



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

Scuola di  
Scienze della Salute Umana

Corso di Laurea in  
BIOTECNOLOGIE

# **Analisi della Robustezza Metabolica nella Simbiosi Pianta-Batterio**

## **Analysis of Metabolic Robustness in Plant- Bacterium Symbiosis**

**Relatore**  
Alessio Mengoni

**Correlatore**  
Marco Fondi

**Candidato**  
Michelangelo Tesi

Anno Accademico 2018/2019



*Questo lavoro di tesi si è servito di un modello computazionale capace di intrappolare la complessità del dialogo metabolico tra le piante leguminose e i batteri azotofissatori.*

*Dedico il lavoro a Simone, che mi ha insegnato che è impossibile fare altrettanto di fronte alla complessità delle emozioni umane.*



# Indice

<b>1. Introduzione.....</b>	<b>2</b>
<b>2. Scopo del lavoro.....</b>	<b>31</b>
<b>3. Materiali e Metodi.....</b>	<b>32</b>
<b>4. Risultati e Discussione.....</b>	<b>43</b>
<b>5. Conclusioni.....</b>	<b>63</b>
<b>6. Bibliografia.....</b>	<b>65</b>
<b>7. Appendice.....</b>	<b>67</b>

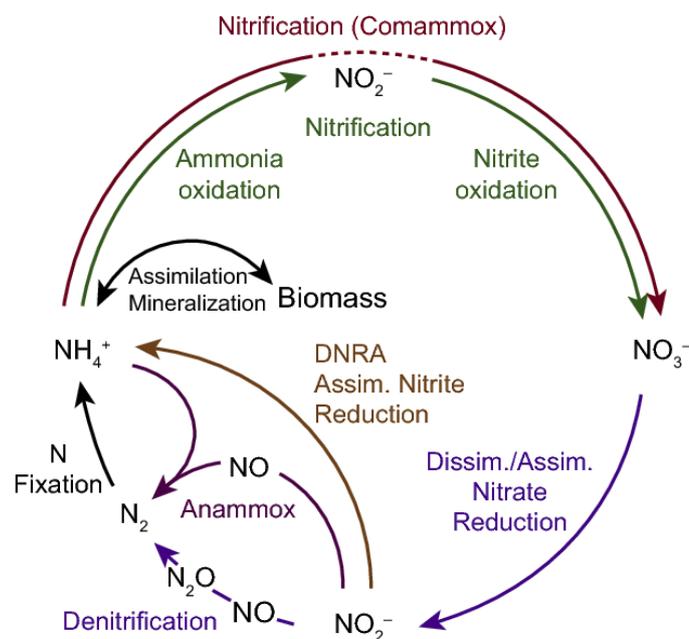
# 1. Introduzione

## 1.1 Il ciclo dell'azoto e la nutrizione azotata delle piante

L'azoto è un elemento indispensabile per tutte le forme di vita in quanto è un componente chiave di proteine, di acidi nucleici e di altri costituenti cellulari. La più grande riserva (benché non l'unica) di azoto nella biosfera è rappresentata dall'azoto biatomico  $N_2$ , che si trova allo stato gassoso e costituisce ben il 78% dell'atmosfera terrestre. Tuttavia, l'azoto che viene incorporato nei composti organici prodotti dal metabolismo delle cellule è azoto in forma ridotta (n.ox.=-3), diverso dall'azoto biatomico. Pertanto, gli organismi viventi hanno bisogno di azoto nella sua forma ridotta, rappresentato da azoto organico e da ammoniaca ( $NH_3$ ), oppure possono servirsi di azoto ossidato, presente all'interno di composti con l'ossigeno, ma solo dopo averlo ridotto ad ammoniaca. Nonostante tutti gli esseri viventi abbiano necessità di azoto, soltanto alcuni procarioti, i cosiddetti azotofissatori o diazotrofi, sono in grado di utilizzare l'azoto biatomico, di per sé assolutamente inerte ovvero fortemente stabile (in virtù del legame triplo che tiene uniti i due atomi di azoto), trasformandolo in composti azotati utilizzabili dagli altri organismi viventi. Questo processo viene indicato come fissazione biologica dell'azoto, per distinguerlo dalla fissazione atmosferica dell'azoto innescata dai fulmini (un fenomeno, quest'ultimo, che porta soltanto un contributo minimo alla quantità totale di azoto fissato). È grazie ai procarioti azotofissatori che forme di azoto biologicamente spendibili vengono distribuite attraverso gli ecosistemi a tutti gli organismi viventi.

L'azotofissazione rappresenta un collo di bottiglia che regola di fatto la quantità di azoto disponibile per gli esseri viventi, ma non è l'unica trasformazione a cui l'azoto va incontro. Infatti questo viene sottoposto a una varietà di trasformazioni chimiche –

biotiche e abiotiche – che alterano il suo stato di ossidazione e lo combinano in numerosi composti diversi. L'enorme quantità di azoto atmosferico è in equilibrio con tutti gli altri composti azotati. Muovendosi tra atmosfera, suolo, acqua e organismi viventi, ciascun atomo di azoto soggiace ciclicamente a numerose trasformazioni e si distribuisce tra specie inorganiche e organiche. Per comprendere il ruolo svolto dall'azoto nella funzionalità di un ecosistema è necessario conoscerne il ciclo biogeochimico, che può essere definito come l'insieme delle vie attraverso le quali tale elemento è riciclato (Figura 1, estratta da<sup>1</sup>).



**Figura 1:** Il ciclo biogeochimico dell'azoto. (DNRA è acronimo per *Dissimilatory Nitrite Reduction to Ammonium*).

I composti dell'azoto più importanti nel rispettivo ciclo biogeochimico sono:

- azoto molecolare o elementare o biatomico  $\text{N}_2$  con n.ox. = 0,
- ione nitrato  $\text{NO}_3^-$  con n.ox. = +5,
- ione nitrito  $\text{NO}_2^-$  con n.ox. = +3,
- ammoniaca  $\text{NH}_3$  e ione ammonio  $\text{NH}_4^+$  con n.ox. = -3,
- azoto organico con n.ox. = -3.

La possibilità di combinarsi in composti con diversi stati di ossidazione, vale a dire con diversi livelli di energia, fa sì che questi composti vengano diffusamente utilizzati dai microrganismi nei processi metabolici di produzione dell'energia necessaria per i loro processi vitali. Queste trasformazioni dell'azoto si articolano in fasi intermedie di ossidazione o di riduzione, nello spettro delle quali microrganismi di vario genere hanno capacità diverse e peculiari.

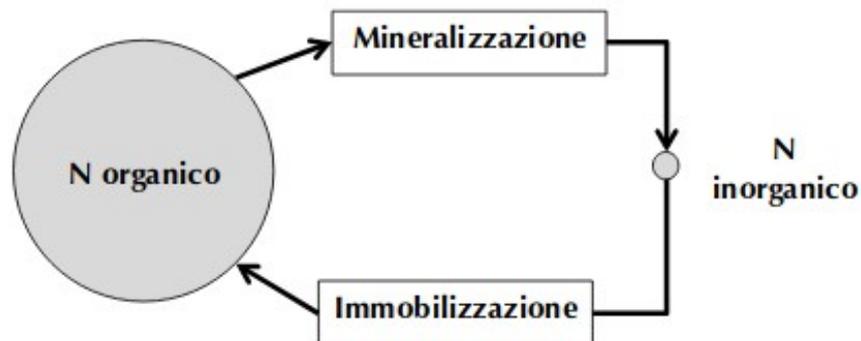
Se consideriamo l'importanza dell'azotofissazione microbica, unitamente a quella delle trasformazioni ossido-riduttive dei composti azotati, concludiamo che il ciclo dell'azoto è in larga parte dipendente dall'attività dei microrganismi. Ai microrganismi, infine, benché non solo a questi, si devono le trasformazioni dell'ammonio ad azoto organico (assimilazione dell'ammonio), e viceversa (ammonizzazione): a maggior ragione diciamo che il ciclo dell'azoto è alimentato prevalentemente dall'attività microbica.

L'uomo, infine, a partire dal secolo scorso incide attivamente sul ciclo dell'azoto. Infatti, agli inizi del Novecento è stato messo a punto il processo di sintesi di Haber-Bosch, grazie al quale si può condurre la reazione  $N_2 + 3H_2 \rightarrow 2 NH_3$  in condizioni artificiali di elevate temperatura e pressione. Il risultato netto del processo industriale di Haber-Bosch è il massiccio aumento del flusso di azoto dalla riserva di azoto biatomico verso altri composti azotati. L'ammoniaca così prodotta a livelli industriali ha numerosi impieghi, il più significativo dei quali è la sintesi di fertilizzanti azotati.<sup>2</sup>

La necessità di una produzione così ingente di fertilizzanti azotati è conseguenza del fatto che l'azoto rappresenta uno dei principali fattori limitanti per la crescita delle piante. Le piante assorbono l'azoto di cui hanno bisogno dal suolo, attraverso le radici. Nel terreno l'azoto si trova sia in forma organica sia in forma inorganica o minerale. L'azoto organico (presente negli organismi viventi, nella sostanza organica in via di decomposizione, e nell'humus) non è direttamente e immediatamente utilizzabile dalle piante. Nel tempo, attraverso il processo di mineralizzazione della sostanza organica, l'azoto organico viene progressivamente trasformato in azoto ammoniacale e successivamente in azoto nitrico, assumibile dalle piante. Fisiologicamente, le radici delle piante hanno la capacità di assorbire l'azoto in forma nitrica, in forma

ammoniacale e in forma di molecole organiche molto semplici, come l'urea e alcuni amminoacidi.<sup>3</sup>

L'azoto minerale del terreno è costituito quasi esclusivamente dall'insieme dell'azoto presente sotto forma di ione nitrato  $\text{NO}_3^-$  (azoto nitrico), di ione ammonio  $\text{NH}_4^+$  (azoto ammoniacale) e in misura nettamente inferiore di ione nitrito  $\text{NO}_2^-$  (azoto nitroso). Le forme organiche sono presenti nel terreno in quantità enormemente più elevate di quelle inorganiche e in media rappresentano più del 95% dell'azoto totale presente nel terreno. La quantità di azoto inorganico del terreno dipende in gran parte dall'andamento nel tempo dei processi di mineralizzazione e di immobilizzazione microbica, che rappresentano rispettivamente i termini di passaggio dell'azoto dalla forma organica a quella inorganica e viceversa (Figura 2, estratta da<sup>3</sup>).

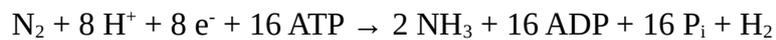


**Figura 2:** Schema delle interazioni tra azoto organico e azoto inorganico del terreno.

## 1.2 L'azotofissazione simbiotica

Abbiamo detto che, tra tutti gli organismi viventi, soltanto alcuni procarioti, i cosiddetti azotofissatori o diazotrofi, sono in grado di utilizzare l'azoto biatomico, trasformandolo in composti azotati utilizzabili dagli altri organismi viventi. Si tratta di procarioti che vivono liberi nel suolo o nelle acque. La classificazione degli azotofissatori non è facile poiché ricadono in raggruppamenti molto diversi tra loro in termini di habitat e di caratteristiche metaboliche. Per quanto filogeneticamente distanti tra loro, tutti gli

azotofissatori sono accomunati dalla capacità di fissare l'azoto atmosferico grazie al medesimo complesso enzimatico: il complesso della nitrogenasi. La nitrogenasi è stata caratterizzata in un gran numero di procarioti, mostrando elevato grado di similarità a livello di sequenza sia nucleotidica sia peptidica e mostrando proprietà del tutto simili per la stragrande maggioranza degli azotofissatori. La reazione catalizzata dalla nitrogenasi è:

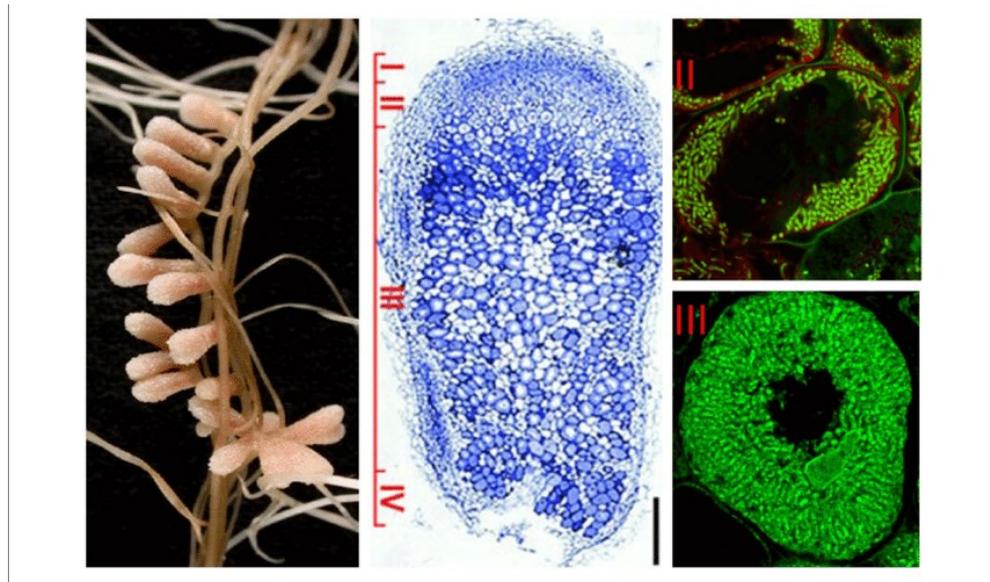


A causa dell'elevata stabilità del triplo legame dell'azoto biatomico, la reazione è estremamente dispendiosa: richiede molta energia (idrolisi di ben 16 molecole di ATP per molecola di  $\text{N}_2$ ) e sei elettroni che devono essere trasferiti fino a ridurre  $\text{N}_2$  a due molecole di  $\text{NH}_3$ , più altri due elettroni necessari alla formazione del prodotto collaterale  $\text{H}_2$ . Un'altra caratteristica comune alla nitrogenasi degli azotofissatori è la sua estrema sensibilità all'ossigeno, il quale ha il potere di inattivare irreversibilmente l'enzima.

Solo alcuni tra gli azotofissatori instaurano simbiosi con piante superiori, un'interazione che si esplica all'interno di strutture neoformate *ad hoc* dalla pianta, chiamate *noduli radicali* (Figura 3, estratta da<sup>4</sup>), in cui i batteri cedono alla pianta l'azoto fissato in cambio della nicchia trofico-protettiva rappresentata dal nodulo. In questo caso si parla di azotofissazione simbiotica.

I microrganismi azotofissatori che sono stati osservati in simbiosi con piante superiori appartengono prevalentemente agli ordini delle *Rhizobiales* (alfaproteobatteri) e delle *Burkholderiales* (betaproteobatteri) – questi ultimi due gruppi collettivamente chiamati rizobi – e alla famiglia delle *Frankiaceae* (attinobatteri). Le piante che partecipano a simbiosi con i rizobi sono la maggior parte delle appartenenti alla famiglia delle leguminose (oltre a poche specie non leguminose), ed è proprio per l'importanza agraria delle leguminose che le simbiosi azotofissatrici che coinvolgono i rizobi sono le più studiate.

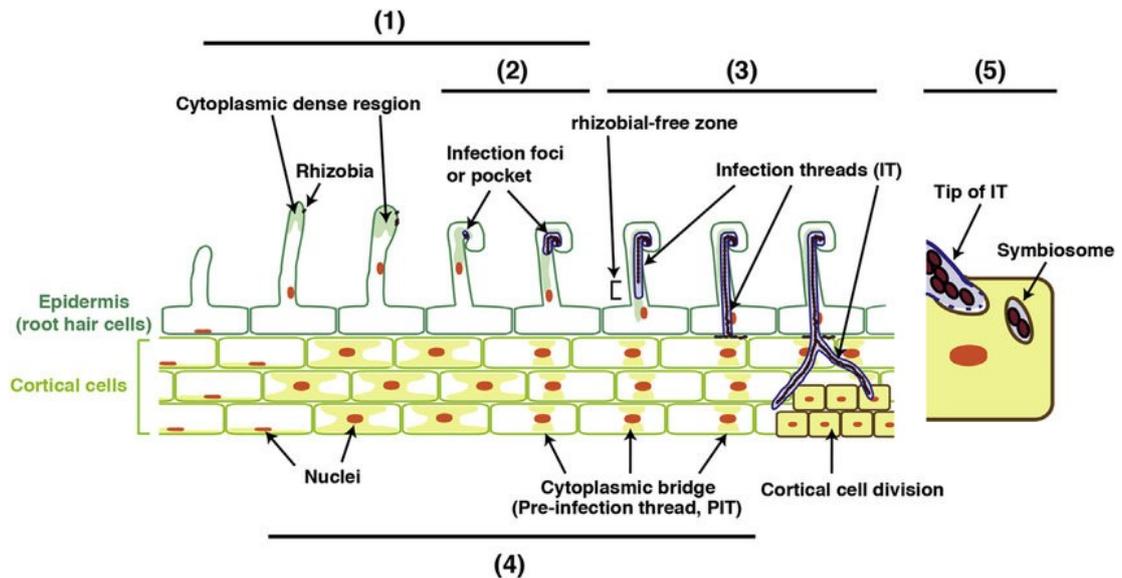
L'instaurarsi della simbiosi tra una leguminosa e il partner batterico è il risultato di una sorta di dialogo molecolare che avviene nella rizosfera, avviato dalla pianta in caso di carenza di azoto nel suolo. La pianta rilascia dalle radici composti detti flavonoidi. Per il rizobio questi composti hanno un significato particolare, quello di induttori specifici dell'espressione di un gruppo di geni, i cosiddetti geni *nod*. L'attivazione di questi geni batterici porta alla sintesi del segnale chiave, cioè di quella molecola che, viaggiando dal batterio alla pianta, innesca attività meristematiche che daranno luogo alla formazione del nodulo in cui i batteri saranno ospitati: una vera neoplasia scatenata da un chitolipooligosaccaride, anche noto come fattore nod.



**Figura 3:** Struttura di noduli radicali fissanti l'azoto formati nell'ambito della simbiosi tra *S. meliloti* e *M. truncatula*. Le diverse zone del nodulo sono indicate nell'immagine in sezione longitudinale: (I) meristema, (II) zona di infezione, (III) zona di fissazione dell'azoto, (IV) zona di senescenza. Le cellule simbiotiche nella zona II contengono i batteri endosimbionti in fase di differenziazione, mentre nella zona III il citoplasma delle cellule ospiti è totalmente occupato da lunghi batteroidi che fissano l'azoto. La fluorescenza verde è data da endosimbionti marcati con Syto9.

