# Genetics and Genomics of Symbiotic Nitrogen Fixation in Legumes: Past, Present and Future

MICHAEL UDVARDI* , CELINE MENS* &
ESTELLE GRUNDY*,**

*Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia Brisbane QLD 4072, Australia. Email: m.udvardi@uq.edu.au
**Integrative Legume Research Group, School of Agriculture and Food Sustainability, The University of Queensland, St Lucia, Brisbane, QLD 4072, Australia

Legumes are a large and diverse family of plants that provide us with food, feed, fuel and feedstocks for industry. They can use atmospheric di-nitrogen for growth, via symbiotic nitrogen fixation (SNF) with bacteria called rhizobia, making them key to sustainable agricultural systems. There are opportunities to increase SNF in legumes to help tackle critical challenges related to the overuse of fertilizer nitrogen in agriculture. The last two decades have seen enormous progress in our understanding of the genetics of SNF, although this is yet to be leveraged to improve SNF in legumes. In principle, two main plant-based approaches exist to improve SNF, one involving genetic engineering and the other using existing natural variation in this complex trait. These approaches are not mutually exclusive and now is an opportune time to attempt to increase SNF in legumes via plant genetics and genomics.

## Introduction

Legumes are a large and diverse family of plants that provide us with sustainable sources of food, feed, fuel and feedstocks for industry (Foyer *et al.* 2016; Semba *et al.* 2021). Well-known legumes include: the common garden pea, which was instrumental in the work of Mendel that established the field of genetics; soybean,

the most important legume on earth in terms of total seed production; other food legumes such as peanut, various types of bean, chickpea, lentil, liquorice, carob; forage species such as alfalfa/lucerne and clover; and many species of trees, such as acacias in Australia. Soybean is an important annual oil seed crop, but another legume, the perennial tree *Pongamia pinnata*, is emerging as a potential source of oil for biodiesel and sustainable aviation fuel, especially as it can be grown on land unsuitable for food crops (Scott *et al.* 2008).

Legumes play key roles in natural ecosystems because of their ability to form nitrogen-fixing symbioses with bacteria called rhizobia, which represent a major source of useful nitrogen (N) for terrestrial ecosystems (Canfield *et al.* 2010; Gou *et al.* 2023). For the same reason, legumes were crucial to the emergence and/or persistence of agriculture in otherwise N-depauperate soils. Indeed, legumes remain essential for N-supply in sustainable agricultural systems, even though they have been sidelined somewhat in conventional agricultural systems by the massive use of synthetic N-fertilizers. Amounting to well over 100 million tonnes of N per year, fertilizer-N helped fuel the Green Revolution and subsequent massive growth in the human population over the past 60 years (Smil 2001). Synthetic fertilizer-N use is a double-edged sword, however, as it generates massive amounts of carbon dioxide during its production, distribution and application, and another, 300-times more potent greenhouse gas, nitrous oxide, when metabolized by soil microbes (Snyder *et al.* 2009; Pan *et al.* 2022). Additionally, about half of N-fertilizers applied to agricultural fields globally are not used by the target crop and are lost to the environment via leaching, erosion and gaseous emissions, which wreaks havoc with natural ecosystems and human health (Sutton *et al.* 2011; Steffen *et al.* 2015). This brings us back to legumes, which by virtue of SNF have much higher intrinsic nitrogen use efficiency (NUE: the ratio of N captured in plant products divided by N-inputs) in cropping systems, e.g., 80% for soybean, compared with other crops such as cereal, which hover around 40% (Zhang *et al.* 2015). Ultimately, this means that less nitrogen is lost from agricultural soils because of legume cultivation than from other crops, such as cereals (Udvardi *et al.* 2021).

Despite their ability to use atmospheric nitrogen for growth, legumes prefer soil nitrogen for energetic reasons, which typically supplies about 30–60% of total plant-N (Herridge *et al.* 2008). Thus, legumes in agriculture derive about 40–70% of their N from the atmosphere, well short of the theoretical maximum of 100% that can be approached under experimental conditions when no N is provided in the growth substrate. Whatever soil-N is removed with harvested plant parts is no longer available to subsequent crops. Of course, the same is true of nitrogen derived from the atmosphere (Ndfa), although some of this N remains in the soil system in the form of crop residue. Thus, legume cultivation results in a net gain of soil-N in many systems (Herridge *et al.* 2008). Nonetheless, given the gap between potential and actual %Ndfa achieved by legumes in agriculture, there are opportunities to increase this net gain further by improving SNF in ways we explore below.

### Genetics and Genomics of SNF

Symbiotic nitrogen fixation (SNF) in legumes is a complex trait that involves the development of specialized organs called nodules, which generally form on roots (Brewin 1991; Ferguson *et al.* 2010). During nodule development, rhizobia bacteria typically enter plant cells via 'infection threads' in epidermal root hair cells that provide access to underlying cortical cells, which accommodate and feed the bacteria as they multiply and eventually differentiate into nitrogen-fixing 'bacteroids'. In exchange for ammonium derived from nitrogen fixation, the plant provides its microsymbionts with all the nutrients required for growth, including organic compounds that are used for energy and biosynthesis and inorganic elements such as iron, sulphur and molybdenum that are required for the production of enzymes such as nitrogenase, which catalyses nitrogen fixation (Udvardi and Poole 2013). Ultimately, ammonium produced by bacteroids is assimilated into amino acids and other compounds by host plant cells before export from nodules to the rest of the plant.

