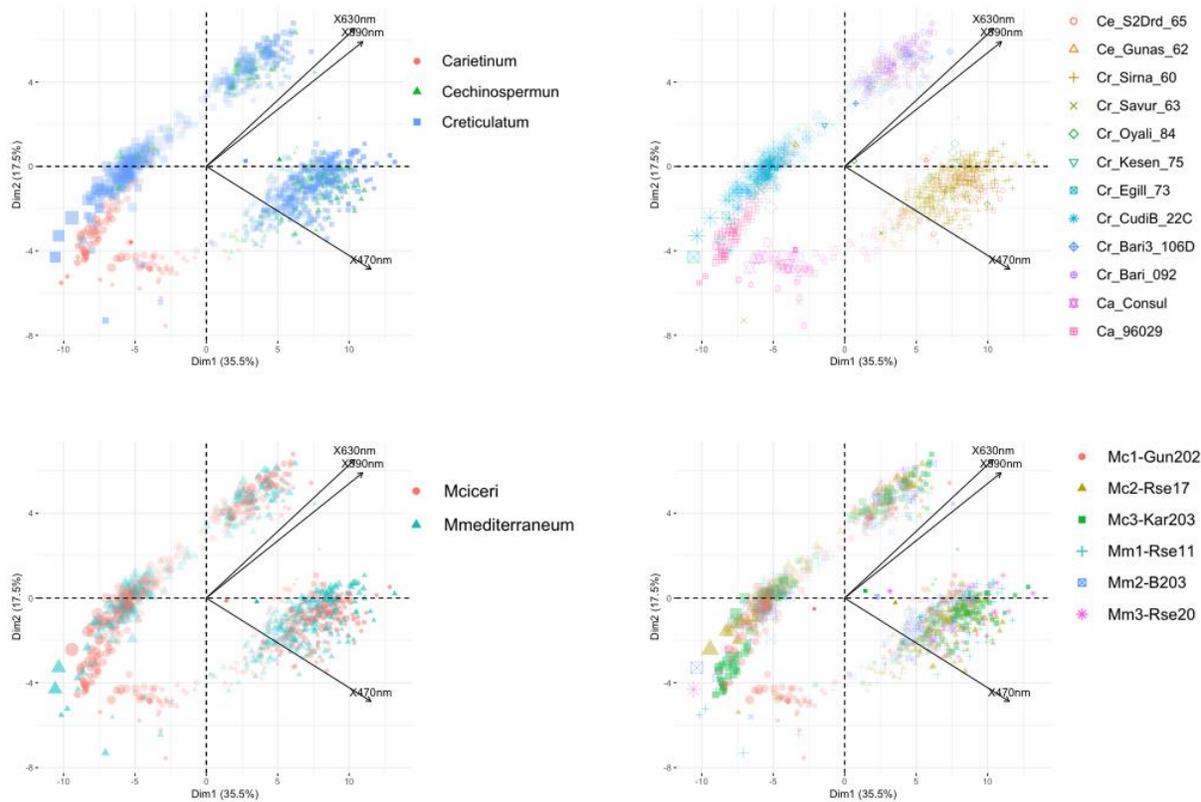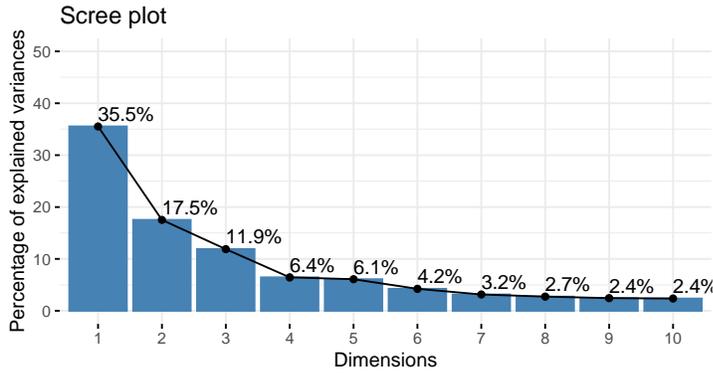
Supplemental Figure 3A.
PCA Biplot for the top 1000 nodules with higher cos2 value (Quality of representation). Size of points represent nodule Area, big pints= 100 mm2, small point until 50 mm2, intensity of color represent the contribution -light =0.05, dark=0.1.
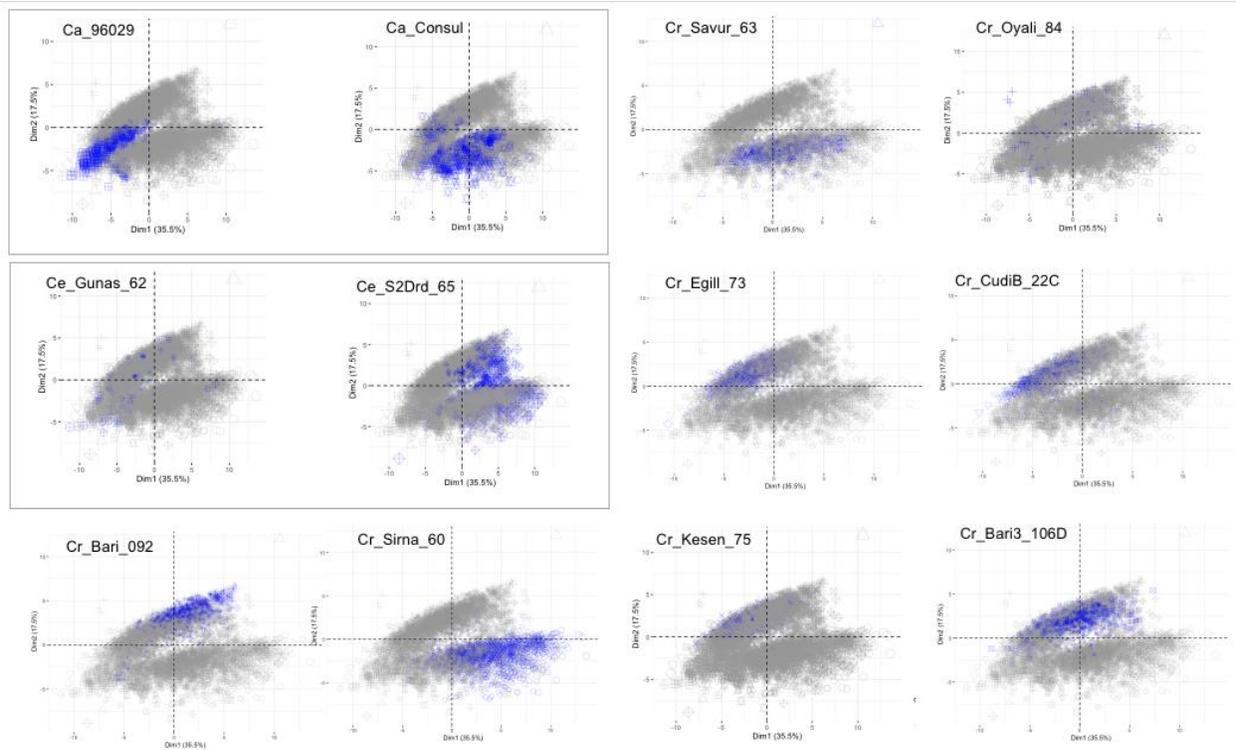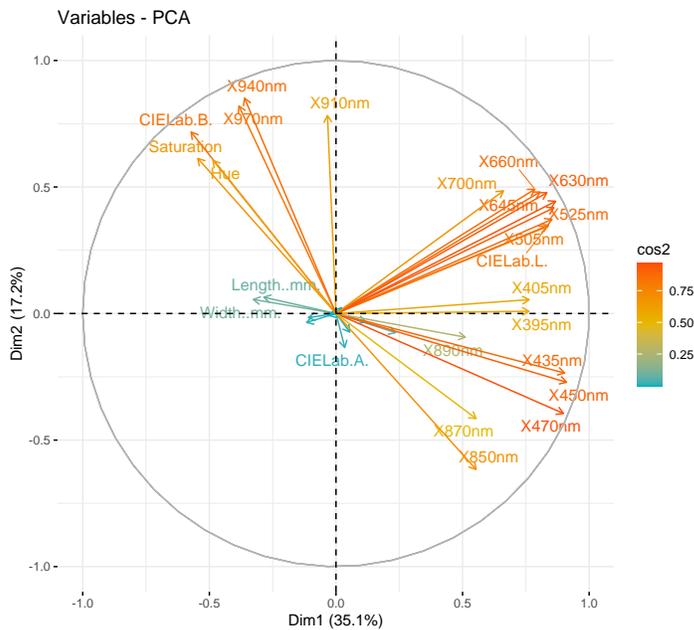
Supplemental Figure 3B.
B. Principal component analysis (PCA) as in Fig. 7C, coloring in blue points from each different *Cicer* genotype.

C

D.

E.
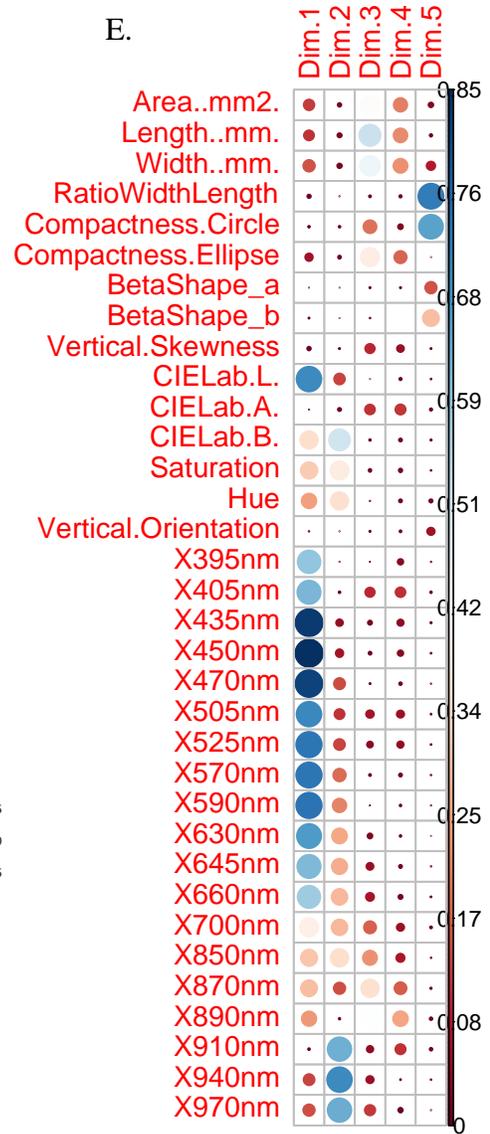
Supplemental Figure 3 CDE.
C. Scree plot of nodule PCA, percentage of variation explained by each dimension. **D.** Variable plots. **E.** Correlation plot of variables and dimensions (for value Cos2 for the variables, 0.05 sig. level)

CHAPTER III

Transcriptome divergence in *Cicer* in response to nitrogen and rhizobia.

ABSTRACT

Biological nitrogen fixation in *Cicer* sp. is an adaptive process that coordinates dual metabolic function in the host and the symbiont. Chickpea wild relatives have strong phenotypes for efficient nitrogen fixation with species-specific interactions. Here we detected common nodule transcriptional signals in efficient host-microbe combination, associated with nitrogen metabolism, while enhanced carbon metabolism characterized less efficient outcomes. Efficiency in N fixation, translated in biomass growth had characteristic transcriptional profile in shoots in overrepresented gene functions in photosynthesis, primary metabolism, secondary metabolism, fatty acid metabolism. The comparative analysis of organs and wild and cultivated host, provide a background to stablish nitrogen nutrition status signals and processes, that could be used as molecular markers for plant status.

INTRODUCTION

Nodule organogenesis is highly specific, mediated by the exchange and recognition of signals between plant host and bacterial symbiont. This cross-kingdom communication is mediated by plant-derived flavonoids, their detection by bacterial *nodD*-family transcription factors, bacterial production of Nod-factor signals, perception of Nod factors by plant receptors and the activation of signaling cascades, the consequent infection of epidermal root hairs via infection threads and the initiation of nodule development from primordia in the root inner cortex. These processes culminate in formation of mature nodules in which highly active carbon and nitrogen metabolism occurs. Development and function of the nodule organ are characterized by transcriptional reprograming that has been extensively studied (hours to days after inoculation) in several legumes, especially in model systems (Moreau et al., 2011, Larrainzar

et al., 2015, Jardinaud & Boivin, 2016). Less is known about how nitrogen status, including variation in the efficiency of nitrogen fixation, impacts plant transcriptional patterns.

During vegetative growth, young and developing plant leaves are a sink for nitrogen (N). Subsequently, during seed development when nitrogen demand is highest, nitrogen can be remobilized from foliar tissue to seed (Xu et al., 2012), although for legumes the most important source of reproductive phase nitrogen is symbiotic nitrogen fixation in root nodules. In addition to being the key source of nitrogen, legume nodules are also a sink tissue for carbon metabolites. Indeed the stoichiometry of C and N is tightly controlled. Photosynthate (in the form of sucrose) is transported to nodules through the phloem, where it is metabolized via glycolysis and the tricarboxylic acid cycle in the plant cytoplasmic compartment. The resulting dicarboxylic acids, succinate and malate, are transported into the bacteroid compartment were further metabolism through the TCA cycle yields the ATP required for conversion of atmospheric dinitrogen to ammonia (NH_4^+) using the nitrogenase enzyme. NH_4 is assimilated in the plant cytoplasmic compartment via the GS/GOGAT cycle into amino acids that are then transported to developing tissues through the xylem to shoots and through the phloem to roots (Bernard & Habash, 2009, McGuiness et al., 2021).

During the transition to flowering and the initiation of pod filling, nitrogen demand increases and plant source sink priorities are modified such that carbon and nitrogen resources are redirected to reproductive tissues. Thus, gene expression related to carbon and nitrogen metabolism is a potential proxy for nitrogen status. This chapter examines transcriptional patterns during this critical period of changing nitrogen demand, analyzing transcripts in young leaves, secondary mature roots, and mature nitrogen-fixing nodules. The studied plant hosts represent an evolutionary gradient in the genus *Cicer* spp., with *C. echinospermun* and *C. reticulatum* being ~100 KYA and ~10KYA distant from cultivated *C. arietinum*. Their transcriptional response is compared under variable inorganic nitrogen status, including deficient (0 mM) and sufficient (10 mM) NH_4NO_3. The quality of nitrogen fixing symbiosis was

75

also varied, using cognate wild *M. mediterraneum* and *M. ciceri* strains that differ in their efficiency depending on host species, as determined in Chapter II of this thesis. *C. echinospermun* is compatible (Nod+) with *M. ciceri* and incompatible (Nod-) with *M. mediterraneum*. By contrast, *C. reticulatum* is compatible with both *Mesorhizobium* species, but fixes N more efficiently with *M. mediterraneum* than with *M. ciceri*.

RESULTS AND DISCUSSION

1. *Cicer* sp. transcriptome distribution under N stress, Inorganic Nitrogen and Symbiosis.

In this study the transcriptome data represent single genotypes of three *Cicer* species, grown following inoculation with evolutionarily diverged symbionts or under fertilization with absent (0mM) or sufficient (10mM) inorganic nitrogen (NH_4NO_3). Aerial (Shoot) transcriptome data are from the youngest fully expanded leaves of an assayed plant. Root transcriptome data derive from secondary roots sampled at 2.5 cM below the shoot/root junction. Nodules correspond to inoculation with symbionts *M. ciceri* or *M. mediterraneum*, with the nodules harvested from the primary root at 2.5 cM below the shoot/root junction. As expected, shoot biomass in *C. arietinum* (Consul) and *C. reticulatum* (Egill) was greatest on inoculation with the cognate *M. mediterraneum* symbiont (Rse19), significantly different from either the zero-nitrogen control or inoculation with *M. cieri* (Rse17) (Fig 1A). Similarly, *C. echinospermun* (Gunas) had greatest shoot biomass growth on inoculation with its cognate symbiont *M. ciceri* (Rse17), significantly different from all other treatments.

Among all genotypes and treatments, the main effect on transcription was at the level of tissue. Thus, the first two axes of an MDS plot partition according to tissue of origin and explain 65% of variation (Fig 1B). The third dimension, which explains 5% variation, partitions according to host species (Figure 1C). The effects of nitrogen treatment were small in this global analysis, which considers all genotypes and treatments together, and only evident in roots and nodules (Fig 1B). Thus, the roots and nodules of plants inoculated with *M. mediterraneum* cluster

76

separately from those receiving inorganic nitrogen treatments or that were inoculated with *M. ciceri*. In an attempt to further resolve nitrogen treatment effects, differential gene expression analysis was conducted separately for individual plant tissues.

2. Nodules

Nodule analyses included 19 samples and identified 20,414 expressed genes. As shown in Fig. 2A, the first dimension (~35% variation) of the MDS plot readily separates according to domestication status, with cultivated separating clearly from wilds. This is interesting because from an evolutionary and genome diversity standpoint cultivated *C. arietinum* (cultivated) and wild *C. reticulatum* are closely related (10 MYA) and equi-distant from wild *C. echinospermum* (100 MYA). Consistent with evolutionary distances, the second dimension (explaining ~23 % variation) resolves all three species, with the greatest distance between the *C. arietinum/C. reticulatum* pair and *C. echinospermum*. Analysis of sample to sample correlation (Fig 2B) confirms separation by host species as the main effect, with the exception of a single *C. echinospermum* sample that is basal to the *C. reticulatum* samples. Interestingly, clustering did not discriminate between *Mesorhizobium* species.

Differential gene expression analysis focused on individual genes, and initial analyses considered similarities and differences in the responses of cultivated *C. arietinum* and wild *C. reticulatum* to *M. ciceri* and *M. mediterraneum*. Ten genes had common patterns of expression in both *C. arietinum* and *C. reticulatum* (Figure 3A) (Supplemental Table 1), all of which were upregulated in nodules inoculated with *M. ciceri* relative to inoculation with *M. mediterraneum*. GO biological process annotations indicate that these *M. ciceri*-upregulated genes have roles in chloroplast function and carbon metabolism: photosynthesis, carbon fixation, photorespiration and the pathway of glyoxylate and dicarboxylate metabolism, the latter of which comprise the proximal source of carbon for bacteroids. These transcriptional patterns suggest increased carbon demand in the *C. arietinum/C. reticulatum* vs *M. ciceri* interaction (Fig 1A), which is

interesting given that *M. ciceri* nodules are less efficient compared to *M. mediterraneum* on these plant hosts.

We also observed numerous genes that were differentially expressed depending on inoculant source, but only in a single plant species (i.e., *C. arietinum* or *C. reticulatum*) (Supplementary table S1). In total, 73 genes (in addition to the 10 common genes, above) were more highly expressed in *M. ciceri* nodules (14 in *C. reticulatum* and 59 in *C. arietinum*). Interestingly, *M. ciceri*-associated transcription in *C. arietinum* and *C. reticulatum* had distinct annotation signatures. In *C. arietinum*, genes more highly expressed in *M. ciceri* nodules were substantially enriched in chloroplast functions, strengthening the conclusion (above) that *M. ciceri* nodules have elevated chloroplast/carbon-related activities. Enriched annotations included GO Cellular Components: photosystem II (score 15.4, ~ 12 genes), "plastid" (21 genes), "thylakoid" (14 genes), and "chloroplasts" (15 genes). Enriched Molecular Function annotations included "metal ion binding" (15 genes), "carbon metabolism pathway" (9 genes), "glyoxylate and dicarboxylate metabolism" (5 genes), and "biosynthesis of secondary metabolites" (8 genes). By contrast, in the case of *C. reticulatum*, genes more highly expressed in *M. ciceri* nodules were not enriched for chloroplast function, but rather contained a preponderance of regulatory proteins involved with transcription, phosphorylation and protein turnover.

Conversely, twenty genes were more highly expressed in *M. mediterraneum* nodules than in *M. ciceri* nodules (7 in *C. reticulatum* and 13 in *C. arietinum*). This set of genes was devoid of chloroplast-related function, but instead populated by transcripts with annotations related to primary and secondary metabolism, transport and development. Some genes are uncharacterized, for example two with MULE transposase domain and zinc finger domain (Gene ID 10585177, and 105852170), domains that have been associated with transposon derived transcription factors (Babu et al., 2006). Examples of annotated genes include Cytochrome P450 oxidoreductase and monooxygenase transcripts, a probable long-chain-alcohol O-fatty-acyltransferase 5 (LOC101503740) relevant in the metabolism of fatty acids, amino acid and other

transporters, molybdenum cofactor, and transferases (glucosyl and acyltransferases) (Supplemental Figure 1). Poulsen & Pødenphant, (2002) observed that Cytochrome P450s were upregulated in *Lotus japonicus* infected with *M. loti*. Here, the chickpea Cytochrome P450s include a homolog of Cytochrome P450 714A1, which is implicated in gibberellin synthesis in Arabidopsis.

Genes with enhanced expression in *M. mediterraneum* nodules also included the CLAVATA3/ESR (CLE)-related protein 4A-2-like (LOC101509380) (Fig 4A). CLEs are plant peptide hormones that activate the autoregulation of nodulation pathway (AON), acting as negative regulators of nodulation (Reid et al., 2011). CLE peptides are post transcriptionally modified, exported to the xylem and transported to aerial tissues (shoot) where they are perceived by the CLAVATA1-like receptor-like kinase. Perception of CLEs by CLAVATA1 leads to an inhibitory shoot-derived signal that is transmitted through the phloem to the root to block subsequent nodule organogenesis (Searle et al., 2003, Reid et al., 2011, Imin et al., 2018, Crook et al., 2016). In the current experiments, CLE-related protein 4A-2-like was exclusively expressed in the nodules of cultivated chickpea (logFC 1.6) (Fig 4B) and higher in *M. mediterraneum* nodules than in *M. ciceri* nodules. It is possible that expression of CLE 4A-2-like is a marker of more efficient symbiosis with *M. mediterraneum* relative to *M. ciceri*. The fact that CLE 4A-2-like was not identified in *C. reticulatum* could indicate differences in the timing of host responses, or it might indicate divergent signaling, between domesticated and wild *Cicer* species.

Nodule differential gene expression was further investigated by including nodule transcriptomes from all three plant species and then identifying differences based on inoculant source. This comparison has two important caveats. Firstly, *C. echinospermun* only forms nodules on inoculation with *M. ciceri*. Secondly, *C. echinospermum* has efficient symbiosis with *M. ciceri*, whereas *C. reticulatum* and *C. arietinum* are less efficient with *M. ciceri* compared to *M. mediterraneum*. With these caveats in mind, the analysis identified 271 genes up-regulated in nodules formed with *M. mediterraneum* and 413 genes up-regulated in nodules formed with *M.*

ciceri (Fig 3B). GO enrichment analysis identified several annotation categories in each of the Molecular Function, Cellular Component and Biological Process ontologies that were enriched depending on the identity of the microbial partner. Thus, transcripts associated with metal ion binding, chlorophyll binding, transcription factor DNA binding, and copper iron binding, were all significantly more abundant in *M. ciceri* nodules (Fig 3C). Enrichment of these Molecular Function categories was mirrored by enrichment of Cellular Component and Biological Process categories that reveal substantial differences in chloroplast-localized and carbon metabolism-related genes, similar to conclusions reached involving only *C. arietinum* and *C. reticulatum* hosts. In contrast to *M. ciceri* nodules, nodules induced by *M. mediterraneum* have patterns of GO enrichment that reflect Molecular Functions related to the cycling of ammonia via amine oxidases, with the implicated enzymes releasing ammonia from three monoamines: phenethylamine, aminoacetone and tryptamine. □-phenethylamine is abundant in certain legume nodules, and, consistent with the current data, varies depending on the identity of the nodulating bacterium (Fujihara et al., 2002). It is tempting to speculate that nitrogen cycling through monoamines provides a transient buffer from otherwise toxic ammonium, which is the immediate product of nitrogen fixation. *M. mediterraneum* nodules were also enriched for the Molecular Function of oxidoreductase. Certain oxidoreductases in soybean root nodules are suggested to have roles in scavenging of H_2O_2 to avoid free radical-induced damage (Dalton et al., 1986). In the Cellular Component ontology, the primary annotation category enriched in *M. mediterraneum* nodules was "integral component of membrane", which agrees with the fact that amine oxidases can be integral membrane proteins (Morris et al., 1997). In any case, the large proportion of membrane proteins is consistent with the complex requirements for transport and metabolism in infected nodule cells, where bacteroids are surrounded by the host symbiosome membrane across which the plant provides ions, metals and carbon compounds to meet the metabolic requirements for bacteroid function (Udvardi & Poole, 2013). The theme of upregulated amine metabolism was also reflected in the Biological Process ontology, where amine metabolic process is one of three equally enriched Biological Process categories.

3. Nodule-specific genes.

Comparisons of nodules vs roots were conducted to identify nodule-specific transcripts considering the four nitrogen treatments. Functionality was inferred by mapping to Gene Ontology terms and to the KEGG (Kyoto encyclopedia of genes and genomes) pathway database. GO and KEGG results were in agreement. Common among all three *Cicer* spp., nodules exhibited substantial elevation of KEGG terms related to "Metabolic Pathways" with 140 genes (~9%) in *C. arietinum* (Fig 5), 130 genes (~10%) in *C. reticulatum* (Fig.6), and 100 genes (~12%) in *C. echinospermun* (Fig.7). The second most enriched category in both roots and nodules of the three species was the "biosynthesis of secondary metabolites" (~5%). In roots of all species, the annotation "plant hormone signal transduction" was enriched (~2% of genes), while in nodules the annotation "starch and sucrose metabolism" was the most enriched category (Supplemental Figure 2 and Supplemental Figure 3).

The data also revealed differences among species. Thus, the pathway of cysteine and methionine metabolism was differentially enriched in *C. echinospermun* roots but not in nodules, while in *C. reticulatum* and in *C. arietinum* the same pathway was upregulated in nodules but not in roots. Interestingly, ATP-binding cassette (ABC) transporters were identified in the gene set list from roots and nodules of cultivated *C. arietinum* (Fig. 5). ABC transporters utilize the energy of ATP to move a variety of specific molecules across biological membranes, and several ABC transporters are required for legume symbioses (Roy et al., 2021). Here, ABC transporters include those implicates in the transport the dicarboxylic acids across the symbiosome membrane.

Similar to conclusions reached when comparing symbiont species, *C. echinospermun* nodules, which were formed exclusively on interaction with *M. ciceri*, were highly enriched relative to roots for the GO Biological Process of carbohydrate metabolism (3.7%, 29 genes). Roots of both *C. echinospermum* and *C. reticulatum* were most enriched for the Biological Process of auxin–activated signaling pathway, while *C. reticulatum* roots were also enriched for annotations

involving amino acid transport, followed by response to oxidative stress, hydrogen peroxide catabolic process, and ethylene-activated signaling pathway.

4. Control (N. deficiency vs treatments) in roots and shoots

To compare the transcriptional profile between nitrogen deficient (control) and nitrogen treatments (inorganic nitrogen or inoculation with *Mesorhizobium*), a contrast matrix of control vs nitrogen treatments was evaluated for roots and shoots independently.