Ai rizobi è necessario raggiungere il nodulo, posto nella cortex radicale. L'ingresso delle cellule batteriche avviene via peli radicali. Una cellula epidermica di giovane pelo in crescita vede l'accumularsi di cellule batteriche al proprio apice e assume, sempre per effetto dei chitolipooligosaccaridi, una tipica deformazione apicale, un incurvamento

dal cui fulcro inizia la crescita verso l'interno di un canale neoformato definito *filo di infezione*, (Figura 4, estratta da<sup>5</sup>) in cui i rizobi avanzano e si moltiplicano alimentati da substrati forniti dalla pianta.



**Figura 4:** Infezione da parte dei rizobi e invasione dell'epidermide radicale e delle cellule corticali. (1) L'attacco dei rizobi ai peli radicali e il fattore Nod accumulato inducono le risposte iniziali che coinvolgono la deformazione e l'incurvamento del pelo radicale. (2) A livello del pelo radicale fortemente incurvato, si verifica la degradazione della parete cellulare e viene indotta l'invaginazione della membrana filo d'infezione nella cellula del pelo radicale. (3) Si verifica un allungamento del filo d'infezione, che è accompagnato da scorrimento del citoplasma e movimento del nucleo. (4) Durante l'infezione da parte dei rizobi ((1) e (3)), nelle cellule corticali della radice si forma un ponte citoplasmatico o un filo di preinfezione per guidare i fili d'infezione in allungamento. (5) Quando un filo d'infezione raggiunge la cellula corticale appena divisa, la membrana del filo d'infezione si rompe e i rizobi vengono rilasciati nella cellula corticale per formare un organello specializzato nel fissaggio dell'azoto, il simbiosoma.

I rizobi che penetrano nel proprio ospite dal filo d'infezione sono rilasciati nel citosol delle cellule che formano il nodulo, dove vanno incontro ad un'estesa modificazione del proprio metabolismo e si differenziano in batteroidi. I batteroidi cessano di dividersi e devolvono all'azotofissazione l'energia che ricevono dalla pianta sotto forma di composti organici. Il batteroide è avvolto da una membrana derivata dal plasmalemma della cellula vegetale, chiamata membrana peribatteroide. Il complesso costituito dal batteroide e dalla membrana peribatteroide viene chiamato simbiosoma.

I due partner della simbiosi inoltre sintetizzano la leg-emoglobina, una proteina del tutto analoga all'emoglobina del sangue degli animali superiori. Essa ha lo scopo di legare l'ossigeno molecolare – mantenendo, nel citoplasma dei batteroidi, una pressione parziale di questo gas che sia compatibile con l'espressione della nitrogenasi – e cederlo selettivamente nel sito in cui si svolge la catena di trasporto degli elettroni, per soddisfare il metabolismo aerobio del batteroide<sup>2</sup>.

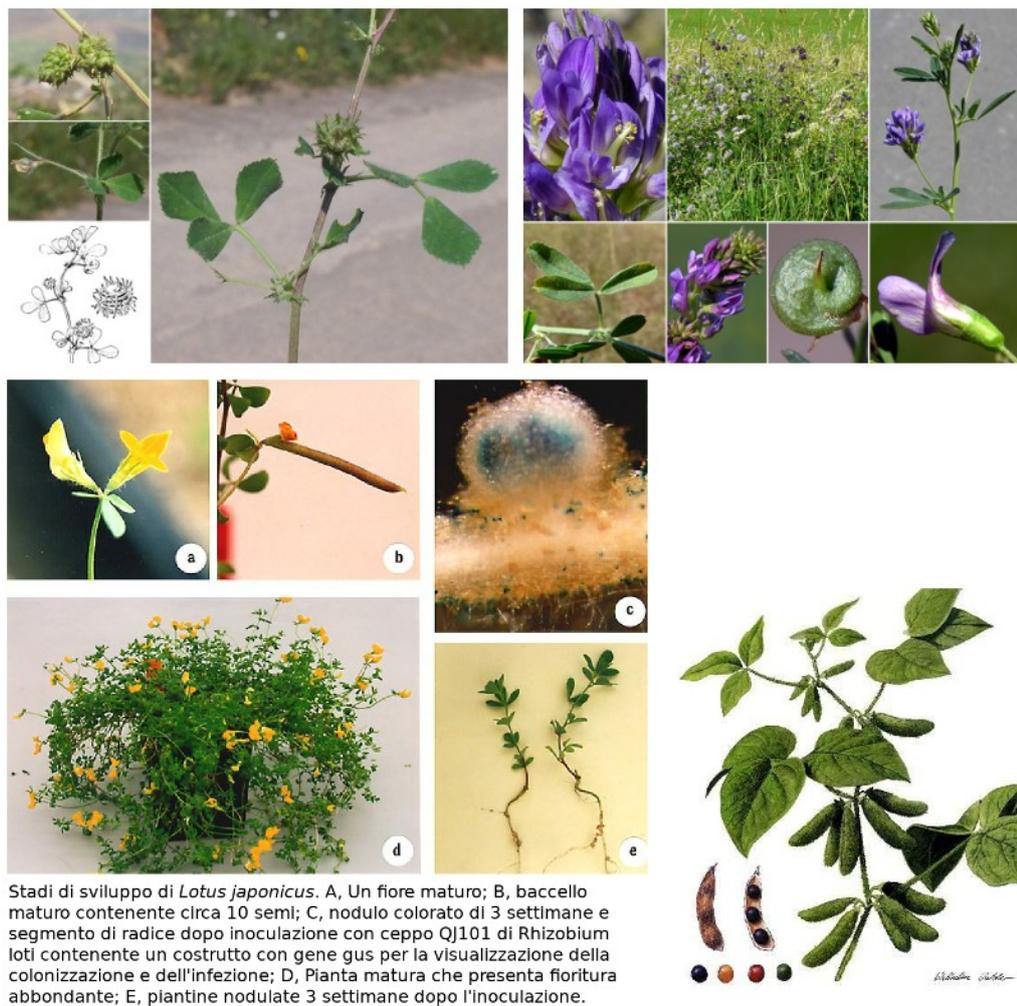
### **1.3 *Medicago truncatula* come leguminosa modello**

La rotazione delle colture con le leguminose ha permesso un consistente aumento della resa agraria sin dall'antichità. Le leguminose saranno inoltre fondamentali per preservare la sostenibilità della produzione agricola, affrontando le sfide dell'aumento della domanda di cibo oltre che di biocarburanti. Le leguminose sono generalmente definite dalla loro tipica struttura florale e dalla capacità (esibita dalla maggioranza di queste) di instaurare simbiosi con i rizobi. Con oltre 18.000 specie, le leguminose si trovano dal circolo polare ai tropici e includono molte colture di importanza agronomica per la produzione di granella e di foraggio e per la silvicoltura. La capacità di oltre l'88% delle leguminose di ottenere azoto dall'aria attraverso i noduli radicali è stata probabilmente un fattore determinante nel successo evolutivo, ecologico ed economico di queste piante.<sup>6</sup>

Negli ultimi 20 anni, il numero crescente di progetti di ricerca sui legumi ha permesso l'emergere di piante modello per le specie di leguminose. I legumi delle colture e dei pascoli sono generalmente non adatti come sistemi modello per la ricerca genetica e genomica. Alcuni legumi coltivati sono tetraploidi (e.g. arachidi), molti hanno grandi genomi (e.g. piselli e fave) e molti sono refrattari alla trasformazione o difficili da rigenerare (e.g. fagioli comuni, piselli e soia). La maggior parte dei legumi da granella ha semi grandi ma relativamente pochi per pianta, e piantine grandi, che impediscono la coltura ad alta densità (e.g. ceci, piselli dagli occhi neri, fagioli verdi, piselli, fagioli e

soia). Alcuni legumi, come la soia, hanno duplicazioni del genoma e alcuni sono autoincompatibili o hanno tempi di generazione lunghi.<sup>7</sup>

Di conseguenza, *Medicago truncatula* (Figura 5, estratta da<sup>8</sup>) e *Lotus japonicus* (Figura 7, estratta da<sup>9</sup>) sono emersi come due principali sistemi modello per la biologia delle leguminose. Inizialmente studiati per indagare le interazioni simbiotiche tra piante e microbi, e in particolare la nodulazione delle leguminose, questi due modelli sono ora ampiamente utilizzati in una varietà di campi della biologia: dalla fisiologia e sviluppo delle piante, alla genetica delle popolazioni e alla genomica strutturale<sup>6</sup>.



**Figura 5** (in alto a sinistra): fotografie di *Medicago truncatula*. **Figura 6** (in alto a destra): fotografie di *Medicago sativa*. **Figura 7** (in basso a sinistra): fotografie di *Lotus japonicus* con didascalie. **Figura 8** (in basso a destra): illustrazione di *Glycine max* (soia).

Mentre *M. truncatula* è una pianta annuale della tribù *Trifolieae* e un parente stretto dell'erba medica e dei trifogli, *L. japonicus* appartiene alla tribù *Loteae* ed è più distante dai legumi coltivati delle zone temperate rispetto a *M. truncatula*. Questa maggiore vicinanza filogenetica alle colture economicamente importanti è fondamentale nella scelta di *M. truncatula* da parte di molti ricercatori e agenzie di finanziamento della ricerca. L'uso di entrambi i legumi modello consente studi genomici comparativi all'interno della famiglia dei legumi, nonché il confronto tra due modelli di sviluppo del nodulo radicale: indeterminato con un meristema nodulare persistente nel caso di *M. truncatula* e determinato in *L. japonicus*. Sfortunatamente, questi due modelli appartengono entrambi allo stesso clade (Galegoidi) che raccoglie i legumi delle zone temperate, mentre la soia e il fagiolo comune sono legumi delle zone tropicali (clade dei Faseolidi). La soia (Figura 8, estratta da<sup>10</sup>) viene quindi proposta come terza leguminosa modello sia per il proprio peso economico sia per la vicinanza filogenetica ad altre colture importanti.<sup>6</sup>

*M. truncatula* è stata proposta per la prima volta come modello da Barker *et al.* (1990) per studiare la simbiosi tra rizobi e leguminose. Ora, è riconosciuta a livello internazionale come modello per tutti gli studi sulle leguminose. *M. truncatula* presenta numerosi vantaggi per la ricerca genomica sulle piante: è autogama, ha un genoma diploide ( $2n = 16$ ) relativamente piccolo (~ 375 Mbp) che è stato sequenziato e annotato, ha un tempo di generazione relativamente breve (circa 4 mesi da seme a seme) e una produzione abbondante di semi.<sup>7</sup>

Come molte piante superiori, *M. truncatula* forma associazioni simbiotiche con una vasta gamma di funghi arbuscolo-micorrizici (AM). In quanto leguminosa, *M. truncatula* è anche in grado di sviluppare noduli radicali con *Sinorhizobium meliloti*, che è una delle specie di rizobio meglio caratterizzate a livello genetico.<sup>6</sup>

Le specie annuali di *Medicago* di origine mediterranea sono state particolarmente importanti nell'agricoltura dell'Australia meridionale a causa del loro ruolo nella rotazione delle colture, in cui le colture di cereali si alternano a quelle di legumi. Il

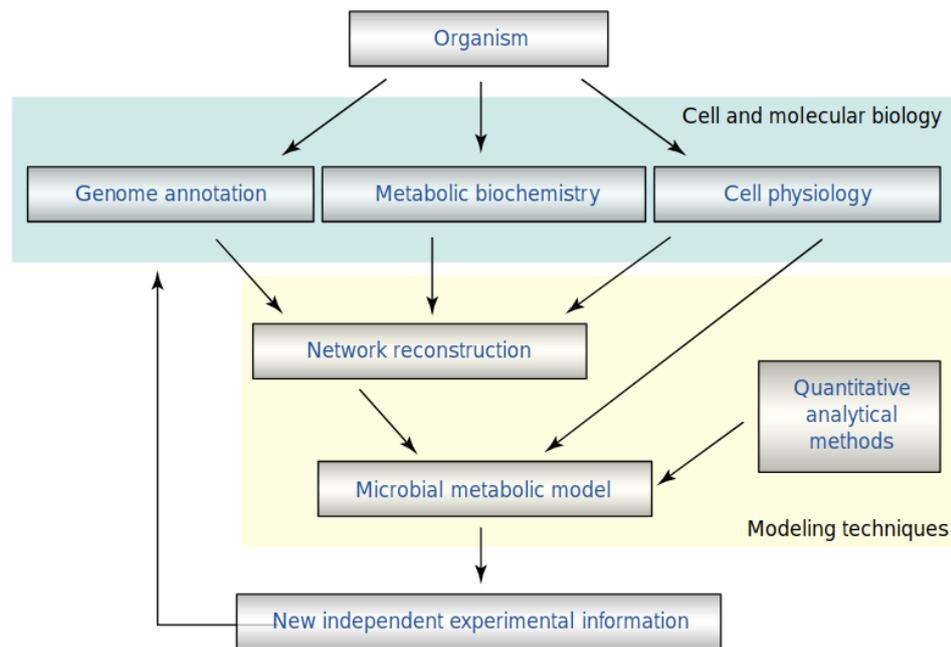
valore di *M. truncatula* fu riconosciuto già nel 1939 e Jemalong era una cultivar comunemente usata. Questo è un eccellente esempio di agricoltura sostenibile, in cui il pascolo di leguminose fornisce foraggio per il bestiame e al contempo azoto simbioticamente fissato disponibile per le colture successive.<sup>11</sup>

Inoltre, *M. truncatula* è strettamente correlata agli altri legumi da foraggio come l'erba medica (*Medicago sativa*, Figura 6, estratta da<sup>12</sup>), il trifoglio rosso (*Trifolium pratense*) e il trifoglio bianco (*Trifolium repens*), i quali quindi potrebbero essere diretti beneficiari degli studi di genomica su *M. truncatula*. I legumi da foraggio sono colture importanti per l'agricoltura sostenibile in tutto il mondo. Oltre a fornire una qualità del foraggio superiore per la produzione animale, e a migliorare la fertilità del suolo attraverso la fissazione dell'azoto, i legumi da foraggio contribuiscono anche al miglioramento della struttura del suolo e al controllo dell'erosione del suolo. L'erba medica (*Medicago sativa*), ad esempio, è cresciuta su oltre 26 milioni di acri e si colloca al terzo posto per superficie coltivata e valore in dollari negli Stati Uniti (USDA Crop Values Summary 2005). Anche i trifogli (*Trifolium* spp.), spesso coltivati insieme a erbe foraggere, sono ampiamente distribuiti.<sup>6</sup>

## **1.4 La modellizzazione metabolica di scala genomica**

Negli ultimi anni sono state sviluppate molte tecnologie sperimentali ad alto rendimento (*high throughput*) che hanno permesso di ottenere sequenze genomiche complete, di valutare l'espressione genica a livello dell'intero genoma, e di esaminare il panorama proteico di cellule e organismi particolari (genoma, trascrittoma, proteoma). È probabile che queste tecnologie sperimentali aumenteranno in termini di velocità e di potenziale nei prossimi anni. Inoltre, miglioreranno le tecnologie di fenotipizzazione ad alto rendimento. Questi sviluppi stanno avendo un profondo impatto sul metodo stesso con cui ci si avvicina alle scienze biologiche. Ad esempio, sta diventando universalmente accettato che le cellule debbano essere viste come sistemi. Tali sistemi rappresentano reti complesse di prodotti genici che interagiscono per produrre funzioni fisiologiche.

Come in molti altri campi della scienza e dell'ingegneria, la generazione su larga scala di set di dati (*datasets*) complessi richiede l'analisi matematica e la simulazione al computer. In passato, tali sforzi hanno rappresentato la curiosità di pochi e sono stati ostacolati dalla mancanza di dati validi su cui poter costruire modelli affidabili. Tuttavia, la costruzione di modelli matematici sta ora assumendo un ruolo centrale in biologia, e il suo uso e importanza sono destinati a crescere. Come si iniziano a costruire tali modelli? La conoscenza dell'elenco dei componenti delle cellule e del modo in cui interagiscono sta diventando sempre più dettagliata, come evidenziato dal gran numero di database metabolici accessibili in internet. Gli effetti di queste interazioni devono quindi essere analizzati e determinati.

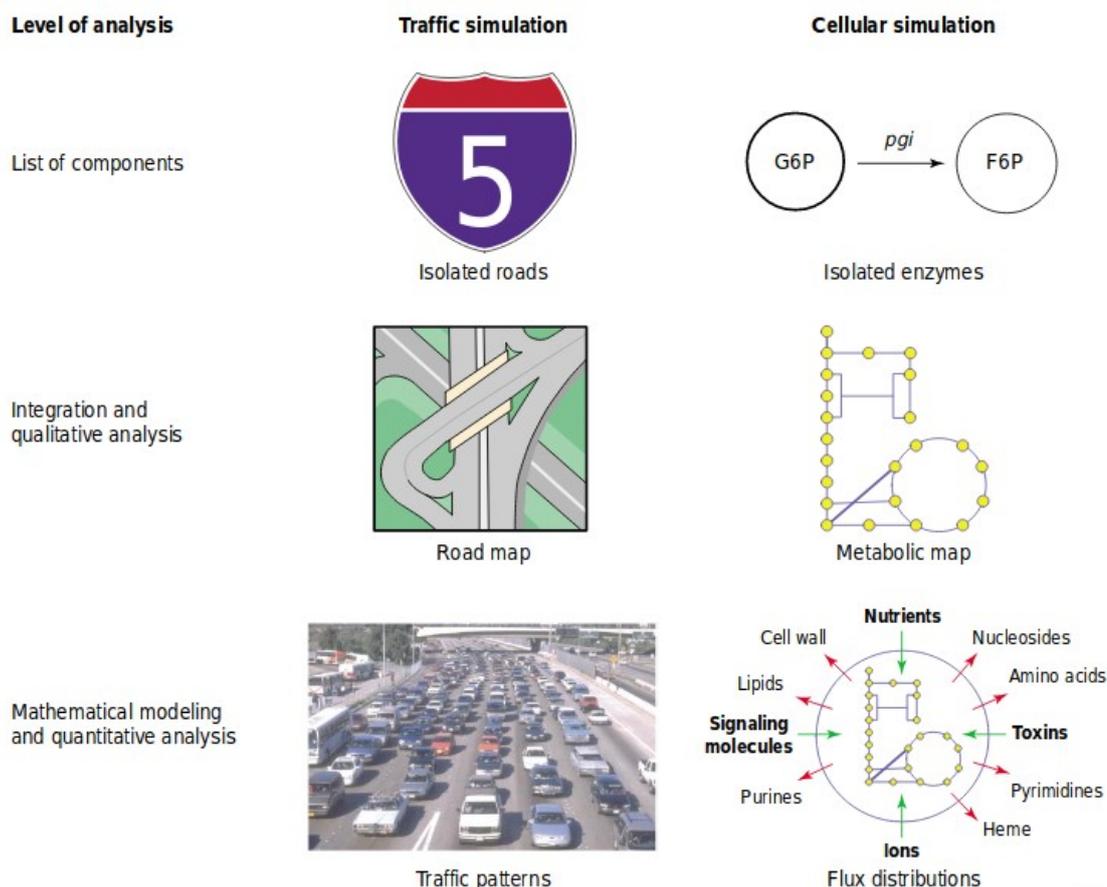


**Figura 9:** Processo integrato di costruzione di modelli metabolici microbici. Tale costruzione richiede una conoscenza approfondita del metabolismo dell'organismo. Dalla sequenza del genoma annotata e dalle caratteristiche biochimiche e fisiologiche sperimentalmente determinate di una cellula, può essere ricostruita la rete delle reazioni metaboliche. Questa rete viene quindi modificata nel contesto di altri vincoli fisiologici per produrre un modello matematico, che può essere utilizzato per generare ipotesi quantitativamente testabili *in silico*. Poiché il modello viene utilizzato per dirigere un piano sperimentale, può essere importante riesaminare ulteriormente le proprietà biologiche dell'organismo.