Thousands of plant and bacterial genes are expressed during nodule development and differentiation and presumably are required for SNF (Colebatch *et al.* 2004; Benedito *et al.* 2008). Indeed, hundreds of plant genes have been shown to be necessary for SNF, via genetic studies (Roy *et al.* 2020). These mostly rely on loss of gene function, from mutations in the DNA or interference in gene expression at the level of RNA (RNAi), to illuminate the roles of genes based on aberrant phenotypes (Szczyglowski *et al.* 1998; Penmetsa and Cook 2000; Ott *et al.* 2005; Tadege *et al.* 2008; Fukai *et al.* 2012; Zhang *et al.* 2022). Genes involved in various aspects of nodule development and function have been characterized genetically, including: (i) initial chemical signalling between the plant and rhizobia, which determines partner compatibility, and subsequently triggers gene expression and cell division in the plant, initiating nodule development; (ii) cell biological processes required for bacterial entry and accommodation in plant cells; (iii) nitrogen-dependent repression and autoregulation of nodule development, which restricts and optimizes the number of nodules produced; (iv) nutrient transport and nodule metabolism that enables metabolic cooperation between the symbiotic partners; (v) nodule oxygen homeostasis to establish the low-oxygen environment required by oxygen-labile nitrogenase; and, finally, (vi) nodule senescence, which can shut down nitrogen fixation in response to increased soil-N, environmental stress, and internal developmental cues, and enables recycling of resources to the rest of the plant (reviewed in Roy *et al.* 2020).

Discovery of genes involved in SNF was largely facilitated by whole genome sequencing, assembly and annotation of the model legumes, *Medicago truncatula* and *Lotus japonicus* (Sato *et al.* 2008; Young *et al.* 2011) and of crops such as soybean (Schmutz *et al.* 2010). Genome sequence information accelerated map-based cloning of genes, identification of mutations caused and 'tagged' by foreign DNA sequences, and selection of genes for 'reverse genetics', in which individuals with defects in specific genes or their expression are isolated before determining the mutant phenotype (Roy *et al.* 2020).

Given the complexity of SNF, the thousands of genes involved and the countless interactions between the products of these genes and the processes they control, it is little wonder that genetic knowledge of SNF has not yet been harnessed to increase SNF effectiveness in crop or pasture legumes. Indeed, with so many 'players', it is hard to know where to 'place your bets', at least when it comes to contemplating engineering approaches. We will come back to this later, but next consider more conventional approaches to plant improvement that use natural variation in traits of interest.

### Natural Variation and Opportunities to Improve SNF

Decades of research have shown that legume SNF effectiveness (quantified as Ndfa) varies with changes in the environment (E), including edaphic or soil factors and climatic conditions (Sulieman and Tran 2016; Santachiara *et al.* 2019), and with the strain of rhizobia chosen as partner (rhizobial genotype, Gr; Mendoza-Suárez *et al.* 2020; Westhoek *et al.* 2021). The resulting knowledge has guided agricultural management practices (M) to optimize SNF, or at least growth and yield of legume crops and forages (GRDC). Few studies have examined the plant genetic (Gp) contribution to SNF effectiveness, although it impacts SNF effectiveness through interactions with the other factors (conceptually, $SNF = Gp \times Gr \times E \times M$). Recent studies with collections of diverse genotypes of several species have found variation in SNF associated with plant genotype and have mapped genetic loci, or Quantitative Trait Loci (QTL) for this complex trait in common bean, soybean and peanut (Kamfwa *et al.* 2019; Yang *et al.* 2019; Bazzer *et al.* 2020; Thilakarathna *et al.* 2021; Nzepang *et al.* 2023; Krueger *et al.* 2024). Interestingly, breeding for increased seed yield of soybean in Canada over 100 years has done little to increase % Ndfa or Ndfa per plant of current versus older varieties (Thilakarathna *et al.* 2021). Given available natural variation in SNF effectiveness within species, there are clearly opportunities to increase %Ndfa and total Ndfa associated with these crops, especially if Ndfa is measured directly or estimated accurately in the process, rather than simply relying on seed biomass/yield or even total seed N as a proxy for SNF.

One approach that we are taking begins with public 'core' or 'mini-core' collections of genotypes that represent much of the genetic diversity of the species, such as the international mungbean mini-core collection of 296 genotypes (Schafleitner *et al.* 2015). These lines are then phenotyped for SNF and related traits and genotyped to identify sequence variation associated with variation in traits of interest. 'Genomic prediction' models can then be generated that predict the performance of individuals based on the specific set of sequence variants, or haplotypes, they possess across the whole genome. These models are based on the phenotypic and genotypic profiles of a 'training set' of lines analysed initially, e.g., a mini-core collection or part thereof. The models can be tested by their ability to predict the performance of lines not in the training set based on their genotypes. Ultimately, such models can be used to select parents for crossing that are likely to

generate offspring with more favourable sets of haplotypes, to optimize performance. This process can be aided by artificial intelligence to select a series of crosses over multiple generations to obtain an optimal 'stack' of haplotypes to increase %Ndfa, total Ndfa, etc., ideally without compromising yield (Hickey *et al.* 2019; Hayes *et al.* 2023). The latter can be avoided by including yield data in prediction models and selecting for multiple desired traits in parallel (Hayes *et al.*, 2023).