In the roots of wild *C. echinospermun* the data revealed three differentially expressed genes not found in *C. reticulatum* or *C. arietinum* (Fig 8A). Two of these genes encode isoforms of polygalacturonases (At1g48100-like), implicated in cell wall modification that may be relevant to cell growth and cell-wall remodeling (Babu & Bayer, 2014). The third *C. echinospermun* gene encodes a cysteine-rich receptor-like protein kinase 25 (CRK) (LOC113785155) (Fig 9). CRKs are a large subfamily of receptor-like kinases (RLK), including those that regulate disease resistance (biotic stress) and cell death in aerial tissues (Yadeta et al., 2017, Quezada et al., 2019) and whose regulation may involve hormonal crosstalk (Wrzaczek et al., 2010). In *C. echinospermun*, CRK LOC113785155 is induced in nitrogen deficient roots compared to nitrogen treatments (nodules or NH_4NO_3) (Fig 9B), suggesting that LOC113785155 may be a marker of nitrogen starvation. In N starved roots of both *C. arietinum* and *C. reticulatum* 45% and 37% of genes, respectively, are annotated as integral components of membranes (Supplementary table S6). Nitrogen starved *C. reticulatum* roots were also enriched for the KEEG pathway of sulfur metabolism, as well as amino acid and purine transporters (Supplementary table S6). Of uncertain relevance, *C. arietinum* roots under rhizobial and N treatment had an abundance of transcripts with the INTERPRO category "transposases" (15%).

Comparing nitrogen starvation vs inorganic or symbiotic treatments in shoots, *C. echinospermun* had fewer differentially expressed genes (24) compared with *C. reticulatum* (381 genes) and *C. arietinum* (168 DE genes). *C. arietinum* shoots were enriched for RNA polymerase pathway, as well as the Molecular Functions oxidoreductase, monooxygenase, and heme binding

among others. Shoot gene expression in response to nitrogen status was largely similar between *C. reticulatum* and *C. arietinum,* although certain Molecular Functions were uniquely observed in *C. reticulatum.* The most abundant Molecular Function in *C. reticulatum* shoots under N starvation is the MAPK signaling pathway (Supplementary table S7), in addition to annotations of transferase, serine/threonine protein kinases, and thiol proteases.

5. Pairwise treatment contrast in roots and shoots

Within each species, a pairwise treatment contrast analysis was performed for roots and shoots. The analysis establishes the status of aerial and root tissues on a per species basis and under two contrasting treatments: (1) nitrogen starvation vs ammonium nitrate, and (2) inoculation with *M. ciceri* vs *M. mediterraneum* (Fig. 11). In wild roots (i.e., either *C. reticulatum* or *C. echinospermum*) there were few differentially expressed genes expressed in the pairwise comparisons (Fig 11). By contrast, shoot transcriptome responses were large and may indicate systemic signaling and metabolic status.

In *C. echinospermum* the largest set of shoot differentially expressed genes (1,638 genes) was observed between plants inoculated with *M. ciceri* (containing nodules) and those inoculated with *M. mediterraneum* (lacking nodules). Pathway enrichment analysis of *C. echinospermun* leaves on *M. ciceri*-inoculated plants reveals upregulation of carbon metabolism, biosynthesis of amino acids, photosynthesis, starch and sucrose metabolism, and nitrogen metabolism (Fig. 12). Within this gene set are eight ferredoxins (Fig 13). Ferredoxins are iron–sulfur (Fe–S)-cluster-containing enzymes involved in electron transfer reactions in chloroplasts, where they are the principal sink for electrons in the production of NADPH (Hanke & Mulo, 2013). In rice the gene *ABNORMAL CYTOKININ RESPONSE 1*(*ABC1*) (ferredoxin-dependent (Fd)-GOGAT) plays a critical role in nitrogen assimilation and carbon–nitrogen balance (He et al., 2020).

Interestingly, the incompatible interaction (inoculation, but no nodules) between *M. mediterraneum* and *C. echinospermun* appears to generate a systemic plant response that is

greater in magnitude that simple nitrogen deficiency. The metabolic processes overrepresented in *C. echinospermun* during incompatible symbiosis with *M. mediterraneum* include cyanoamino acid metabolism, galactose metabolism, phenylpropanoid biosynthesis and zeatin biosynthesis (Fig 12A). In soybean (Agtuca et al., 20220), zeatin biosynthesis (cytokinin) was also differentially enriched (based on direct analysis of metabolites) in functional nodules compared with nodules formed by rhizobia unable to fix nitrogen.

By contrast, in the case of *C. reticulatum* shoots the largest set of differentially expressed genes (1,476 genes) was observed in the comparison of nitrogen deficient (0 mM N) vs inorganic nitrogen (10 mM NH_4NO_3)-treated plants (Fig. 14). Under conditions of non-limiting nitrogen, the leaves of *C. reticulatum* were enriched in transcripts annotated as Cellular Compartment "ribosome" (12 % of differentially expressed genes) and Molecular Function "photosynthesis", "plant hormone signal transduction", "fatty acids metabolism" and "biosynthesis of unsaturated fatty acids" (Fig. 12). *C. reticulatum* nitrogen starved leaves were enriched for transcripts characterized as "plant pathogen interaction", "flavonoid biosynthesis", "isoflavonoid biosynthesis", "ubiquitin mediated proteolysis" and "endocytosis" (Fig. 14). Interestingly, in legumes (iso)flavonoids have been recognized for their role as root signaling molecules, including specificity factors in host-microbe communication in the rhizosphere (Liu & Murray, 2016). Flavonoid metabolite levels are typically enhanced in nodules and in rhizobia-inoculated roots based on comparison to non-inoculated roots (Gifford et al., 2018). By contrast, in foliar tissues, flavonoid levels are often associated with biotic and abiotic stress.

Although small in comparison to shoots, *C. reticulatum* roots did exhibit differential response to inorganic nitrogen. We detected four up-regulated genes, including uncharacterized LOC101490725 and a cationic peroxidase 1-like gene (101492294, logFC = 5.3). In addition to possibly removing H_2O_2 under stressful conditions, cationic peroxidases can catalyze the degradation of indole-3-acetic acid (IAA) (Grambow & Langenbeck-Schwich, 1983), and certain peroxidases have a role in roots lignin polymerization (Marjamaa et al., 2009). Two additional

genes (101490986 and 101491298, logFC -2.5) were annotated as linoleate 9S-lipoxygenase-like enzymes that perform the oxygenation of polyunsaturated fatty acids, the initial step in biosynthesis of oxylipins such as jasmonates. JA is a phytohormone and signaling molecule, and in aerial tissues JA is associated with stress responses, with elicitation by phytopathogens and wounding (Bostock et al., 2014). In roots JA has several roles in root-biotic interactions (Gutjahr & Paszkowski, 2009). Exogenous JA can suppress nodule formation in a dose-dependent manner (Sun et al., 2006), while JA can also stimulate the expression of Nod factors in rhizobia, leading to enhanced nodulation (Ferguson & Mathesius, 2014).

CONCLUSION

The capacity to form efficient nodule symbiosis in legumes is an adaptive trait. In environments with poor soil nitrogen, symbiotic plants have a competitive advantage for survival and reproduction. As demonstrated in Chapter II of this thesis, symbiotic function can be greater with native, co-evolved combinations of plants and their rhizobia. Thus wild *Cicer* species vary in the efficiency of biological nitrogen fixation depending on the available species of *Mesorhizobium*. The current dataset examined gene transcription across an evolutionary gradient of plants and microbes, comparing *C. echinospermun*, *C. reticulatum* and *C. arietinum* hosts treated with native and non-native *Mesorhizobium* strains or with and without sufficient inorganic nitrogen. Analysis of the bulk transcriptome of nodules, roots and leaves revealed responses to nitrogen that are both common and different among plant species.

A majority of previous transcriptional studies focused on gene expression during early nodule development, when massive transcriptional reprograming leads to nodule organogenesis, and responses occur as early as minutes to hours after inoculation (Larrainzar et al., 2015, Limpens et al., 2013). Transcriptional reprograming upon activation of the symbiotic pathway is a conserved feature of root symbiosis (Hocher et al., 2011). Studies in model legumes have identified tissue- and cell-specific transcriptional responses beginning with Nod factor

perception, progressing through nodule development and function, and ending with nodule senescence (Mergaert et al., 2020, Puppo et al., 2005, Van de Velde et al., 2006). Transcript and metabolite analysis reveal numerous nodule-specific genes that are expressed during nodule development (e.g., (Jardinaud et al., 2022, Mergaert et al., 2020) and in mature nodule cells of the fixation zone (Agtuca et al., 2020, Jardinaud & Boivin, 2016). The current study extends these previous analyses by focusing on transcriptional patterns during the later stage of nodule development, in particular during plant reproduction which is the period of greatest nitrogen demand for seed filling (Voisin et al., 2003). Plant growth is the outcome of a tight association between nitrogen metabolism and carbon metabolism, with each organ and developmental stage varying characteristically in their source-sink relationships (Voisin et al., 2003, Udvardi & Poole, 2013, Larrainzar et al., 2014, Copeland et al., 1995).

C. reticulatum and *C. arietinum*, two closely-related plant hosts with compatible nodulation with *M. ciceri* and *M. mediterraneum,* have commonalities in their symbiont-associated transcription. *M. ciceri* nodules have elevated plastid/carbon-related activities, while *M. mediterraneum* nodules have enhanced expression of regulatory proteins (transcription, phosphorylation and protein turnover), membrane localized genes, as well as upregulation of genes for amine metabolism (nitrogen cycling). Interestingly, comparisons involving *C. echinospermum* and *M. ciceri* also reveal a substantial enrichment of chloroplast-localized and carbon metabolism-related genes. Thus, enhanced transcription of genes related to carbon metabolism and plastid-related functions is characteristic of *M. ciceri* nodules, irrespective of the plant host species.

Nitrogen nutrition is a key factor in plant growth regulation. One might expect that the different nitrogen fixation phenotypes observed in *Cicer* species (Chapter II) would be correlated with whole-plant level C/N balance. Indeed, the current data reveal enhanced expression of nitrogen metabolism genes in nodules formed by the efficient *M. mediterraneum* symbiont on *C. arietinum* and *C. reticulatum*. Similarly, transcriptional response in leaves reveal increased

expression of carbon metabolism genes under conditions of sufficient nitrogen, including expression of transcripts for GS/GOGAT reactions that link C/N metabolism (Hodges, 2002). For example, the leaf transcriptome of *C. echinospermun* plants inoculated with efficient *M. ciceri* was substantially enriched for KEGG pathways related to photosynthesis, carbon metabolism, nitrogen metabolism, and secondary metabolism (Fig 7 and Fig. 13).

The uptake of nutrients based on N status is feedback regulated. Under high N status the influx of NO_3 is downregulated by controlling transporter activity (O'Brien et al., 2016). Nutrient status is also a signal that can determine plant growth processes, such as root system architecture, leaf development, seed dormancy, and flowering time, and thus responses to nitrogen nutrition can overlap with hormone signaling (Vidal et al., 2020, O'Brien et al., 2016, Coruzzi & Bush, 2001). In Arabidopsis roots nitrate (NO_3) acts as a tissue-specific signaling molecule, with pericycle NO_3 regulating auxin signaling that in turn modulates lateral rot formation (Vidal et al., 2020). Here we detected nitrogen-related variation in several hormone pathways, including auxin synthesis, JA signaling, and cytokinin signaling in both shoots and roots.

Our annotation-based analysis of transcriptome data is not without limitations. Firstly, function is inferred entirely based on the logic of gene expression and homology-determined function, both of which are indirect criteria. Moreover, the relevance of transcripts without annotations cannot be evaluated. Nevertheless, the fact that many common gene expression patterns were observed among plant species, or in response to related treatments within plant species, adds credence to the relevance of the recorded patterns. Future studies might test gene expression under additional conditions to strengthen correlations with symbiosis and nitrogen status, or reverse genetic tools such as CRISPR could be used as a genetic test of functional relevance.

In addition to inferring functional responses to nitrogen treatments, this transcript dataset has been instrumental to develop improved annotation of PacBio HiFi genomes for each of the three *Cicer* species analyzed here. The data has proven especially useful for the identification and annotation of a family of ~300 nodule cysteine rich (NCR) peptides, which are short and below the length thresholds of many ab initio gene prediction tools. NCR peptides can modulate nodule development in a species-specific manner (Montiel et al., 2016) and are critical in bacteroid differentiation (Van de Velde et al., 2010). NCR's genes are diverse and vary in numbers depending on the plant species. >600 NCR genes are present in the *Medicago truncatula* genome, most with nodule-specific patterns of expression. In chickpea (cultivated *Cicer*), previous automated gene prediction efforts identified only 63 genes in *C. arietinum*. With the aid of this transcriptome data set we identified at least 125 NCR peptides, and >300 across all three *Cicer* species. This data opens the way to explore the role of these signaling peptides in *Cicer* symbioses, including the regulation of nodule development and nitrogen fixation. This data set may also useful in genetic studies of nitrogen fixation specificity, for example genetic mapping that was initiated in Chapter II and the selection of candidate genes in future QTL analysis studies.

MATERIALS AND METHODS

Biological material and experimental design.

A single genotype of each *Cicer* species [*Cicer arietinum* (CDC Consul), *C. reticulatum* (Egill_73), and *C. echinospermum* (Gunas_62)] was selected for mRNA analysis. Each genotype received four independent treatments consisting either of bacterial species (*M. mediterraneum*-Rse19 or *M. ciceri*-Rse17) or inorganic nitrogen (10 mM and 0 mM NH_4NO_3). Plants were grown individually in 3 gallon pots containing sterilized sand under greenhouse conditions, with nutrition maintained using nitrogen-free fertilizer as detailed in Chapter II. Each treatment was replicated 10 times. Plants were harvested at mid-pod set, when nitrogen demand is known to be high.

Tissue collection was performed between 8 and 11 am, with tissue immediately frozen in liquid nitrogen and stored at -80 °C until extraction. Shoot tissue was sampled by selecting 3 new but fully expanded leaves per plant. Fully intact roots and nodules were harvested simultaneously, with pots immersed in buckets filled with water to gently release the roots from the sand substrate. Secondary root tissue was harvested at 2.5 cm below the shoot/root junction and collected into 5 ml Eppendorf tubes. Nodules were collected from the same root location described for sampling of secondary roots.

Note: Air quality during 2-3 weeks before harvesting was poor due to California wildfires (Aug 2018) (Carr fire, Mendocino fire).

Sample processing and RNA extraction

Frozen tissue was finely ground in liquid nitrogen using sterile mortars and pestles. RNA extraction was performed using the Invitrogen TRIzol kit, according manufacturer's instructions. Isolated RNA was suspended in sterile, RNAse-free water.

RNA concentration was estimated using both a Nanodrop UV spectrophotometer and also by fluorometry using a Qubit fluorometer. The RNA integrity number (RIN), which is a measure of RNA quality, was determined using an Agilent Bioanalyzer. mRNA libraries were constructed by contract with Novogene Inc. Sequence data was obtained as 150 bp paired-end reads using the

Illumina platform PE150, targeting 20M reads per sample. Raw reads were processed using QC Trimmomatic 3.8.

Analysis

Each set of reads was mapped against the reference genome of *C. arietinum* (GCF_000331145.1_ASM33114v1_genomic). Mapping was performed with HiSat2-index-align-2.1 with parameters Trim from 5' end of read 7, minimum intron length 20, and maximum intron length 500,000. The alignment rate range is given in Table 1.

Read quantification. Gene-based read counting

The output mapped BAM files were read in Rstudio 2021.09.2 using feature counts, not allowing multiple overlap, and counting multimapping reads. The EdgeR-limma-voom protocol for RNAseq analysis was followed (Law et al., 2014). Normalization factors consider library size as a scaling factor per library, with mean library size of 23.5 million reads. The total mappable genes were 26,661. Genes with >10 read counts in a treatment were considered for further analysis. Normalization was according to the trimmed mean of M-values (TMM) method, implemented in edgeR. Differential gene expression was determined for multiple comparisons across tissues and treatments using "Decide test" per contrast, and using adjust.method = "BH" (p.value = 0.05) with a threshold log-fold change of 1. Exceptionally, comparisons of root and nodule data among species used a more conservative p-value of 0.01 and a threshold log-fold change of 3.

Differentially expressed genes enrichment analysis.

The resulting gene lists, which represent the set of significantly up and down regulated per contrasting condition (entrezID), were used for GeneOntology (GO) enrichment analysis using the online tool DAVID (https://david.ncifcrf.gov). DAVID implements functional annotation clustering that reduces redundancy, and a grouping algorithm that assumes that similar annotations have similar gene members.

REFERENCES

Agtuca BJ, Stopka SA, Evans S, *et al.*, 2020. Metabolomic profiling of wild-type and mutant soybean root nodules using laser-ablation electrospray ionization mass spectrometry reveals altered metabolism. *Plant J* **103**, 1937-58.

Babu MM, Iyer LM, Balaji S, Aravind L, 2006. The natural history of the WRKY–GCM1 zinc fingers and the relationship between transcription factors and transposons. *Nucleic Acids Research* **34**, 6505-20.

Babu Y, Bayer M, 2014. Plant Polygalacturonases Involved in Cell Elongation and Separation-The Same but Different? *Plants (Basel, Switzerland)* **3**, 613-23.

Bernard SM, Habash DZ, 2009. The importance of cytosolic glutamine synthetase in nitrogen assimilation and recycling. *New Phytol* **182**, 608-20.

Bostock RM, Pye MF, Roubtsova TV, 2014. Predisposition in plant disease: exploiting the nexus in abiotic and biotic stress perception and response. *Annu Rev Phytopathol* **52**, 517-49.

Copeland L, Lee HS, Cowlishaw N, 1995. Carbon metabolism in chickpea nodules. *Soil Biology and Biochemistry* **27**, 381-5.

Coruzzi G, Bush DR, 2001. Nitrogen and carbon nutrient and metabolite signaling in plants. *Plant Physiol* **125**, 61-4.

Crook AD, Schnabel EL, Frugoli JA, 2016. The systemic nodule number regulation kinase SUNN in Medicago truncatula interacts with MtCLV2 and MtCRN. *Plant J* **88**, 108-19.

Dalton DA, Russell SA, Hanus FJ, Pascoe GA, Evans HJ, 1986. Enzymatic reactions of ascorbate and glutathione that prevent peroxide damage in soybean root nodules. *Proc Natl Acad Sci U S A* **83**, 3811-5.

Ferguson BJ, Mathesius U, 2014. Phytohormone Regulation of Legume-Rhizobia Interactions. *Journal of Chemical Ecology* **40**, 770-90.

Fujihara S, Terakado J, Takenaka M, Yoneyama T, 2002. Specific occurrence of β-phenethylamine in root nodules formed from Bradyrhizobium-legume symbiosis. *Plant and Soil* **238**, 123-32.

Gifford I, Battenberg K, Vaniya A, *et al.*, 2018. Distinctive Patterns of Flavonoid Biosynthesis in Roots and Nodules of Datisca glomerata and Medicago spp. Revealed by Metabolomic and Gene Expression Profiles. *Front Plant Sci* **9**, 1463.

Grambow HJ, Langenbeck-Schwich B, 1983. The relationship between oxidase activity, peroxidase activity, hydrogen peroxide, and phenolic compounds in the degradation of indole-3-acetic acid in vitro. *Planta* **157**, 132-7.

Gutjahr C, Paszkowski U, 2009. Weights in the balance: jasmonic acid and salicylic acid signaling in root-biotroph interactions. *Mol Plant Microbe Interact* **22**, 763-72.

Hanke GUY, Mulo P, 2013. Plant type ferredoxins and ferredoxin-dependent metabolism. *Plant, Cell & Environment* **36**, 1071-84.

He L, Li M, Qiu Z, *et al.*, 2020. Primary leaf-type ferredoxin 1 participates in photosynthetic electron transport and carbon assimilation in rice. *The Plant Journal* **104**, 44-58.

Hocher V, Alloisio N, Auguy F, *et al.*, 2011. Transcriptomics of actinorhizal symbioses reveals homologs of the whole common symbiotic signaling cascade. *Plant Physiol* **156**, 700-11.

Hodges M, 2002. Enzyme redundancy and the importance of 2-oxoglutarate in plant ammonium assimilation. *J Exp Bot* **53**, 905-16.

Imin N, Patel N, Corcilius L, Payne RJ, Djordjevic MA, 2018. CLE peptide tri-arabinosylation and peptide domain sequence composition are essential for SUNN-dependent autoregulation of nodulation in Medicago truncatula. *New Phytol* **218**, 73-80.

Jardinaud M-F, Carrere S, Gourion B, Gamas P, 2022. Symbiotic Nodule Development and Efficiency in the Medicago truncatula Mtefd-1 Mutant Is Highly Dependent on Sinorhizobium Strains. *Plant and Cell Physiology*, pcac134.

Jardinaud MF, Boivin S, 2016. A Laser Dissection-RNAseq Analysis Highlights the Activation of Cytokinin Pathways by Nod Factors in the Medicago truncatula Root Epidermis. **171**, 2256-76.

Larrainzar E, Gil-Quintana E, Seminario A, Arrese-Igor C, González EM, 2014. Nodule carbohydrate catabolism is enhanced in the Medicago truncatula A17-Sinorhizobium medicae WSM419 symbiosis. *Frontiers in Microbiology* **5**.
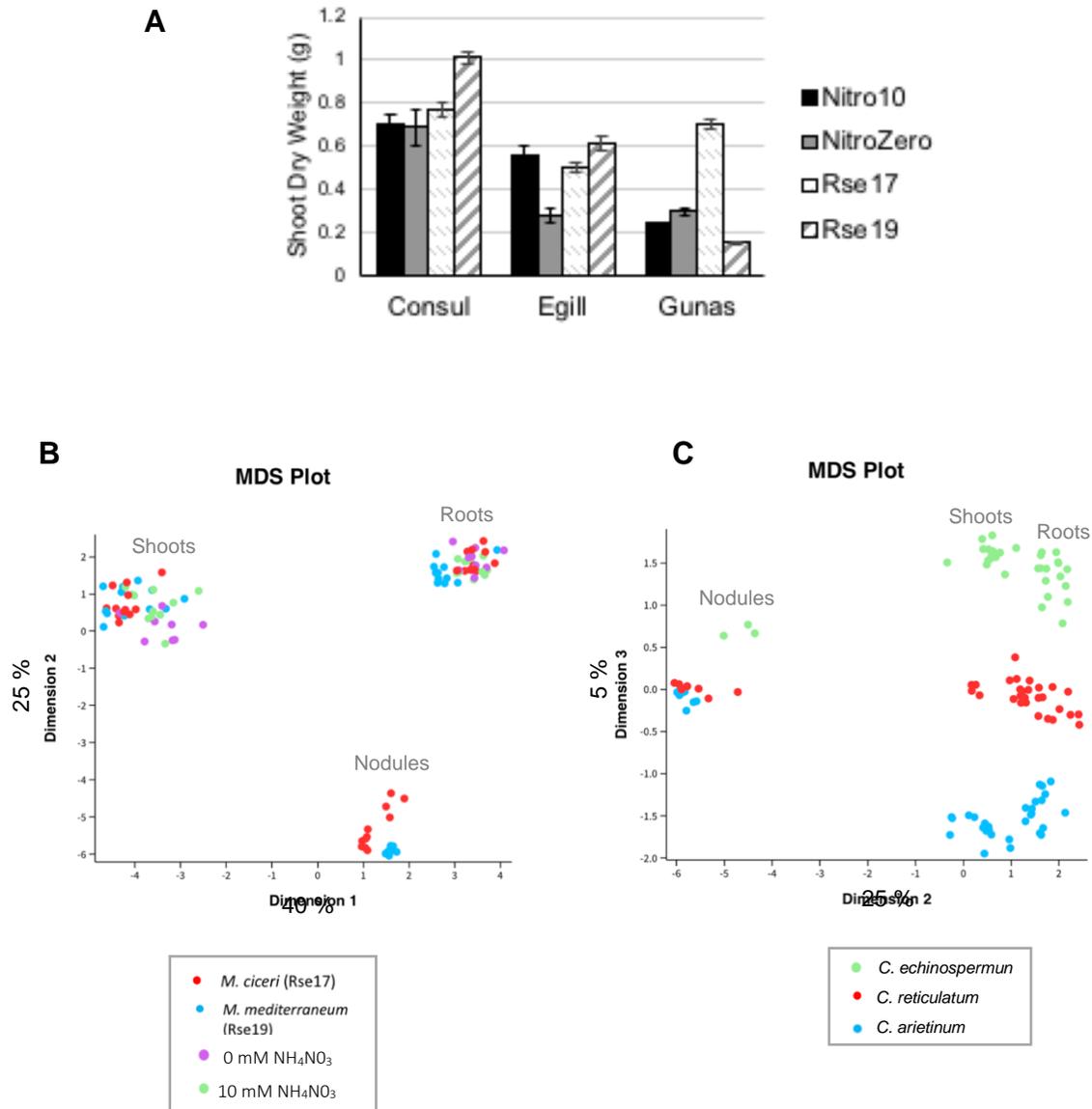
Larrainzar E, Riely BK, Kim SC, *et al.*, 2015. Deep Sequencing of the Medicago truncatula Root Transcriptome Reveals a Massive and Early Interaction between Nodulation Factor and Ethylene Signals. *Plant Physiol* **169**, 233-65.

Law CW, Chen Y, Shi W, Smyth GK, 2014. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**.

Limpens E, Moling S, Hooiveld G, *et al.*, 2013. Cell- and tissue-specific transcriptome analyses of Medicago truncatula root nodules. *PLoS One* **8**.

Liu CW, Murray JD, 2016. The Role of Flavonoids in Nodulation Host-Range Specificity: An Update. *Plants (Basel, Switzerland)* **5**.

Marjamaa K, Kukkola EM, Fagerstedt KV, 2009. The role of xylem class III peroxidases in lignification. *J Exp Bot* **60**, 367-76.

Mcguiness PN, Reid JB, Foo E, 2021. The influence of ethylene, gibberellins and brassinosteroids on energy and nitrogen-fixation metabolites in nodule tissue. *Plant Sci* **305**, 110846.

Mergaert P, Kereszt A, Kondorosi E, 2020. Gene Expression in Nitrogen-Fixing Symbiotic Nodule Cells in Medicago truncatula and Other Nodulating Plants. *Plant Cell* **32**, 42-68.

Montiel J, Szucs A, Boboescu IZ, Gherman VD, Kondorosi E, Kereszt A, 2016. Terminal Bacteroid Differentiation Is Associated With Variable Morphological Changes in Legume Species Belonging to the Inverted Repeat-Lacking Clade. *Mol Plant Microbe Interact* **29**, 210-9.

Moreau S, Verdenaud M, Ott T, *et al.*, 2011. Transcriptional reprogramming during root nodule development in Medicago truncatula. *PLoS One* **6**.