Qui descriviamo il processo di costruzione di modelli matematici del metabolismo, mostrato schematicamente in Figura 9 (estratta da<sup>13</sup>) (il processo illustrato è riferito ad un metabolismo microbico, ma è altrettanto valido per organismi multicellulari). Dalla sequenza del genoma annotata e dalle caratteristiche biochimiche e fisiologiche determinate sperimentalmente per un dato organismo, la rete delle reazioni metaboliche può essere ricostruita (*metabolic network reconstruction*), per quanto è nota. La rete metabolica ricostruita viene quindi analizzata utilizzando varie tecniche matematiche di modellizzazione. Questi metodi di analisi quantitativa consentono la simulazione della crescita dell'organismo (o della colonia microbica) e del comportamento *in silico*.

I modelli matematici e la loro simulazione al computer ci consentono di esaminare la funzione integrata della rete metabolica ricostruita. Una rete ben definita di per sé non è sufficiente per descrivere quantitativamente il comportamento di un sistema, come mostrato in Figura 10 (estratta da<sup>13</sup>). In questa immagine viene tracciata un'analogia tra la simulazione delle condizioni del traffico in una città e la simulazione del comportamento di una rete metabolica. Il primo passo per entrambe le situazioni è generare un elenco dei componenti funzionali del sistema. Per la simulazione del traffico, questo potrebbe essere rappresentato da un elenco di tutte le strade principali della città, insieme ai luoghi che sono collegati da queste strade. Per una cellula, i prodotti genici vengono individuati e caratterizzati grazie al genoma annotato e ai dati sperimentali generati *de novo* o disponibili in letteratura.

In entrambe le situazioni, il passaggio successivo è determinare come sono collegati questi componenti funzionali. Queste informazioni possono essere integrate in una "mappa", ovvero una mappa stradale nel primo caso o una mappa metabolica ricostruita nel secondo. Una volta che una rete è stata descritta in modo sufficientemente dettagliato, è possibile formulare alcune previsioni qualitative. Ad esempio, una mappa stradale viene utilizzata per determinare il percorso nel viaggio da un luogo a un altro. Le distanze relative possono essere confrontate tra più percorsi possibili. Allo stesso modo, la rete metabolica ricostruita può essere utilizzata per studiare la connettività dei metaboliti e altre caratteristiche della struttura della rete.



**Figura 10:** Un'analogia tra la simulazione delle condizioni del traffico urbano e la simulazione di una cellula mediante l'analisi dei sistemi. Per entrambe le simulazioni, il primo passo è generare un elenco di tutti i componenti rilevanti (ad es. strade o prodotti genici) del sistema, dopo di che deve essere determinata e specificata l'integrazione di questi componenti. A questo punto si possono fare alcune previsioni qualitative sulle prestazioni del sistema. Infine, la modellizzazione matematica viene utilizzata per analizzare quantitativamente il sistema quando risponde a una serie di fattori ambientali o a un cambiamento nella rete. Abbreviazioni e spiegazioni: F6P, fruttosio-6-fosfato; G6P, glucosio-6-fosfato; IGGI (fosfogluccosio isomerasi) è l'enzima che catalizza la reazione.

La mappa stradale completa, tuttavia, presenta dei limiti. Ad esempio, sebbene sia possibile accertare la possibilità di guidare da una destinazione all'altra, il tempo di viaggio effettivo non è noto. Il tempo di viaggio dipende in parte dalle condizioni del traffico, che a loro volta dipendono dalla strada, dall'ora del giorno, dal tempo e da numerosi altri fattori che concorrono. Non è possibile ottenere tutti i dati necessari per specificare ciascun fattore che contribuisce. Tuttavia, stimando o approssimando molte di queste condizioni, creando così un modello per le condizioni del traffico che sia valido e realistico almeno in una particolare situazione, è possibile ottenere un calcolo

ragionevole del tempo di viaggio. Tale calcolo non poteva essere effettuato avendo a disposizione la sola mappa stradale.

Allo stesso modo, senza includere ulteriori informazioni, la mappa metabolica ricostruita di un organismo è limitata nella sua capacità di generare previsioni quantitative sul fenotipo. Il comportamento di una cellula dipende da molti fattori ambientali come la temperatura, oltre che dagli elementi di regolazione che rendono la cellula responsiva all'ambiente – come cascate di trasduzione del segnale e reti di modulazione dell'espressione genica – i quali non sono inclusi nel modello metabolico.

Inoltre il flusso di metaboliti che “attraversano” una reazione chimica nell'unità di tempo è principalmente dipendente dalla cinetica dell'enzima che catalizza quella particolare reazione (a meno che non si tratti di una reazione spontanea). Proseguendo nell'analogia con il traffico urbano, è come se dicessimo che il numero di autovetture che percorrono una strada in un'ora dipende dalla natura particolare di quella strada (ad es. condizione dell'asfalto o limite di velocità) e non soltanto dalla sua funzionalità all'interno della mappa stradale (ad esempio il fatto di connettere due principali punti d'interesse della città). Proprietà come la stechiometria sono relativamente facili da stabilire, mentre le proprietà cinetiche del metabolismo sono in genere molto più difficili da ottenere. Sono stati sviluppati diversi approcci alla modellizzazione cellulare dinamica, tuttavia è chiaro che affinché la costruzione di modelli dinamici dettagliati abbia successo su scala genomica, si devono compiere molti progressi nella stima dei parametri cinetici degli enzimi<sup>13</sup>.

#### **1.4.1 Applicazioni dei modelli metabolici di scala genomica**

Benché l'obiettivo di sviluppare un modello cellulare completamente specifico richiederà l'inclusione di parametri cinetici, metodi computazionali di analisi dei flussi e altri approcci hanno dimostrato avanzate capacità predittive dei modelli metabolici e hanno avuto successo in una moltitudine di applicazioni. Negli ultimi 20 anni, i modelli metabolici hanno raccolto un'attenzione crescente da parte della ricerca. Nel 1999 fu riportato il primo modello metabolico, un modello di *Haemophilus influenzae*<sup>14</sup>, e l'anno

successivo quello di *Escherichia coli*<sup>15</sup>. Nel 2003 fu pubblicato un modello metabolico di lievito<sup>16</sup>, dando così inizio a una nuova era per la biologia dei sistemi. Ad oggi, esistono modelli metabolici per un'ampia varietà di cellule, comprese quelle di microrganismi, piante e mammiferi. In alcuni casi, i modelli metabolici possono rappresentare l'intero tessuto o il corpo di un organismo pluricellulare<sup>17</sup>.

Gli utilizzi e le applicazioni dei modelli metabolici di scala genomica possono essere raggruppati in tre principali categorie:

1. Contestualizzazione di dati *high-throughput*,
2. Progettazione razionale nell'ingegneria metabolica,
3. Scoperta di proprietà sconosciute della rete.

**Contestualizzazione di dati *high-throughput*:** Mentre la biologia evolve sempre più nella direzione di una disciplina “*big data*”, una sfida emergente è stata determinare come organizzare, ordinare, mettere in relazione e contestualizzare tutti i dati *high-throughput* ora disponibili. I modelli metabolici di scala genomica sono stati utilizzati come impalcature per l'integrazione sistematica dei dati omici sfruttando la caratteristica dei modelli metabolici di contenere in forma strutturata l'informazione sulla relazione tra geni, enzimi e reazioni metaboliche (*Gene-Protein-Reaction relationship*).

I dati *high-throughput* possono essere sovrapposti a un modello metabolico di scala genomica in diversi modi. Se si tratta di dati di espressione genica (dati di espressione di proteine, dati di flusso <sup>13</sup>C, o dati di HPLC derivata dai profili di secrezione del prodotto), un modo altamente funzionale di utilizzare un modello metabolico per la contestualizzazione dei dati, è imporre direttamente vincoli al modello sulla base dei valori nel *dataset* sperimentale. Ad esempio, se un *dataset* sperimentale indica che gli enzimi glicolitici sono altamente attivi in una determinata condizione, allora il flusso nel modello può essere incanalato attraverso la glicolisi vincolando *in silico* i relativi flussi, forzando così il flusso attraverso le reazioni attivate e consentendo la valutazione dei cambiamenti nelle distribuzioni globali dei flussi.

**Progettazione razionale nell'ingegneria metabolica:** L'ingegneria metabolica prevede l'uso della tecnologia del DNA ricombinante per alterare selettivamente il metabolismo cellulare e migliorare in modo mirato una funzione cellulare target. La complessità delle reti metaboliche, intrecciata in più strati di regolazione di ogni enzima a livello trascrizionale, proteico e di substrato, rende estremamente difficile prevedere gli effetti dell'ingegneria metabolica e spesso causa conseguenze indesiderate o risultati non ottimali, quando la base delle decisioni ingegneristiche è rappresentata da intuizioni o conoscenze parziali del network metabolico. Gli svantaggi intrinseci dell'utilizzo di strumenti come questi ultimi nella guida dell'ingegneria metabolica hanno motivato l'uso di modelli metabolici di scala genomica in quella che è stata definita “*systems metabolic engineering*”. L'uso di modelli metabolici di scala genomica rappresenta una grande evoluzione per il settore, in cui le analisi a livello di sistema sono state per la prima volta portate al livello di determinare strategie ingegneristiche ottimali. Nel 2003, ad esempio, è stato pubblicato un metodo chiamato OptKnock che impiega un programma di ottimizzazione a due livelli per cercare reazioni che, “spente” tramite knock-out, inducano una sovrapproduzione di un metabolita desiderato, ma senza compromettere la crescita ottimale della coltura microbica.

**Scoperta di proprietà sconosciute della rete:** Con lo sviluppo di potenti strumenti di biologia molecolare nell'ultimo mezzo secolo, una mentalità riduzionista ha dominato la pratica della biologia. Tuttavia, gli scienziati hanno da tempo compreso l'importanza del pensiero olistico nello studio dei sistemi biologici, poiché reti cellulari complesse possono generare fenomeni emergenti che non sarebbero rilevabili da approcci riduzionisti. I geni condizionatamente essenziali (cioè i “letali sintetici”) non sarebbero osservati, per esempio, se i geni fossero studiati esclusivamente in isolamento. Negli ultimi anni, alcuni degli stessi strumenti alla base della biologia riduzionista sono stati estesi a metodi high-throughput, consentendo di:

- osservare l'intero patrimonio genetico dell'organismo (sequenziamento del genoma),
- “fotografare” insieme tutte le peculiarità genetiche che contraddistinguono l'organismo dagli altri individui della sua specie (genotipizzazione high-throughput),

- studiare simultaneamente numerose proprietà della rete esibite in una determinata condizione ambientale o di sviluppo (studio dell'espressione genica high-throughput),
- testare la rete in un colpo solo sotto numerose condizioni ambientali (fenotipizzazione high-throughput).

Tutti questi dati a livello di intero organismo hanno permesso e allo stesso tempo hanno richiesto la produzione di modelli metabolici di scala genomica, i quali a loro volta hanno consentito l'analisi dei fenomeni emergenti focalizzandosi sull'intera rete piuttosto che su singole vie o geni, e sono state sviluppate molte tecniche computazionali per sondare le proprietà della rete. Questi tipi di analisi a livello di rete saranno fondamentali per svelare completamente le complesse relazioni genotipo-fenotipo nelle cellule.

I modelli metabolici di scala genomica hanno reso possibile l'osservazione di proprietà altrimenti inaccessibili del network: questo è stato il contributo più diretto dei modelli metabolici alla nostra comprensione del metabolismo. Le proprietà metaboliche come l'esistenza di loop, l'utilizzo preferenziale di un pathway, la connettività dei metaboliti e la ridondanza dei pathways sono state tutte studiate in modelli metabolici di scala genomica usando metodi computazionali. Molte di queste analisi di rete vengono eseguite attraverso varianti di una tecnica computazionale, la FBA (vedi avanti). Uno dei principali fini per il quale sono state utilizzate le analisi di rete è il miglioramento delle annotazioni del genoma esistenti, come nel processo di analisi dei *gap* guidato dalla FBA per la costruzione di modelli. Tuttavia, sono stati sviluppati altri metodi più sistematici per migliorare le annotazioni del genoma, basati sulle analisi fornite da un modello metabolico di scala genomica. Uno di questi metodi ha utilizzato un modello metabolico di scala genomica di *S. cerevisiae* per ricavare annotazioni di geni metabolici condizionali, ottenendo una maggiore precisione rispetto alle annotazioni di ontologia genica nel determinare la funzione dei geni. Un'altra importante applicazione degli strumenti di analisi della rete è stata la scoperta di geni co-regolati, attraverso un processo computazionale chiamato *flux coupling analysis*. Questo metodo è stato recentemente validato attraverso i profili metabolici derivati da NMR (*Nuclear Magnetic Resonance*) di knockouts a singolo gene di lievito. I set di reazioni accoppiate

sono stati utilizzati per vari scopi, inclusa la previsione di nuovi bersagli farmacologici in *Mycobacterium tuberculosis*<sup>18</sup>.

#### **1.4.2 Ricostruzione di un modello metabolico di scala genomica**

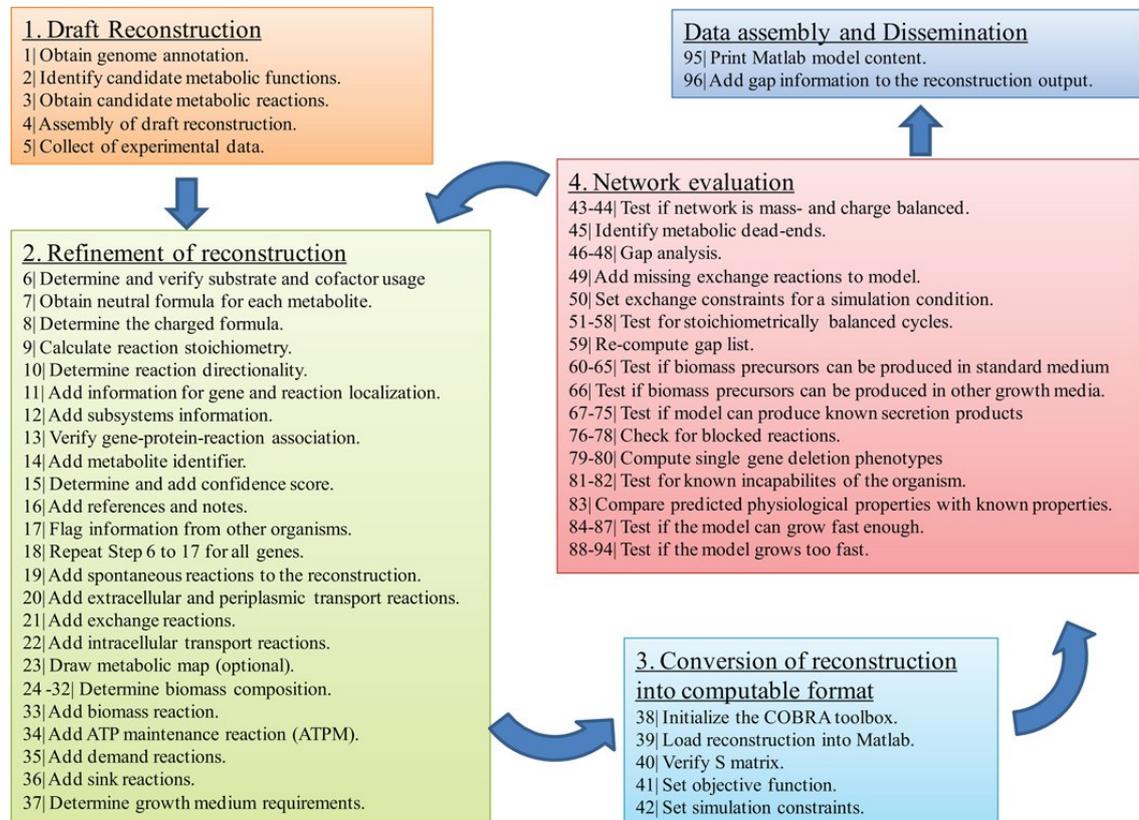
Il processo di ricostruzione della rete metabolica consiste di quattro fasi principali seguite dal suo uso potenziale nella fase 5 (Figura 11, estratta da<sup>19</sup>):

1. Creazione di una ricostruzione “draft” (grezza)
2. Perfezionamento della ricostruzione manuale
3. Conversione della ricostruzione in un modello matematico
4. Valutazione e correzione della rete
5. Utilizzo futuro

Ricostruire un network metabolico richiede una grande quantità di informazioni: in primis un genoma annotato, ma anche dati fenotipici e di chimica analitica. Non sempre queste informazioni sono disponibili o sufficientemente accurate per l'organismo oggetto di studio, ed in questi casi è possibile utilizzare specifici strumenti bioinformatici che permettono, a partire da un genoma, di realizzare in modo del tutto automatico un'annotazione genica e conseguentemente un primo modello draft. Questo processo è ovviamente soggetto ad errori, il che rende ancor più necessaria la seconda fase – quella più lunga – che consiste nella revisione del modello stesso. Per completare questa fase è necessario consultare estensivamente la letteratura e controllare ogni via metabolica, prestando attenzione alle reazioni, alla carica dei metaboliti, alla stechiometria e alla direzionalità delle reazioni, al mantenimento dell'ATP, alle reazioni di trasporto intra- ed extra-cellulare e all'esatta composizione della biomassa. A questi scopi è consigliabile prendere come riferimento organismi simili dal punto di vista filogenetico a quello che si vuole ricostruire.