Attempts to harness novel variability in legume species for plant breeding are being aided by pan genome projects. A pan genome is the collective set of DNA sequences of all or a representative subset of individuals of a species, much of which is shared between all individuals (the core genome), while a significant fraction, ranging from 4.8% in narrow-leafed lupin to 49.9% in soybean, for example (Garg *et al.* 2022; Liu *et al.* 2020), is present in some but not all members of the species (the ancillary genome representing presence–absence variations (PAV) and gene copy number variation (CNV)). Initial attempts to assemble pan genomes relied on a single, high-quality reference sequence, which was used to identify core and ancillary DNA sequences in other individuals of the species, mainly using relatively short sequences of a few hundred base pairs from high-throughput sequencing instruments. Such an approach leaves a lot of ancillary genome sequences 'orphaned' and unable to be positioned on specific chromosomes. Very recently, however, high-throughput, long-read (thousands to tens of thousands of base pairs) sequencing has made it possible to assemble, *de novo*, entire genomes of multiple, diverse individuals of a species (Sharma *et al.* 2022). For example, we recently sequenced eight diverse genomes of mungbean, using HiFi PacBio sequencing, which each assembled into the 11 chromosomes of mungbean with cumulative sequences ranging from 512 to 577 Mbp covering more than 99% of expected conserved genes (Mens *et al.* unpublished). The resulting mungbean pan-genome will aid our efforts to improve SNF in this species, via predictive plant breeding, as outlined above.

## Control of SNF and Opportunities to Engineer It

Biological nitrogen fixation, which takes place exclusively in prokaryotes (eubacteria and archaebacteria), requires biological energy (adenosine triphosphate (ATP) and reducing equivalents (e-), see reaction below) and oxygen levels within the organism that are orders of magnitude lower than ambient levels, because nitrogenase enzyme complexes are rapidity inactivated by free oxygen.

Nitrogenase reaction: $N_2 + 8\,H^+ + 8e^- + 16\,ATP \rightarrow 2NH_3 + H_2 + 16\,ADP + 16\,P_i$

where ADP is adenosine diphosphate and $P_i$ is a phosphate. The oxygen-labile nature of nitrogenase causes a conundrum for many nitrogen-fixing diazotrophs because they require oxygen for respiratory metabolism and energy production. This 'paradox' has been resolved in multiple, interesting ways via evolution. These include barriers to reduce oxygen diffusion into cells, rapid respiration to maintain steady-

state levels of oxygen low enough for nitrogenase to function, and spatial or temporal separation of nitrogen fixation from photosynthesis, which generates oxygen in photoautotrophic diazotrophs (Robson and Postgate 1980). In nitrogen-fixing legume nodules, the plant and resident rhizobia work together to achieve nanomolar concentrations of oxygen in infected cells, as opposed to the equilibrium level of 250 micromolar oxygen in sterile water. This is achieved via a combination of gaseous diffusion resistance in the outer cell layers of nodules, high concentrations of the oxygen-binding and transporting protein leghemoglobin, and high rates of respiration in plant and rhizobial cells within nodules (Ott *et al.* 2005; Layzell and Hunt 1990; Appleby 1984; Dakora and Aitkins 1989; Bryan *et al.* 1988). Leghemoglobins are the most abundant plant proteins in nodules, and the corresponding gene families in legumes have expanded relative to those in non-legumes, presumably because extra copies were selected for their crucial role in nodule oxygen homeostasis. Indeed, the leghemoglobin genes are the archetypal nodulin genes, or genes expressed specifically in nodules. It remains to be seen whether some or all legumes have evolved optimal levels of leghemoglobin for maximal SNF. It will be interesting to determine if there are differences in leghemoglobin levels in different genotypes of the same species and whether these correlate with SNF performance. If there is latitude to further optimize leghemoglobin levels in some species, this could be done by plant breeding using natural variation, and/or by engineering projects that increase the number or activity of leghemoglobin genes. The latter could conceivably be done by adding genes to plant genomes or by genome editing existing genes to increase their expression, for example.