Morris NJ, Ducret A, Aebersold R, Ross SA, Keller SR, Lienhard GE, 1997. Membrane amine oxidase cloning and identification as a major protein in the adipocyte plasma membrane. *J Biol Chem* **272**, 9388-92.

O'brien JA, Vega A, Bouguyon E, *et al.*, 2016. Nitrate Transport, Sensing, and Responses in Plants. *Mol Plant* **9**, 837-56.

Puppo A, Groten K, Bastian F, *et al.*, 2005. Legume nodule senescence: roles for redox and hormone signalling in the orchestration of the natural aging process. *New Phytologist* **165**, 683-701.

Quezada EH, García GX, Arthikala MK, Melappa G, Lara M, Nanjareddy K, 2019. Cysteine-Rich Receptor-Like Kinase Gene Family Identification in the Phaseolus Genome and Comparative Analysis of Their Expression Profiles Specific to Mycorrhizal and Rhizobial Symbiosis. *Genes (Basel)* **10**.

Reid DE, Ferguson BJ, Hayashi S, Lin Y-H, Gresshoff PM, 2011. Molecular mechanisms controlling legume autoregulation of nodulation. *Ann Bot* **108**, 789-95.

Roy S, Breakspear A, Cousins D, *et al.*, 2021. Three Common Symbiotic ABC Subfamily B Transporters in Medicago truncatula Are Regulated by a NIN-Independent Branch of the Symbiosis Signaling Pathway. *Molecular Plant-Microbe Interactions®* **34**, 939-51.

Searle IR, Men AE, Laniya TS, *et al.*, 2003. Long-Distance Signaling in Nodulation Directed by a CLAVATA1-Like Receptor Kinase. *Science* **299**, 109-12.

Sun J, Cardoza V, Mitchell DM, Bright L, Oldroyd G, Harris JM, 2006. Crosstalk between jasmonic acid, ethylene and Nod factor signaling allows integration of diverse inputs for regulation of nodulation. *The Plant Journal* **46**, 961-70.

Udvardi M, Poole PS, 2013. Transport and metabolism in legume-rhizobia symbioses. *Annu Rev Plant Biol* **64**, 781-805.

Van De Velde W, Pérez-Guerra JC, De Keyser A, *et al.*, 2006. Aging in legume symbiosis: a molecular view on nodule senescence in Medicago truncatula. *Plant Physiol* **141**.

Van De Velde W, Zehirov G, Szatmari A, *et al.*, 2010. Plant Peptides Govern Terminal Differentiation of Bacteria in Symbiosis. *Science* **327**, 1122-6.

Vidal EA, Alvarez JM, Araus V, *et al.*, 2020. Nitrate in 2020: Thirty Years from Transport to Signaling Networks. *Plant Cell* **32**, 2094-119.

Voisin AS, Salon C, Jeudy C, Warembourg FR, 2003. Root and Nodule Growth in Pisum sativum L. in Relation to Photosynthesis: Analysis Using (13)C-labelling. *Ann Bot* **92**, 557-63.

Wrzaczek M, Brosché M, Salojärvi J, *et al.*, 2010. Transcriptional regulation of the CRK/DUF26 group of receptor-like protein kinases by ozone and plant hormones in Arabidopsis. *BMC Plant Biol* **10**, 95.

Xu G, Fan X, Miller AJ, 2012. Plant nitrogen assimilation and use efficiency. *Annu Rev Plant Biol* **63**, 153-82.

Yadeta KA, Elmore JM, Creer AY, *et al.*, 2017. A Cysteine-Rich Protein Kinase Associates with a Membrane Immune Complex and the Cysteine Residues Are Required for Cell Death. *Plant Physiol* **173**, 771-87.
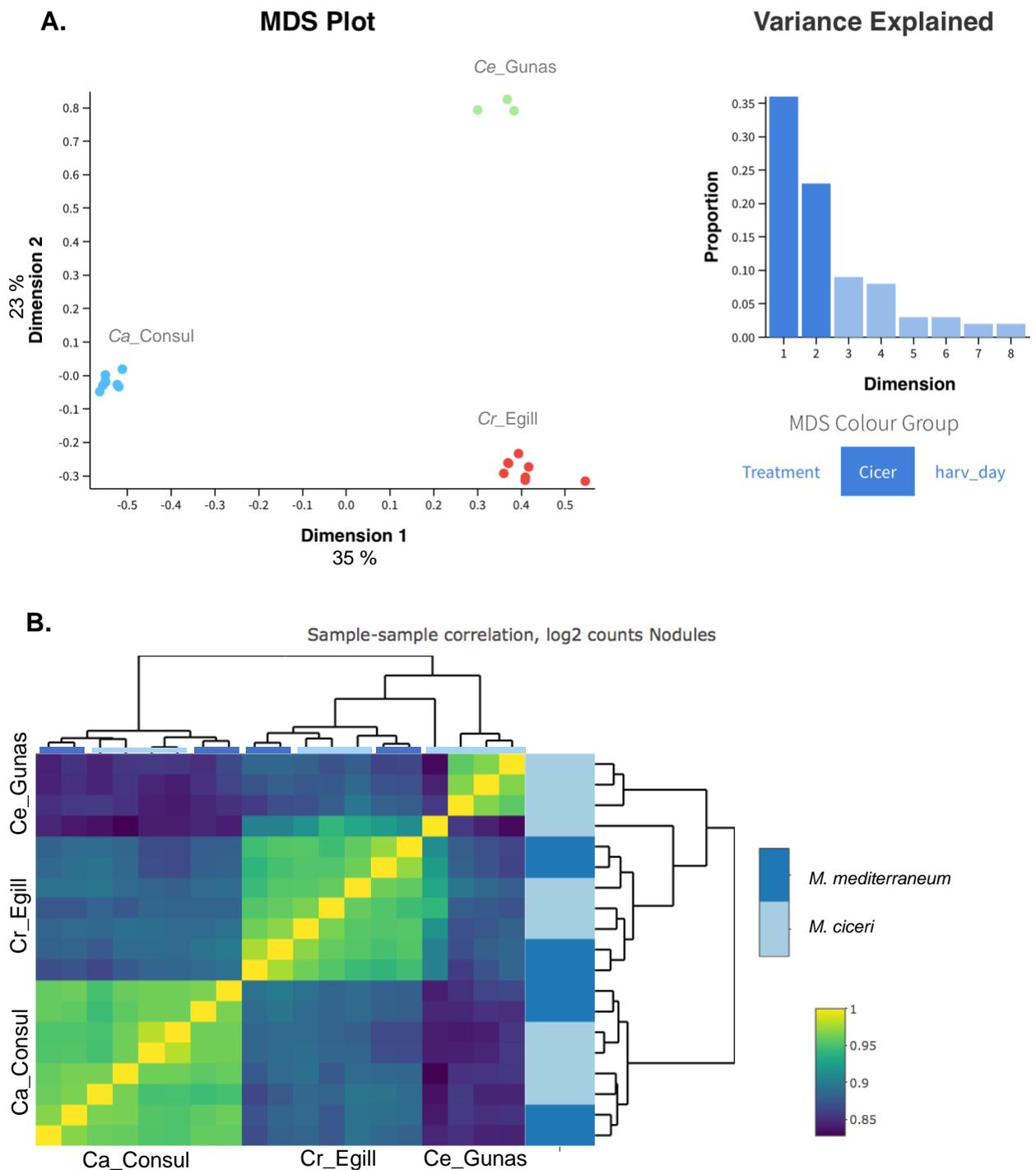
Cicer spp.	Genotype	Treatment	Sample_ID	HiSAT2 Overall aligment rate
C. arietinum	Consul	MmR19	C_Mm.Ln.109	96.30%
			C_Mm.Ln.113	96.92%
			C_Mm.Ln.115	96.30%
			C_Mm.Ln.116	97.24%
		McR17	C.Mc.Ln.73	96.72%
			C.Mc.Ln.75	97.31%
			C.Mc.Ln.76	96.36%
			C.Mc.Ln.81	96.48%
C. reticulatum	Egill73	MmR19	E_Mm.Ln.121	93.49%
			E_Mm.Ln.124	92.34%
			E_Mm.Ln.126	93.81%
			E_Mm.Ln.128	93.12%
		McR17	E.Mc.Ln.85	93.13%
			E.Mc.Ln.89	90.27%
			E.Mc.Ln.91	93.02%
			E.Mc.Ln.92	92.20%
C. echinospermun	Gunas62	McR17	G.Mc.Ln.103	91.77%
			G.Mc.Ln.108	90.68%
			G.Mc.Ln.97	90.99%

Table 1. Overall alignment rate for nodule samples.

Figure 1.A Plant biomass in response to N treatments. *C. arietinum* (Consul), and *C. reticulatum* (Egill) respond with higher biomass to the treatment with *M. mediterraneum* (Rse19), while for *C. echinospermun* (Gunas) responds to inoculant *M. ciceri* (Rse17).

B. Global transcriptome view for shoots and roots under N starvation, NH4NO3 (10mM), or symbiosis. Symbiosis status with either *M. ciceri* or *M. mediterraneum*, MDS plot representing gene expression of 22609 genes from 107 samples of chickpea, first (40 %) and second dimension (25%), MDS plot color by treatment, nodule global transcriptome is differentiated in the first dimension by rhizobial treatment effect. **C.** Second (25%) and third dimension (5%), color by *Cicer* sp.

Figure 2. A. Nodule transcriptome MDSPlot. **B.** Sample to sample correlation, hierarchical cluster analysis. Green squares compared transcriptional profile with samples of the same species. Rhizobia treatment annotated on the right side.

A. CariMcvsMm CretiMcvsMm

up
down

20331
20394

B. MciceriVsMmedi

C. Nodule GO terms

UP_Mc(1)_v_Mm   %   DOWN_Mc_v_Mm(-1)

Figure 3. A. Venn diagram for Nodules significantly up- and down-regulated genes (P-value >0.05, log fold change1). Contrast *M. ciceri* (1) vs. *M. mediterraneum* (-1). **B.** Volcano plot of DE genes (413) identified in nodules of *C. echinospermun*, *C. reticulatum*, and *C. arietium*. Contrasting conditions, rhizobia source *M. mediterraneum* Down (-1), over *M. ciceri* (1). **C.** GO Terms Enriched in genes found with Differential expression in nodules. Blue and Orange bar are counts of DE genes, gray bar is percentage per category. In B and C, contrasting the expression across the three *Cicer* sp. (contrast -1for Mm, vs 1Mc).
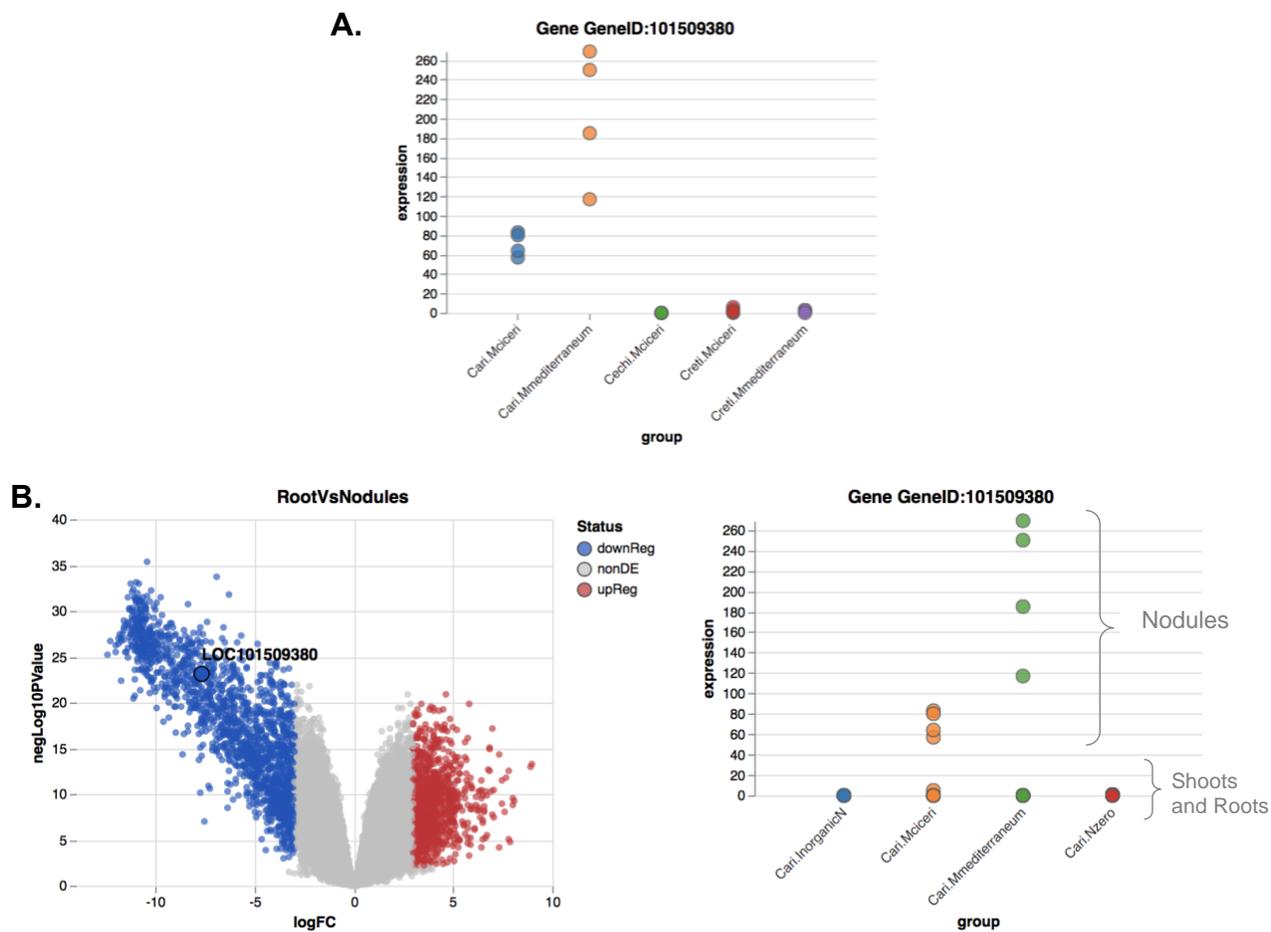
Fig 4. Plant peptide hormone expressed only in *C. arietinum*, not in the wilds. **A.** Nodule expression profile of gen LOC101509380 CLAVATA (CLV)/ EMBRYO SURROUNDING REGION (ESR)-RELATED PROTEIN (CLE) upregulated in *M. mediterraneum* nodules vs *M. ciceri*. **B.** Volcano plot (contrast Roots vs. Nodules), and expression profile of gene LOC101509380 (logFC-7.6) in all tissues, samples and conditions of *C. arietinum*, indicating that this gene is exclusively expressed in the nodules. Expression=logCPM
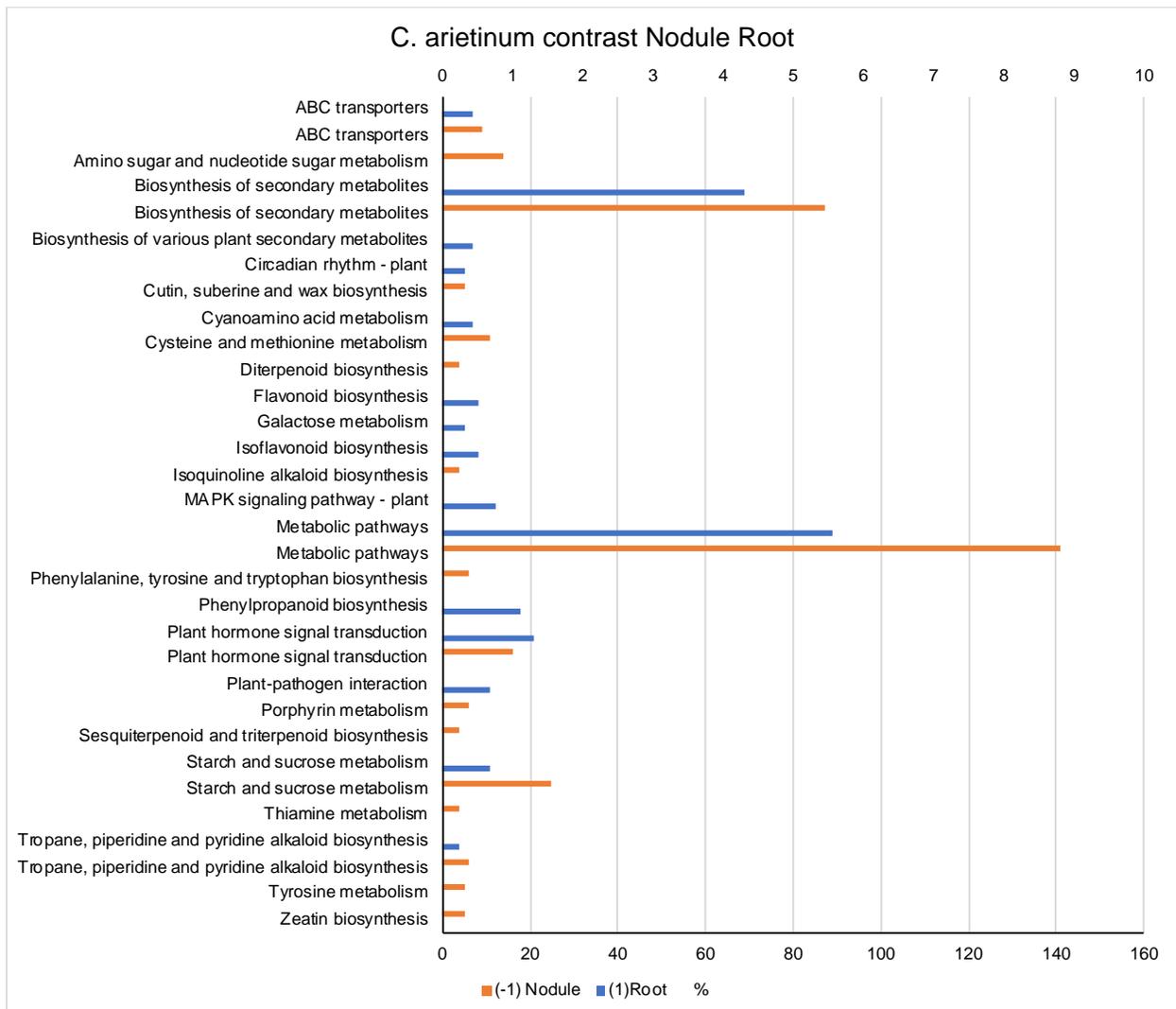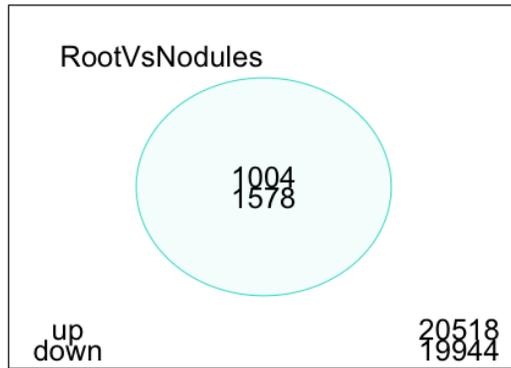
Fig 5. *C. arietinum* Contrast DE genes Root vs Nodules. Overrepresented genes in roots (1) over nodules (-1).
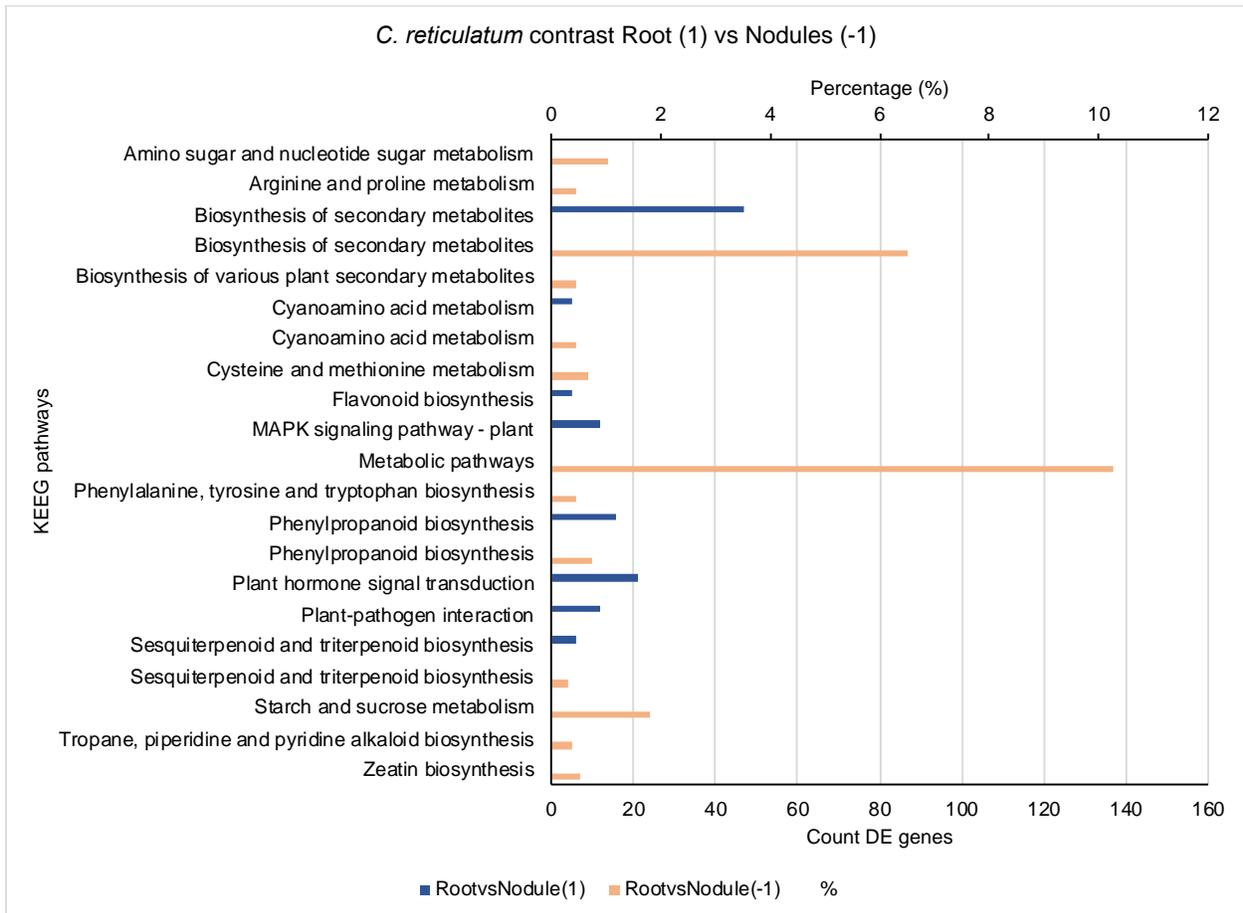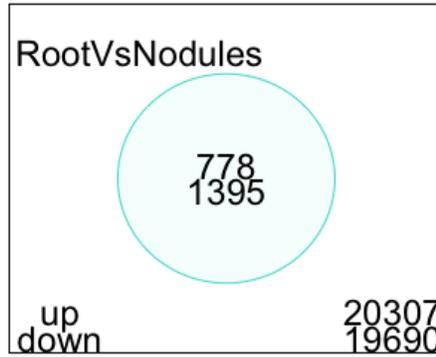
Fig 6. *C. reticulatum* Contrast DE genes Root vs Nodules

Fig 7. *C. echinospermun* contrast DE genes Root vs Nodules

A.