La terza fase della realizzazione di un modello metabolico consiste nella conversione dello stesso in un formato matematico sul quale si possano svolgere simulazioni. I sistemi di modellizzazione matematica di un network metabolico possono essere molteplici, e si differenziano tra loro per la scala a cui possono essere applicati e per il

grado di accuratezza che forniscono. Possono suddividersi in sistemi di modellazione cinetica per mezzo di ODEs (equazioni differenziali ordinarie) e sistemi di modellazione “constraint-based”. Il sistema di simulazione discusso in questa tesi è un sistema “constraint-based” denominato “Flux balance analysis”, i dettagli del quale saranno spiegati tra poco.

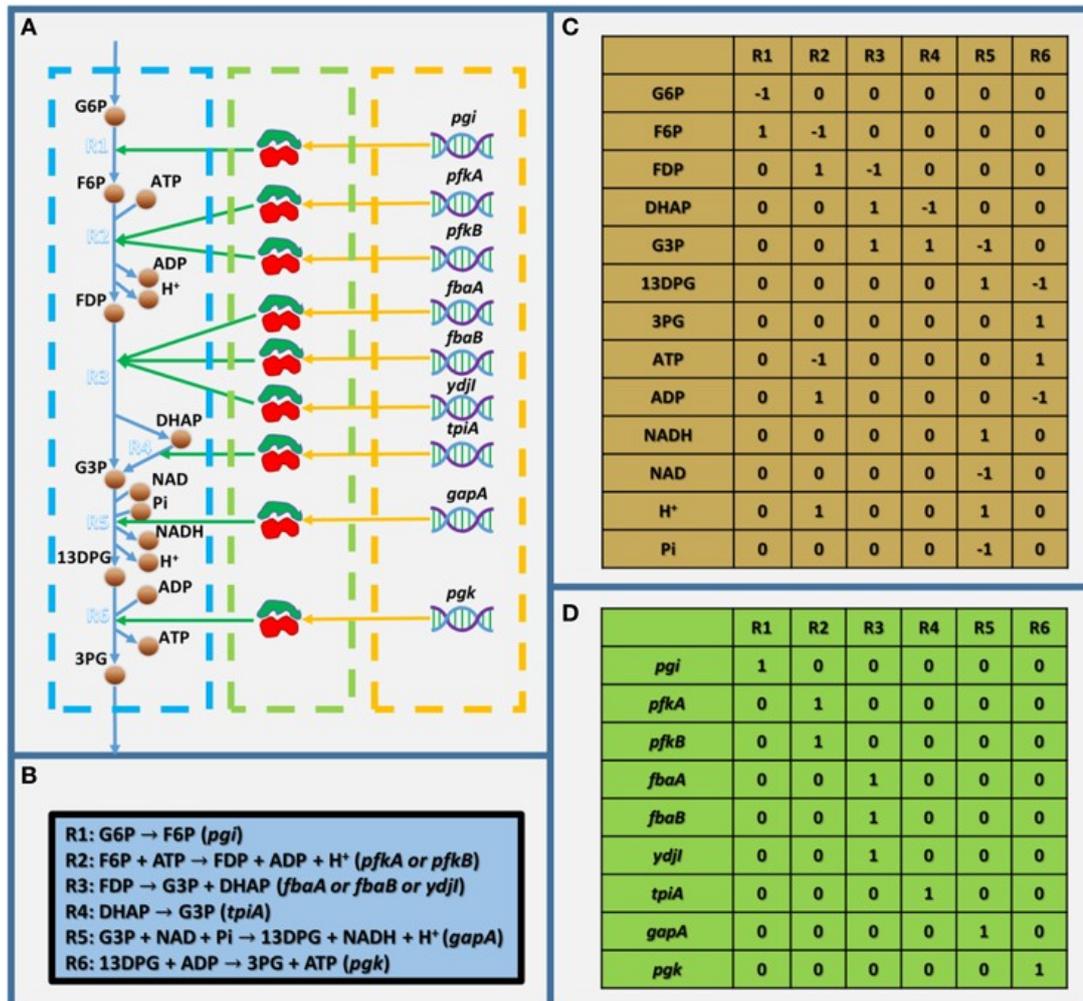


**Figura 11:** Panoramica della procedura per ricostruire iterativamente le reti metaboliche.

La quarta fase consiste nell'analisi e nel controllo della qualità del network ricostruito. In particolare ci si occupa della ricerca e del riempimento dei “gaps” (“buchi” nel network che tipicamente si devono alla mancanza di un certo numero di geni, sfuggiti al processo di annotazione automatica) e della correzione di errori compiuti nelle fasi precedenti. Se si hanno a disposizione dati sperimentali riguardanti l’organismo di interesse, è necessario continuare a rifinire il modello fino ad ottenere la massima aderenza possibile con la realtà.

Al termine di questa fase il modello può essere usato con profitto nelle simulazioni, ed è pronto per essere divulgato<sup>19</sup>.

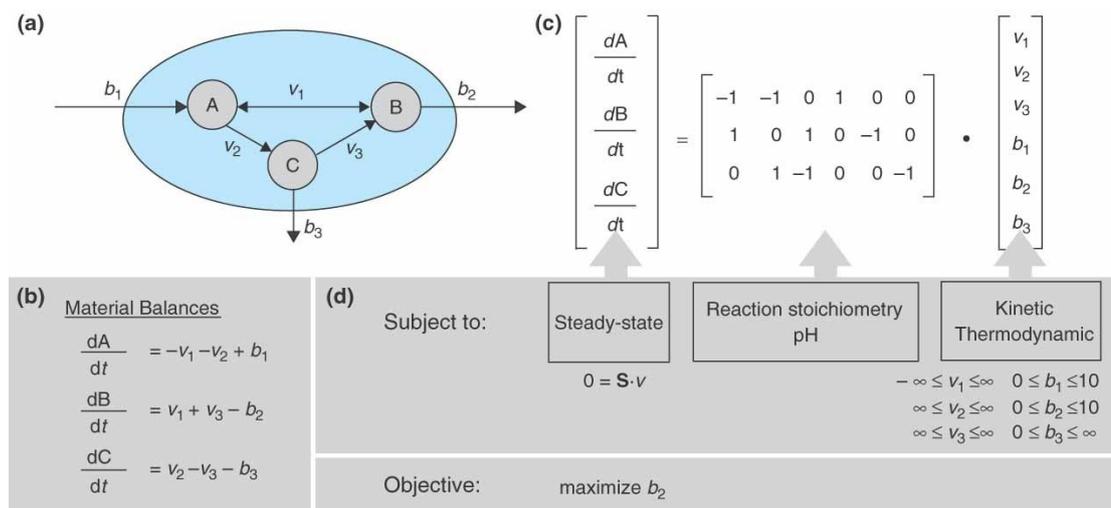
### 1.4.3 La modellizzazione basata su vincoli e la Flux Balance Analysis



**Figura 12:** Modello semplificato che mostra la struttura di base dei modelli metabolici di scala genomica: il modello analizzato è la "fase preparatoria" della glicolisi. (A) Illustrazione del modello semplificato, (B) equazioni biochimiche all'interno del modello semplificato, (C) matrice stechiometrica del modello semplificato e (D) matrice dell'associazione gene-reazione. In (A), i riquadri blu, verde e arancione tratteggiati indicano rispettivamente le reazioni metaboliche, gli enzimi e i geni. G6P, D-glucosio-6-fosfato; F6P, D-fruttosio-6-fosfato; FDP, D-fruttosio-1-6-bisfosfato; G3P, gliceraldeide-3-fosfato; 13DPG, 3-fosfo-D-gliceril-fosfato; 3PG, 3-fosfo-D-glicerato; e Pi, fosfato.

Sebbene esistano diversi approcci alla modellizzazione metabolica, tutti condividono un requisito fondamentale: una matrice stechiometrica basata su una rete metabolica ricostruita. Ogni colonna della matrice stechiometrica corrisponde a una reazione

chimica o di trasporto, con valori non-nulli che identificano i metaboliti che partecipano alla reazione e i coefficienti stechiometrici che corrispondono a ciascun metabolita (Figura 12, estratta da<sup>17</sup>). La matrice contiene anche la direzionalità delle reazioni: i substrati e i prodotti nella matrice hanno rispettivamente coefficienti negativi e positivi. Considerando le righe della matrice anziché le colonne, ogni riga della matrice stechiometrica può anche essere considerata come l'elenco delle reazioni a cui partecipa un determinato metabolita. Questa interpretazione è utile quando si bilancia la massa per ciascun metabolita nella rete.



**Figura 13:** (a) Illustrazione di una piccola rete di reazioni composta da tre metaboliti (A, B e C), tre reazioni di trasporto e tre reazioni enzimatiche.  $v_i$  indica il flusso attraverso la reazione  $i$  e  $b_j$  rappresenta il flusso attraverso la proteina di trasporto  $j$ . (b) Le equazioni del bilanciamento di massa sono mostrate per ciascun metabolita. (c) Una matrice stechiometrica è popolata secondo l'Equazione 1. (d) Ipotesi, vincoli, e una funzione obiettivo sono elencati per il sistema.

La conservazione della massa (Figura 13, estratta da<sup>20</sup>, riquadro b) per tutti i metaboliti è espressa da un sistema di equazioni differenziali scritte per tutte le concentrazioni  $c$  di metaboliti, come segue:

$$\frac{dc(t)}{dt} = S \cdot v \quad (1)$$

dove  $S$  è la matrice stechiometrica e  $v(t)$  il vettore dei flussi di reazione.

Si noti che il metabolismo opera su una scala temporale molto più rapida rispetto agli eventi di regolazione o di divisione cellulare. È quindi spesso ragionevole supporre che le dinamiche metaboliche abbiano raggiunto uno stato quasi stabile o pseudo-stabile, in cui le concentrazioni dei metaboliti non cambiano. Questo porta all'equazione di bilanciamento del metabolita:

$$S \cdot v(t) = 0 \quad (2)$$

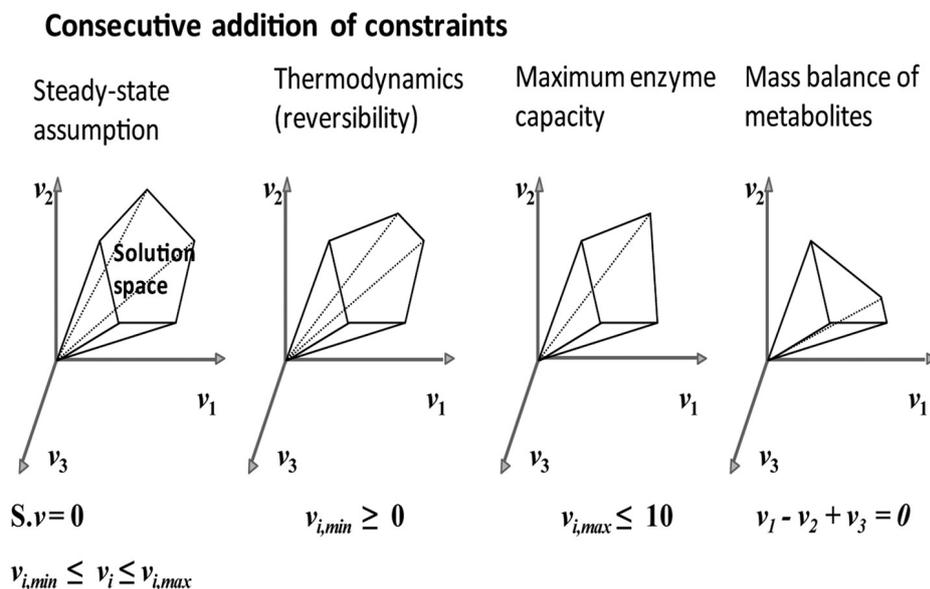
L'equazione (2) è un sistema omogeneo di equazioni lineari. Richiede che ciascun metabolita venga consumato nella stessa quantità con cui viene prodotto, e costituisce la base per un'ulteriore analisi dei flussi metabolici basata sulla matrice stechiometrica.

**Vincoli sui flussi delle reazioni:** La matrice stechiometrica può essere annotata includendo ulteriori informazioni importanti sulle reazioni o sui metaboliti. Le annotazioni sulla matrice più comuni includono la reversibilità di ogni reazione e il compartimento cellulare in cui si verifica ciascuna reazione. Più in generale, i flussi di reazione sono vincolati sulla base di costanti cinetiche, rapporti di concentrazione misurati o stimati, o per riflettere l'impostazione sperimentale (Figura 13d). I limiti superiore e inferiore possono essere applicati ai flussi di reazioni individuali ( $v_{min} \leq v \leq v_{max}$ ) e le direzioni di reazione possono essere definite semplicemente impostando  $v_{min}=0$  o  $v_{max}=0$  rispettivamente per le reazioni che possono procedere solo in avanti o solo indietro. L'annotazione aggiuntiva della matrice potrebbe includere informazioni più dettagliate sulla cinetica delle reazioni. Ad esempio, è possibile calcolare l'assorbimento dei metaboliti o i tassi di secrezione senza difficoltà in diversi casi; questi possono essere usati come parte del modello metabolico.

**Collegamento con le informazioni genomiche:** La matrice stechiometrica può anche essere esplicitamente collegata ai dati del genoma e dell'espressione genica per l'uso in alcune applicazioni (Figura 12D): si può collegare ogni reazione chimica o di trasporto alle proteine e ai geni che ne consentono il verificarsi. Queste relazioni (*Gene-Protein-Reaction relationship*) sono spesso complicate, poiché le proteine sono spesso costituite da più subunità, a volte più enzimi catalizzano la stessa reazione chimica e alcuni enzimi catalizzano più di una reazione.

**La FBA consente di identificare soluzioni particolari del sistema:** Il comportamento metabolico della popolazione di cellule è definito da una particolare combinazione di valori dei flussi delle reazioni del network. La distribuzione dei flussi nella rete metabolica rappresenta quindi ciò che si vuole ottenere quando si utilizza il modello per eseguire una simulazione computazionale del comportamento della popolazione di cellule. La soluzione è una grandezza  $n$ -dimensionale, dove  $n$  è il numero delle reazioni comprese nella rete, e lo spazio delle soluzioni è rappresentato da tutti i possibili valori assumibili dalla distribuzione dei flussi. Tale spazio  $n$ -dimensionale è illimitato fintanto che non vengono posti vincoli sul sistema, ma si restringe (Figura 14, estratta da<sup>21</sup>) allorché si impongono:

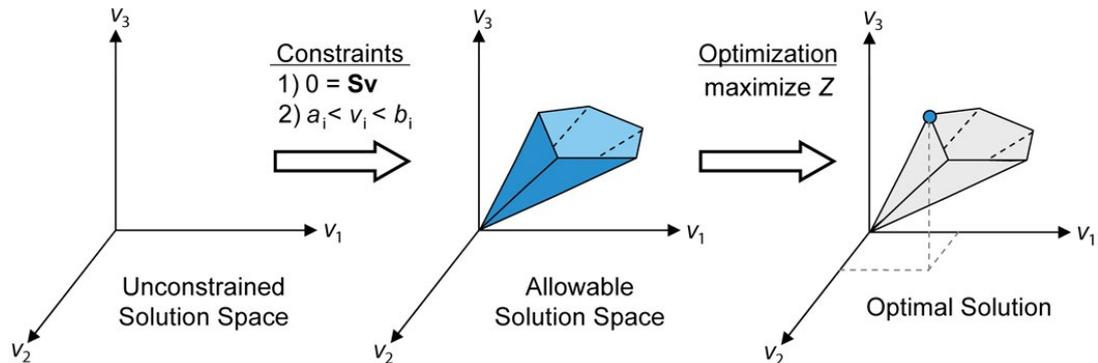
1. Il bilanciamento della massa di ogni metabolita,
2. L'assunzione di pseudo-stabilità del sistema,
3. I limiti inferiore e superiore sui flussi per ogni reazione della rete.



**Figura 14:** Aggiunta dei vincoli nella procedura di ricostruzione e analisi basate sui vincoli. L'aggiunta dei vincoli al modello riduce lo spazio delle soluzioni eliminando le soluzioni biologicamente impossibili.

Poiché questi vincoli possono essere rappresentati come equazioni e disequazioni lineari, i metodi di programmazione lineare (LP) possono essere utilizzati per identificare punti con valori ottimali massimizzando oppure minimizzando una determinata **funzione obiettivo**. Il termine *flux balance analysis* indica l'applicazione di

metodi di LP per analizzare i flussi in condizioni di bilanciamento dei metaboliti e pseudo-stabilità del sistema (*flux balance*).



**Figura 15:** Le basi concettuali della modellizzazione basata sui vincoli e dell'FBA. Senza vincoli, la distribuzione dei flussi in una rete biologica può trovarsi in qualsiasi punto dello spazio delle soluzioni. Quando i vincoli di bilanciamento della massa imposti dalla matrice stechiometrica  $S$  (1) e i vincoli di capacità enzimatiche imposti dai limiti inferiore e superiore ( $a_i$  e  $b_i$ ) (2) vengono applicati alla rete, viene definito uno spazio delle soluzioni consentito. La rete può acquisire qualsiasi distribuzione di flussi all'interno di questo spazio. Attraverso l'ottimizzazione di una funzione obiettivo, l'FBA è in grado di identificare un'unica distribuzione di flussi ottimale che si trova ai margini dello spazio delle soluzioni consentito.

Nel caso più semplice, viene ottimizzato un singolo flusso di reazione. Molte diverse funzioni obiettivo sono state testate per la loro utilità nel predire il comportamento fenotipico, basandosi su modelli di reti su scala ridotta. Una pratica comune è quella di definire una reazione di crescita artificiale che drena i componenti chimici del peso secco della cellula, nelle rispettive proporzioni, e produce un metabolita virtuale che rappresenta la biomassa cellulare. Questa reazione può essere utilizzata come funzione obiettivo in determinate condizioni, a seconda del tipo di cellula e delle condizioni ambientali. L'idea alla base di questo approccio è che qualsiasi organismo, in natura, si è evoluto in modo tale da ottimizzare il proprio set di reazioni metaboliche per ottenere la crescita più rapida possibile. Ma la funzione obiettivo usata in una simulazione FBA può essere qualunque reazione o set di reazioni si desidera dichiarare come tale, sulla base di quello che ci si aspetta essere il comportamento reale del sistema biologico che si sta simulando. La dichiarazione di una funzione obiettivo permette quindi di

identificare la distribuzione ottimale dei flussi metabolici nel network all'interno dello spazio  $n$ -dimensionale definito dai vincoli imposti (Figura 15, estratta da<sup>22</sup>)<sup>20</sup>.