Typically, diazotrophs control the expression of nitrogenase genes in several ways, including repression by oxygen and repression by alternative sources of N, to avoid wasting resources and energy on a process that either cannot proceed or is not needed (Roberts and Brill 1981; Merrick and Edwards 1995). The curious thing about rhizobia is that although they retain oxygen control over nitrogenase gene expression (Fischer 1994; Rutten and Poole 2019), they appear to have lost the capacity to repress nitrogenase expression in response to alternative sources of N (Udvardi *et al.* 1992; Udvardi and Day 1997). This is great for legumes, as they are able to 'corner' rhizobia into fixing nitrogen and into releasing ammonia to host cells in nodules, by controlling oxygen levels and a strict diet of carbon compounds they deliver to their captive rhizobia (Ott *et al.* 2005; Schulte *et al.* 2021). Despite the apparent loss of bacterial N-control over SNF, plants impose their own layer of N-control on the process, by repressing nodule development and/or nodule persistence when sufficient N is available in the soil (Figure 1). Control of nodulation is affected both locally and systemically via a system involving rhizobia- and/or nitrate-responsive signalling peptides, called CLAVATA3/endosperm-surrounding region-related (CLE) peptides that interact with leucine-rich receptor-like kinases in the root or shoot during autoregulation of nodulation (AON) or nitrate-dependent regulation of nodulation respectively (as reviewed in Ferguson *et al.* 2019). This results in lower abundance of the micro-RNA miR2111, which under conditions allowing for
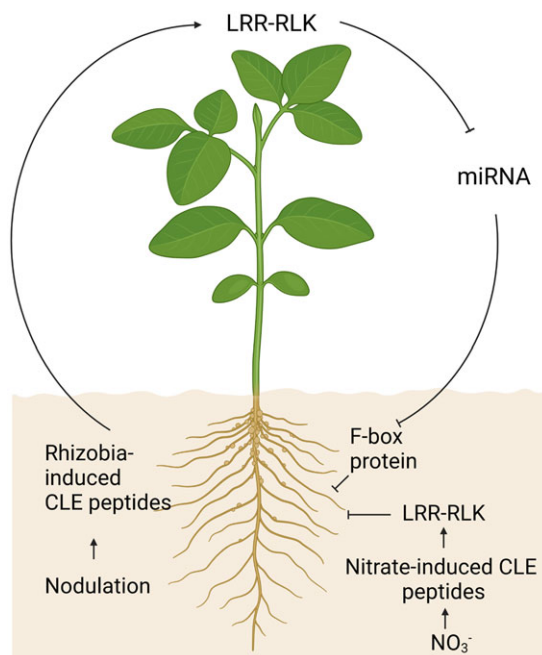
**Figure 1.** Autoregulation and nitrate repression of nodulation in legumes. CLE peptides produced in roots and/or nodules in response to rhizobia (e.g., GmRIC1/2, MtCLE13/35, LjCLE-RS1/2) or to soil nitrate (e.g., GmNIC1, MtCLE35, LjCLE-RS2) act locally within roots in the case of nitrate repression, or systemically through the autoregulation of nodulation (AON) pathway, via CLE receptors belonging to the leucine-rich repeat receptor-like kinase (LRR-RLK) family to repress further nodulation. In the case of AON at least, perception of the CLEs in the shoot results in a decrease of active miRNA (miR2111) that targets transcripts of the F-box protein, too much love (TML) for degradation in the roots. TML is part of the 26S proteasome pathway and targets an unknown positive regulator of nodulation for degradation, thereby inhibiting nodulation.

nodulation targets the mRNA of too much love (TML, a root regulator in *Lotus japonicus*) for degradation (Tsikou *et al.* 2018; Zhang *et al.* 2021). This F-box protein in turn targets an unknown regulator of cell division and nodule development for degradation, resulting in inhibition of nodulation (Takahara *et al.* 2013). The miR2111 was first characterized as a regulator of phosphate starvation suggesting a more general role in plant nutrition. In this way, the presence or addition of sufficient soil-N inhibits nodule development, sparing the associated plant resources for growth elsewhere.

Legumes are also able to put the brakes on metabolic activity and energy consumption of mature nodules in response to sufficient soil-N, via nitrogen-responsive genes that orchestrate nodule senescence. Senescence is a genetically controlled developmental process that allows plants to recycle cellular constituents and reuse them for growth elsewhere. Nodule senescence is triggered by several

environmental and endogenous plant signals (Swaraj and Bishnoi 1996; Kazmierczak *et al.* 2020; Wang *et al.* 2023), including signals related to soil-N or reproductive development when the demand for N and other nutrients to support seed production is high and vegetative tissues are being sacrificed to supply this demand. Recently, genes encoding transcription factors that regulate nodule senescence (NAC, acronym derived from three reported proteins that contain a highly conserved domain in their N-terminal region) have been described (Wang *et al.* 2023; Yu *et al.* 2023).

Knowledge of mechanisms and genes underpinning N-repression of nodulation and activation of nodule senescence opens the possibility of optimizing nodule numbers and/or prolonging their longevity to increase SNF, especially in agricultural soils with relatively high levels of available soil-N. For instance, by reducing or nullifying the activity of specific genes involved in N-regulation of nodule development, such as one or more of the CLE genes, via genome editing, it may be possible to maintain high levels of nodulation and SNF under conditions of relatively high soil-N. This could potentially spare soil N for subsequent crops, forcing the legume to utilize atmospheric-N instead. While this might be expected to reduce legume productivity slightly, the overall economic, environmental and social benefits to the system, including reduced need for fertilizer-N and concomitant losses to the environment might well outweigh the costs.

Although there are other potential targets for optimizing and increasing SNF, such as control of carbon import and distribution by sugar and organic-acid transporters in nodules and control of nitrogen removal from nodules via amino acid or ureide transporters (Udvardi and Poole 2013; Tegeder 2014) the quantitative nature of SNF, the complexity of gene interaction networks underlying SNF and the environmental challenges that plants face in agricultural systems suggest that the potential for large increases in SNF in the field from changes in one or two genes will be small in most cases. A possible exception to this might be engineering/editing of genes controlling nodule development and longevity in response to soil N. Conceivably, uncoupling nodule development and nitrogen fixation from control by soil-N, which would undermine the competitiveness of legumes in natural ecosystems, could lead to large increases in SNF under high soil-N conditions under non-competitive cropping conditions.