Carietinum Cechinospermun

18
10

0
0

0
3

0
0

1
10

0
0

8

44
105

up
down

Creticulatum

21086
21021

B. Roots Top 500

Color Key and Histogram

1   2   3
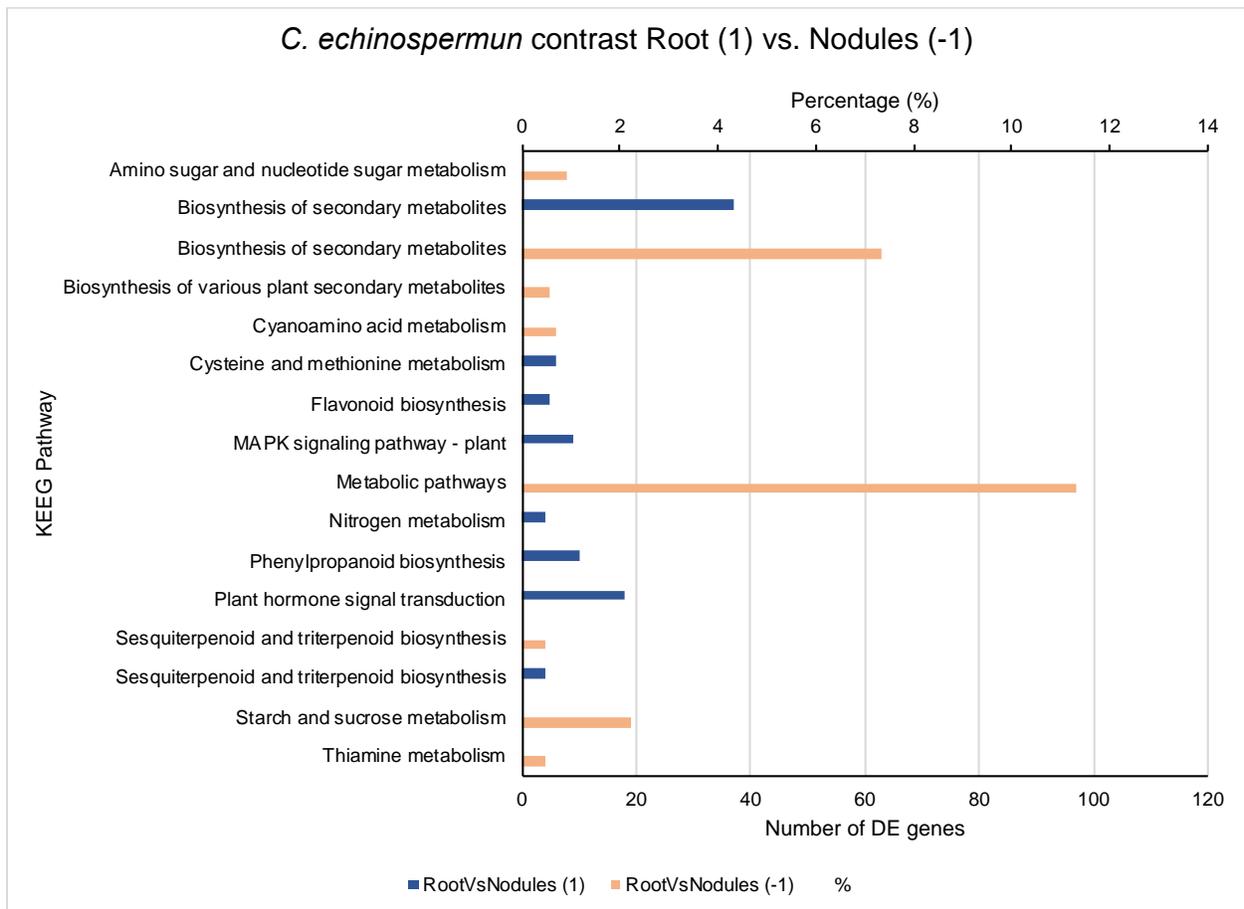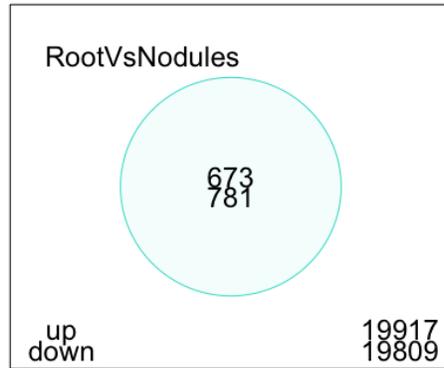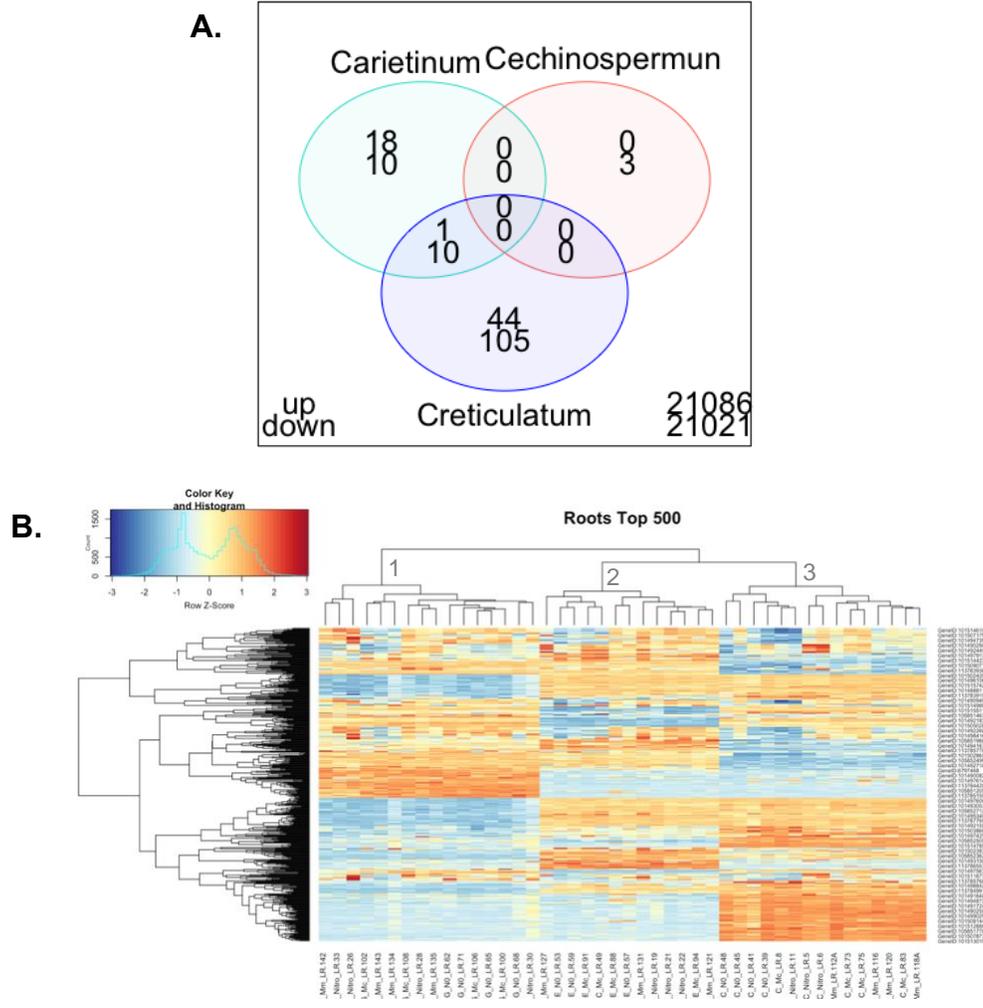
Fig 8A. Root transcriptional profiling of control vs treatments. The control group is nitrogen deficiency status, and the treatments include the response to inorganic Ammonium Nitrate (10mM), and the symbiosis status. **B.** Unsupervised hierarchical cluster analysis for top 500 genes Root, first level cluster discriminate by species (1- *C. echinospermun*, 2- *C. reticulatum*, and 3-*C. arietinum*).
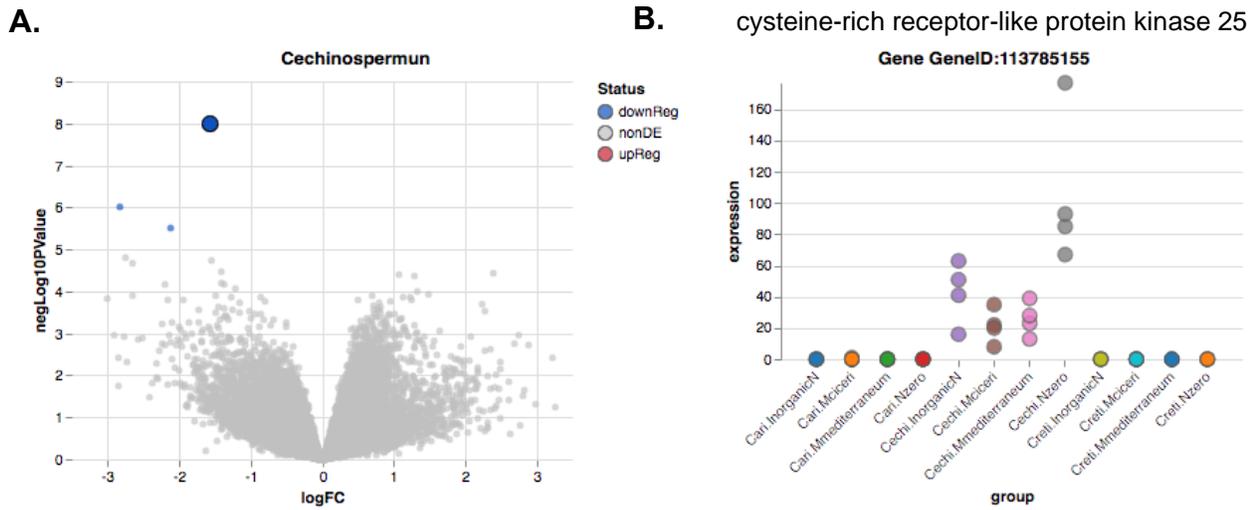
A. Volcano plot depicting DE genes in roots of C. echinospermun when N deficient treatment (downReg) was contrasted with roots under N and rhizobia treatments. **B.** Expression profile for cysteine-rich receptor-like protein kinase 25 is exclusive of wild relative *C. echinospermun*. Expression=logCPM
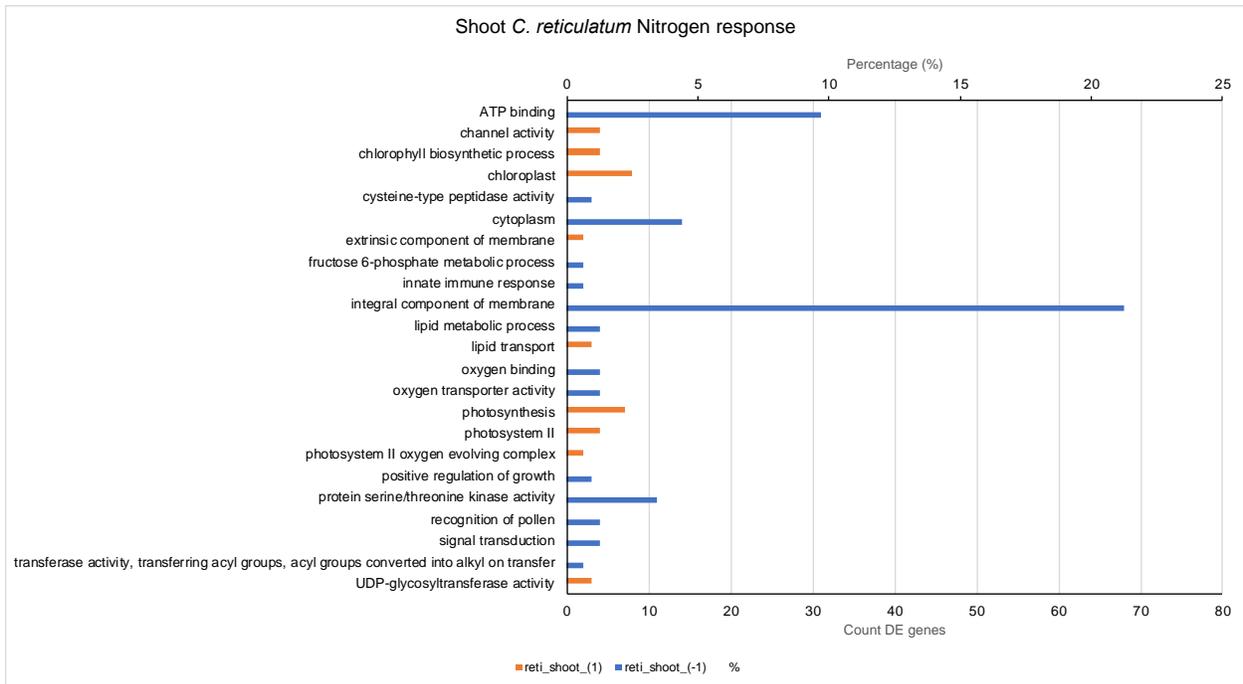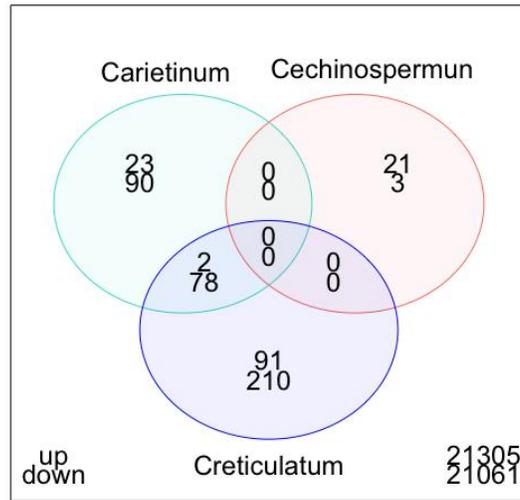
Fig 10A. Leaves transcriptional profiling of control vs treatments. The control group is nitrogen deficiency status, and the treatments include the response to inorganic ammonium Nitrate (10mM), and the symbiosis status. B. GO Terms Enriched in genes found with Differential expression in *C. reticulatum* shoots. Orange bar, terms up-regulated in the rhizobia and inorganic N treatments (1), Blue bar terms enriched in nitrogen deficient shoots (-1). Gray bar (second axes) is percentage
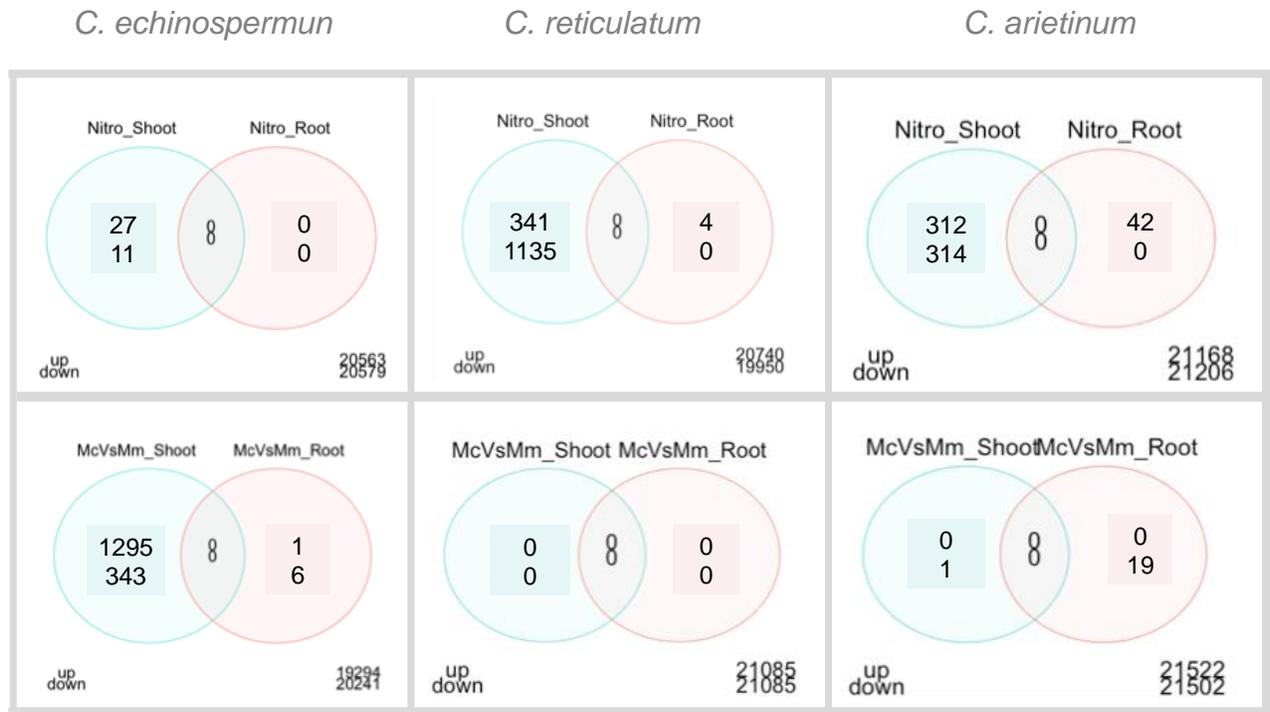
Fig 11. Pairwise treatment contrast in roots and leaves per species. Top panels contrast the nitrogen deficiency status with N sufficient status. Bottom panels
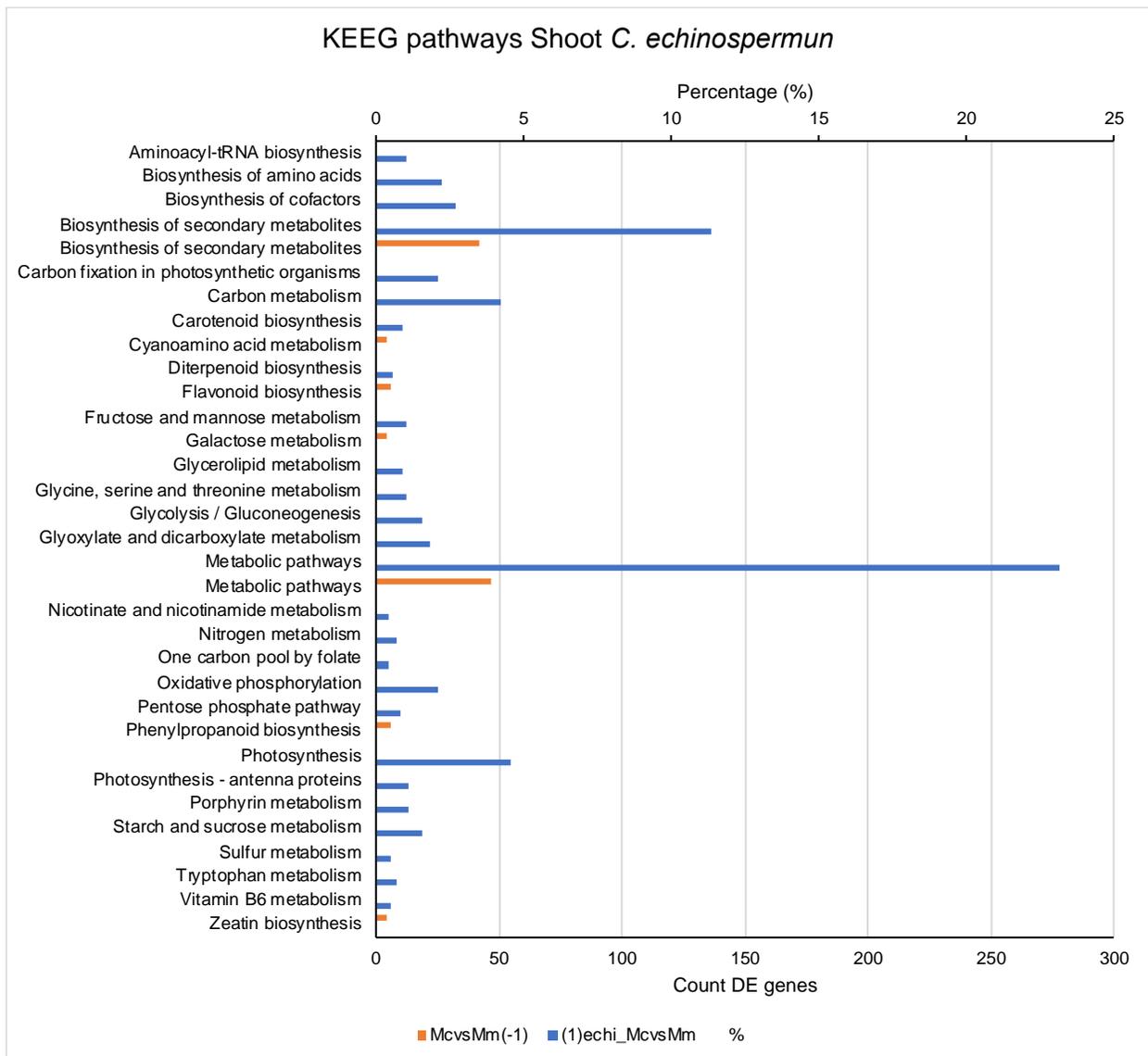
Fig 12. KEEG pathway annotations DE genes *C. echinospermun* shoots under symbiosis. Contrast Upregulated (1) with *M. ciceri* treatment vs (-1) *M. mediterraneum.*
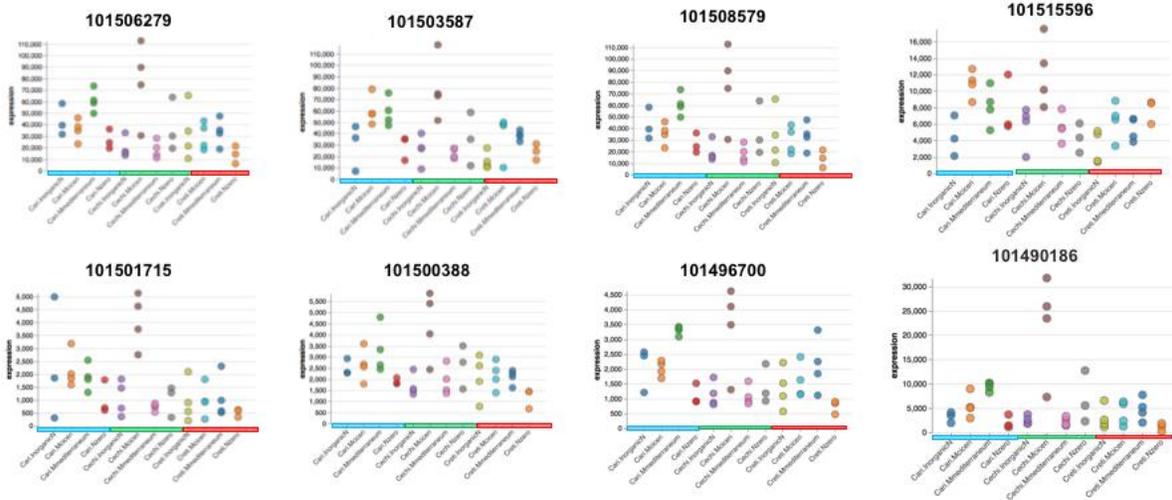
Fig 13. Expression profile of eight genes with "ferredoxin" term up-regulated in leaves of *C. M. ciceri*. Blue- *C. arietinum,* Green- *C. echinospermun,* Red- *C. reticulatum echinospermun* with compatible symbiont. Expression=logCPM
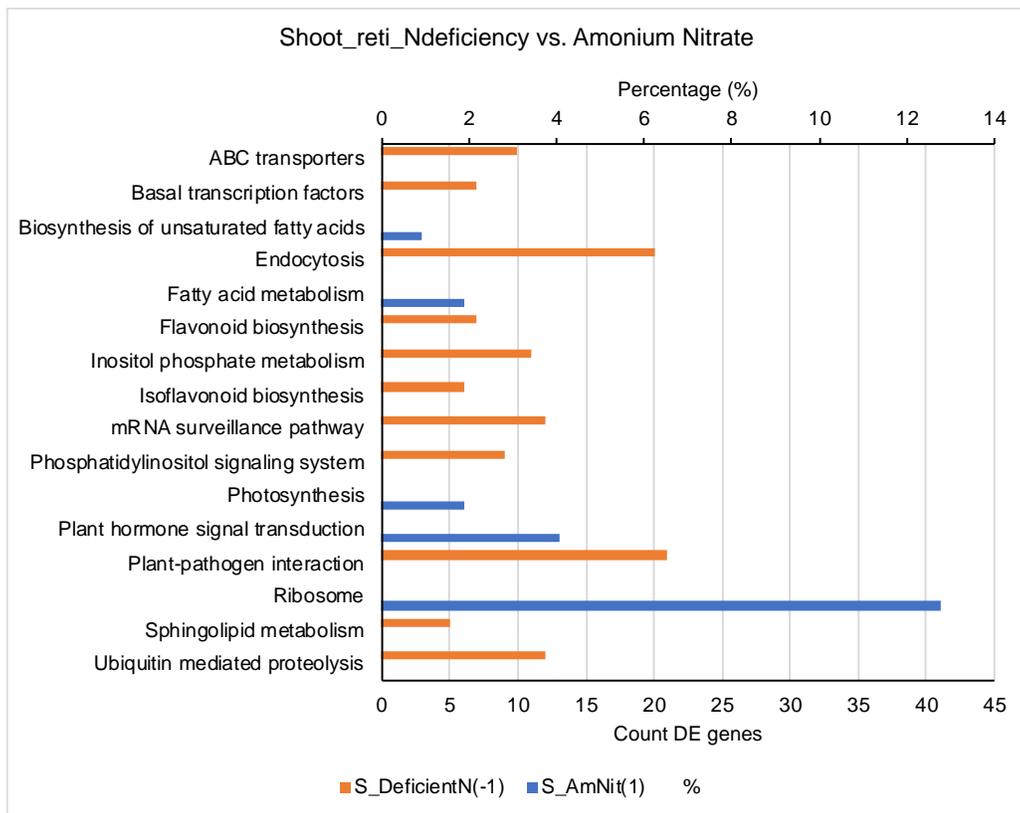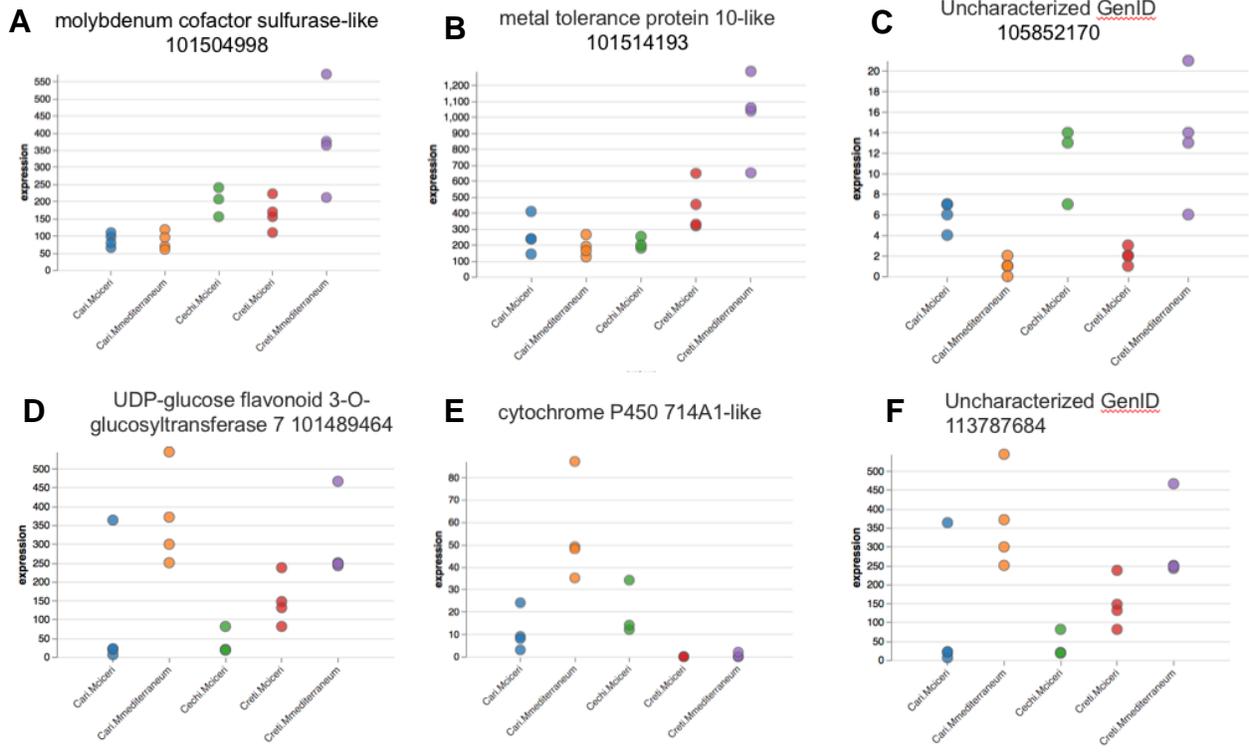
108

Fig 14. GO terms annotation1476 DE genes *C. reticulatum* shoots under N deficiency vs ammonium nitrate.