## 1.5 ViNE – Virtual Nodule Environment

Nonostante l'importanza del metabolismo per la fissazione simbiotica dell'azoto, c'è stato un uso limitato della modellizzazione metabolica nello studio dei rizobi e della fissazione simbiotica dell'azoto. Ad oggi, sono stati riportati modelli metabolici di scala genomica di varia qualità per solo tre rizobi: *Sinorhizobium meliloti*, *Rhizobium etli* e *Bradyrhizobium diazoefficiens*. Attualmente, *Medicago truncatula* e *Glycine max* (la soia) sono gli unici legumi con modelli metabolici di scala genomica pubblicati. Ad eccezione del modello metabolico di scala genomica della soia, questi modelli sono stati utilizzati nelle analisi preliminari sulla fissazione simbiotica dell'azoto, fornendo risultati generalmente coerenti con le aspettative. Tuttavia, fino ad oggi tutte le analisi presentano due importanti limiti. Le simulazioni con i modelli di rizobi ignorano il metabolismo delle piante, mentre le simulazioni con il modello metabolico di scala genomica di *M. truncatula* hanno sfruttato una ricostruzione metabolica di *S. meliloti* molto approssimativa. Inoltre, tutte le simulazioni si sono concentrate sullo stadio finale della fissazione simbiotica dell'azoto e non hanno preso in considerazione le diverse fasi della preliminare progressione dello sviluppo del nodulo, in cui il metabolismo rimane scarsamente compreso.

Queste lacune hanno motivato diCenzo *et al.* (2019)<sup>23</sup> a produrre una rappresentazione olistica *in silico* del metabolismo integrato dell'olobionte costituito da una pianta di *M. truncatula* nodulata da *S. meliloti*. Tale modello è stato chiamato **Virtual Nodule Environment** (ViNE). Questa ricostruzione combinata multi-compartimento tiene conto dell'attività metabolica dei tessuti radicali e aerei, insieme a un nodulo costituito da cinque zone di sviluppo.

Nel medesimo studio viene presentata una caratterizzazione inaugurale del ViNE, che include analisi sulle proprietà metaboliche specifiche per ogni zona, sul rapporto tra

l'azotofissazione e la crescita della pianta e sull'uso di dicarbossilati come fonte di carbonio da parte dei batteroidi. In futuro ci si aspetta che tale modello metabolico fornirà una potente piattaforma per la generazione di ipotesi volta a comprendere e valutare quantitativamente la fissazione simbiotica dell'azoto, nonché a guidare i tentativi di ingegnerizzare la fissazione simbiotica dell'azoto verso una maggiore efficienza simbiotica.

Riportiamo quindi una sintesi del processo di ricostruzione del modello metabolico.

### **Preparazione di una ricostruzione metabolica migliorata di *Sinorhizobium meliloti***

Una nuova ricostruzione metabolica di *S. meliloti* è stata costruita utilizzando come base di partenza l'esistente ricostruzione metabolica "core" iGD726, la quale è stata integrata attraverso il modello metabolico di scala genomica iGD1575 di, e attraverso fonti di letteratura. La rifinitura manuale del modello ha prodotto la ricostruzione finale, denominata iGD1348, e contenente 1348 geni, 1407 reazioni (1164 associate ad almeno un gene) e 1160 metaboliti.

### **Aggiornamento della ricostruzione della rete metabolica di *Medicago truncatula***

La ricostruzione della rete metabolica di *M. truncatula* si è basata su un modello metabolico di scala genomica già pubblicato. Tale modello disponibile è stato costruito sulla versione Mt3.5v5 di *M. truncatula*, una versione ormai obsoleta dato che in seguito alla pubblicazione il genoma di *M. truncatula* è stato risequenziato per due volte, affinando l'annotazione genomica. Pertanto, è stato necessario aggiornare il modello metabolico all'ultima versione disponibile dell'annotazione genomica, cioè la versione 5.0. Una volta aggiornata la ricostruzione, è stata perfezionata (ad esempio, è stata associata una spesa di ATP alla maggior parte delle reazioni di trasporto, dato che nella ricostruzione pubblicata erano erroneamente modellizzate come reazioni di diffusione facilitata). La ricostruzione modificata contiene 2522 geni, 2920 reazioni (1722 associate ad almeno un gene) e 2742 metaboliti. Infine, si è sfruttata la versione definitiva della ricostruzione metabolica per produrre un modello specifico per il tessuto della radice e un modello specifico per i tessuti aerei della pianta.

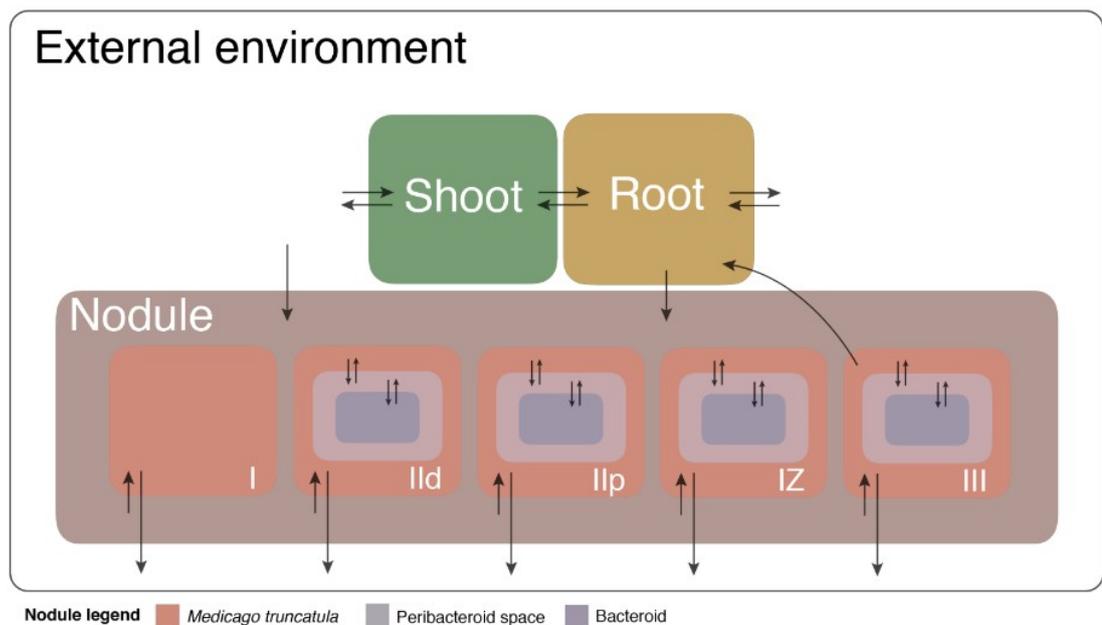
### **Ricostruzione del metabolismo di una pianta di *M. truncatula* nodulata**

Una volta importati in MATLAB in formato COBRA il modello di *M. truncatula* e il modello di *S. meliloti*, e aggiunte tutte le reazioni di trasporto (e non solo di trasporto) necessarie a modellizzare gli scambi metabolici tra la cellula vegetale e il simbiosoma in essa contenuto, quattro copie del modello integrato *M. truncatula* – *S. meliloti* sono state preparate per rappresentare quattro distinte zone di sviluppo del nodulo: zona II distale, zona II prossimale, interzona II-III e zona III di azotofissazione (si rimanda alla Figura 3 per l'illustrazione della suddivisione del nodulo in zone). Inoltre, è stata inclusa una versione del modello di *M. truncatula* precedente all'integrazione con *S. meliloti* per rappresentare la zona I (meristema apicale).

Successivamente, tutte le reazioni di scambio di *S. meliloti* e tutte le reazioni di trasporto di *M. truncatula* sono state eliminate in ciascuno dei cinque modelli. Eccezione è stata fatta per la zona nodulare III, dove l'importazione di  $N_2$  e l'esportazione di  $H_2$  da parte di *M. truncatula* non sono state rimosse. A questo modello, una reazione irreversibile che converte i protoni dello spazio peribatteroide in protoni periplasmatici di *S. meliloti* è stata aggiunta esclusivamente nella zona nodulare III. Successivamente, si sono aggiunte le reazioni di scambio tra il nodulo e la radice e tra il nodulo e l'ambiente esterno. Infine, sono state aggiunte reazioni per trasferire singolarmente asparagina e glutammina dal citoplasma di *M. truncatula* della zona nodulare III (la zona di azotofissazione) al tessuto radicale. La Figura 16 (estratta da<sup>23</sup>) riassume la struttura del modello.

Una serie di reazioni di biomassa sono state aggiunte al modello combinato. Una reazione di biomassa specifica per ogni zona nodulare è stata aggiunta alla zona II distale, alla zona II prossimale e all'interzona II-III, in ciascuna combinando le biomasse di *M. truncatula* e *S. meliloti* con un rapporto 75 : 25. La biomassa della zona I consiste solo della biomassa di *M. truncatula*. Nessuna reazione di biomassa è stata aggiunta alla zona III poiché lo scopo di questa zona era di fissare l'azoto. Successivamente, è stata preparata una reazione globale della biomassa del nodulo combinando la biomassa della zona I, della zona II distale, della zona II prossimale e dell'interzona II-III in un rapporto 5 : 45 : 45 : 5. È stata inoltre preparata una reazione

di biomassa vegetale combinando la biomassa della parte aerea e quella della radice in un rapporto di 66,7 : 33,3. Infine, è stata preparata una reazione globale della biomassa che combinava la biomassa vegetale con la biomassa nodulare in un rapporto di 98 : 2. La reazione globale della biomassa è stata impostata come funzione obiettivo durante tutte le simulazioni FBA, se non diversamente indicato. Il tasso totale di utilizzo dell'ossigeno da parte della pianta e dei batteroidi della zona nodulare III è stato limitato sulla base di dati sperimentali pubblicati.



**Figura 16:** Rappresentazione visiva del ViNE. Riassunto schematico della struttura generale della pianta *M. truncatula* nodulata da *S. meliloti*. Il modello contiene tre tessuti vegetali (parte aerea, radice, nodulo) con il nodulo suddiviso in cinque zone di sviluppo. Le frecce indicano le reazioni di trasporto con la direzione rappresentativa della direzionalità delle reazioni di trasporto. La scala della figura non ha significato.

### Sovrapporre al modello i dati di espressione genica delle zone nodulari

Lo spazio delle reazioni di ciascuna zona nodulare è stato vincolato in base ai dati di RNA-seq zona-specifici di *M. truncatula* – *S. meliloti* di Roux e collaboratori<sup>24</sup>. Tali dati di sequenziamento dell'RNA sono stati sottoposti ad analisi statistiche, con la finalità di stabilire una distinzione tra geni altamente espressi o comunque indispensabili per la crescita e geni scarsamente espressi o non espressi. I geni classificati nella seconda categoria sono stati eliminati dal modello (soltanto quelli relativi al nodulo o ai batteri, e

non quelli relativi alla radice o alla parte aerea della pianta), e con essi le rispettive reazioni metaboliche o di trasporto.

### **Procedure di modellizzazione**

Il software impiegato per la ricostruzione del modello metabolico e per le simulazioni di Flux Balance Analysis è lo stesso che si è utilizzato per condurre le analisi su ViNE che sono oggetto di questa tesi. Si veda in “Materiali e Metodi” per i dettagli.

## **2. Scopo del lavoro**

Questo lavoro di tesi si è servito del modello metabolico di scala genomica denominato ViNE pubblicato da diCenzo *et al.* che ricostruisce il metabolismo di una pianta di *Medicago truncatula* nodulata dal batterio *Sinorhizobium meliloti*. Sfruttando tale modello metabolico, si sono effettuati due studi. Il primo è consistito nella ricerca di reazioni batteriche le cui variazioni di flusso avessero un particolare impatto sul metabolismo della pianta e quindi sul suo tasso di crescita, e nella classificazione delle reazioni batteriche sulla base dell'effetto delle loro variazioni di flusso sul tasso di crescita della pianta. Questa analisi, chiamata *Robustness Analysis*, si è potuta svolgere soltanto in seguito alla messa a punto di un codice che eseguisse l'analisi in maniera esaustiva. Nel secondo studio, si è analizzato l'effetto di valori sub-ottimali di efficienza di azotofissazione, parzialmente compensati da una maggiore nodulazione, sul tasso di crescita della pianta.

### 3. Materiali e metodi

Nello studio di diCenzo *et al.*<sup>23</sup>, per produrre il modello integrato ViNE, per manipolarlo e per eseguire le simulazioni di Flux Balance Analysis, sono stati utilizzati MATLAB versione R2016b e COBRA Toolbox commit 9b10fa1. Quasi tutti i calcoli di *linear programming* implicati nelle simulazioni di FBA sono stati eseguiti con il solver iLOG CPLEX Studio 12.7.1. In questo lavoro di tesi, si è utilizzato il medesimo software, al fine di standardizzare i risultati ottenuti.

#### **MATLAB**

MATLAB<sup>25</sup> (MATrix LABoratory) è un ambiente per il calcolo numerico e l'analisi statistica, e comprende un omonimo linguaggio di programmazione di alto livello sviluppato da MathWorks. MATLAB permette di manipolare matrici, visualizzare graficamente dati e funzioni, sviluppare algoritmi, creare interfacce utente, ed è in grado di interfacciarsi con programmi scritti in altri linguaggi – tra cui C, C++, Java, Fortran e Python.

#### **COBRA Toolbox**

Il toolbox COBRA<sup>26</sup> è un pacchetto in linguaggio di programmazione MATLAB (disponibile anche in Python) che permette di effettuare predizioni quantitative di comportamenti cellulari utilizzando un approccio di “CONstraint-Based Reconstruction and Analysis” (COBRA).

L'approccio COBRA permette di creare ed analizzare modelli metabolici senza aver necessità di un gran numero di parametri determinati sperimentalmente riguardo alla fisiologia dell'organismo che si vuole ricostruire: vengono definiti una serie di vincoli biofisici (suddivisione in compartimenti, conservazione della massa e direzionalità delle reazioni) che impongono dei limiti al numero di stati in cui il sistema può trovarsi e lo rendono risolvibile in maniera univoca da tecniche di calcolo come la FBA.

La prima versione del pacchetto è stata sviluppata nel 2007 da Bernhard Palsson e dal suo gruppo della UC San Diego 66, e permetteva inizialmente di effettuare calcoli di FBA per analizzare i flussi metabolici in un network e per identificare geni essenziali. Col tempo sono state aggiunte – e continuano ad essere implementate – nuove funzioni quali FBA dinamica, identificazione e rimozione di loop, creazione di sub-network, flussomica e simulazione di knock-out genici.

### **3.1 Come è stata eseguita la Robustness Analysis**

L'analisi di robustezza (*Robustness Analysis*, RA) è un metodo che si serve della FBA per indagare il rapporto tra il flusso attraverso una data reazione e il flusso attraverso la funzione obiettivo. L'analisi rivela quanto la funzione obiettivo sia sensibile alle variazioni di flusso di una particolare reazione. Il flusso attraverso tale reazione viene variato, e per ciascuno dei valori testati viene ottimizzata la funzione obiettivo e registrato il suo valore. In altre parole, il flusso ottimizzato della reazione obiettivo viene studiato come una funzione del flusso attraverso la data reazione.

Lo studio di RA svolto in questo lavoro di tesi è consistito nella ricerca di reazioni batteriche le cui variazioni di flusso avessero un particolare impatto sul metabolismo della pianta e quindi sul suo tasso di crescita, e nella classificazione delle reazioni batteriche sulla base dell'effetto delle loro variazioni di flusso sul tasso di crescita della pianta. L'obiettivo è stato quello di eseguire la RA su ciascuna delle reazioni del metabolismo dei rizobi che invadono il nodulo (zona II distale, zona II prossimale e interzona II-III) e di quelli che si sono già differenziati in batteroidi azotofissatori (zona III). (Il codice è riportato per intero in Appendice).

Inizialmente, per eseguire la RA, si è sfruttata l'apposita funzione nativa offerta nel COBRA Toolbox, cioè la funzione `robustnessAnalysis`. Di seguito l'utilizzo della funzione:

- *Inputs:*
  - `model` – struttura rappresentata da un modello COBRA

- `controlRxn` – Reazione d’interesse di cui si vuole testare la variazione di flusso
- *Optional Inputs:*
  - `nPoints` – Numero di punti in cui viene diviso l’intervallo testato, ovvero numero dei valori di flusso testati della reazione di controllo (default = 20)
  - `plotResFlag` – traccia il grafico dei risultati (default `true`)
  - `objRxn` – Reazione obiettivo, da massimizzare (di default viene assunta la funzione obiettivo attualmente impostata nel modello)
  - `objType` – Massimizzazione (`‘max’`) o minimizzazione (`‘min’`) della funzione obiettivo (default = massimizzazione)
- *Outputs:*
  - `controlFlux` – valori di flusso all’interno della gamma del massimo e del minimo per la reazione d’interesse
  - `objFlux` – Valore ottimale della funzione obiettivo per ogni valore di flusso della reazione di controllo

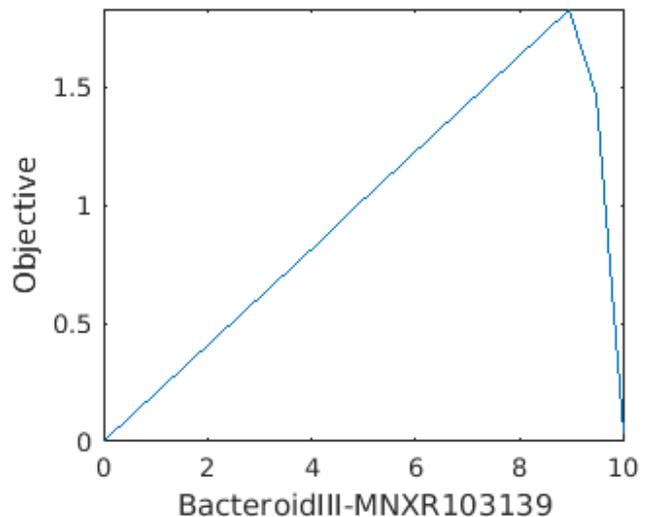
La funzione `robustnessAnalysis` individua un intervallo di valori di flusso della reazione di controllo, compresi tra un estremo sinistro e un estremo destro (cioè minimo e massimo), su cui eseguire la RA. Suddivide tale intervallo in  $n$  punti, a ciascuno dei quali corrisponde un valore di flusso della reazione di controllo. Per ciascuno di questi valori di flusso, la funzione esegue le seguenti operazioni:

1. Imposta sia il limite sinistro che il limite destro del flusso esattamente al valore da testare: in questo modo viene forzato quel preciso valore di flusso attraverso la reazione di controllo;
2. Ottimizza la funzione obiettivo – che in questo caso equivale a dire che massimizza il flusso attraverso la reazione di biomassa complessiva del modello (o tasso di crescita della pianta nodulata);
3. Registra il valore di flusso testato alla  $i$ -esima posizione di un vettore dedicato (`controlFlux`), e registra il valore ottimale della funzione obiettivo alla  $i$ -esima posizione di un altro vettore apposito (`objFlux`).