## Looking Forward

To summarize, two decades of genetic and genomic research have given us deep insight into the biochemistry and molecular and cell biology of legume nodule development and symbiotic nitrogen fixation. This, together with the recent development of legume pan genomes, the demonstration of natural variation in SNF effectiveness within species and advances in genome editing, genome-based predictive breeding and artificial intelligence to optimize haplotype stacking, opens two complementary pathways to improve SNF in crop and pasture legumes
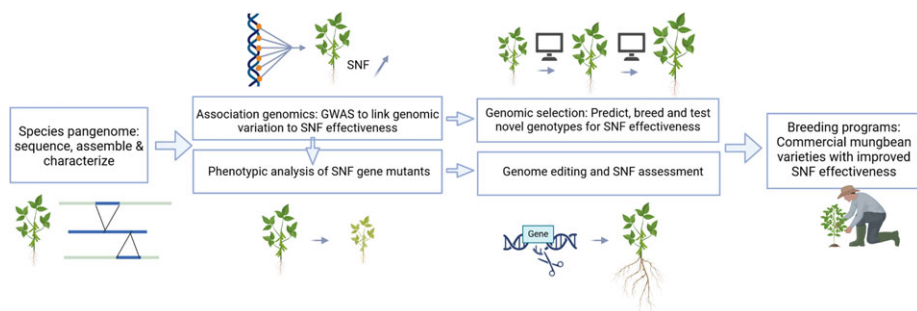
**Figure 2.** Pathways to increase SNF in legumes. Pan genomes reveal the full genetic potential of a species, including the complete set of genes and variation in DNA sequence and content amongst individuals. This enables discovery of genes involved in SNF, via genome wide association studies (GWAS) of all genes and genetic studies of specific genes. The potential of such genes and their natural alleles to improve SNF can be tested by genomic selection of parents that contribute desired sets of alleles or haplotypes to offspring. In parallel, genome editing can generate novel genetic variation that may be incorporated into breeding programmes, along with optimal haplotype stacks from genomic selection, to increase SNF.

(Figure 2). Starting with a high-quality pan genome representing diverse germplasm (seeds, plants, or plant parts useful in crop breeding) available to plant breeders, one pathway to SNF enhancement (upper pathway in Figure 2) involves phenotyping hundreds of genetically diverse individuals to determine their cumulative nitrogen fixation activity. Integrative measures of SNF, such as mass-spectrometric measurements of $^{15}N/^{14}N$ ratio and %N in tissues combined with biomass measures, to determine total Ndfa are key to this. Together with genomic sequence information for each individual, they will enable associations to be made between DNA sequence and phenotypic variation across the whole genome. In particular, genome-wide association studies (GWAS) can identify specific genes and gene variants that have the greatest impact on SNF, which will complement existing knowledge of SNF genetics and identify new targets for detailed genetic analysis and, potentially, genome editing (lower pathway in Figure 2). Association genomics also informs genomic prediction models that can guide the genomic selection of best parents to bring together optimal haplotypes for SNF in progeny. After multiple generations of progeny genotyping, selection and crossing, individuals with desired genotypes/ haplotypes are tested for SNF effectiveness, with improved genotypes entering commercial breeding programmes. In parallel, genome editing may produce genetic and phenotypic novelty not present in natural populations, thus complementing the genome-enabled breeding pathway. Such novel material can also enter commercial breeding pipelines, most easily so if genome editing is done in genotypes that are similar to, if not identical to, advanced stage breeding material. Thus, there is work to do for both molecular biologists and molecular plant breeders, and a need for dialogue between the two to ensure efforts are coordinated and aligned with breeding objectives. For early-career scientists just entering this area of R&D, there has never

been a better time to tackle the task of increasing SNF in legumes to help solve the wicked problem of feeding the growing population without destroying the environment upon which all life depends.

## References

**Appleby CA** (1984) Leghemoglobin and Rhizobium respiration. *Annual Review of Plant Physiology* **35**(1), 443–478.

**Bazzer SK, Ray JD, Smith JR, Fritschi FB and Purcell LC** (2020) Mapping quantitative trait loci (QTL) for plant nitrogen isotope ratio (δ 15 N) in soybean. *Euphytica* **216**, 1–15.

**Benedito VA, Torres-Jerez I, Murray JD, Andriankaja A, Allen S, Kakar K, Wandrey M, Verdier J, Zuber H and Ott T** (2008) A gene expression atlas of the model legume Medicago truncatula. *The Plant Journal* **55**(3), 504–513.

**Brewin NJ** (1991) Development of the legume root nodule. *Annual Review of Cell Biology* **7**(1), 191–226.

**Bryan JK, Hunt S, Glenn EW, Walsh KB, Roy HP, David TC and Layzell DB** (1988) Regulation of $O_2$ concentration in soybean nodules observed by in situ spectroscopic measurement of leghemoglobin oxygenation. *Plant Physiology* **87**(2), 296–299.

**Canfield DE, Glazer AN and Falkowski PG** (2010) The evolution and future of Earth's nitrogen cycle. *Science* **330**(6001), 192–196.

**Colebatch G, Desbrosses G, Ott T, Krusell L, Montanari O, Kloska S, Kopka J and Udvardi MK** (2004) Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in Lotus japonicus. *The Plant Journal* **39**(4), 487–512.

**Dakora FD and Aitkins CA** (1989) Diffusion of oxygen in relation to structure and function in legume root nodules. *Australian Journal of Plant Physiology* **16**(1), 131–140.

**Ferguson BJ, Indrasumunar A, Hayashi S, Lin MH, Lin YH, Reid DE and Gresshoff PM** (2010) Molecular analysis of legume nodule development and autoregulation. *Journal of Integrative Plant Biology* **52**(1), 61–76.