Supplemental figure 1. Expression profile of up-regulated genes in nodules formed with *M. mediterraneum* symbiont. (A-C) *C. reticulatum* (D-F) *C. arietinum.* Expression=logCPM
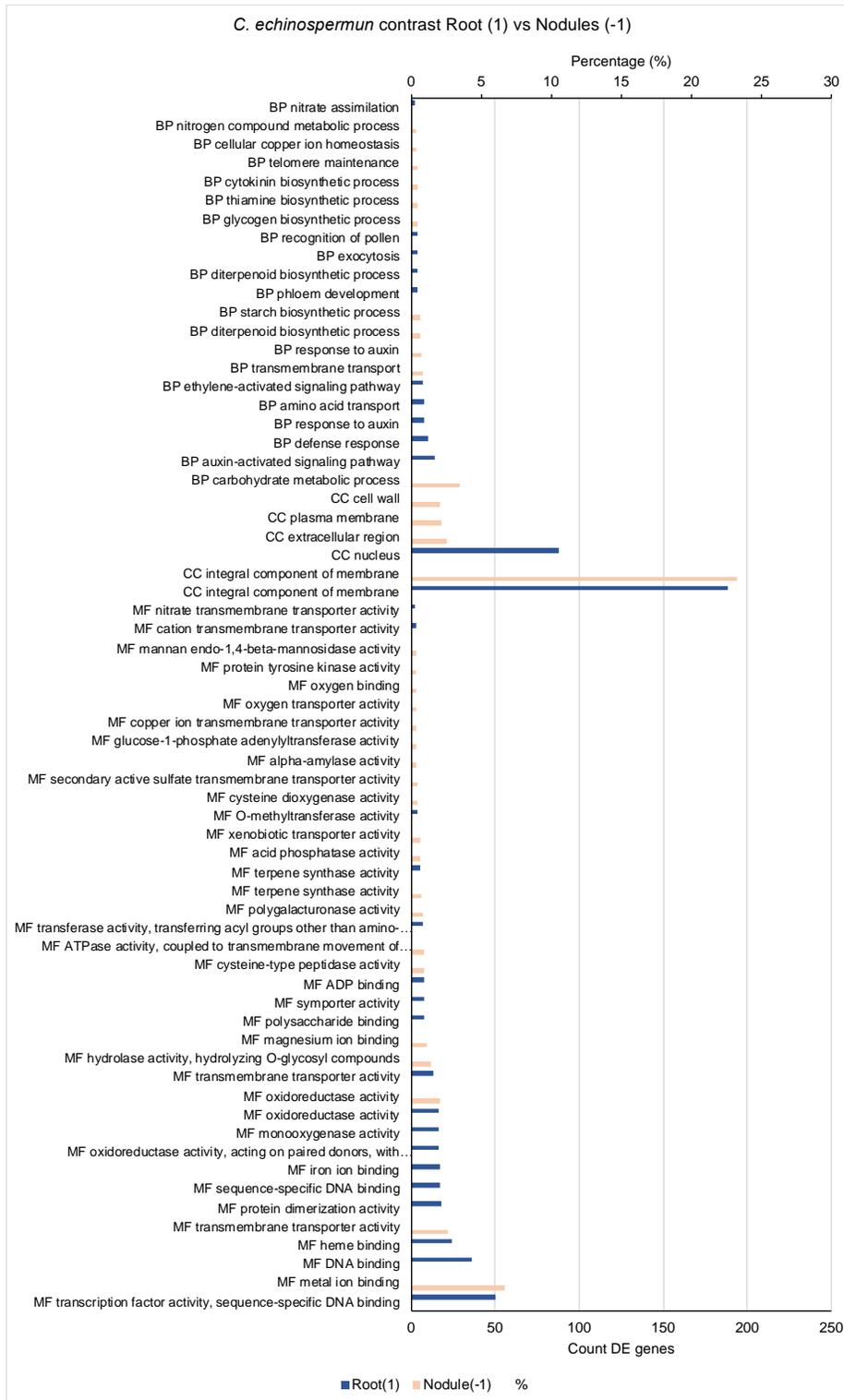
Supplemental Fig. 2. Enrichment GO terms Chart- C. echinospermun Roots vs. Nodules
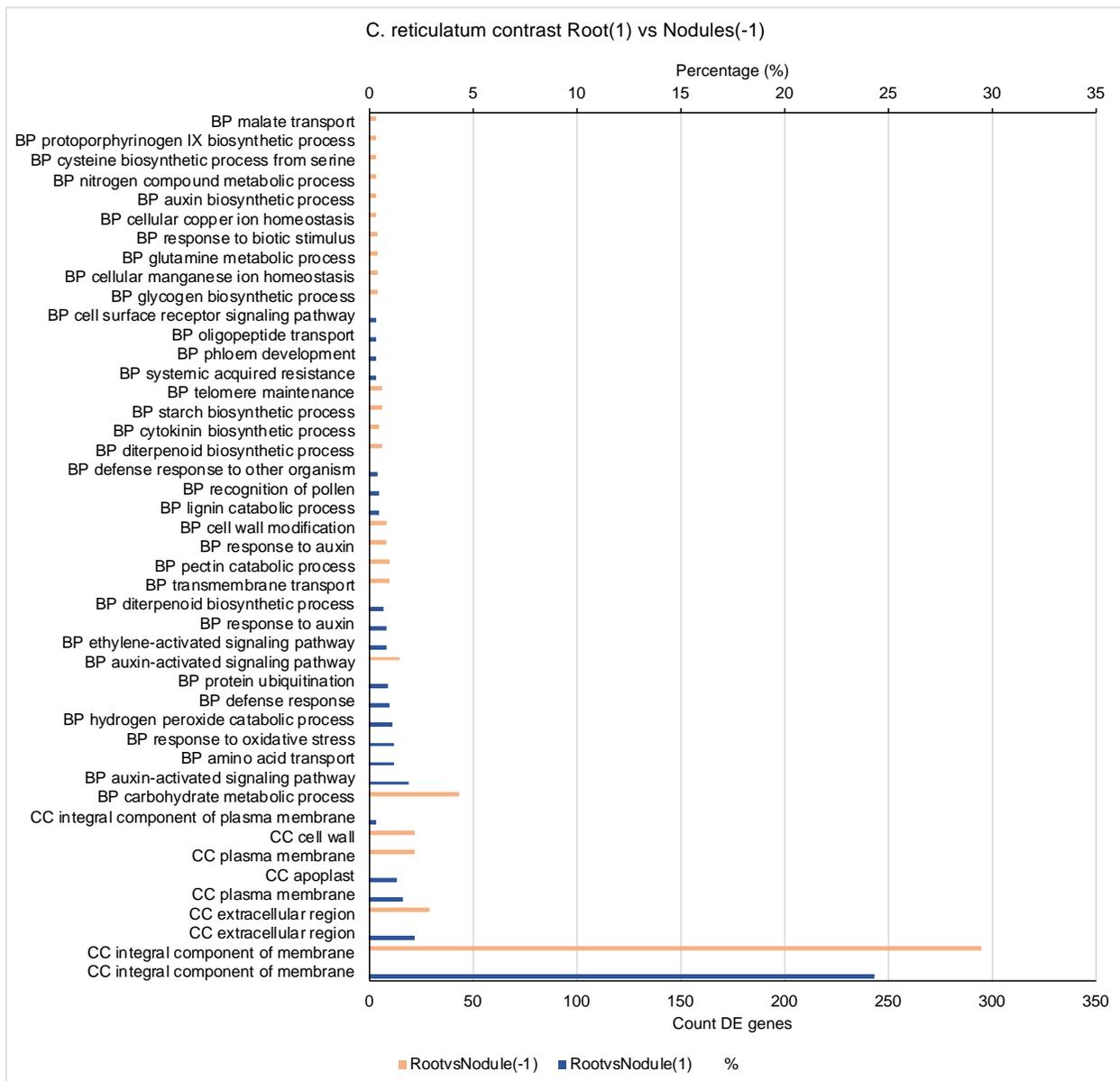
Supplemental figure 3. Enrichment GO terms (Biological process and Cellular Compartment) Chart- C. reticulatum Roots vs. Nodules

C. reticulatum

Supplemental Figure 5. Roots *C. reticulatum* DE genes associated with JA Signaling

CHAPTER IV.

Genomic features in symbiosis island
and host-microbe- interaction loci in *Mesorhizobium* spp.

Collaborators:

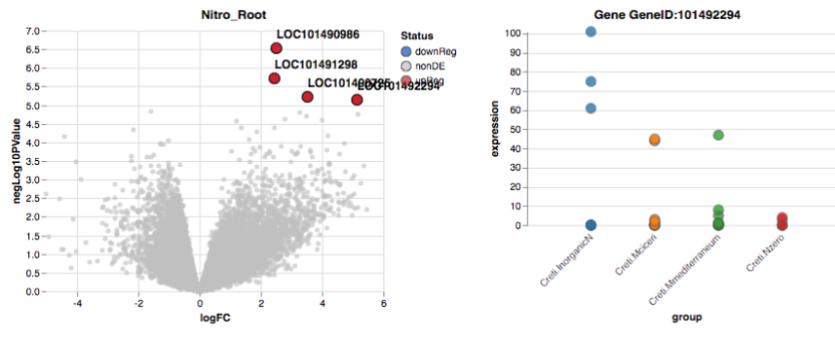L. M. Perilla-Henao, A. Greenlon., G. Robinson, Hien P. Nguyen, N. Incer V., N. Carrasquilla, B. Riley, S. Hussain, U. Ahmad, M. Atif, K. Riaz, D. R. Cook.

ABSTRACT

Chapter IV presents data on the evolution of bacterial symbiosis in chickpea, by examining the genomic relationship between strains associated with chickpea cultivation in Pakistan and strains at the wild center of origin in South-Eastern Turkey, focusing in particular on the provenance of genes involved in Type III secretion and Nod factor synthesis. The data reveal 4 new species associated with chickpea in Pakistan, all derived from a narrow diversity within the genus *Mesorhizobium*. Phylogenetic and haplotype analyses demonstrate that the symbiosis island is composed of ancient haplotypes that have reassorted based on horizontal gene transfer to construct the current diversity of Pakistani agricultural strains. Given that these Pakistani new species occupy the hot and dry Thal desert region, these strains are candidates for traits to adapt chickpea-*Mesorhizobium* symbiosis to extreme environments.

INTRODUCTION

Symbiosis for nitrogen fixation in chickpea (*Cicer arietinum*) is a narrow host range interaction, restricted to bacteria of the genus *Mesorhizobium* (*Phylobacteriaceae*, Proteobacteria) (Jarvis et al., 1997, Laranjo et al., 2012, Kamboj et al., 2010, Perret et al., 2000, Broughton et al., 2000). The Subcommittee on the taxonomy of *Rhizobium* and *Agrobacterium* recognizes at least 48 named species (de Lajudie et al., 2019) within *Mesorhizobium*, while many additional clades were evident based on whole-genome analyses but lack formal nomenclature (Greenlon et al.,

2019). The genus is diverse in its geographical distribution, species, and gene content (Greenlon et al., 2019). At least 28 genospecies are recognized as chickpea symbionts in agricultural systems, compared to only two related *Mesorhizobium* species that occur in natural habitats in association with the crop's immediate wild progenitor, *Cicer reticulatum* (Greenlon et al., 2019). Thus the overwhelming majority of the symbiont's genomic diversity is observed at global scales in agricultural situations. Local diversity of the symbiont can be high but also unique, thus few genospecies have broad geographic distribution (Greenlon et al., 2019). Indeed, Greenlon et al (2019) speculate that geographically-restricted diversity of the symbiont may facilitate adaptation of the microbe and the crop to local agronomic factors (Greenlon et al., 2019).

Within chickpea symbionts, genome size is relatively similar, in the range of 6 Mb, with ~6,500 predicted genes per genome. Conversely, gene content is highly variable. Considering 15 finished genomes of chickpea symbionts, pangenome analysis reveals high genome plasticity, with a core genome of 1,217 genes, a highly variable accessory genome, and pangenome of 42,874 genes. Further analysis including draft genomes revealed a striking value of 171,982 orthologous groups for the chickpea-associated *Mesorhizobium* pangenome (Greenlon et al., 2019). Despite a high gene-content variation at the genus level, Greenlon found that gene sharing is correlated with core genome nucleotide distance (Greenlon et al., 2019). In contrast comparisons at the full genome level, 10% of the *Mesorhizobium* genome has low diversity that is correlated with host of origin and contains genes for symbiosis. Interestingly, this low-diversity symbiosis island exhibits higher rates of gene flow than the high diversity background (Greenlon et al., 2019, Haskett et al., 2016b, Sullivan & Ronson, 1998).

In contrast to many other rhizobia that carry symbiosis genes on extrachromosomal elements (Flores et al., 1998), the *Mesorhizobium* symbiosis island is a chromosomally-encoded integrative conjugative element (ICE) (Haskett et al., 2016b, Sullivan & Ronson, 1998). ICE elements encoded their own conjugation machinery, with insertion at the 3' end of a tRNA gene (Juhas et al., 2009). Insertion and excision occur via a site-specific recombination system that

utilizes *att* sites which flanking the integrative element (Haskett et al., 2017). Different *Mesorhizobium* species have distinct excision and integration patterns, generating a monopartite island (continuous unit) or tripartite island (separated fragments) (Haskett et al., 2017, Haskett et al., 2016b). The ICE element of *M. Loti* ICESym*Ml*[TA] has been well characterized as a monopartite element, while the *Mesorhizobium* chickpea symbionts have distinct integration patterns that correlate with phylogenetic clades (200/433 monopartite vs. 181/433 tripartite) (Greenlon et al., 2019). Such structural variation in *Mesorhizobium* ICEs may affect plant interactions, given that the specific gene content of ICE elements determines host range (Haskett et al., 2016a, Haskett et al., 2016b, Sullivan et al., 2002).

Despite the relative conservation of the symbiosis island, phylogenetic evidence suggests high levels of small-scale horizontal gene transfer (HGT) within the symbiosis island, potentially contributing to genomic variation important to symbiosis (Greenlon et al., 2019). The functional gene clusters in the symbiosis island are associated with plant host interaction enabling conjugative transfer, quorum sensing, interference with ethylene synthesis (ACC deaminase), nitrogenase enzyme production (Nif genes), nitrogen fixation (Fix genes), nod factor biosynthesis (Nod genes) and Type 3 Secretion System (T3SS) genes (Nascimento et al., 2012, Sullivan et al., 2002, Teulet et al., 2022).

One of the major specificity factors in legume-rhizobia interactions is the bacterial Nod factor ligand. Host recognition of cognate Nod factor structures results in intracellular accommodation of the bacterium and the initiation of nodule organogenesis. Plant host-secreted flavonoids act as inducers of transcriptional activator NodD proteins, which bind the *nod* box and regulate expression of *nod* genes. Nodulation genes are arranged in multiple coordinately regulated operons, including structural genes and regulatory *nod* genes. The common nod genes (*nod*ABC) are conserved and are responsible for the synthesis of the core lipooligosaccharide structure (Kobayashi et al., 2004), which can vary in the length of glucosamine units and the length and saturation of the lipid moiety. NodA is an acylation factor, NodB is a chitooligosacharide deacetylase, and NodC is the *N*-acetylglucosaminyltransferase. A number of additional gene products (NodE, NodF, NodG, NodH, NodL, NodP, Q, NodS, NodX, NodZ, NolK,

116

NolO) are enzymes that add different chemical substituents that are host-specific modifications on the acyl chain or at the terminal reducing and non-reducing ends of the glucosamine residues (Caetano-Anollés & Gresshoff, 1992, Broughton et al., 2000, Perret et al., 2000). Nod factors released by rhizobia in the rhizosphere are inducers of host root nodulation responses. Nod factors are recognized by a high affinity LysM receptor complex on the root hair cell surface, activating the common symbiotic pathway and leading to intracellular infection, nodule organogenesis and nitrogen fixation (Oldroyd & Downie, 2008, Oldroyd et al., 2011), as detailed in previous chapters.

A distinct type of bacterial "machine" at the interface of prokaryote-eukaryote host interactions is the Type 3 Secretion System (T3SS). The T3SS complex - or injectosome- spans the bacterial cell membranes and connects its tip directly in a host cell, transporting bacterial proteins (Type III effectors or T3Es) from the cytoplasm to the extracellular environment (Chang et al., 2014). T3SS are found in the genome of gram-negative bacteria with different lifestyles and habitats. The structural elements of the injectosome are highly conserved across gram-negative bacteria, representing nine protein families namely SctC, SctJ, SctN, SctQ, SctR, SctS, SctT, SctU, and SctV – following the unifying nomenclature for secretion and cellular translocation (sct) and the suffix of the *Yersinia* system (Tampakaki, 2014). In plant pathogenic bacteria, these orthologous genes were originally designated as hypersensitive response and pathogenicity (*hrp*) genes. Bacterial plant pathogens carrying *hrp* inactivating mutations are unable to infect plants, without detrimental effects on bacterial growth *in vitro*. Genes for T3SS in rhizobia are called *rhc* (Rhizobium conserved), and the translocated cargo proteins are called Nops (nodulation outer proteins). Phylogenetics of common T3SS genes in rhizobia reveal that the T3SS forms a distinct group to those of pathogenic bacteria (Tampakaki, 2014, Teulet et al., 2022). While homology-based searches can be used for conserved elements, bioinformatic predictions of translocated effectors is complicated by high rates of divergence and presence-absence variation (Staehelin & Krishnan, 2015, Kimbrel et al., 2013).

In several symbionts of soybean (*Glycine max*), functional T3 effectors (T3Es or Nops) have been identified using immunodetection techniques, or under control of flavonoid-inducible

promoters, or with fusions of adenylate cyclase translocation reporter systems, or in heterologous *N. benthamiana* using the T3SS of *P. syringae* pv. tomato DS3000, as reviewed by (Staehelin & Krishnan, 2015). The species for which *bona fide* T3Es have been identified include *Sinorhizobium* (*Rhizobium*, *Ensifer*) *fredii* NGR234, USDA207, *Bradyrhizobium japonicum* USDA110, *B. elkanii* USDA61, *Mesorhizobium loti* MAFF303099, among others. Some characterized effectors have expression during early symbiotic stages, while others are expressed in the developed nodule, however the biochemical function is known for only few symbiosis effector proteins (Miwa & Okazaki, 2017, Staehelin & Krishnan, 2015). In the soybean-*B. japonicum* symbiosis, regulatory and structural gene T3SS mutants can extend the range of symbiosis in a host gene-specific manner (*Rj2/Rfg1* and *Rj4*) (Tsukui et al., 2013), leading to a model in which certain T3SS functions can limit host range when recognized by host NBS-LRR genes (Yang et al., 2010). Interestingly, in the symbiosis between *Bradyrhizobium* Strain DOA9 and *Aeschynomene indica*, the T3Es effector gene *ernA* promotes Nod factor-independent nodulation, activating the signaling pathway downstream of Nod factor perception (Teulet et al., 2019). In summary, the injection of T3Es can both inhibit and promote nodulation phenotypes (Staehelin & Krishnan, 2015, Teulet et al., 2022, Chang et al., 2014, Yang et al., 2010).

There is growing recognition that some plant traits, including tolerance to abiotic stress, may arise through interaction with rhizobacteria (Dimkpa et al., 2009, de Zelicourt et al., 2013) – an extension of the "holobiont" hypothesis (Brucker & Bordenstein, 2013). In this regard, genes that govern symbiotic establishment might be especially significant. In the case of legumes, stress-tolerant symbiotic nitrogen fixation (SNF) is a topic of longstanding interest (Imran et al., 2015, Dimkpa et al., 2009, de Zelicourt et al., 2013), motivated by the possibility that stress tolerant rhizobia might confer adaptations for crop productivity under arid conditions. While symbiotic nitrogen fixation (SNF) provides access to reduced nitrogen (N) and satisfies a primary constraint to crop productivity, abiotic stress negatively impacts SNF. Low soil moisture affects root microbiome composition (Xu et al., 2018) and SNF at multiple levels, including signaling during initiation of nodule organogenesis and the maintenance of nodule metabolism (Aranjuelo et al., 2014).

Pakistan's Thal Desert region represents one of the most extreme environments for chickpea production, accounting for a remarkable ninety percent of Pakistan's chickpea cultivation. The Thal Desert is typified by low rainfall, high temperatures and soils of high sand content (Ahmad et al., 2019). The majority of Pakistan's chickpea farmers are small holders, sowing the crop under rainfed conditions during the fall or winter with low agronomic inputs. As a consequence of low agronomic inputs and environmental extremes, profitable cultivation is possible only in years with sufficient residual soil moisture (Sani et al., 2018). Indeed, during the period 2013 to 2017 chickpea yields in Pakistan were half of that in other South Asian countries and less than 30% of those obtained in more intensively managed chickpea-growing regions (Merga & Haji, 2019). Low rates of production, despite high cultivation acreage and demand, underlie Pakistan's first rank for global chickpea imports in 2018 and 2019 (FAO stats, 2020).

While Pakistan's chickpea landraces represent a genetically narrow crop germplasm, there is evidence that long-standing selection on a subset of allelic variation may have adapted the crop to locally arid conditions (Sani et al., 2018). Here we sought to extend those studies by collecting and characterizing chickpea's endemic *Mesorhizobium* symbiont. Whole genome analyses demonstrate that Pakistani symbionts represent unique genomic diversity, including the identification of new species. More detailed analyses of symbiosis genes, including those involved in production of Nod factor and T3SS signals, were used to understand the provenance of the symbiosis ICE island among Pakistani strains. The characterized microbes provide a genomic resource representing symbionts endemic to arid agricultural soils and potentially contain adaptations to arid agricultural environments.

RESULTS

To identify and describe the diversity of chickpea *Mesorhizobium* symbionts adapted to arid conditions, in 2017 we sampled 20 agricultural soils in the Punjab province of Pakistan with longstanding histories of chickpea production (Figure 1). Symbiotic bacteria were trapped by sowing chickpea into collected soils under greenhouse conditions at the University of California,

119

Davis. To reduce potential contamination, soil as well as the surface of seeds were sterilized and irrigated with sterile water over the course of 37 days, at which point plants had reached the flowering stage and roots were harvested and rinsed. Nodules were collected, surface-sterilized and mechanically crushed prior to plating on nutrient media from which single colonies were isolated. Colonies were screened by means of PCR using *Mesorhizobium*-specific *Nod*C primers, with positive colonies re-purified and cultured in liquid media for DNA extraction. Genomic DNA was sequenced and the genomic data was assembled to obtain 11 draft *Mesorhizobium* genomes (Supplementary table 2). Average genome N50 values were 64 kb (range 45-79 kb), with 62% GC content and estimated genome sizes ranging from 5,581,658 bp to 7,051,457 bp. Based on PROKKA annotation, the number of coding sequences (CDS) ranged from 5,406 to 6,814 (Supplementary table 2).

The identity of individual strains was initially determined by sequence comparison to the international collection of type strain genomes. The Type Strain Genome Server (TYGS), which compares input whole genome sequences and 16S rDNA sequences to type strains, indicated that the closest bacterial type strains were members of the genus *Mesorhizobium* but was unable to confidently assign species, i.e., the identified strains were equidistant from *Mesorhizobium muleiense, M. wenxiniae, M. temperatum, M. delmotii, M. prunaredense, M. mediterraneum, M. metallidurans, M. helmanticense, M. sanjuanii, M. sangaii, M. waimense, M. tamadayense, M. intechi, M. erdmanii, M. japonicum, M. sophorae,* and *M. carmichaelinearum* (Supplementary Figure 2).

More detailed comparisons to type strains were calculated using intergenomic distances estimated as digital DNA-DNA hybridization (dDDH) values using the Genome BLAST Distance Phylogeny (GBDP) method (Meier-Kolthoff et al., 2013). dDDH is an indirect method that calculates a theoretical DNA-DNA hybridization value among genomes. Although potentially error-prone, dDDH has value because it permits comparison of modern genomic data to historic, *in vitro*-determined genome-genome hybridization values. Queries with the Pakistani *Mesorhizobium* genomes revealed that none of the type strain genomes yielded pairwise dDDH

120

values exceeding 70% (Supplementary table 3), indicating that the genomes sequenced in this study represent novel species relative to the international type strain collection.

A dendrogram depicting the calculated intergenomic distances among the Pakistani *Mesorhizobium* genomes and the set of closely related type strains genomes is shown in Figure 2. The 11 Pakistani *Mesorhizobium* genomes represent four independent clades, which we assign as New Species 1-4 (Figure 2). Strain *Mesorhizobium* s68_pk9, which is a singleton representative of a novel species, was uniquely recovered from sandy soils in the Khoshab district and is basal to the clade of *M. temperatum* SDW018 and *M. delmoti* SFM4623 (Figure 2). Four isolates, s62_pk201, s67_pk5, s74_pk44 and s75_pk146, form a coherent clade that is sister to *M. muleiense* CGMCC 1.11022 (Figure 2); these isolates were geographically disperse, being trapped from separate soil samples originating on 3 farms in the Bhakkar and Jhang districts (Figure 2, Supplemental table 1). In three cases, pairs of strains sampled from the same soil had highly similar genomes and are likely to be clones, i.e., s56_pk21 and s56_pk22 (Khoshab district), s78_pk47 and s78_pk49 (Bhakkar district), and s64_pk3 and s64_pk4 (Jhang district) (Figure 2). Although sample size is too low for confident statistical analysis, it is evident that *Mesorhizobium* diversity is geographically disperse, such that individual districts can harbor multiple species and individual species can occur across multiple districts.

dDDH analyses are not only indirect, they also limit the range of possible comparisons to type strain genomes present in international repositories. Importantly, as demonstrated by Greenlon et al., (2019), the majority of *Mesorhizobium* genomes lack formal taxonomic nomenclature and thus are not represented in the type strain collections. As of September 2022, there were 1,211 genome sequences in the NCBI genome database assigned to the genus *Mesorhizobium* of which only 142 are assigned to 51 named species. A more comprehensive approach to classifying bacterial organisms is to used whole genome data to identify speciation cutoffs and assess strain similarities; such genomic approaches are agnostic of human-assigned species names. Using the broad diversity of *Mesorhizobium* genomes available at NCBI, we calculated the pairwise average nucleotide identity (ANI) among the 11 Pakistani genomes and a representative set of 237 *Mesorhizobium* genomes, including all species assigned by Greenlon

and colleagues. Using 95% ANI as the lower boundary, the data nominate 45 species-level assignments (S4. Supplementary table 4). According to these criteria the Pakistani isolates represent four novel species of *Mesorhizobium*, which agrees with the less direct dDDH method (above).

To evaluate phylogenetic relationships among strains we compared 400 conserved protein coding genes among the set of 237 genomes, confirming that Pakistani strains form distinct groups within a wider diversity of *Mesorhizobium* (Figure 3A and 3B) and supporting species assignments made using 95% genome-wide ANI. Three of the Pakistani species nest phylogenetically within *Mesorhizobium* Clade 5, which previously included three distinct species (5A, 5B and 5C) as described in Greenlon et al (2019). We assign the new species New Sp. 1, New Sp.2, and New Sp.3 in clades 5D, 5E and 5F respectively. Within Clade 5, the native symbionts of wild *Cicer reticulatum* occur exclusively in 5A, predominantly in a coherent subgroup that is distinct from the majority (97%) of agricultural strains which have origins in India, Ethiopia and Morocco. The fact that the seven Pakistani Clade 5 strains represent three new species, distinct from ~125 previously characterized strains, indicates remarkable novelty within the Pakistani *Mesorhizobium* community and underscores the relationship between geography and strain diversity noted by Greenlon.

The fourth Pakistani species, which includes four isolates from two agricultural districts (Bhakkar and Jhang), represents an additional novel source of *Mesorhizobium* diversity on chickpea. Different from the other Pakistani strains, New Sp. 4 forms a group that is sister to Greenlon's Clades 5 and 6 strains and an in-group relative to Clade 9 (Figure 3 and Greenlon's supplemental Figure 1), which is notable because Clades 5 and 6 contain the majority of strains that associate with the crop's wild progenitor species, *C. reticulatum*.