A titolo di esempio, vengono mostrate la Tabella 1 e la Figura 17 che rappresentano l'output della RA eseguita con la funzione nativa `robustnessAnalysis` su una reazione (`BacteroidIII_MN XR103139`) scelta a caso dal network metabolico dei batteroidi della zona nodulare III. Dalla tabella si può vedere che il valore 0,000 è stato il più piccolo valore testato, ovvero rappresenta l'estremo sinistro dell'intervallo testato, mentre 10,024 è stato il più grande valore testato, cioè rappresenta l'estremo destro dell'intervallo testato.

controlFlux	objFlux
0,000	0,000
0,528	0,108
1,055	0,215
1,583	0,323
2,110	0,431
2,638	0,539
3,166	0,646
3,693	0,754
4,221	0,862
4,748	0,969
5,276	1,077
5,803	1,185
6,331	1,292
6,859	1,400
7,386	1,508
7,914	1,616
8,441	1,723
8,969	1,831
9,497	1,462
10,024	0,025

**Tabella 1:** La colonna sinistra contiene il vettore dei valori di flusso testati della reazione di controllo, mentre la colonna destra contiene i corrispondenti valori di flusso della reazione obiettivo ottimizzata.



**Figura 17:** Il grafico mostra la reazione obiettivo ottimizzata in funzione del valore imposto attraverso la reazione di controllo. L'unità di misura del flusso, sia in ascissa che in ordinata, è quella impostata nel modello.

In che modo la funzione nativa `robustnessAnalysis` individua l'intervallo da testare?

La funzione `robustnessAnalysis` stabilisce i due estremi in questo modo:

1. Cambia momentaneamente la funzione obiettivo, impostando la reazione di controllo come funzione obiettivo.
2. Minimizza la nuova funzione obiettivo e registra questo valore come estremo sinistro dell'intervallo da testare.

3. Massimizza la nuova funzione obiettivo e registra questo valore come estremo destro dell'intervallo da testare.
4. Cambia nuovamente la funzione obiettivo, ripristinando la funzione obiettivo originale (in questo caso il tasso di crescita del modello).
5. Infine esegue la vera e propria RA, come spiegato precedentemente.

Durante lo svolgimento del lavoro di RA sul modello, si è trovato che questo criterio utilizzato dalla funzione nativa `robustnessAnalysis` per stabilire l'intervallo da testare rappresentava un notevole ostacolo per la corretta esecuzione della RA su tutte le reazioni metaboliche da studiare. In sostanza, la funzione nativa `robustnessAnalysis` si è rivelata inadeguata ad eseguire la RA in modo esaustivo.

Per sopperire a questa carenza, si è svolta l'intera RA da capo, ma stavolta utilizzando una funzione scritta *de novo* per essere adatta a qualsiasi reazione studiata. L'esecuzione della RA di sé per sé è rimasta inalterata: si divide l'intervallo tra l'estremo sinistro e l'estremo destro in  $n$  punti, e così si trovano  $n+1$  punti, ciascuno corrispondente a un valore di flusso per il quale quindi massimizzare il tasso di crescita del modello. La variazione, rispetto alla funzione nativa del COBRA Toolbox, consiste nella determinazione dell'intervallo da testare.

A cominciare dal valore di flusso ottimale per la reazione di controllo, si cerca a destra approssimativamente il primo valore per cui il tasso di crescita è azzerato, e si imposta tale valore di flusso come estremo destro dell'intervallo. La stessa operazione si ripete a sinistra del valore di flusso ottimale.

Ecco in dettaglio la strategia che si è usata per ogni reazione per identificare i due estremi dell'intervallo dei valori  $X$  su cui eseguire l'analisi di robustezza: per prima cosa, si ottimizza il modello, massimizzando il tasso di crescita, e si registra il flusso ottimale della reazione di controllo.

Idealmente, si vuole trovare il primo valore di flusso, a destra del flusso ottimale, che azzeri il tasso di crescita. Dato che questa operazione sarebbe molto onerosa dal punto

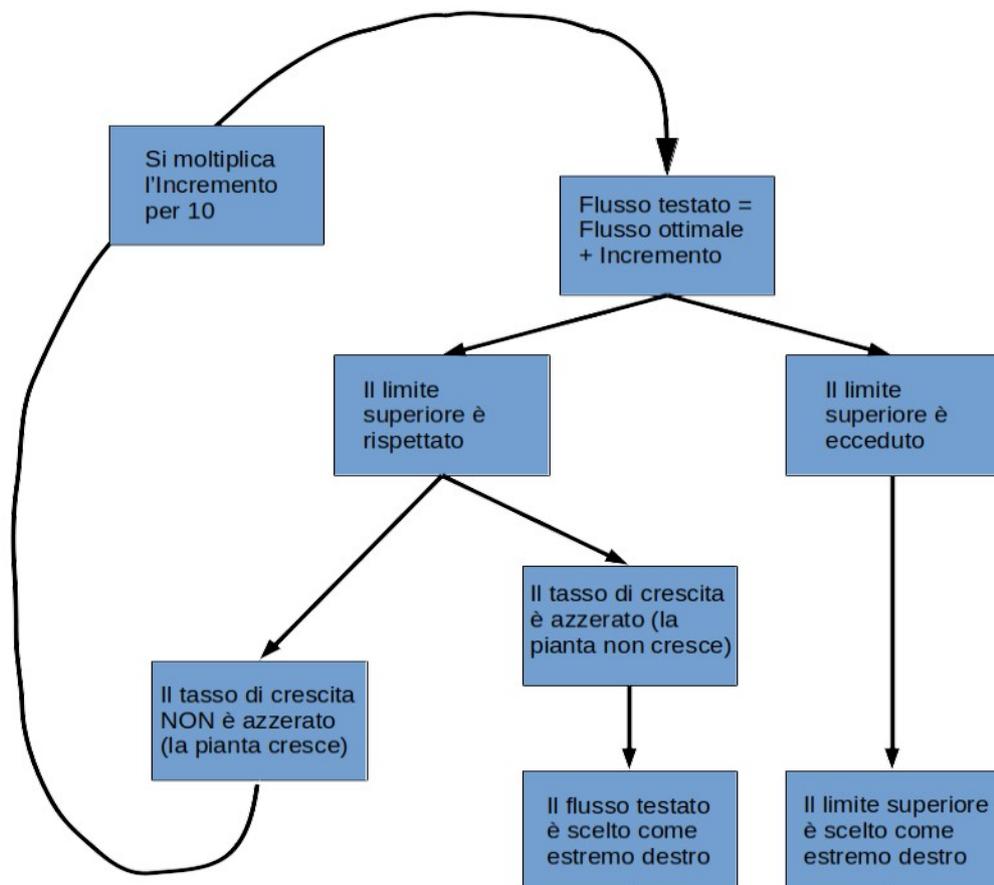
di vista computazionale, e non indispensabile, si è preferita una ricerca per approssimazione. Si cerca quindi soltanto l'ordine di grandezza del primo valore di flusso che azzerava la crescita. Per farlo, si è utilizzato il seguente loop `while`. (Il loop `while` ripete una data operazione  *fintanto*  che una particolare condizione è rispettata, dopodiché si interrompe).

Per definizione, il flusso ottimale corrisponde ad un picco della funzione obiettivo. Pertanto, allontanandosi dal valore di flusso ottimale, la funzione obiettivo certamente non crescerà (quindi, o rimane costante o diminuisce). Il loop `while` funziona in questo modo: si definisce un incremento iniziale molto piccolo (sperimentalmente si è determinato che  $10^{-5}$  è un intervallo sufficientemente piccolo), e quindi si apre il loop `while`. All'interno del loop `while`, innanzitutto, si definisce il valore del flusso da testare come somma del flusso ottimale e dell'incremento scelto (`forcedFlux = optFlux + increment`). A questo punto, si impone che il limite inferiore del flusso di reazione e il suo limite superiore siano esattamente uguali a `forcedFlux`, e si ottimizza il modello, massimizzando il tasso di crescita. *In primis* ci si assicura che tale flusso testato (`forcedFlux`) non ecceda il limite superiore che normalmente è imposto per quella data reazione: se il flusso testato è superiore o uguale a quello che normalmente è il limite superiore di flusso per la reazione di controllo, allora si è già trovato l'estremo destro, e cioè il limite superiore stesso, e quindi si conclude il loop `while`. Una volta verificato che il flusso testato non ecceda il limite superiore di flusso per la reazione di controllo, si procede così:

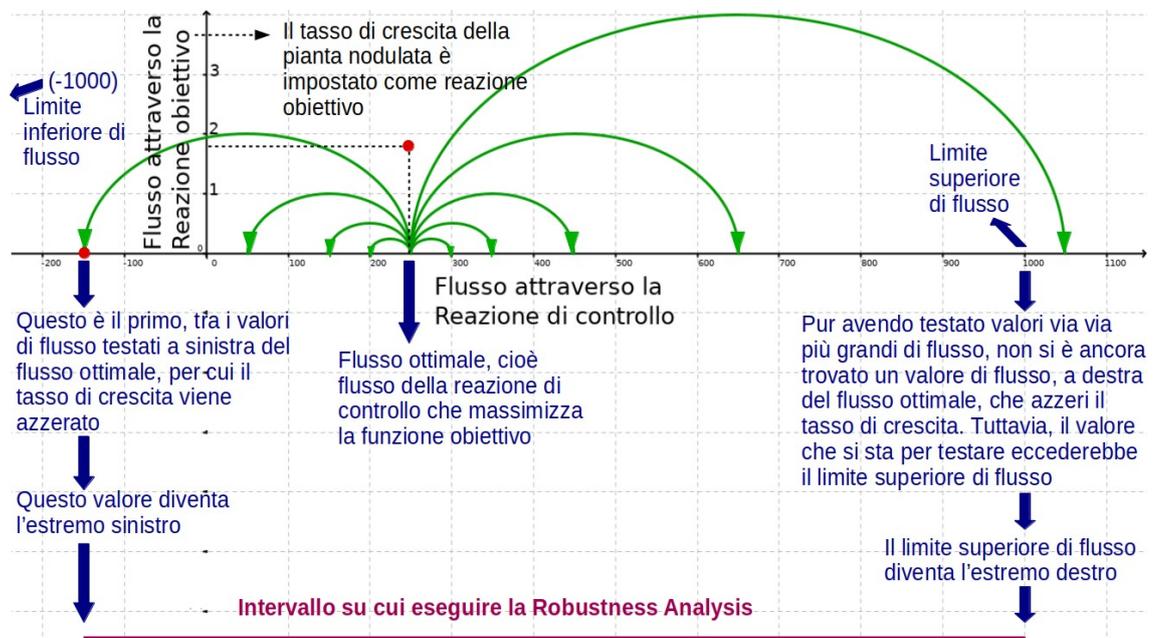
1. Se il tasso di crescita NON è maggiore di zero, significa che si è trovato un valore che azzerava il tasso di crescita, e quindi si sceglie tale valore come estremo destro dell'intervallo.
2. Se invece il tasso di crescita è ancora maggiore di zero, significa che è necessario spostarsi di nuovo sull'asse X per trovare un valore che azzeri il tasso di crescita. Pertanto, si moltiplica l'incremento per 10 e si ripete l'operazione codificata nel loop `while`. Se nel round precedente il flusso testato era `forcedFlux = optFlux +  $10^{-5}$` , adesso sarà `forcedFlux = optFlux +  $10^{-4}$` . Se il tasso di crescita è azzerato (oppure se il flusso testato eccede il limite superiore), allora si sceglie questo valore come estremo destro e si conclude il

loop while, altrimenti si moltiplica di nuovo l'incremento per 10 e si ripete il loop while.

Con questo loop while, si trova l'estremo destro dell'intervallo su cui eseguire la RA. Successivamente, si utilizza una procedura simile per trovare anche l'estremo sinistro. Una volta impostati l'estremo sinistro e l'estremo destro, si esegue la RA nell'intervallo tra questi due valori. Questa funzione scritta *de novo* ha permesso di eseguire correttamente la RA su tutte le reazioni d'interesse. La Figura 18 e la Figura 19 riassumono schematicamente e graficamente la strategia utilizzata per determinare l'intervallo su cui eseguire la RA.



**Figura 18:** Riassunto schematico della strategia utilizzata per determinare l'estremo destro dell'intervallo su cui eseguire la *Robustness Analysis*. L'estremo sinistro viene individuato attraverso una procedura analoga.



**Figura 19:** Esempificazione schematica della strategia utilizzata per trovare i due estremi dell'intervallo su cui eseguire la *Robustness Analysis*. Nell'esempio, si testano valori di flusso via via minori, a sinistra del flusso ottimale, fino a trovare un valore che azzeri il tasso di crescita. Questo valore viene scelto come estremo sinistro dell'intervallo. Quindi, si testano valori di flusso via via maggiori, a destra del flusso ottimale, nel tentativo di trovare anche da questo lato un valore di flusso che azzeri il tasso di crescita. Tuttavia, l'operazione viene ripetuta fin quando non si trova che il prossimo valore da testare eccederebbe il limite superiore di flusso della reazione di controllo. In circostanze come questa, il limite superiore di flusso viene scelto come estremo destro dell'intervallo.

Una volta messo a punto il codice per eseguire la RA, lo si è usato per classificare tutte le reazioni studiate sulla base dell'esito della RA. Si sono suddivise tutte le reazioni studiate in tre classi, che si sono ritenute sufficienti a riassumere la diversità dei risultati osservati. Le reazioni sono state classificate in reazioni *costanti*, reazioni *competitive* e reazioni *sinergiche*. Si veda la sezione "Risultati e Discussione" per la spiegazione e il commento delle tre classi. La figura 20 illustra il criterio utilizzato per la classificazione.

```

se  $\text{maxObj} - \text{minObj} \leq 0.2 \cdot \text{maxObj}$ 
    Allora classifica la reazione tra le reazioni "costanti"
altrimenti
    se  $\text{alphaObj} \geq 0.95 \cdot \text{maxObj}$ 
        Allora classifica la reazione tra le reazioni "competitive"
    altrimenti
        Allora classifica la reazione tra le reazioni "sinergiche"
    fine
fine

```

**LEGENDA:**

**maxObj**  
Massimo valore della funzione obiettivo tra quelli nel dataset della RA

**minObj**  
Minimo valore della funzione obiettivo tra quelli nel dataset della RA

**alphaObj**  
Valore della funzione obiettivo predetto imponendo il più piccolo in valore assoluto tra i valori della reazione di controllo compresi nel dataset della RA

**Figura 20:** Pseudocodice che spiega il criterio usato per la classificazione delle reazioni sulla base dei risultati di RA, con rispettiva legenda.

## 3.2 Come è stato eseguito lo studio dell'efficienza di azotofissazione

Nella seconda parte di questo lavoro di tesi, si è cercato di valutare gli effetti di valori sub-ottimali di efficienza di azotofissazione, compensati da un'aumentata nodulazione, sul tasso di crescita della pianta. Si è assunto come valore ottimale di efficienza di azotofissazione (cioè mmol di N<sub>2</sub> fissati per grammo di nodulo) quello predetto dalla simulazione FBA con impostazioni predefinite. Quindi, si sono scelti 100 valori sub-ottimali di efficienza di azotofissazione, ciascuno rappresentato da una percentuale del valore ottimale: 1% del valore ottimale, 2% del valore ottimale, 3% del valore ottimale, e così via fino ad arrivare al 100% del valore ottimale. Per ogni valore sub-ottimale simulato, il tasso di nodulazione (cioè grammo di nodulo per grammo di pianta) è stato aumentato per compensare la minore efficienza di azotofissazione. Infine, per ognuno dei 100 valori, dopo aver adeguatamente aumentato il tasso di nodulazione, si è simulata la crescita del modello massimizzando il tasso di crescita della pianta. Si è così ottenuto uno studio del tasso di crescita della pianta in funzione dell'efficienza di azotofissazione. (L'intero codice è riportato in Appendice).

La seguente strategia è stata utilizzata per le simulazioni in cui l'efficienza di azotofissazione veniva variata mentre il tasso di nodulazione veniva ottimizzato per compensare la minore efficienza di azotofissazione. Innanzitutto, il modello è stato ottimizzato con le impostazioni predefinite, poi la biomassa del nodulo è stata rimossa dalla reazione complessiva della biomassa ed è stata forzata attraverso una reazione *sink* (cioè una reazione virtuale che drena un reagente ma non ha alcun prodotto, e serve per impedire l'accumulo del reagente stesso) a una velocità pari alla sua sintesi predetta nella soluzione di FBA iniziale.

È stato quindi impiegato il loop seguente. È stata impostata l'efficienza desiderata di azotofissazione e il tasso di produzione di biomassa nodulare è stato aggiornato per compensare la differenza di efficienza rispetto alle impostazioni predefinite (ad esempio, se l'efficienza veniva ridotta del 50%, la biomassa nodulare era raddoppiata). Inoltre, il tasso massimo di assorbimento di ossigeno della zona III e il flusso di azotofissazione sono stati adeguati in base al rapporto aggiornato tra la produzione di biomassa del nodulo e della pianta.

Il tasso di produzione della biomassa vegetale è stato quindi ottimizzato. In base al flusso di azotofissazione predetto nella soluzione del passaggio precedente:

- il tasso di produzione di biomassa è stato aggiornato,
- sono stati aggiornati i limiti inferiori delle reazioni del costo di mantenimento di tutte le zone nodulari,
- il tasso massimo di assorbimento di ossigeno e il flusso di azotofissazione della zona III sono stati aggiornati.

Il tasso di produzione di biomassa vegetale è stato nuovamente ottimizzato. Questo loop è stato ripetuto fino a quando l'efficienza di azotofissazione non rientrava entro lo 0,5% del valore desiderato.

Si è scoperto che questo loop identificava solo massimi locali per il tasso di produzione di biomassa vegetale. Per trovare il massimo globale, è stato impiegato un secondo loop. In ogni loop, il tasso di nodulazione è stato leggermente aumentato rispetto alla soluzione precedente. Successivamente, il loop precedente è stato nuovamente

impiegato per identificare il massimo locale. Questa operazione è stata ripetuta fino a quando il massimo locale seguito ad un leggero aumento del tasso di nodulazione era inferiore al massimo locale precedente; si è assunto che questo significasse che la precedente iterazione rappresentava il massimo globale. Sono stati quindi registrati i risultati per il massimo globale.