**Ferguson BJ, Mens C, Hastwell AH, Zhang M, Su H, Jones CH, Chu X and Gresshoff PM** (2019) Legume nodulation: The host controls the party. *Plant, Cell & Environment* **42**(1), 41–51. doi: 10.1111/pce.13348

**Fischer H-M** (1994) Genetic regulation of nitrogen fixation in rhizobia. *Microbiological Reviews* **58**(3), 352–386.

**Foyer CH, Lam H-M, Nguyen HT, Siddique KH, Varshney RK, Colmer TD, Cowling W, Bramley H, Mori TA and Hodgson JM** (2016) Neglecting legumes has compromised human health and sustainable food production. *Nature Plants* **2**(8), 1–10.

**Fukai E, Soyano T, Umehara Y, Nakayama S, Hirakawa H, Tabata S, Sato S and Hayashi M** (2012) Establishment of a Lotus japonicus gene tagging population using the exon-targeting endogenous retrotransposon LORE1. *The Plant Journal* **69**(4), 720–730.

**Garg G, Kamphuis LG, Bayer PE, Kaur P, Dudchenko O, Taylor CM, Frick KM, Foley RC, Gao L, Lieberman E, Edwards D and Singh KB** (2022) A pan-genome and chromosome-length reference genome of narrow-leafed lupin (*Lupinus angustifolius*) reveals genomic diversity and insights into key industry and biological traits. *The Plant Journal* **111**(5), 1252–1266.

**Gou X, Reich PB, Qiu L, Shao M, Wei G, Wang J and Wei X** (2023) Leguminous plants significantly increase soil nitrogen cycling across global climates and ecosystem types. *Global Change Biology* **29**(14), 4028–4043.

**Hayes BJ, Chen C, Powell O, Dinglasan E, Villiers K, Kemper KE and Hickey LT** (2023) Advancing artificial intelligence to help feed the world. *Nature Biotechnology* **41**, 1188–1189. doi: 10.1038/s41587-023-01898-2

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

**Hickey LT, Hafeez AN, Robinson H, Jackson SA, Leal-Bertioli SCM, Tester M, Gao C, Godwin ID, Hayes BJ and Wulff BBH** (2019) Breeding crops to feed 10 billion. *Nature Biotechnology* **37**(7), 744–754. doi: 10.1038/s41587-019-0152-9

**Kamfwa K, Cichy KA and Kelly JD** (2019) Identification of quantitative trait loci for symbiotic nitrogen fixation in common bean. *Theoretical and Applied Genetics* **132**, 1375–1387.

**Kazmierczak T, Yang L, Boncompagni E, Meilhoc E, Frugier F, Frendo P, Bruand C, Gruber V and Brouquisse R** (2020) Legume nodule senescence: a coordinated death mechanism between bacteria and plant cells. *Advances in Botanical Research* **94**, 181–212.

**Krueger CB, Ray JD, Smith JR, Dhanapal AP, Arifuzzaman M, Gao F and Fritschi FB** (2024) Identification of QTLs for symbiotic nitrogen fixation and related traits in a soybean recombinant inbred line population. *Theoretical and Applied Genetics* **137**(4), 89.

**Layzell DB and Hunt S** (1990) Oxygen and the regulation of nitrogen fixation in legume nodules. *Physiologia Plantarum* **80**(2), 322–327.

**Liu Y, Du H, Li P, Shen Y, Peng H, Liu S, Zhou G-A, Zhang H, Liu Z and Shi M** (2020) Pan-genome of wild and cultivated soybeans. *Cell* **182**(1), 162–176. e113.

**Mendoza-Suárez MA, Geddes BA, Sánchez-Cañizares C, Ramírez-González RH, Kirchhelle C, Jorrin B and Poole PS** (2020) Optimizing rhizobium-legume symbioses by simultaneous measurement of rhizobial competitiveness and N2 fixation in nodules. *Proceedings of the National Academy of Sciences* **117**(18), 9822–9831.

**Mens C**, *et al.* (Unpublished) For access to mungbean genome sequences please contact the corresponding author.

**Merrick M and Edwards R** (1995) Nitrogen control in bacteria. *Microbiological Reviews* **59**(4), 604–622.

**Nzepang DT, Gully D, Nguepjop JR, Zaiya Zazou A, Tossim H-A, Sambou A, Rami J-F, Hocher V, Fall S and Svistoonoff S** (2023) Mapping of QTLs associated with biological nitrogen fixation traits in peanuts (Arachis hypogaea l.) using an interspecific population derived from the cross between the cultivated species and its wild ancestors. *Genes* **14**(4), 797.

**Ott T, van Dongen JT, Gu C, Krusell L, Desbrosses G, Vigeolas H, Bock V, Czechowski T, Geigenberger P and Udvardi MK** (2005) Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. *Current Biology* **15**(6), 531–535.

**Pan S-Y, He K-H, Lin K-T, Fan C and Chang C-T** (2022) Addressing nitrogenous gases from croplands toward low-emission agriculture. *NPJ Climate and Atmospheric Science* **5**(1), 43.

**Penmetsa RV and Cook DR** (2000) Production and characterization of diverse developmental mutants of Medicago truncatula. *Plant Physiology* **123**(4), 1387–1398.

**Roberts G and Brill W** (1981) Genetics and regulation of nitrogen fixation. *Annual Reviews in Microbiology* **35**(1), 207–235.