Greenlon et al (2019) established that the core genomes of chickpea mesorhizobia are highly diverse relative to their symbiotic ICE. Thus, we asked whether the *Mesorhizobium* Pakistani genomes vary in their symbiotic island properties, initially focusing on integrase genes whose patterns of presence-absence are indicative of ICE structure, i.e., monopartite or tripartite configuration. We found that all four Pakistani species lack integrase *Int*S1, while uniformly

possessing *Int*S2 and *Int*G and varying in the presence/absence of *Int*M (Table 1). We also recorded presence/absence variation for *Int*M in several type species of *Mesorhizobium* (Table 1). These patterns are consistent with a multi-partite (most likely tripartite) structure of symbiotic islands in the Pakistani *Mesorhizobium* species, similar to all of Greenlon's Clade 5 strains. Interestingly, Greenlon's Clade 6 is composed exclusively of monopartite strains, having only *Int*S1. Thus the tripartite integrase configuration in New Sp. 4, which is basal to the group of Clade 5 and Clade 6, may indicate an ICE origin from Clade 5.

In addition to variation in chromosomal organization of the symbiotic island inferred from integrase gene content, we further explored gene content presence/absence and SNP variation within the ICE regions, focusing on genes that encode proteins for symbiotic interaction, namely the Nod and T3SS gene clusters. Prior efforts determined that certain Ethiopian genomes contain two ICE regions: one which is orthologous to other chickpea symbionts, and a second, non-orthologous and vestigial ICE region containing phylogenetically distinct Nod and T3SS clusters (Sup. Figure 2) (Nguyen *et al*, personal communication; Greenlon et al., 2019). Pakistani strains only carry a single ICE element, with Nod and T3SS clusters that are orthologous across all chickpea symbionts.

Despite their orthologous nature, we observe substantial variation in the phylogenetic status of these gene clusters, not only among Pakistani *Mesorhizobium* strains but also throughout the analyzed set of chickpea symbionts genomes (11 Pakistani + 60 comparator strains). Considering the Nod and T3SS regions separately (Figure 4), phylogenetic analyses reveal broad coherence with core genome phylogenetic assignments. Thus, most strains from genome groups 6 and 7 form coherent clades based on Nod of T3SS variation that is distinct from Clade 5 strains. However, there are numerous exceptions to these broad relationships (Figure 5), highlighting the likely role of horizontal gene transfer in shaping variation within the symbiosis island. For example, wild strains Oyali 205 and Oyali 204 have core genomes that nest within Clade 6A, however their Nod gene regions have an alternate ancestry within Clade 5. Interestingly the phylogenetic signal from the T3SS gene clusters of Oyali 205 and Oyali 204 agrees with the core genome assignment to Clade 6A and thus is discordant with the Nod-determined

phylogenies (Figure 4). The fact that sub-regions with the ICE symbiosis island can have discordant histories suggests horizontal gene transfer at the scale of regulons, as predicted by Greenlon et al (2019).

Similar analysis of Pakistani strains reveals that their Nod and T3SS gene clusters are most closely related to those of Clade 5 symbiosis islands (Sup. Fig 3 and Sup Fig 4). Interestingly, this is even the case for the Nod and T3SS gene clusters of New Species 4, for which the core genome nests external to the clade of Clade 5 and 6 strains. Thus, New Sp. 4 Nod and T3SS regulons appear to originate from Clade 5, in agreement with the prediction from integrase gene patterns, above. The distinctiveness of Pakistani Nod and T3SS gene clusters from those of Clade 6 is notable, because, at the crop's center of origin, strains of Clade 5 and Clade 6 are the most frequent native symbionts of the wild crop progenitor, *C. reticulatum*. The observation of only Clade 5-related ICEs among Pakistani strains may indicate a narrow historical introduction, or biased persistence after introduction. Similarly, the absence of Nod and T3SS clusters related to *Mesorhizobium* clade 7 (*M. ciceri*) supports the common understanding that commercial inoculants are not used in Pakistani chickpea production, because *M. ciceri*, which is the native symbiont of *C. echinospermum*, is the predominant species in commercial inoculant preparations.

Despite a common origin from Clade 5, the T3SS and Nod regions of Pakistani strains have complex phylogenetic histories, typically more closely related to non-Pakistani strains of diverse geographic origins than they are to one another. There are examples of Pakistani strains of the same species having phylogenetically distinct gene clusters, and of strains of different Pakistani species having gene clusters that are similar (Supplementary Figures 3 and 4). We tested the congruence of Nod and T3SS by calculating Robinson-Foulds distances, finding an RF value of 0.79 which indicates a medium to high incongruence among the two topologies. The incongruence detected by Robinson-Foulds is a quantitative statement about different evolutionary histories of the Nod and T3SS regions. As noted above, these patterns simultaneously reflect both the horizontal transfer of sub-regions within the symbiotic ICE and broad conservation of genomic subgroups (Supplemental Figures 3 and 4).

These phylogenetic patterns trace to ancestral variation from wild strains of Turkish origin that are themselves complex. For example, the strains of New Sp. 1 are widely distributed in the T3SS phylogeny (Supplementary Figure 4), less closely related to one another than they are to wild strains with disperse geographic origins. These patterns are consistent with the acquisition of different T3SS regulons into a single native Pakistani *Mesorhizobium* species from different wild strain-derived ICE elements. Similarly, the T3SS and Nod regions of New Sp. 3 (Pakistan) are co-phyletic with those of Dereichi 202 and Kurtoni 202 (Turkey) and distinct from other Pakistani agricultural strains. Complex patterns are also found in the global collection of agricultural trains. Thus the subgroup containing *M. ciceri* strains (group 7A) is well conserved and sister to a conserved groups of strains from Ethiopia that fall within a monophyletic group of genomes in Clades 1-4 (Figure 3 and 5), suggesting that the T3SS and Nod regions within the Ethiopian strains derived from common *M. ciceri* ancestry. Even among wild strains of Turkish origin phylogenetic signal of core and ICE gene clusters can be non-coherent. Thus the *M. mediterraneum* genome group (6A) from Turkey is well conserved, yet several strains with 6A genomes have Nod and/or T3SS regions with alternate phylogenetic assignments, as described above for Oyali-205 and Oyali 204. The fact that Pakistani strains have Nod and T3SS signals that are co-mingled within complex phylogenetic patterns that trace to wild progenitor strains indicates that T3SS and Nod haplotypes are ancient and stable, while also dynamic. Structural variation among these conserved haplotypes is depicted graphically in Supplemental Figures 5 and 6.

Genetic variation in the functional regions for Nod factor biosynthesis and conserved T3SS machinery was also investigated using linkage disequilibrium (LD) metrics, in which a high r^2 indicates that physically linked sites are also genetically linked. Conversely, low r^2 values between physically-linked sites indicates a history of frequent recombination. Within the set of 248 genomes we observe a higher frequency of genetically-unlinked SNPs in the T3SS cluster as compared with the Nod cluster (Fig 6). The Nod cluster has two main regions, one with high LD values (1 to ~10000 pb), and the second with low LD values (~10000 to ~16000), corresponding to the NodABC genes. The LD analysis for the T3SS cluster shows a mosaic-like pattern, with only

the first region of 2000 bp with a high LD value r2=1, followed by regions with low LS (0.002). The LD mosaicism, when compared with Sup. Fig 6. do not correspond with the low (~1000 to ~18000) or high (~18000 to ~4500) nucleotide identity regions across the T3SS alignments (Supplemental figure 6).

DISCUSSION

Chickpea (*Cicer arietium*, also known as Gram) is an important pulse crop throughout the Indian subcontinent, including Pakistan which ranks third worldwide for consumption and cultivated area. We report new diversity of nitrogen-fixing symbionts associated with arid soils where chickpea is grown in the Thal Desert, Pakistan, and where the sampled fields are considered marginal land for agriculture. Agricultural systems in the region rely on low-input, and the crop develops on residual soil moisture after annual rainfall (Ahmad et al., 2019, Sani et al., 2018). Nevertheless, chickpea production in these areas is part of traditional practices, a cash crop, and a source of protein nutrition for small-holder farmers.

Under the prevailing traditional cultivation practices, Pakistani farmers rarely use rhizobial inoculants, instead relying on microbial strains indigenous to local soils (Sani et al., 2018). Here we identify four new species belonging to the genus *Mesorhizobium* that survive in arid soils in Pakistan (Figure 2, Figure 3), suggesting that microbial diversity in the area has been largely unexplored. Prior efforts have isolated strains from Pakistani environments, but in those cases the methods were limited to traditional approaches for strain assignment (Latif et al., 2013, Zaheer et al., 2019) and therefore lacked the resolving power of whole genome analysis. Indeed, comparison of 16S sequences were inadequate for species differentiation in *Mesorhizobium* (Sup Fig. 1). Even multi-gene phylogenies involving *atp*D, *dna*J, *gln*A, *gyr*B, and *rec*A, grossly underestimate bacterial genomic diversity (Lorite et al., 2018, Benjelloun et al., 2019, Laranjo et al., 2012, Zahran, 2001).

The most comprehensive analysis of global diversity in the genus *Mesorhizobium* showed that 805 genomes were represented broadly in 10 phylogenetically clades (Greenlon et al., 2019).

Here we use representative genomes from each previously-described *Mesorhizobium* clade to evaluate the relationship with the sequenced strains from Pakistani farm fields. Both genome-wide nucleotide distance using ANI (Jain et al., 2018) and phylogenetic analyses involving 400 conserved orthologs (Segata et al., 2015) confirm the novel diversity of the Pakistani *Mesorhizobium* collection (Figure 3). Despite a high rate of novel species, the Pakistani strains fall within a narrow subset of the known diversity of chickpea's symbiont community – namely, within a monophyletic group composed of Greenlon's Clades 5, 6, 9 and 10. Within this group, three of the Pakistani species are contained within Greenlon's Clade 5, while New Sp. 4 falls within a previously unrecognized clade designated Clade 11, that is sister to Greenlon's Clades 5 and 6 and internal to Clade 9. Notably, we did not identify strains related to *M. ciceri* (Clade 7), which is the most common source of commercial inoculum for cultivated chickpea. Moreover, Pakistani isolates represent different rhizobial species compared to the ones found in agricultural soils in the neighboring country of India (Greenlon et al., 2019), where chickpea has also been grown traditionally.

The identification of four *Mesorhizobium* species reflects variation in the background *Mesorhizobium* genome, with all species sharing an orthologous symbiosis island that confers nodulation on the chickpea host. It is well established that the capacity of nodulation and nitrogen fixation is encoded in a ~500kb genomic region that is transferred horizontally as conjugative integrative element (ICE). In some species, ICE integration occurs as a single co-linear segment referred to as a monopartite symbiotic island, while in other species the symbiotic island integrates as three separate pieces in different regions of the genome, so-called tripartite ICE elements (Greenlon et al., 2019, Haskett et al., 2016b). The repertoire of integrase gene paralogs is indicative of ICE structure. The four new Pakistani species lack *Int*S1 and instead have orthologs of *Int*S2 and *Int*G, suggesting a tripartite ICE organization (Greenlon et al., 2019). A fourth integrase, *Int*M which is implicated in γ fragment integration (Haskett et al., 2016b), was variable among strains, being absent in New Sp. 3 as well as in two strains within New Sp. 1 (Figure 2). It is unknown whether genomes lacking *Int*M have novel ICE structure. To discern among the two

scenarios, it would require additional sequencing using long reads technologies such as PacBio or Nanopore.

The Nod factor ligand and T3SS effectors are critical in structuring *Mesorhizobium*-host interactions (Haskett et al., 2016b, Greenlon et al., 2019, Tampakaki, 2014) and thus their provenance and diversity are of interest. Comparisons of functional gene clusters of Nod factor and T3SS for the new species indicates that the most probable origin of the symbiotic island among Pakistani strains was *Mesorhizobium* Clade 5, and more specifically the wild-nodulating *C. reticulatum* strains from Turkey. Interestingly, both phylogenetic analyses and haplotype structure suggest multiple, independent introductions of these plant-associated signaling regulons from distinct wild strain diversity. Consistent with this interpretation, analysis of linkage disequilibrium within ICE regions encoding the Nod factor biosynthesis or conserved T3SS machineries indicate a high degree of small-scale horizontal gene transfer. Interestingly, topological congruence among T3SS and Nod region is maintained in phylogenetic clades 7 and 6 of Turkish origin. By contrast, strains with background genome origins in Clade 5 have substantial variation in Nod and T3SS phylogenies (Figure 5).

The current results agree with a model in which historical introduction of chickpea into new agricultural areas has simultaneously expanded the diversity of chickpea symbionts. According to this model, new symbionts emerge when resident soil bacteria in the genus *Mesorhizobium*, adapted to local environments, acquire the symbiotic island via horizontal gene transfer. The symbiosis island is assumed to originate from wild strains, introduced with the crop but that do not persist. Thus the genetic capacity for chickpea symbiosis invades native microbial populations via the machinery of conjugal transfer and site-specific integration. In the case of Pakistan, the native *Mesorhizobium* diversity is likely the product of hundreds-of-thousands of years of natural selection, followed by thousands of cultivation cycles and crop-rhizobia interactions. Thus these native strains, now carrying chickpea-associated ICE elements, are candidates for traits that govern survival and symbiosis under the arid conditions of the Thal desert.

In addition to the rhizosphere-host interactions, microbial adaptation to soil environments is driven by physical- chemical, and environmental properties (Tecon & Or, 2017). Studies in the hyper-arid Atacama Desert show that indigenous microbial communities have cycles of activity after rare precipitation events (Schulze-Makuch et al., 2018) and that microbiome composition is driven by adaptations to withstand extreme dryness and UV radiation. An important aspect is that soil microbial composition is spatially and temporarily dynamic, in particular in response to rainfall events (Tecon & Or, 2017). It is likely that similar variability in community survival and composition occurs with seasonal changes in the Thal Desert region of Pakistan. Moreover, long-standing agricultural practices, over millennia, have likely further impacted soil biophysical properties, including microbial activity. Thus the novel *Mesorhizobium* diversity reported here may represent only a fraction of a larger community of novel and diverse chickpea rhizosphere microorganisms, potentially extending the range of functional microbial traits for resilience to dry and arid conditions. Such microbial diversity could contain adaptation to seasonal or persistent local conditions, including extreme high temperatures, low and variable precipitation, UV, wind, humidity, and minimal agricultural inputs.

Other researchers have explored extreme environments to identify microbial traits for adaptation to abiotic stress in *Mesorhizobium*, including tolerance to low pH, heat and salt stress. Interestingly, Zaw et al. (Zaw et al., 2022) found that the local soil environment was a better predictor of trait values than was phylogenetic placement of strains. While the four new species of *Mesorhizobium* identified here are promising candidates as inoculants, it will be important to test their performance, i.e., symbiotic efficiency and tolerance to abiotic stress, under a range of conditions: *in vitro*, in greenhouses using controlled inoculation experiments, and ultimately under field conditions.

CONCLUSION

We found that chickpea symbionts associated with arid soils from Pakistan are novel lineages, representing at least four novel species. This diversity is distinct to previously reported

agricultural collections of *Mesorhizobium* spp. Interestingly, the Pakistani symbiont diversity are related to species found in nodules of crop's wild progenitor, *C. reticulatum*, suggesting a narrower introduction scenario compared to other chickpea agricultural regions in the world.

MATERIAL AND METHODS

Soil collection

During the chickpea growing season in 2017, 22 farms were visited. 20 cm depth soil was collected from independent farms located in three districts of the Punjab province, Pakistan. Air-dried samples of 100 grams of soil were triple bagged and shipped to University of California-Davis, USA.

Rhizobia trapping

Each bag containing 100 grams of soil was used as substrate to grow chickpea under greenhouse conditions in Davis, California, USA. Soil was transferred to newly sterilized pots, with sterile coffee filters at the bottom to contain soil. Into each soil, three chickpea seeds (cultivar ICCV96029) were planted and irrigated daily with sterile DI water without fertilizer application. For two out of 22 soils there was no seedling germination, while the remaining 20 soils had from one to three plants flowering at harvest time (37 days after sowing). Root systems were carefully removed from the soil and washed with sterile DI water. Nodules were separated from roots and rinsed several times with DI water, after which they were either processed fresh or stored in sterile tubes containing silica gel desiccant.

Bacterial strain isolation

Individual fresh nodules were surface sterilized, rinsed and crushed with sterile DI water. Nodule extracts were diluted two-fold and dispensed in parallel on petri plates containing yeast mannitol extract agar (YM) or Lupin Agar (LA) and incubated at 28 °C for 12-15 days. YM media contained D-glucose 3g/l, Mannitol 2g/L, Yeast extract 1g/L, K_2HPO_4 0.5 g/L, $MgSO_4.7H_2O$ 0.2g/l,

NaCl 0.1 g/L, CaSO$_4$.2H$_2$O 0.05 g/L, NH4Cl 0.1 g/L and Agar 15 g/L, pH 6.8. LA media contained Mannitol 5g/L, D-glucose 5g/L, Yeast Extract 1.25 g/L, MgSO$_4$&H2O 0.8 g/L, NaCl 0.1 g/L, CaCl$_2$H2O 0.2 g/L, plus K$_2$HPO$_4$ stock 20ml, KH$_2$PO$_4$ stock 20 ml, FeSO$_4$ stock, 10 ml, and 1 ml of trace elements at pH 6.8. The stocks for LA media were prepared as follows: K$_2$HPO$_4$ 0.87g/l, KH$_2$PO$_4$ 0.68 g/L, FeSO$_4$ 0.5 g/L, and trace elements stock of Na$_2$B$_4$O$_4$ 2.34 g/L, MnSO$_4$4H$_2$O 2.03 g/L, ZnSO$_4$7H2O 0.22 g/L, CuSO$_4$5H$_2$O 0.08 g/L, Na$_2$MoO$_4$2H$_2$O 0.13g/L. When colonies were visible they were selected and purified by serial streaking on fresh plates of the same type of media (YE or LA). Purified single colonies were selected and transferred to 3ml of liquid YM or LA media, grown for 72h at 28 °C. Simultaneously, colony polymerase chain reaction (PCR) was performed using primers specific for the gene *Mesorhizobium "ciceris" Nod*C gene (Forward-TGTCATCATCCCCTGCTACA, Reverse- GTTGTTGGCGCAAATAAGGT). Only isolates with positive PCR reactions corresponding to the expected size 680 bp were selected for DNA extraction and library preparation.

Genomic DNA extraction, library preparation, and whole genome sequencing (WGS).

For DNA manipulations, strains were grown in YME liquid cultures. Turbid cultures were centrifuged and the pellets resuspended in 200 ul of fresh YME. Genomic DNA was extracted using the Qiagen blood and tissue DNA extraction kit, following manufacturer instructions with final elution in 60 ul. DNA quality was evaluated using a Nanodrop spectrophotometer, retaining samples with 260/230 ratio lower than 2.0. Thirty ul of genomic DNA was diluted with 20 ul of elution buffer (Quiagen) and subject to clean-up with an equal volume (50 ul) of room temperature Ampure XP beads (Agencourt). Beads with bound DNA were retrieved by exposure to a magnet for 10 minutes (until clear), after which the supernatant was removed and beads were washed 2 times with fresh 80% ethanol. Dry beads were treated with elution buffer and DNA was recovered in the eluate. DNA was quantified using Qubit BR DNA.

Genomic DNA libraries were prepared for sequencing using the KAPA HyperPlus Library Preparation kit according with manufacturer instructions, except that barcodes were incorporated using customized stubby adaptors (Genome center, University of California- Davis) and 5 cycles of

PCR amplification. Libraries were quantified using a Qubit instrument and reagents. Library quality was determined using the Agilent Bioanalyzer to assess fragment size distribution. Each sample was normalized to 6nM before pooling. Pooled libraries (average insert size 594 bp) were submitted for paired-end sequencing (PE 150) on an IlluminaHiSeq4000 at the UC Davis Genome Center.

Assembly and annotation of 11 *Mesorhizobium* genomes.

Retrieved reads where trimmed for low-quality using Trimmomatic 3.8. The genome was assembled using SPADES_3.12.0 with default parameters and Kmer 35. Both steps were performed using Cyverse platform (Merchant et al., 2016). Assembled genomes were annotated using PROKKA 1.14.5 (parameters: minimum contig size 200, genetic code 11).

Phylogenetic inference, average nucleotide identity (ANI) analysis.

Whole genome-based taxonomic analysis was performed by uploading the 11 sequences to the Type (Strain) Genome Server (TYGS), a free bioinformatics platform available under https://tygs.dsmz.de (Meier-Kolthoff & Göker, 2019). The analysis provide genome-based taxonomy using the Genome BLAST Distance Phylogeny approach (GBDP) for genomic sequence-based species delimitation (Meier-Kolthoff et al., 2013).

Average nucleotide identity (ANI) was determined by pair-wise analysis among strains using FastANI (version 1.2, default settings) (Jain et al., 2018). Pair-wise comparisons included the eleven strains isolated here, and also a larger dataset of 216 previously-assembled high quality draft *Mesorhizobium* genomes (Greenlon et al., 2019), many of which lack formal taxonomic nomenclature. Pairwise ANI values from FastANI were re-formatted using custom scripts and used to cluster genomes at 95 percent identity in Mothur using the cluster command (version v.1.44.2) (Schloss et al., 2009).

The amino acid sequences of 400 conserved microbial proteins were used to determine evolutionary relationships using PhyloPlAn (Segata et al., 2013). Phylogenetic analysis was performed in two ways. First, the 11 newly sequenced genomes were compared with PacBio finished genomes for *Mesorhizobium* strains representing each of the ten phylogenetic clades

identified in the genus by Greenlon *et al.* (2019). Then the PhyloPlAn method was used to determine relationships among the 11 new genomes, 216 *Mesorhizobium* Illumina short read assemblies from Greenlon et al (2019), and 20 previously-sequenced *Mesorhizobium* genomes from other studies (Supplemental Table 4).

Symbiotic island structural properties.

To determine the symbiotic island properties and predict the probable organization of the symbiotic island, we followed the criteria stablished by Greenlon et al. (2019), where a monopartite symbiotic island has one homolog of tyrosine recombinase IntS2 (Haskett et al., 2016b) and a tripartite symbiotic island has one or more homologs of integrases IntS1, IntG and IntM. Using BLASTn searches with each integrase gene as query, a presence/absence matrix was developed (Table 1). A presence hit had an e-value zero, with query coverage 100% and pairwise identity ranging from 85.4% to 100%.

Nod and T3SS conserved cluster analysis.

The fourteen PacBio Genomes from *Mesorhizobium* sequenced by Greenlon et al., 2019 were used as benchmark to identify the T3SS gene clusters (Sup. Fig 2). Homology-based searches using the validated T3SS clusters were then performed against a broader set of *Mesorhizobium* draft genomes (Supplementary Table 5), including the eleven strains sequenced here and several metagenomes that were sequenced directly from nodule DNA. Completeness of the symbiosis island region was assessed by comparison to the reference PacBio genomes. A common ~35 Kpb region containing the nine conserved genes for the T3SS apparatus was selected for multiple alignments (Fig 4B). In most cases the nine conserved genes were in a complete contig. For the few genomes in which the target genes were in more than one contig, we concatenated adjacent contigs. The Nod cluster was delimited from NodD to NodABC, with a common region of ~10 Kbp (Fig 4A).

Nucleotide multiple sequence alignments were performed with Geneious alignment-Blosum62 (Gap open penalty 12, Gap extension 3). Trees were computed with RAxML 8.2.11

133

(Stamatakis, 2014), Nucleotide Model GTR GAMMA Rapid hill-climbing search for best-scoring ML tree with1000 bootstrap replicates. Tree topology congruency was tested using weighted Robinson-Fould distance using R packages "phangorn" and "TreeDist". The multiple alignments of the regions containing the conserved T3SS machinery genes and the NodABCD cluster were used for generating variant call format (VCF), using VCF tools for calculating r^2 to asses linkage disequilibrium among variant sites.

REFERENCES

Ahmad A, Khan MR, Shah SHH, *et al.*, 2019. Agro-ecological zones of Punjab, Pakistan. In. *FAO document.* Rome, 68.

Aranjuelo I, Arrese-Igor C, Molero G, 2014. Nodule performance within a changing environmental context. *Journal of Plant Physiology* **171**, 1076-90.

Benjelloun I, Thami Alami I, Douira A, Udupa SM, 2019. Phenotypic and Genotypic Diversity Among Symbiotic and Non-symbiotic Bacteria Present in Chickpea Nodules in Morocco. *Frontiers in Microbiology* **10**.

Broughton WJ, Jabbouri S, Perret X, 2000. Keys to Symbiotic Harmony. *Journal of Bacteriology* **182**, 5641-52.

Brucker RM, Bordenstein SR, 2013. The hologenomic basis of speciation: gut bacteria cause hybrid lethality in the genus Nasonia. *Science* **341**, 667-9.

Caetano-Anollés G, Gresshoff PM, 1992. Plant genetic suppression of the non-nodulation phenotype of Rhizobium meliloti host-range nodH mutants: gene-for-gene interaction in the alfalfa-Rhizobium symbiosis? *Theoretical and Applied Genetics* **84**, 624-32.

Chang JH, Desveaux D, Creason AL, 2014. The ABCs and 123s of bacterial secretion systems in plant pathogenesis. *Annu Rev Phytopathol* **52**, 317-45.

De Lajudie PM, Andrews M, Ardley J, *et al.*, 2019. Minimal standards for the description of new genera and species of rhizobia and agrobacteria. *International Journal of Systematic and Evolutionary Microbiology* **69**, 1852-63.

De Zelicourt A, Al-Yousif M, Hirt H, 2013. Rhizosphere Microbes as Essential Partners for Plant Stress Tolerance. *Molecular Plant* **6**, 242-5.

Dimkpa C, Weinand T, Asch F, 2009. Plant–rhizobacteria interactions alleviate abiotic stress conditions. *Plant, Cell & Environment* **32**, 1682-94.

Flores M, Mavingui P, Girard L, *et al.*, 1998. Three replicons of Rhizobium sp. Strain NGR234 harbor symbiotic gene sequences. *J Bacteriol* **180**, 6052-3.

Greenlon A, Chang PL, Damtew ZM, *et al.*, 2019. Global-level population genomics reveals differential effects of geography and phylogeny on horizontal gene transfer in soil bacteria. *Proc Natl Acad Sci U S A* **116**, 15200-9.

Haskett T, Wang P, Ramsay J, *et al.*, 2016a. Complete Genome Sequence of Mesorhizobium ciceri Strain CC1192, an Efficient Nitrogen-Fixing Microsymbiont of Cicer arietinum. *Genome Announc* **4**.

Haskett TL, Ramsay JP, Bekuma AA, Sullivan JT, O'hara GW, Terpolilli JJ, 2017. Evolutionary persistence of tripartite integrative and conjugative elements. *Plasmid* **92**, 30-6.

Haskett TL, Terpolilli JJ, Bekuma A, *et al.*, 2016b. Assembly and transfer of tripartite integrative and conjugative genetic elements. *Proceedings of the National Academy of Sciences*, 201613358.

Imran A, Mirza MS, Shah TM, Malik KA, Hafeez FY, 2015. Differential response of kabuli and desi chickpea genotypes toward inoculation with PGPR in different soils. *Front Microbiol* **6**, 859.

Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S, 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications* **9**, 5114.