Per esaminare l'effetto dell'efficienza di azotofissazione sulla crescita della pianta in assenza di limitazione dell'ossigeno nella zona III, è stato impiegato lo stesso processo con l'eccezione che il tasso massimo di assorbimento di ossigeno nella zona III è stato mantenuto costante a 1000  $\mu\text{mol/h}$  per grammo di peso secco della pianta.

Successivamente si è condotta un'analisi più complessa. Dato che il tasso di nodulazione viene accresciuto via via che l'efficienza di azotofissazione diminuisce, per tutti i valori sub-ottimali testati di efficienza di azotofissazione si ottengono tassi di nodulazione superiori al 2%, che è il valore preimpostato nel modello sulla base di dati sperimentali. Quando si testano valori di efficienza di azotofissazione molto ridotti, si ottengono tassi di nodulazione assai elevati necessari a compensare la ridotta azotofissazione, fino ad ottenere valori inverosimilmente alti, ad esempio del 25%. In un caso del genere, il 25% della biomassa dell'olobionte sarebbe costituita dai soli noduli: un valore assolutamente irrealistico. Per evitare risultati del genere, si è ripetuto l'intero studio, imponendo stavolta un tetto di nodulazione, ovvero imponendo che la biomassa nodulare non potesse superare una percentuale stabilita della biomassa totale, anche a scapito dell'effetto di compensazione. Questa analisi con tetto di nodulazione è stata quindi ripetuta per cinque diversi valori del tetto di nodulazione (3%, 6%, 9%, 12%, 30%). Questa grandezza, quindi, è stata trattata come una seconda variabile indipendente – a fianco dell'altra rappresentata dall'efficienza di azotofissazione – che influisce sul valore della variabile dipendente, cioè il tasso di crescita della pianta. I risultati di questo studio in due variabili sono stati rappresentati mediante un grafico 3D (si veda in “Risultati e Discussione”). Inoltre, per ogni valore testato del tetto di nodulazione, si è tracciato un grafico del tasso di crescita della pianta in funzione dell'efficienza di azotofissazione.

## 4. Risultati e discussione

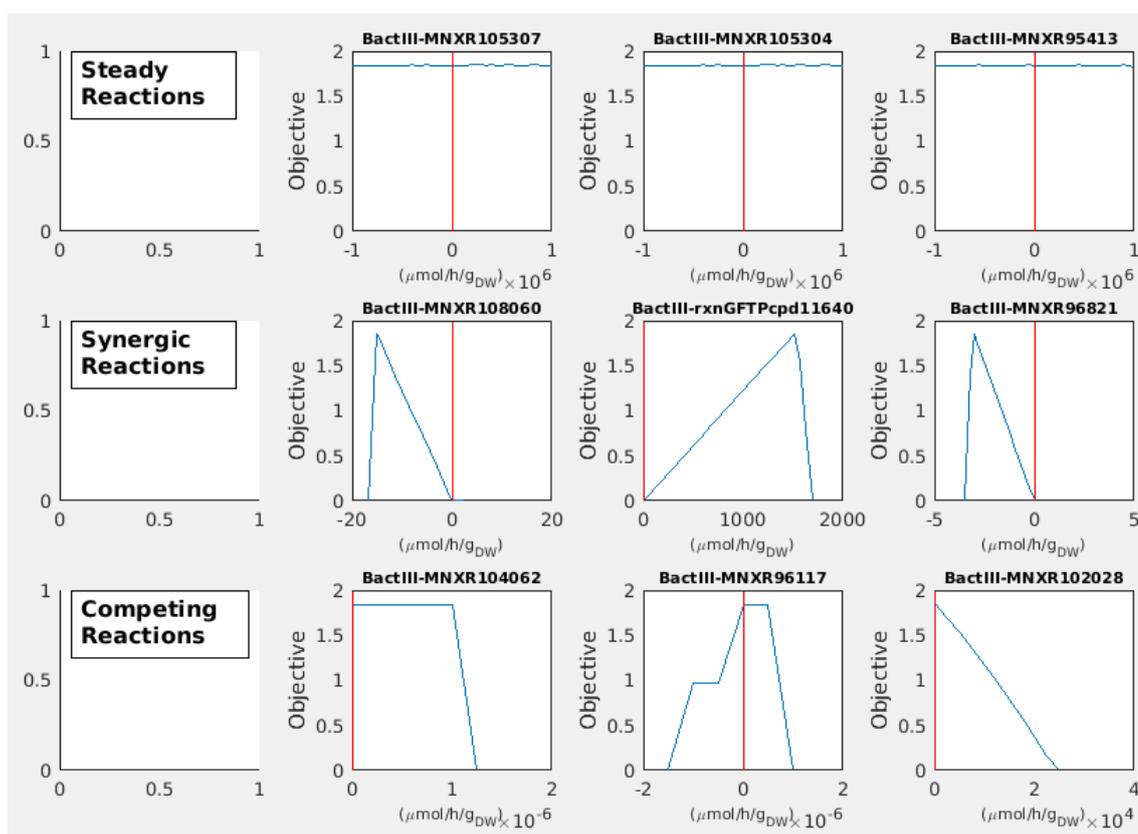
### 4.1 Classificazione delle reazioni batteriche in base alla *Robustness Analysis*

Una volta implementato il codice adatto all'esecuzione della Robustness Analysis, si è svolta tale analisi su tutte le reazioni batteriche di tutte le zone di sviluppo nel nodulo – ad eccezione della zona I (meristema apicale), nella quale sono assenti i batteri. In ciascuna delle zone nodulari II distale, II prossimale, interzona II-III e zona III si è eseguita l'analisi di robustezza su ogni singola reazione del network metabolico batterico. Dato che ad ogni zona nodulare sono associate centinaia di reazioni batteriche, si è prodotta una vasta mole di dati. Per estrapolare informazioni utili da tale dataset, si è scelto di suddividere tutte le reazioni in tre classi che si sono ritenute sufficienti a riassumere la diversità dei risultati osservati. Le reazioni sono state classificate in reazioni *costanti*, reazioni *competitive* e reazioni *sinergiche*.

Le tre classi di reazione sono:

- Reazioni costanti (*steady*)  
Le variazioni di flusso della reazione di controllo non hanno alcun effetto sul tasso di crescita della pianta. Che la reazione sia attiva oppure no, non viene predetto alcun effetto sul tasso di crescita della pianta.
- Reazioni competitive (*competing*)  
Il tasso di crescita della pianta ha un picco quando il flusso attraverso la reazione di controllo è nullo – ovvero quando tale reazione non è attiva. All'aumentare del valore assoluto del flusso di reazione, il tasso di crescita può rimanere costante per un determinato intervallo, o diminuire gradualmente, oppure andare immediatamente a zero; in ogni caso non può aumentare.
- Reazioni sinergiche (*synergic*)

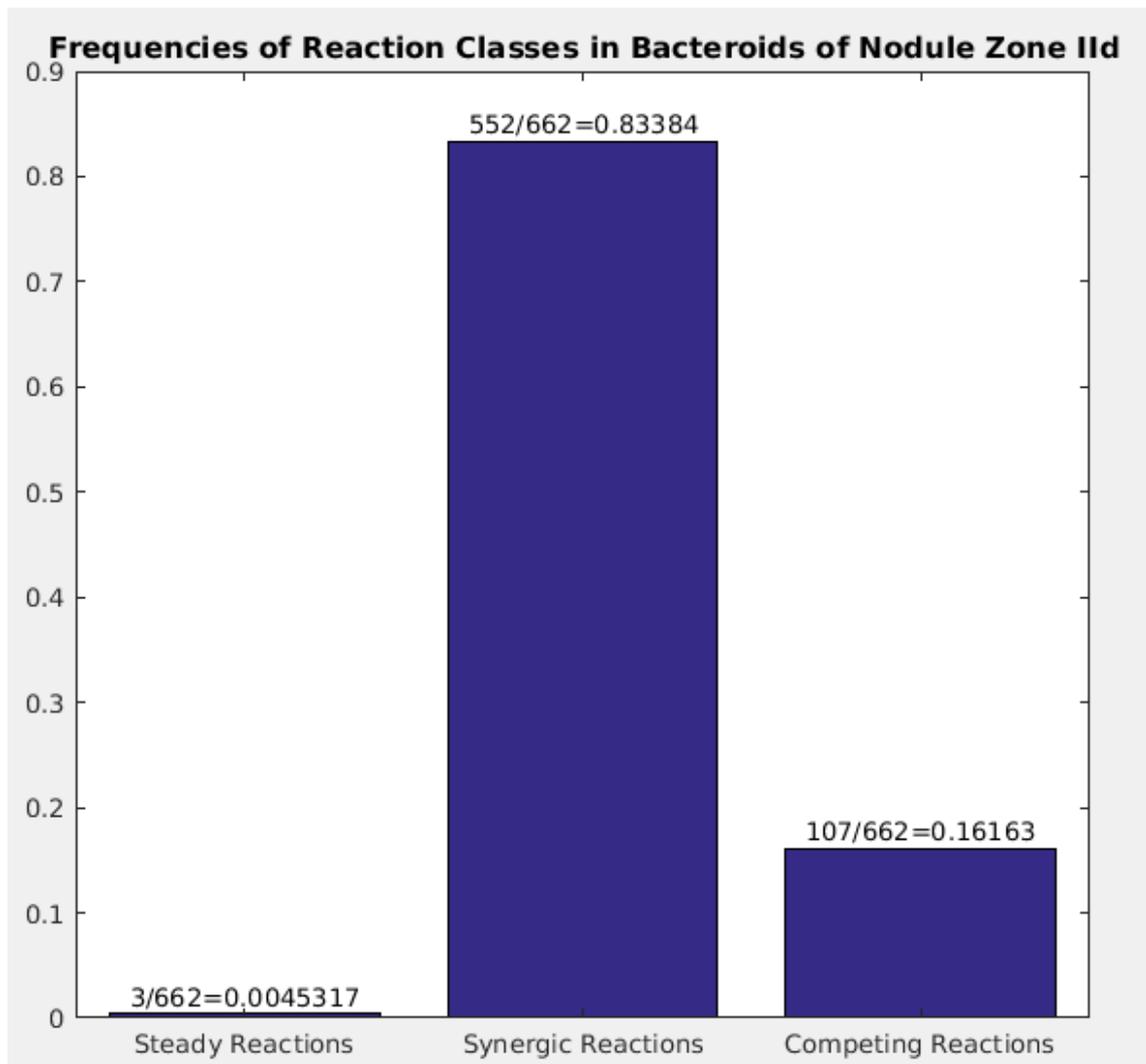
È indispensabile che la reazione di controllo sia attiva affinché la pianta possa crescere: quando il flusso attraverso la reazione di controllo è nullo, anche il tasso di crescita della pianta è nullo. Via via che il flusso della reazione aumenta, cresce linearmente anche il valore del tasso di crescita della pianta, fino ad arrivare al valore ottimale del flusso della reazione, oltre il quale si registra una rapida caduta del tasso di crescita fino al suo azzeramento.



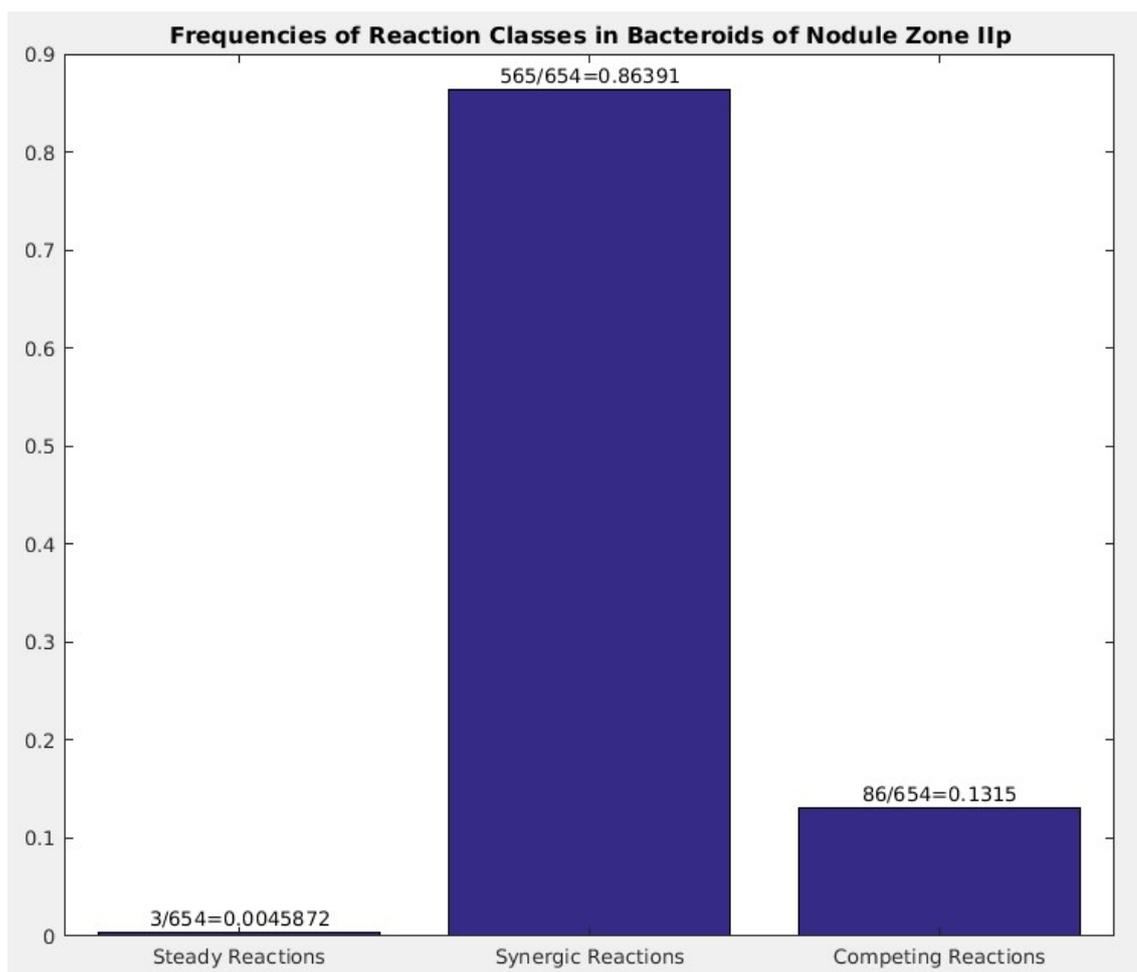
**Figura 21:** Campione casuale di reazioni estratto dal dataset della zona nodulare III per ciascuna delle tre classi di reazioni. In ogni grafico, in ascissa si trova il flusso della reazione di controllo e in ordinata il flusso attraverso la reazione obiettivo (rappresentata dal tasso di crescita della biomassa vegetale). La linea rossa verticale indica il punto per cui il flusso attraverso la reazione di controllo è nullo. L'unità di misura è  $\mu\text{mol/h/gDW}$ . La scritta in grassetto sopra ad ogni grafico è l'identificativo della reazione di controllo. Meritevole di attenzione è anche la differenza che si osserva tra le varie reazioni relativamente alla scala sull'asse X.

La figura 21 riporta, a titolo di esempio, un campione casuale di reazioni estratto dal dataset della zona nodulare III per ciascuna delle tre classi di reazioni.

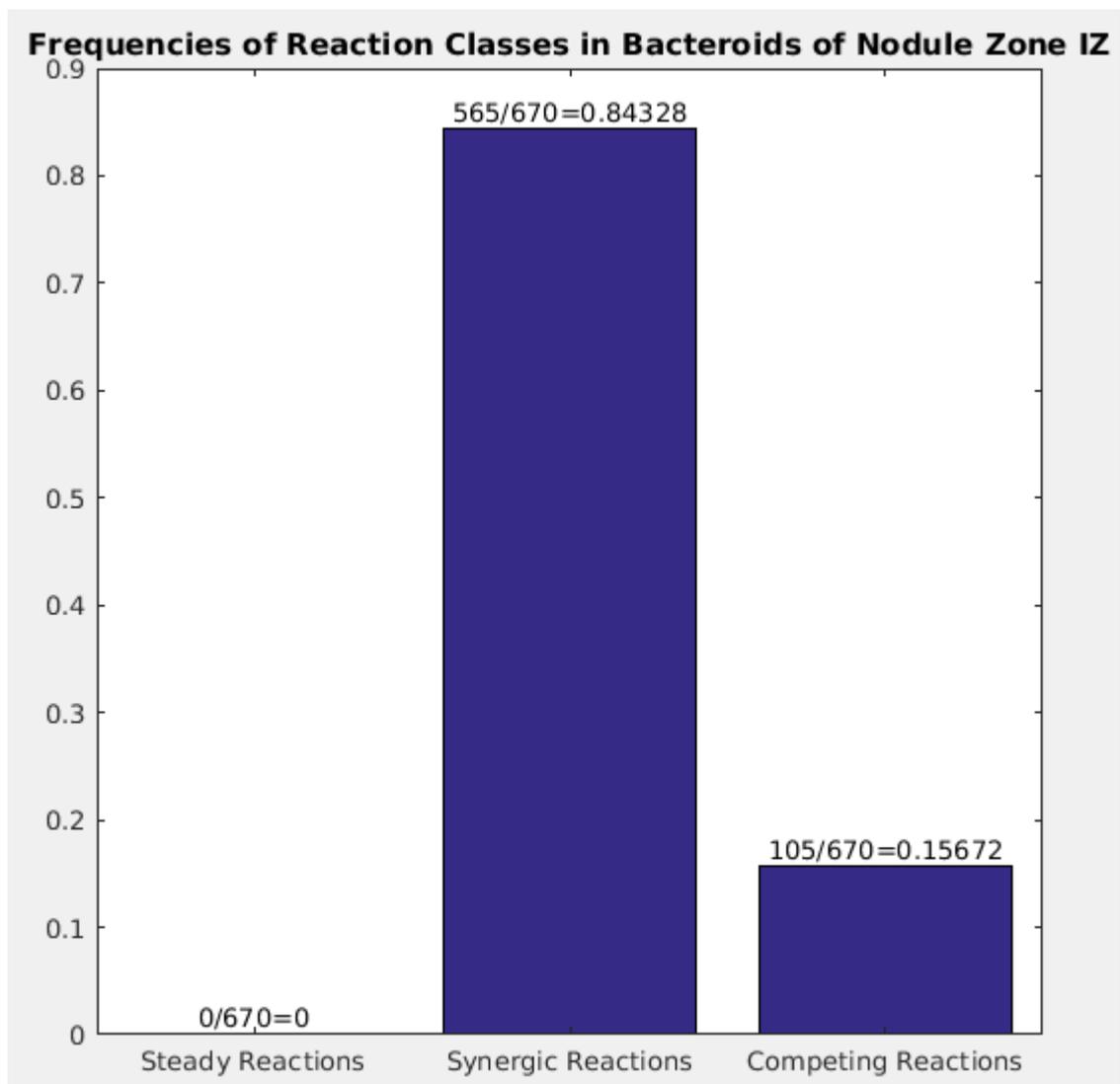
Una volta automatizzata la classificazione delle reazioni, si è calcolata la percentuale di reazioni appartenenti ad ogni classe sul totale delle reazioni comprese in una data zona nodulare. Si sono rappresentati i risultati attraverso un istogramma delle frequenze delle tre classi di reazioni per ciascuna delle zone nodulari analizzate. I quattro istogrammi prodotti sono riportati in Figura 22, Figura 23, Figura 24 e Figura 25.