**Robson RL and Postgate JR** (1980) Oxygen and hydrogen in biological nitrogen fixation. *Annual Reviews in Microbiology* **34**(1), 183–207.

**Roy S, Liu W, Nandety RS, Crook A, Mysore KS, Pislariu CI, Frugoli J, Dickstein R and Udvardi MK** (2020) Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. *The Plant Cell* **32**(1), 15–41.

**Rutten PJ and Poole PS** (2019) Oxygen regulatory mechanisms of nitrogen fixation in rhizobia. *Advances in Microbial Physiology* **75**, 325–389.

**Santachiara G, Salvagiotti F and Rotundo JL** (2019) Nutritional and environmental effects on biological nitrogen fixation in soybean: a meta-analysis. *Field Crops Research* **240**, 106–115.

**Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A and Kawashima K** (2008) Genome structure of the legume, Lotus japonicus. *DNA Research* **15**(4), 227–239.

**Schafleitner R, Nair RM, Rathore A, Wang Y, Lin C, Chu S, Lin P, Chang J and Ebert AW** (2015) The AVRDC – The World Vegetable Center mungbean (*Vigna radiata*) core and mini-core collections. *BMC Genomics* **16**(344), 1–11. doi: [10.1186/s12864-015-1556-7](10.1186/s12864-015-1556-7)

**Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ and Cheng J** (2010) Genome sequence of the palaeopolyploid soybean. *Nature* **463**(7278), 178–183.

**Schulte CCM, Borah K, Wheatley RM, Terpolilli JJ, Saalbach G, Crang N, de Groot DH, Ratcliffe RG, Kruger NJ, Papachristodoulou A and Poole PS** (2021) Metabolic control of nitrogen fixation in rhizobium-legume symbioses. *Science Advances* **7**(31), eabh2433.

**Scott PT, Pregelj L, Chen N, Hadler JS, Djordjevic MA and Gresshoff PM** (2008) Pongamia pinnata: an untapped resource for the biofuels industry of the future. *Bioenergy Research* **1**, 2–11.

**Semba RD, Ramsing R, Rahman N, Kraemer K and Bloem MW** (2021) Legumes as a sustainable source of protein in human diets. *Global Food Security* **28**, 100520.

**Sharma P, Masouleh AK, Topp B, Furtado A and Henry RJ** (2022) De novo chromosome level assembly of a plant genome from long read sequence data. *The Plant Journal* **109**(3), 727–736.

**Smil V** (2001) *Enriching the Earth: Fritz Haber, Carl Bosch, and the Transformation of World Food Production*. The MIT Press.

**Snyder CS, Bruulsema TW, Jensen TL and Fixen PE** (2009) Review of greenhouse gas emissions from crop production systems and fertilizer management effects. *Agriculture, Ecosystems & Environment* **133**(3-4), 247–266.

**Steffen W, Richardson K, Rockström J, Cornell SE, Fetzer I, Bennett EM, Biggs R, Carpenter SR, De Vries W and De Wit CA** (2015) Planetary boundaries: Guiding human development on a changing planet. *Science* **347**(6223).

**Sulieman S and Tran L** (2016) *Legume Nitrogen Fixation in a Changing Environment*. Springer,

**Sutton MA, Oenema O, Erisman JW, Leip A, van Grinsven H and Winiwarter W** (2011) Too much of a good thing. *Nature* **472**(7342), 159–161.

**Swaraj K and Bishnoi N** (1996) Physiological and biochemical basis of nodule senescence in legumes: a review. *Plant Physiology and Biochemistry* **23**(2), 105–116.

**Szczyglowski K, Shaw RS, Wopereis J, Copeland S, Hamburger D, Kasiborski B, Dazzo FB and de Bruijn FJ** (1998) Nodule organogenesis and symbiotic mutants of

the model legume Lotus japonicus. *Molecular Plant-Microbe Interactions* **11**(7), 684–697.

Tadege M, Wen J, He J, Tu H, Kwak Y, Eschstruth A, Cayrel A, Endre G, Zhao PX and Chabaud M (2008) Large-scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume Medicago truncatula. *The Plant Journal* **54**(2), 335–347.

Takahara M, Magori S, Soyano T, Okamoto S, Yoshida C, Yano K, Sato S, Tabata S, Yamaguchi K, Shigenobu S, Takeda N, Suzaki T and Kawaguchi M (2013) TOO MUCH LOVE, a novel Kelch repeat-containing F-box protein, functions in the long-distance regulation of the legume-Rhizobium symbiosis. *Plant and Cell Physiology* **54**(4), 433–447. doi: 10.1093/pcp/pct022

Tegeder M (2014) Transporters involved in source to sink partitioning of amino acids and ureides: opportunities for crop improvement. *Journal of Experimental Botany* **65**(7), 1865–1878.

Thilakarathna MS, Torkamaneh D, Bruce RW, Rajcan I, Chu G, Grainger CM, Szczyglowski K, Hill B and Raizada MN (2021) Testing whether pre-pod-fill symbiotic nitrogen fixation in soybean is subject to drift or selection over 100 years of soybean breeding. *Frontiers in Agronomy* **3**, 725813.