Jarvis BDW, Van Berkum P, Chen WX, *et al.*, 1997. Transfer of Rhizobium loti, Rhizobium huakuii, Rhizobium ciceri, Rhizobium mediterraneum, and Rhizobium tianshanense to Mesorhizobium gen. nov. *International Journal of Systematic and Evolutionary Microbiology* **47**, 895-8.

Juhas M, Van Der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW, 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS microbiology reviews* **33**, 376-93.

Kamboj DV, Bhatia R, Pathak DV, K. SP, 2010. Role of nodD gene product and flavonoid interactions in induction of nodulation genes in Mesorhizobium ciceri. *Physiol. Mol. Biol. Plants* **16**, 77.

Kimbrel JA, Thomas WJ, Jiang Y, *et al.*, 2013. Mutualistic Co-evolution of Type III Effector Genes in Sinorhizobium fredii and Bradyrhizobium japonicum. *PLOS Pathogens* **9**, e1003204.

Kobayashi H, Naciri-Graven Y, Broughton WJ, Perret X, 2004. Flavonoids induce temporal shifts in gene-expression of nod-box controlled loci in Rhizobium sp. NGR234. *Mol Microbiol* **51**, 335-47.

Laranjo M, Young JP, Oliveira S, 2012. Multilocus sequence analysis reveals multiple symbiovars within Mesorhizobium species. *Syst Appl Microbiol* **35**, 359-67.

Latif S, Khan S, Naveed M, Mustafa G, Bashir T, Mumtaz AS, 2013. The diversity of Rhizobia, Sinorhizobia and novel non-Rhizobial Paenibacillus nodulating wild herbaceous legumes. *Arch Microbiol* **195**, 647-53.

Lorite MJ, Estrella MJ, Escaray FJ, *et al.*, 2018. The Rhizobia-Lotus Symbioses: Deeply Specific and Widely Diverse. *Frontiers in Microbiology* **9**.

Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M, 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* **14**, 60.

Meier-Kolthoff JP, Göker M, 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nature Communications* **10**, 2182.

Merchant N, Lyons E, Goff S, *et al.*, 2016. The iPlant Collaborative: Cyberinfrastructure for Enabling Data to Discovery for the Life Sciences. *PLOS Biology* **14**, e1002342.

Merga B, Haji J, 2019. Economic importance of chickpea: Production, value, and world trade. *Cogent Food & Agriculture* **5**, 1615718.

Miwa H, Okazaki S, 2017. How effectors promote beneficial interactions. *Current Opinion in Plant Biology* **38**, 148-54.

Nascimento FX, Brígido C, Glick BR, Oliveira S, 2012. ACC deaminase genes are conserved among Mesorhizobium species able to nodulate the same host plant. *FEMS Microbiology Letters* **336**, 26-37.

Oldroyd GE, Downie JA, 2008. Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu Rev Plant Biol* **59**, 519-46.

Oldroyd GE, Murray JD, Poole PS, Downie JA, 2011. The rules of engagement in the legume-rhizobial symbiosis. *Annu Rev Genet* **45**, 119-44.

Perret X, Staehelin C, Broughton WJ, 2000. Molecular Basis of Symbiotic Promiscuity. *Microbiology and Molecular Biology Reviews* **64**, 180-201.

Sani S, Chang PL, Zubair A, *et al.*, 2018. Genetic Diversity, Population Structure, and Genetic Correlation with Climatic Variation in Chickpea (Cicer arietinum) Landraces from Pakistan. *Plant Genome* **11**.

Schloss PD, Westcott SL, Ryabin T, *et al.*, 2009. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology* **75**, 7537-41.

Schulze-Makuch D, Wagner D, Kounaves SP, *et al.*, 2018. Transitory microbial habitat in the hyperarid Atacama Desert. *Proceedings of the National Academy of Sciences* **115**, 2670-5.

Segata N, Börnigen D, Morgan XC, Huttenhower C, 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nature Communications* **4**, 2304-.

Staehelin C, Krishnan HB, 2015. Nodulation outer proteins: double-edged swords of symbiotic rhizobia. *Biochem J* **470**, 263-74.

Stamatakis A, 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-3.

Sullivan JT, Ronson CW, 1998. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Proc Natl Acad Sci U S A* **95**, 5145-9.

Sullivan JT, Trzebiatowski JR, Cruickshank RW, *et al.*, 2002. Comparative Sequence Analysis of the Symbiosis Island of Mesorhizobium loti Strain R7A. *Journal of Bacteriology* **184**, 3086-95.

Tampakaki AP, 2014. Commonalities and differences of T3SSs in rhizobia and plant pathogenic bacteria. *Front Plant Sci* **5**, 114.

Tecon R, Or D, 2017. Biophysical processes supporting the diversity of microbial life in soil. *FEMS microbiology reviews* **41**, 599-623.

Teulet A, Busset N, Fardoux J, *et al.*, 2019. The rhizobial type III effector ErnA confers the ability to form nodules in legumes. *Proc Natl Acad Sci U S A* **116**, 21758-68.

Teulet A, Camuel A, Perret X, Giraud E, 2022. The Versatile Roles of Type III Secretion Systems in Rhizobia-Legume Symbioses. *Annual Review of Microbiology* **76**, null.

Tsukui T, Eda S, Kaneko T, *et al.*, 2013. The Type III Secretion System of Bradyrhizobium japonicum USDA122 Mediates Symbiotic Incompatibility with Rj2 Soybean Plants. *Applied and Environmental Microbiology* **79**, 1048-51.

Xu L, Naylor D, Dong Z, *et al.*, 2018. Drought delays development of the sorghum root microbiome and enriches for monoderm bacteria. *Proceedings of the National Academy of Sciences* **115**, E4284.

Yang S, Tang F, Gao M, Krishnan HB, Zhu H, 2010. R gene-controlled host specificity in the legume–rhizobia symbiosis. *Proceedings of the National Academy of Sciences* **107**, 18735-40.

136

Zaheer A, Malik A, Sher A, *et al.*, 2019. Isolation, characterization, and effect of phosphate-zinc-solubilizing bacterial strains on chickpea (Cicer arietinum L.) growth. *Saudi journal of biological sciences* **26**, 1061-7.

Zahran HH, 2001. Rhizobia from wild legumes: diversity, taxonomy, ecology, nitrogen fixation and biotechnology. *J Biotechnol* **91**, 143-53.

Zaw M, Rathjen JR, Zhou Y, Ryder MH, Denton MD, 2022. Rhizobial diversity is associated with inoculation history at a two-continent scale. *FEMS Microbiol Ecol* **98**.

TABLES

	intS2 1191 bp		intG 924 bp		intM 1362 bp		intS1	
	Query cover	Per. Ident	Query cover	Per. Ident	Query cover	Per. Ident	Query cover	Per. Ident
Mesorhizobium tamadayense DSM 28320			100%	82.74%				
Mesorhizobium waimense ICMP 19557	98%	50%	99%	69.13%			100%	86.36%
Mesorhizobium sophorae ICMP 19535	100%	84%						
Mesorhizobium sangaii DSM 100039	92%	31%			94%	39%		
Mesorhizobium erdmanii USDA 3471	99%	85%	81%	32.67%	96%	63%		
Mesorhizobium carmichaelinearum ICMP 18942					87%	40%		
Mesorhizobium intechi BD68T	99%	85%	99%	69.45%				
Mesorhizobium japonicum MAFF 303099	99%	85%	99%	69.77%				
Mesorhizobium metallidurans STM 2683	100%	84%	99%	87.58%	94%	39%		
Mesorhizobium helmanticense CSLC115N	99%	86%	81%	33.47%				
Mesorhizobium sanjuanii BSA136	99%	85%	81%	32.67%				
Mesorhizobium mediterraneum USDA 3392	100%	91%	100%	90.88%	94%	39.08%	100%	98.64%
Mesorhizobium wenxiniae WYCCWR 10195	100%	90%	100%	90.55%				
Mesorhizobium temperatum SDW018	98%	92%	100%	88.60%				
Mesorhizobium delmotii STM4623	92%	30%			97%	95%		
Mesorhizobium prunaredense STM4891	99%	84%	100%	82.11%				
Mesorhizobium muleiense CGMCC 1.11022	100%	91%	100%	90.23%				
New sp1- s62pk201	100%	87.4%	100%	97.1%				
New sp1- s67pk5	100%	88%	100%	97.1%				
New sp1- s74pk44	100%	99.8%	100%	99.6%	100%	99.9%		
New sp1- s75pk146	100%	99.8%	100%	99.6%	100%	99.9%		
New sp2- s68pk9	100%	99.8%	100%	99.6%	100%	99.9%		
	intS2 1191 bp		intG 924 bp		intM 1362 bp		intS1	

	Query cover	Per. Ident	Query cover	Per. Ident	Query cover	Per. Ident	Query cover	Per. Ident
New sp3- s56pk21	100%	99.2%	100%	85.5%				
New sp3- s56pk22	100%	99.2%	100%	85.4%				
New sp4- s78pk47	100%	99.8%	100%	99.6%	100%	99.9%		
New sp4-s78pk49	100%	99.8%	100%	99.6%	100%	99.9%		
New sp4-s64pk3	100%	87.9%	100%	97.1%	98.9%	72.4%		
New sp4-s64pk4	100%	87.9%	100%	97.1%	98.9%	72.4%		

Table 1. BLAST analysis of four integrases indicating query coverage and percentage of identity.

Gene	Product	Min.	Max.	Length	Direction
nodD	LysR family transcriptional regulator	161	1078	918	Forward
	TIM barrel protein	1127	1930	804	Reverse
	LacI family DNA-binding transcriptional regulator	2095	3156	1062	Reverse
	Dicarboxylate/amino acid:cation symporter	3432	4718	1287	Reverse
	Sulfotransferase domain-containing protein	5108	5869	762	Reverse
	ABC transporter permease	6153	6941	789	Reverse
nodI	Nodulation factor ABC transporter ATP-binding protein NodI	6945	7859	915	Reverse
nodC	Chitooligosaccharide synthase NodC	8121	9446	1326	Reverse
nodB	Chitooligosaccharide deacetylase NodB	9462	10121	660	Reverse
nodA	NodA family N-acyltransferase	10118	10708	591	Reverse

Table 2.
Annotations and relative positions in multiple alignment for Nod cluster genes

Gene	Product	Min.	Max.	Length	Direction
	Porin	15	1154	1140	Forward
	Acyltransferase family protein	1644	3515	1872	Reverse
CpaD	CpaD family pilus assembly lipoprotein	3600	4163	564	Reverse
	Type II and III secretion system protein family protein	4173	5555	1383	Reverse
	Response regulator transcription factor	5666	6352	687	Reverse
	Hypothetical protein	6960	7448	489	Forward
tdh	L-threonine 3-dehydrogenase	8029	9063	1035	Reverse
	Adenosine kinase	9129	10127	999	Reverse
metK	Methionine adenosyltransferase	10243	11418	1176	Reverse
	Serine hydroxymethyltransferase	11476	12789	1314	Reverse
ArsR	ArsR family transcriptional regulator	12851	13195	345	Reverse
NodX	Nodulation protein NodX	14252	16189	1938	Reverse
NolW	Nodulation protein NolW- Secretin	16491	17189	699	Reverse
NolB	Nodulation protein NolB	17376	17885	510	Forward
sctJ	Type III secretion inner membrane ring lipoprotein SctJ	17895	18758	864	Forward
NolU	Nodulation protein NolU- Inner rod	18760	19398	639	Forward
sctL	Type III secretion system stator protein SctL	19395	20018	624	Forward
sctN	Type III secretion system ATPase SctN	20075	21373	1299	Forward
	Hypothetical protein	21349	21861	513	Forward
sctQ	Type III secretion system cytoplasmic ring protein SctQ	21951	22979	1029	Forward
sctR	Type III secretion system export apparatus subunit SctR	22972	23637	666	Forward
sctS	EscS/YscS/HrcS family type III secretion system export apparatus protein	23667	23915	249	Forward
sctT	Type III secretion system export apparatus subunit SctT	23924	24742	819	Forward
sctU	EscU/YscU/HrcU family type III secretion system export apparatus switch protein	24739	25776	1038	Forward
NopP	Effector protein NopP	26133	26933	801	Forward
	Hypothetical protein	27476	29482	2007	Reverse
	Hypothetical protein	30937	31266	330	Forward
	Hypothetical protein	31698	32588	891	Forward
sctV	Type III secretion system export apparatus subunit SctV	32821	34884	2064	Forward
	Tetratricopeptide repeat protein	34938	35471	534	Forward
MucR	MucR family transcriptional regulator	35938	36408	471	Forward
	Cold-shock protein	36631	36843	213	Reverse
NopA	Nodulation protein NopA- pilus	<31453	31581	>129	Forward

Table 3. Annotated genes in analyzed T3SS cluster region for *Mesorhizobium.* Conserved structural genes in grey.
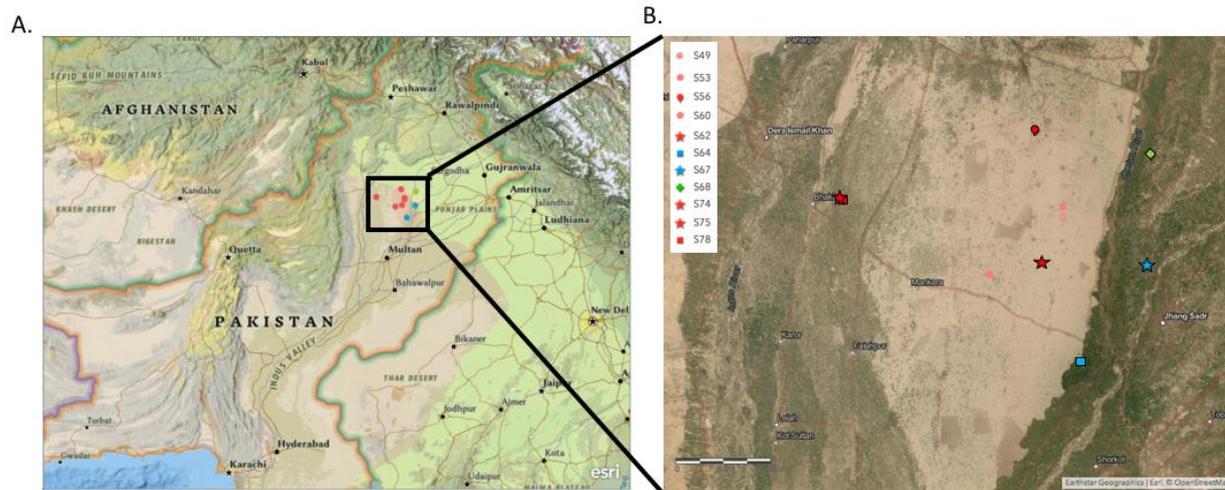
Figure 4.1. Location of agricultural soil in chickpea farms in Pakistan, Punjab province. **A.** Country-level map of Pakistan, colored points represent the different districts: Bhakkar (red), Jhang (blue), Khoshab (green). **B.** Close-up of sampling area, each point shows a soil, filling colors represent districts. Light red circles indicate soils were no rhizobia was recovered and sequenced. Other shapes indicate the novel species of *Mesorhizobium*; novel sp. 1(ρ), novel sp. 2 (◊), novel sp. 3 (∗), novel sp. 4 (☐).Latitude and longitude in supplemental table S1. Maps were generated using ArcGIS Online.
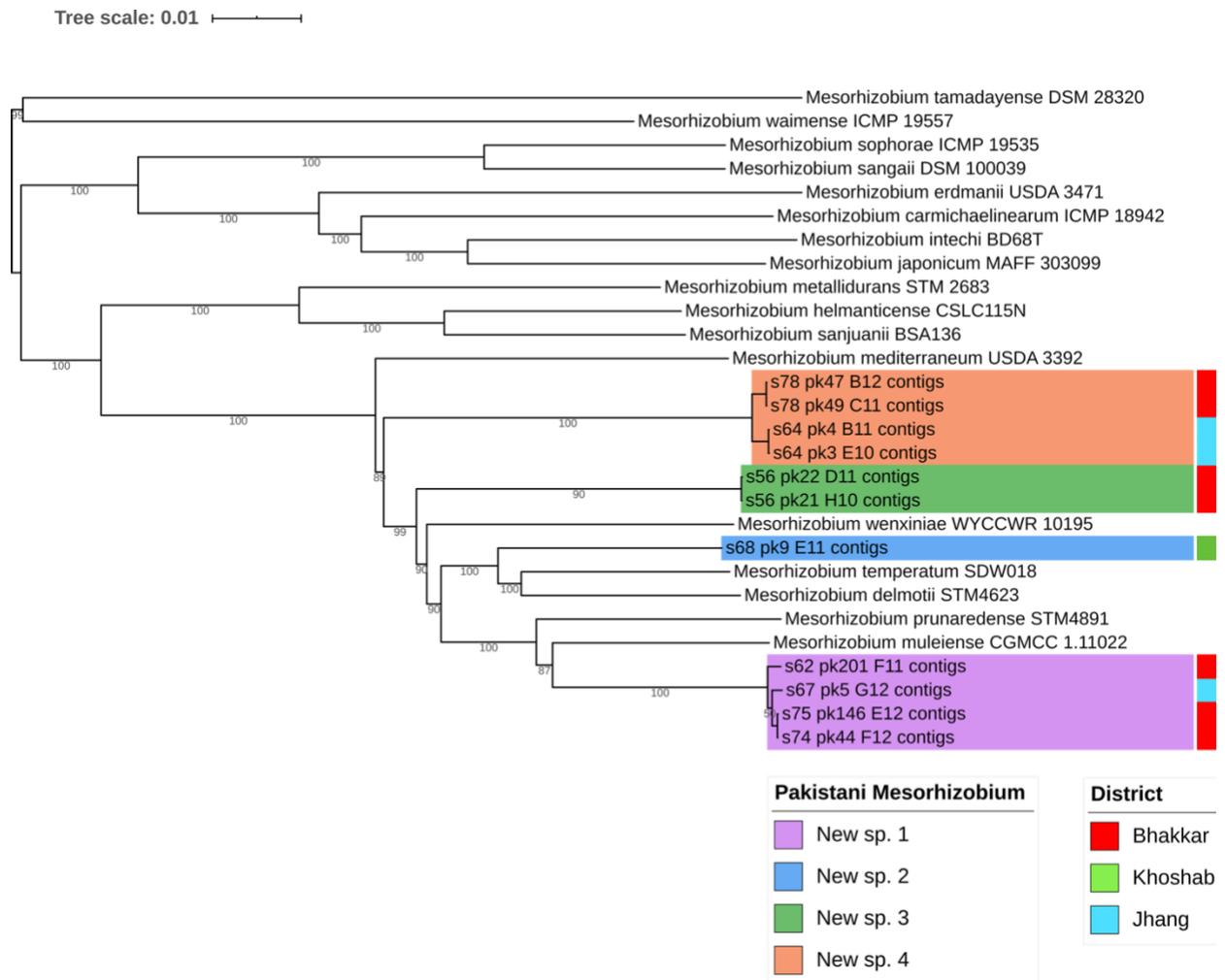
Figure 4.2. Whole genome phylogram of intergenomic distances, among 11 Pakistani genomes and 17 type strains, using the Genome Blast Distance Phylogeny approach (GBDP). Tree was inferred with FASTME 2.1.4. Branches show pseudo-bootstrap support values from 100 replications. Leaf labels represent the supported new species clusters based on dDDH metrics, when compared with type strains database. Full dDDH pairwise comparisons in supplemental table 3.
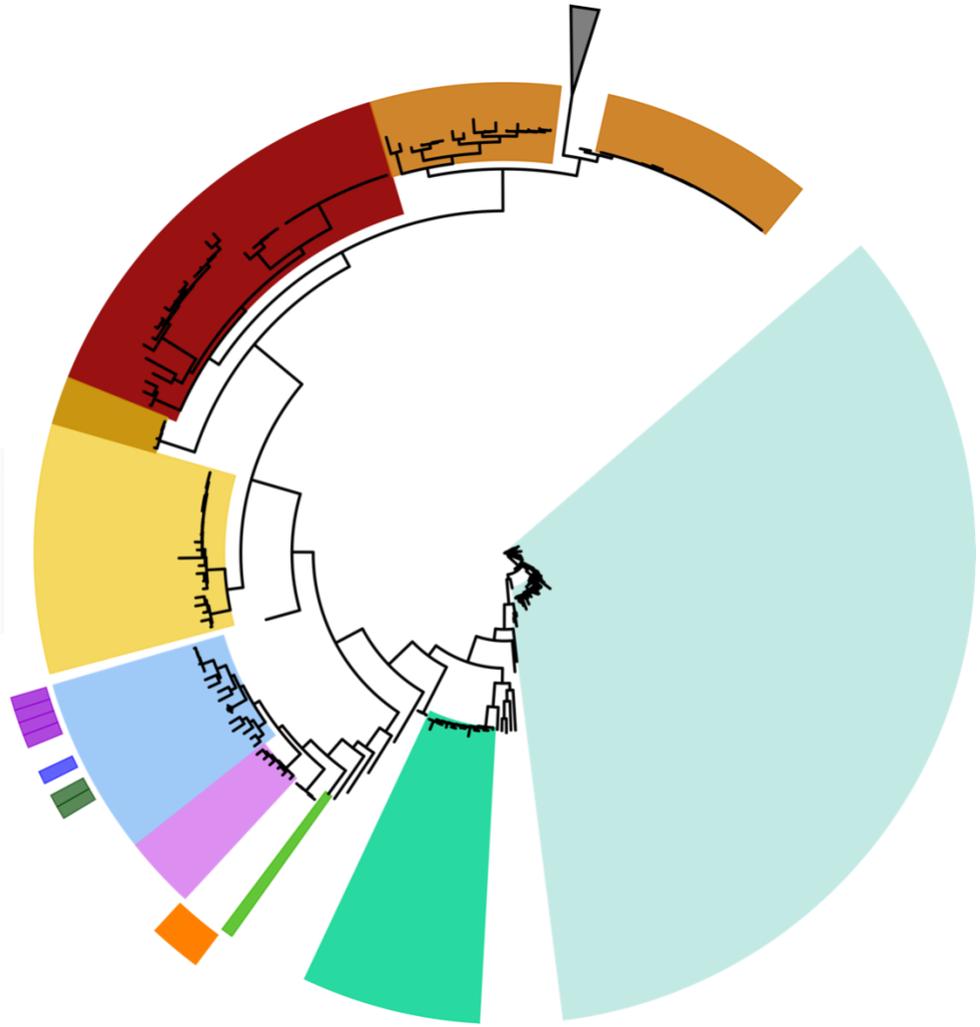
A.

Tree scale: 0.1 ⊢—⊣

ANI-95 Clades
- Clade 1
- Clade 3
- Clade 2
- Clade 4
- Clade 5
- Clade 6
- Clade 7
- Clade 8
- Clade 9

Pakistani Mesorhizobium
- New sp.1
- New sp. 2
- New sp. 3
- New sp. 4

B.

145
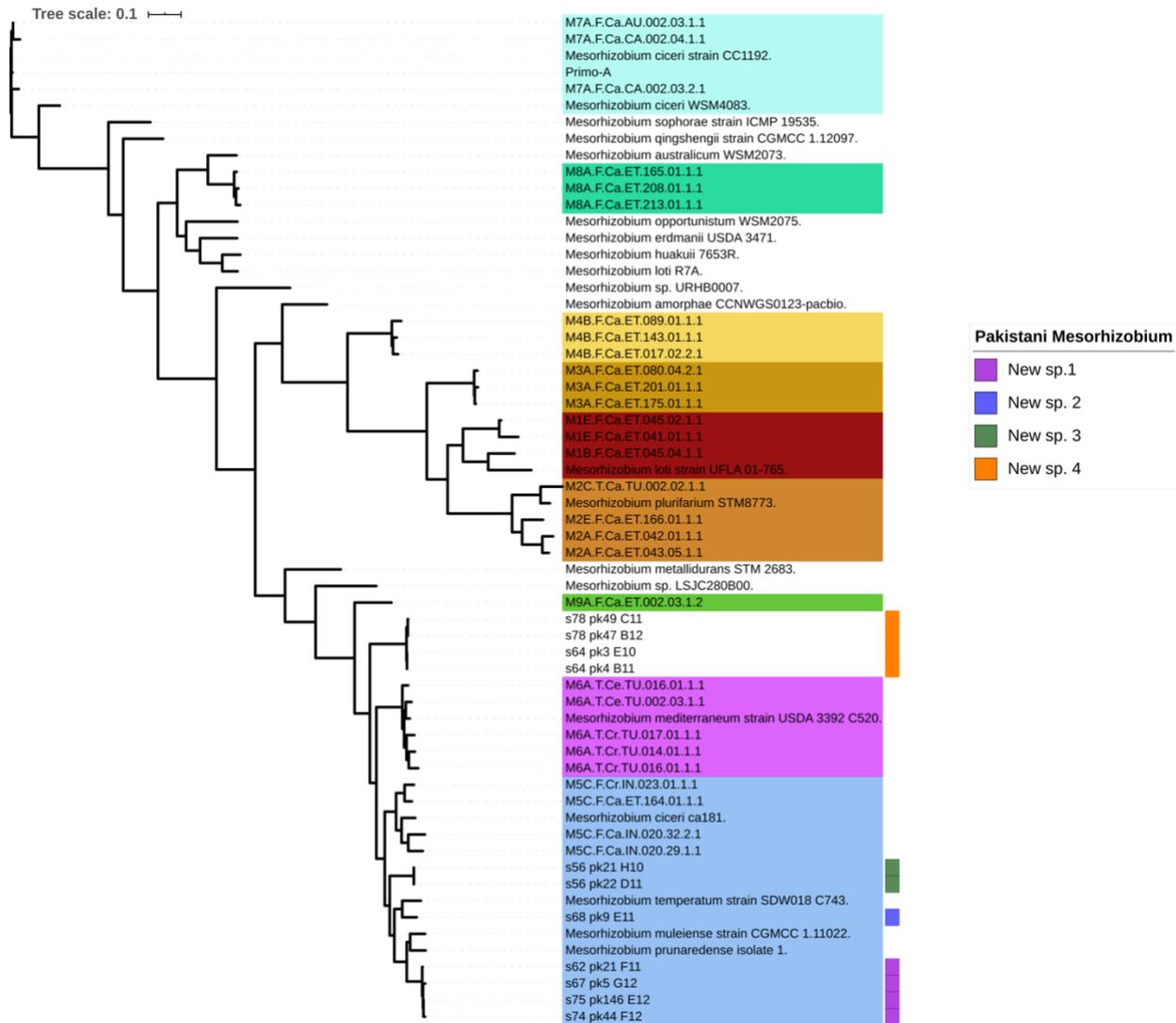
Figure 4.3. Phylogenetic reconstruction of 400 conserved prokaryotic genes (PhyloPlan) for 237 *Mesorhizobium* genomes, including representative strains from the 10 clades identified by Greenlon et al., 2019. **A.** Circular representation of tree constructed with. Grey triangle represent collapsed **B.** Linear representation Trimmed tree, with few genomes per clade of Mesorhizobium genomes, Solid colors represent the closest related clade numbers of the genus Mesorhizobium, based on the ANI-95 analysis of Greenlon et al., 2019. New species from Pakistan, are marked in outer band.
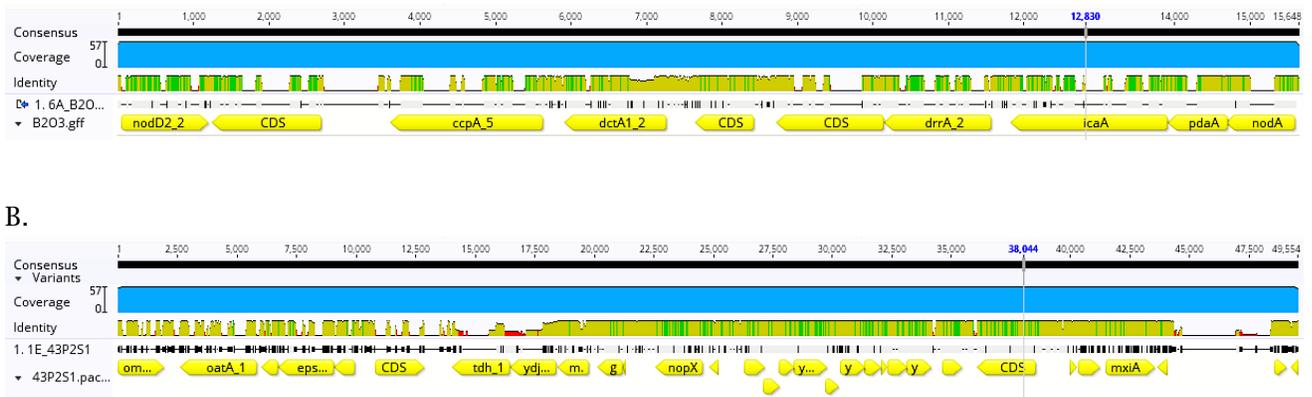
A.