Tsikou D, Yan Z, Holt DB, Abel NB, Reid DE, Madsen LH, Bhasin H, Sexauer M, Stougaard J and Markmann K (2018) Systemic control of legume susceptibility to rhizobial infection by a mobile microRNA. *Science* **362**(6411), 233–236. doi: 10.1126/science.aat6907

Udvardi M, Below FE, Castellano MJ, Eagle AJ, Giller KE, Ladha JK, Liu X, Maaz TM, Nova-Franco B and Raghuram N (2021) A research road map for responsible use of agricultural nitrogen. *Frontiers in Sustainable Food Systems* **31**(5), 660155.

Udvardi M and Poole PS (2013) Transport and metabolism in legume-rhizobia symbioses. *Annual Review of Plant Biology* **64**, 781–805.

Udvardi MK and Day DA (1997) Metabolite transport across symbiotic membranes of legume nodules. *Annual Review of Plant Biology* **48**(1), 493–523.

Udvardi MK, Lister DL and Day DA (1992) Isolation and characterization of a ntrC mutant of Bradyrhizobium (Parasponia) sp. *ANU289. Journal of General Microbiology* **138**(5), 1019–1025.

Wang X, Qiu Z, Zhu W, Wang N, Bai M, Kuang H, Cai C, Zhong X, Kong F and Lü P (2023) The NAC transcription factors SNAP1/2/3/4 are central regulators mediating high nitrogen responses in mature nodules of soybean. *Nature Communications* **14**(1), 4711.

Westhoek A, Clark LJ, Culbert M, Dalchau N, Griffiths M, Jorrin B, Karunakaran R, Ledermann R, Tkacz A, Webb I, James EK, Poole PS and Turnbull LA (2021) Conditional sanctioning in a legume–*Rhizobium* mutualism. *Proceedings of the National Academy of Sciences* **118**(19), e2025760118

Yang Q, Yang Y, Xu R, Lv H and Liao H (2019) Genetic analysis and mapping of QTLs for soybean biological nitrogen fixation traits under varied field conditions. *Frontiers in Plant Science* **10**, 419631.

Young ND, Debellé F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KF, Gouzy J and Schoof H (2011) The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**(7378), 520–524.

Yu H, Xiao A, Wu J, Li H, Duan Y, Chen Q, Zhu H and Cao Y (2023) GmNAC039 and GmNAC018 activate the expression of cysteine protease genes to promote soybean nodule senescence. *The Plant Cell* **35**(8), 2929–2951.

**Zhang M, Su H, Gresshoff PM and Ferguson BJ** (2021) Shoot-derived miR2111 controls legume root and nodule development. *Plant, Cell & Environment* **44**(5), 1627–1641. doi: 10.1111/pce.13992

**Zhang M, Zhang X, Jiang X, Qiu L, Jia G, Wang L, Ye W and Song Q** (2022) iSoybean: a database for the mutational fingerprints of soybean. *Plant Biotechnology Journal* **20**(8), 1435–1437. doi:https://doi.org/10.1111/pbi.13844

**Zhang X, Davidson EA, Mauzerall DL, Searchinger TD, Dumas P and Shen Y** (2015) Managing nitrogen for sustainable development. *Nature* **528**(7580), 51–59.

## About the Authors

**Michael Udvardi** is Professor of Legume Genomics with the Queensland Alliance for Agriculture and Food Innovation at the University of Queensland, Australia. He earned his PhD in plant biochemistry from the Australian National University in 1989. He is primarily interested in how plants obtain nitrogen for growth, either as mineral nitrogen from the soil or from atmospheric di-nitrogen via symbiotic nitrogen fixation in bacteria. He has contributed to our understanding of symbiotic nitrogen fixation in legumes, especially of transport and metabolism in root nodules, using biochemical, molecular, genetic and genomic methods. He was amongst the first to characterize ammonium and nitrate transporters in plants. He was part of a large international team that sequenced and analysed the genome of the model legume, *Medicago truncatula*. Currently, his group focuses on the development of pan-genomic resources to accelerate breeding of tropical pulses, including mungbean and pigeonpea. Dr Udvardi has published over 200 papers in refereed scientific journals. He was Elected Fellow of the American Association for the Advancement of Science in 2012 for his contributions to our understanding of legume biology, especially symbiotic nitrogen fixation. In 2023, he was awarded the Adam Kondorosi Academia Europaea Award for Advanced Research in recognition of the impact of his research on plant–microbe interactions and plant science and his generous service to the scientific community.

**Celine Mens** is a Postdoctoral Research Fellow with the Queensland Alliance for Agriculture and Food Innovation at the University of Queensland, Australia. Dr Mens obtained her PhD in Plant Molecular Biology at the University of Queensland in 2022, focused on the characterization of molecular signals in Autoregulation of Nodulation and nitrate-dependent regulation of nodulation in model legumes such as soybean and *M. truncatula*. She is now a postdoc in Michael Udvardi's lab working on genome assembly and annotation, as well as quantifying symbiotic nitrogen fixation in diverse mungbean populations.

**Estelle Grundy** is a postdoctoral research fellow working for Professor Michael Udvardi with the Queensland Alliance for Agriculture and Food Innovation at the University of Queensland (UQ), Australia. She recently completed her PhD thesis on

the functional characterization of defence genes in soybean nodulation at the Integrative Legume Research Group (School of Agriculture and Food Sustainability, UQ) headed by A/Prof. Brett Ferguson. Her current research is centred on understanding the genetic basis of symbiotic nitrogen fixation (SNF) in mungbean by exploring the diversity of SNF across diverse genotypes.