Figure 4.4. Typical gene organization in *Mesorhizobium* genomes (Supplemental table 5), symbionts of Chickpea. For gene clusters associated with host interaction, indicating regions of high and low identity in multiple alignments. **A.** Nod cluster, represented between the NodD transcriptional regulator and NodA, gene annotations table 2. **B.** Region with conserved structural T3SS genes (gene annotations table 3).
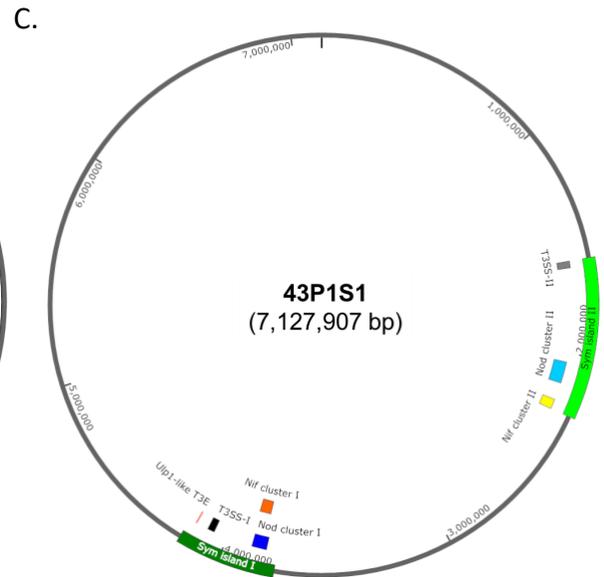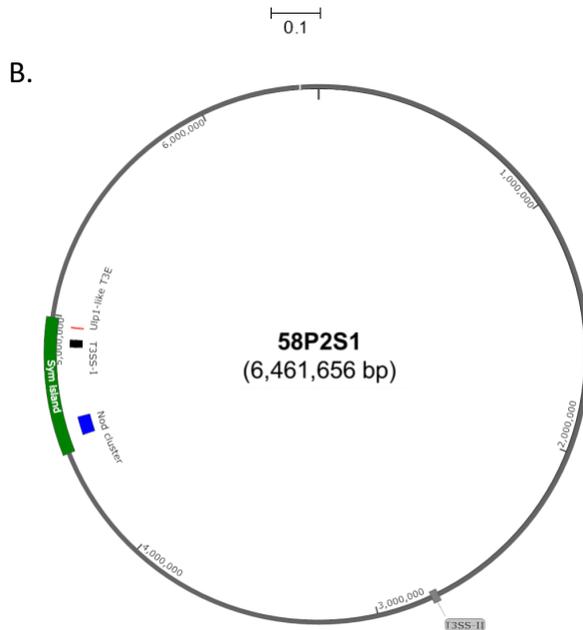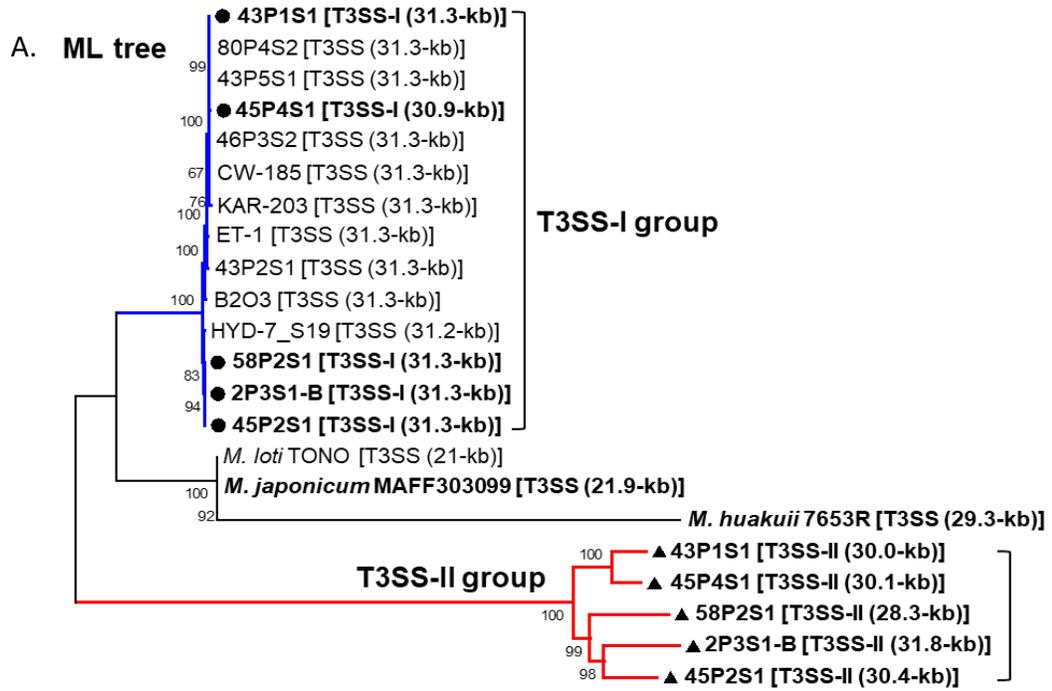
Figure 4.5.
Co-plot for dendrograms obtained for equivalent Nod Factor cluster (Left) and T3SS Cluster (right). The normalized and weighted RF distance for the two topologies is 0.796. Shaded colors represent congruent clustering in the two regions. Asterisks of the same color in each side indicate incongruent clustering among Nod region and T3SS region.

Figure 4.6. Linkage disequilibrium, r² calculated for variant sites of independently obtained multiple alignments for each region, Nod Factor cluster (Left) and T3SS Cluster (right). LD. Axis represent the nucleotide order and positions corresponding
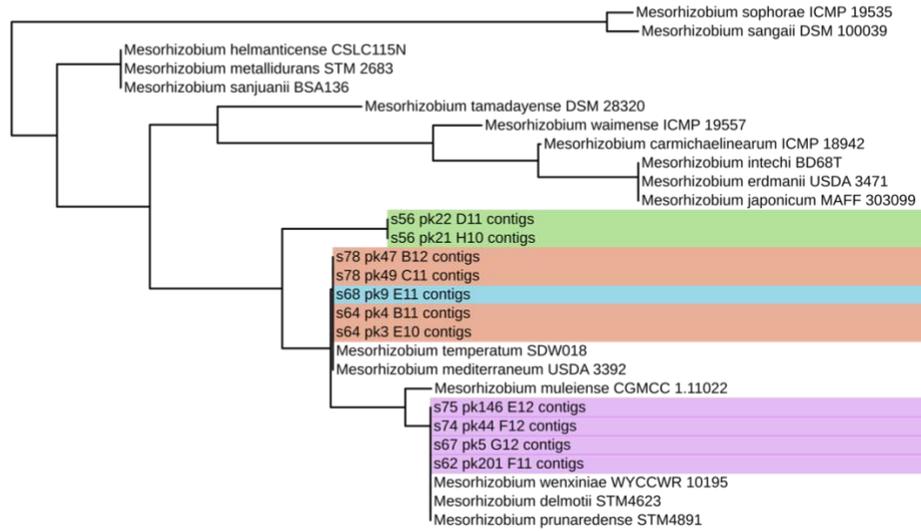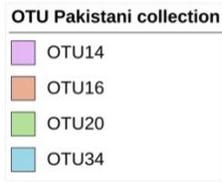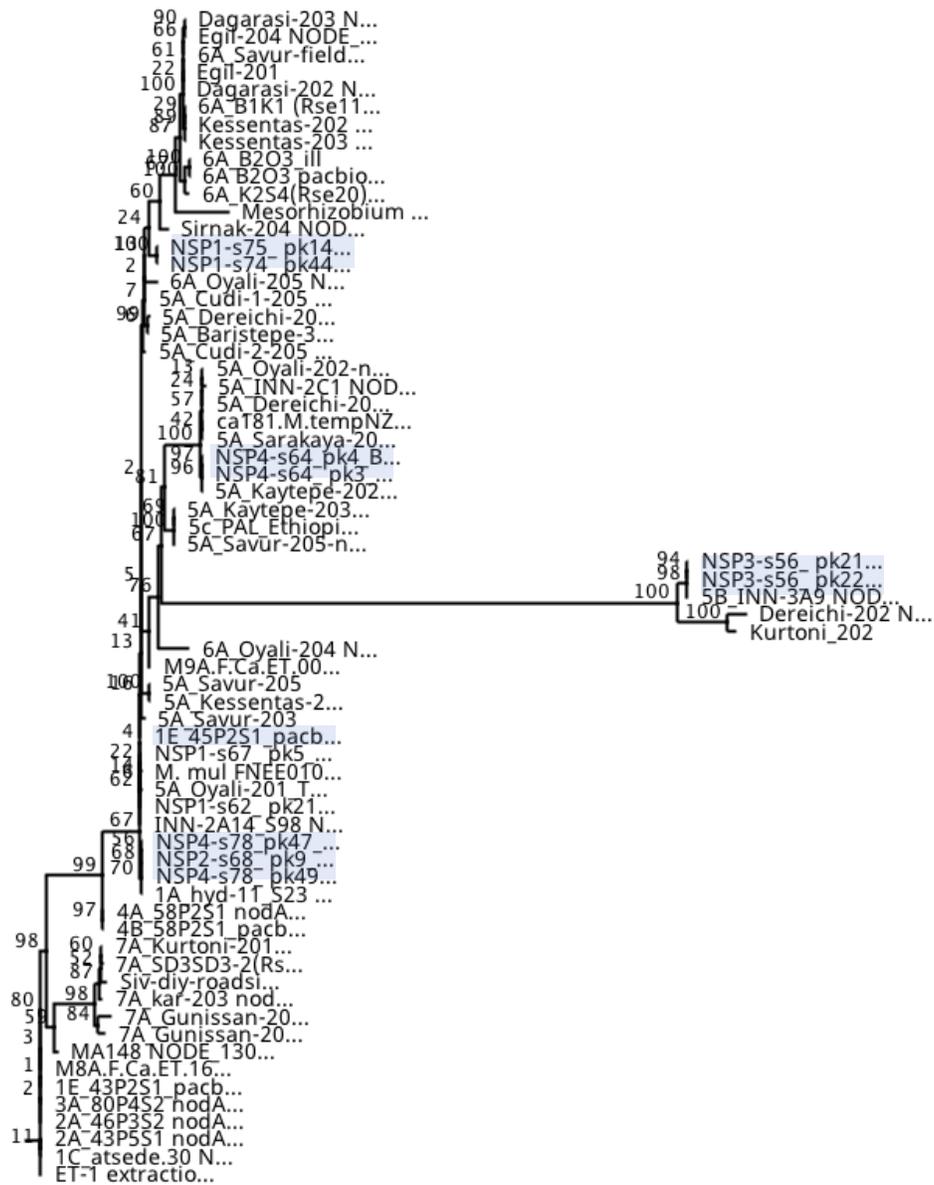
Supplemental Figure 4.1.
Tree of 16S rDNA gene of 11 Mesorhizobium Pakistani strains, compared with Mesorhizobium type strains. Solid colors represent OTU (Operational Taxonomic Units) assignments (Species level) based on our ANI-95 genome-wide analysis.

A. **ML tree**

● 43P1S1 [T3SS-I (31.3-kb)]
80P4S2 [T3SS (31.3-kb)]
43P5S1 [T3SS (31.3-kb)]
● 45P4S1 [T3SS-I (30.9-kb)]
46P3S2 [T3SS (31.3-kb)]
CW-185 [T3SS (31.3-kb)]
KAR-203 [T3SS (31.3-kb)]
ET-1 [T3SS (31.3-kb)]
43P2S1 [T3SS (31.3-kb)]
B2O3 [T3SS (31.3-kb)]
HYD-7_S19 [T3SS (31.2-kb)]
● 58P2S1 [T3SS-I (31.3-kb)]
● 2P3S1-B [T3SS-I (31.3-kb)]
● 45P2S1 [T3SS-I (31.3-kb)]
M. loti TONO [T3SS (21-kb)]
M. japonicum MAFF303099 [T3SS (21.9-kb)]
M. huakuii 7653R [T3SS (29.3-kb)]

T3SS-I group

T3SS-II group

▲ 43P1S1 [T3SS-II (30.0-kb)]
▲ 45P4S1 [T3SS-II (30.1-kb)]
▲ 58P2S1 [T3SS-II (28.3-kb)]
▲ 2P3S1-B [T3SS-II (31.8-kb)]
▲ 45P2S1 [T3SS-II (30.4-kb)]

0.1

B.

58P2S1
(6,461,656 bp)

C.

43P1S1
(7,127,907 bp)

151

Supplemental Figure 4.2.
Chickpea *Mesorhizobium* symbionts display divergent symbiotic island context. **A.** Maximum Likelihood tree from fourteen Pacbio genome *Mesorhizobium* chickpea symbionts. T3SS gene clusters including *M. japonicum* MAFF303099, *M. loti* TONO, and *M. huakuii* 7653R. Bootstrap values are expressed as percentages of 1,000 replications. The chickpea symbionts possessing at least two T3SSs are highlighted by black dots (for T3SS-I) or triangles (for T3SS-II).
**B.** 58P2S, typically recruiting two T3SSs and one *nod* gene cluster. **C.** 43P1S1; two T3SSs and two *nod* gene clusters. The symbiosis islands (dark green), T3SSs (T3SS-I group, black; T3SS-II group, grey), Nod gene clusters (blue), and Ulp1-like T3 effectors (T3E, red) identified outside T3SSs are shown.
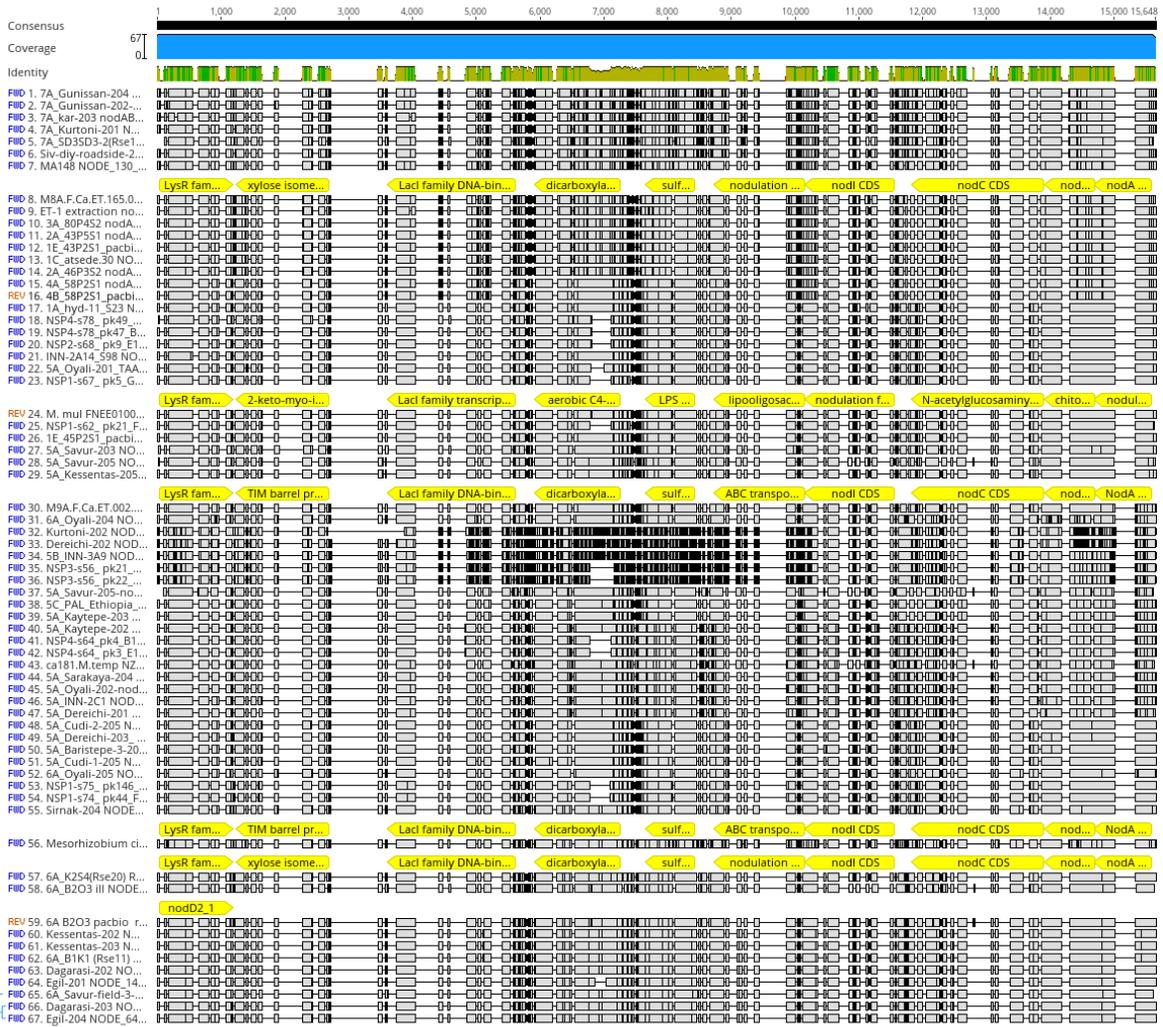
Supplemental Figure 4.3.
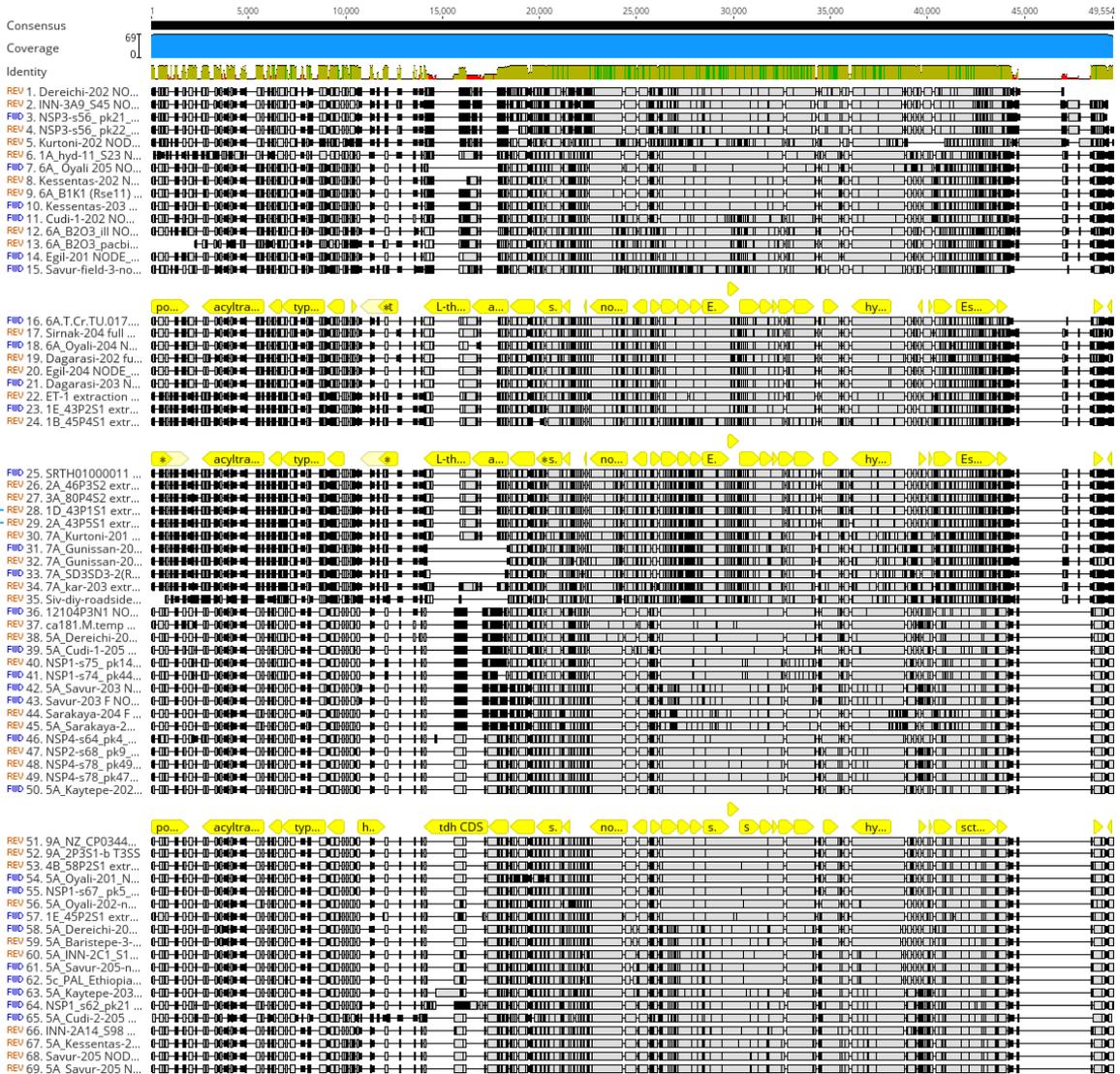Nod Cluster region. RAxML Tree 1000 bootstrap for 67 genomes of chickpea symbionts.
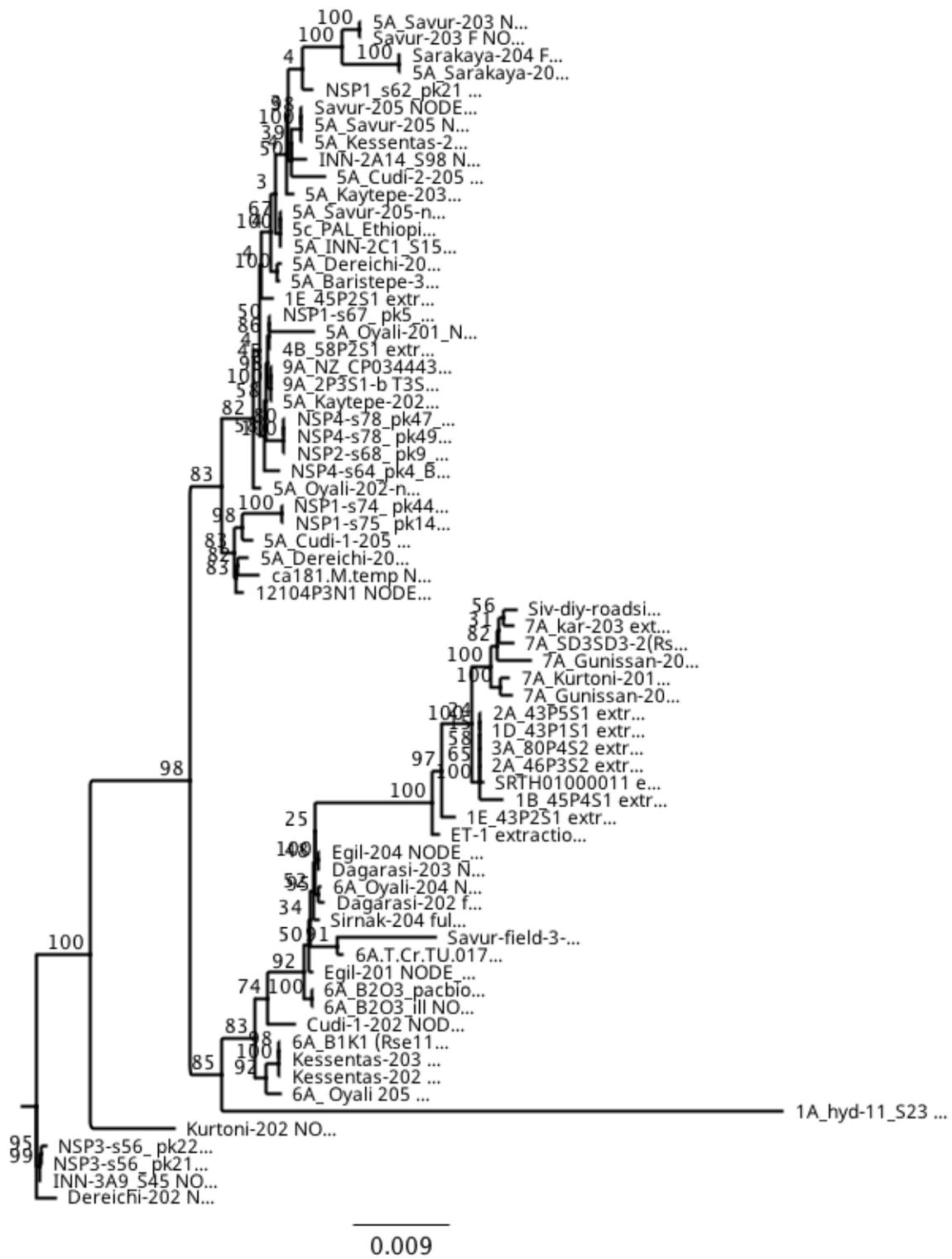
Supplemental Figure 4.4.
RAxML Tree 1000 bootstrap of T3SS region for 69 genomes of chickpea symbionts.

Supplemental Figure 4.5.
Multiple alignment Nod region (10 genes) with few examples of annotated sequences

Supplemental Figure 4.6.
Multiple alignment T3SS region (33 genes), with few examples of annotated sequences