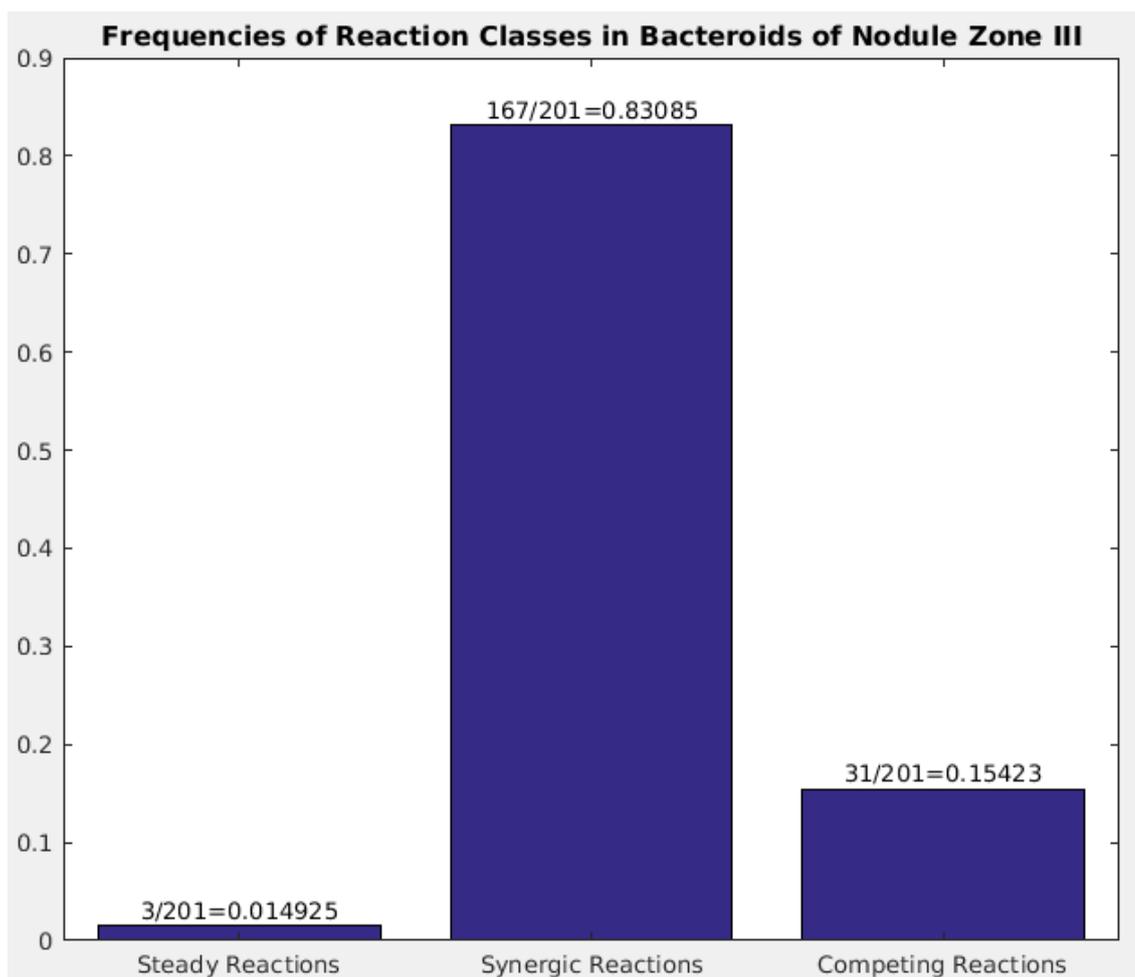
**Figura 22:** Istogramma delle frequenze delle tre classi di reazioni nella zona nodulare II distale



**Figura 23:** Istogramma delle frequenze delle tre classi di reazioni nella zona nodulare II prossimale



**Figura 24:** Istogramma delle frequenze delle tre classi di reazioni nell'interzona II-III



**Figura 25:** Istogramma delle frequenze delle tre classi di reazioni nella zona nodulare III

**Le reazioni competitive sono in accordo con il risultato atteso?** La prima considerazione da fare rispetto ai risultati di classificazione è il confronto tra l’osservato e l’atteso. Si ricordi dall’introduzione (paragrafo “1.5 ViNE – Virtual Nodule Environment”) che:

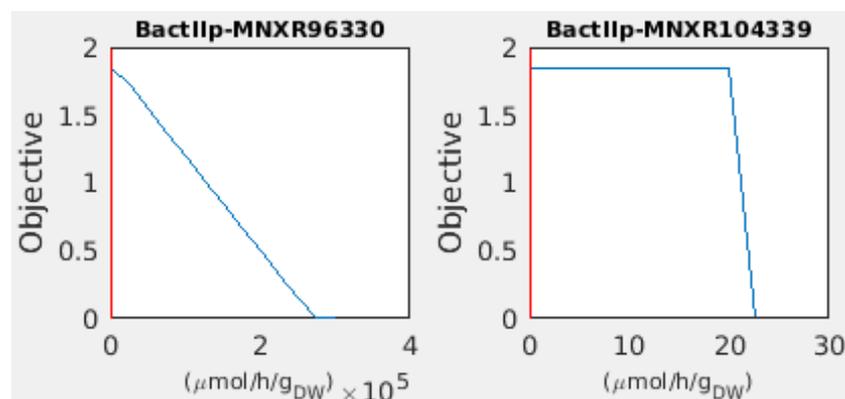
*Lo spazio delle reazioni di ciascuna zona nodulare è stato vincolato in base ai dati di RNA-seq zona-specifici di *M. truncatula* – *S. meliloti* di Roux e collaboratori. Tali dati di sequenziamento dell’RNA sono stati sottoposti ad analisi statistiche, con la finalità di stabilire una distinzione tra geni altamente espressi o comunque indispensabili per la crescita e geni scarsamente espressi o non espressi. I geni classificati nella seconda categoria sono stati eliminati dal modello (soltanto quelli relativi al nodulo o ai batteri, e non quelli relativi*

alla radice o alla parte aerea della pianta), e con essi le rispettive reazioni metaboliche o di trasporto.

Questo significa che ci si aspetta che ogni reazione presente nel modello (cioè che non sia stata eliminata dal modello sulla base dei dati di trascrittomica) porti un flusso non-nullo. Questa condizione è in effetti soddisfatta per tutte le reazioni costanti e per tutte le reazioni sinergiche. Dalla simulazione FBA, tuttavia, risulta che alcune reazioni siano inattive. Tali reazioni non coincidono esattamente con le reazioni competitive. Infatti, l'osservazione di numerosi campioni casuali di reazioni competitive estratti dai dataset della RA ha permesso di distinguere questa classe in due sotto-categorie:

- Reazioni competitive per le quali il tasso di crescita è massimale soltanto quando il flusso attraverso la reazione di controllo è nullo – chiamiamole arbitrariamente reazioni *di tipo alfa*;
- Reazioni competitive per le quali il tasso di crescita è massimale fintanto che il flusso attraverso la reazione di controllo rimane compreso in un intervallo di valori che include lo zero – chiamiamole arbitrariamente reazioni *di tipo beta*.

La figura 26 illustra attraverso due esempi la differenza tra le reazioni competitive di tipo alfa e le reazioni competitive di tipo beta.



**Figura 26:** La reazione di sinistra è competitiva di tipo alfa, mentre quella di destra è competitiva di tipo beta.

Quando si effettua una simulazione FBA massimizzando il tasso di crescita, per le reazioni di tipo alfa viene predetto un flusso nullo, mentre per le reazioni di tipo beta

viene predetto un qualche valore di flusso (non necessariamente uguale a zero) compreso nell'intervallo che massimizza il tasso di crescita. Le reazioni competitive di tipo alfa, pertanto, rappresentano una discordanza tra ciò che ci si aspetta sulla base della conoscenza del sistema biologico e ciò che si osserva nella simulazione. Questo limite dovrà essere imputato ad uno tra tre possibili fattori:

- Un errore nella produzione dei dati di trascrittomici,
- Un errore nell'interpretazione dei dati di trascrittomici o nell'imposizione di tali dati sul modello,
- Un'incompletezza del modello.

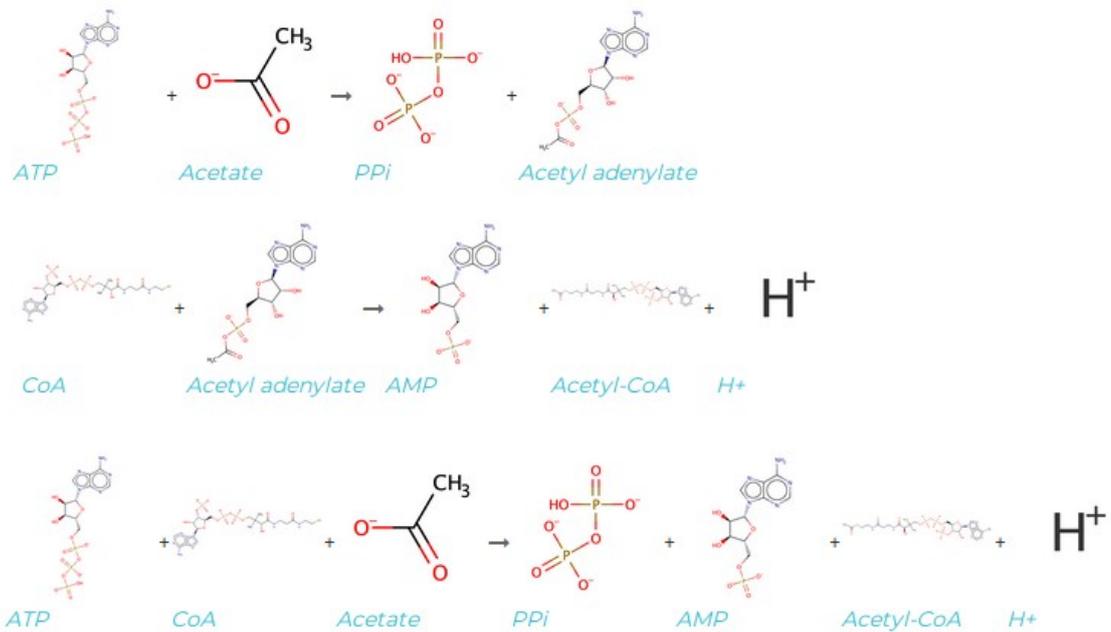
Le reazioni competitive nel loro complesso rappresentano il 13-16% delle reazioni classificate. Il disaccordo tra la previsione del modello e ciò che ci si aspetta su base biologica, quindi, si limita soltanto ad una quota (le reazioni di tipo alfa) di quel 13-16%.

**Le reazioni costanti risultano da un *energyless loop*.** Le reazioni costanti sono soltanto tre nella zona II distale, tre nella zona II prossimale, zero nell'interzona II-III e tre nella zona III. Essendo in numero così limitato, si sono potute considerare una per una, e si è scoperto che sono esattamente le stesse tre reazioni metaboliche sia nella zona II distale, sia nella zona II prossimale, sia nella zona III. Queste tre reazioni, illustrate in Figura 27, sono rappresentate da:

1. MNXR105307 – Primo passaggio della reazione catalizzata dall'enzima 6.2.1.1, ovvero l'Acetil-CoA Sintetasi,
2. MNXR105304 – Secondo e ultimo passaggio della medesima reazione,
3. MNXR95413 – Ancora la reazione catalizzata dall'enzima 6.2.1.1, cioè l'Acetil-CoA Sintetasi, qui riassunta in un'unica trasformazione.

In altre parole, la reazione MNXR105307+MNXR105304 e la reazione MNXR95413 rappresentano la stessa reazione, nel primo caso articolata in due passaggi e nel secondo caso svolta in un solo passaggio complessivo. Questo significa che la stessa reazione è stata erroneamente inserita due volte nel modello, e ciò consente il formarsi di un *energyless loop*, ossia un ciclo in cui la stessa reazione si svolge allo stesso tempo in

avanti e indietro, senza alcun consumo netto di energia. Questo ciclo rende non validi i risultati della RA eseguita su queste tre particolari reazioni.



**Figura 27:** Le tre reazioni classificate come costanti. La prima reazione è il primo dei due passaggi catalizzati dall'enzima Acetil-CoA Sintetasi. La seconda reazione è il secondo passaggio catalizzato dallo stesso enzima. La terza reazione è ancora la reazione catalizzata dalla Acetil-CoA Sintetasi, riassunta in un'unica trasformazione.

**Le reazioni sinergiche rappresentano il risultato atteso.** Le reazioni sinergiche sono reazioni essenziali per la crescita del modello, in accordo con ciò che ci si aspetta in seguito al vincolamento del modello con i dati di trascrittomici. Infatti, è indispensabile che la reazione di controllo sia attiva affinché la pianta possa crescere: quando il flusso attraverso la reazione di controllo è nullo, anche il tasso di crescita della pianta è nullo. Via via che il flusso della reazione aumenta, cresce linearmente anche il valore del tasso di crescita della pianta, fino ad arrivare al valore ottimale del flusso della reazione. Superato il valore ottimale di flusso della reazione di controllo, la reazione drena troppi reagenti che invece dovrebbero essere incanalati verso altre reazioni, e inoltre determina un accumulo di prodotti superflui, con l'effetto di ostacolare la crescita della pianta.

Infine, è probabile che le reazioni competitive di tipo beta siano da assimilare a reazioni sinergiche "ridondanti". In altre parole, si tratterebbe di reazioni che non sono

indispensabili alla crescita, e che pertanto possono avere un flusso, contribuendo così alla crescita del modello, oppure possono restare inattive, venendo compensate da altre reazioni metaboliche. Soltanto quando durante la RA viene forzato un flusso eccessivo attraverso la reazione, allora si compromette la crescita della pianta, esattamente come accade per le reazioni sinergiche.

## **4.2 Risultati osservati nello studio sull'efficienza di azotofissazione**

Per cercare di emulare il fenomeno noto come *autoregolazione della nodulazione* che si osserva nelle leguminose, si sono eseguite simulazioni in cui l'efficienza di azotofissazione veniva variata e il tasso di nodulazione veniva ottimizzato di conseguenza per compensare la riduzione dell'efficienza di azotofissazione. Si sono testati numerosi valori sub-ottimali di efficienza di azotofissazione, compensati da un'aumentata nodulazione, per valutare il loro effetto sul tasso di crescita della pianta.

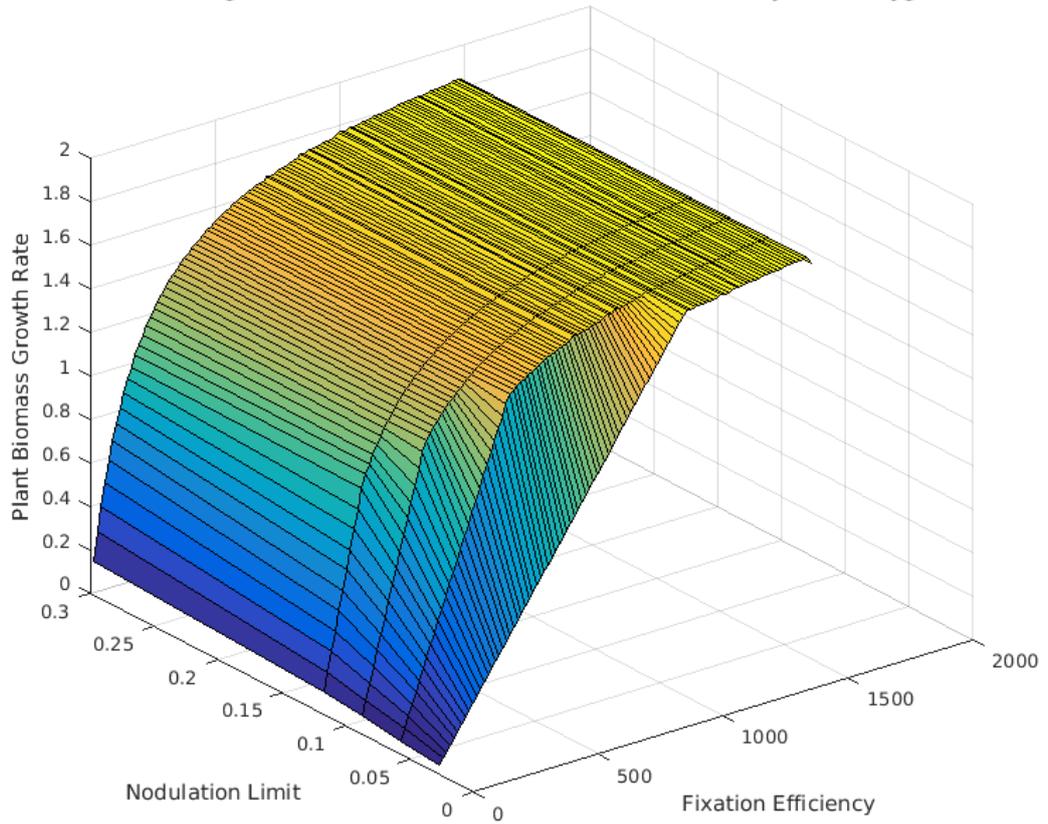
Successivamente, si è aggiunta un'altra variabile nell'analisi. Dato che il tasso di nodulazione viene accresciuto via via che l'efficienza di azotofissazione diminuisce, per tutti i valori sub-ottimali testati di efficienza di azotofissazione si ottengono tassi di nodulazione superiori al 2%, che è il valore preimpostato nel modello sulla base di dati sperimentali. Quando si testano valori di efficienza di azotofissazione molto ridotti, si ottengono tassi di nodulazione assai elevati necessari a compensare la ridotta azotofissazione, fino ad ottenere valori inverosimilmente alti, ad esempio del 25%. In un caso del genere, il 25% della biomassa dell'olobionte sarebbe costituita dai soli noduli: un valore assolutamente irrealistico.

Per evitare risultati del genere, si è ripetuto l'intero studio, imponendo stavolta un tetto di nodulazione, ovvero imponendo che la biomassa nodulare non potesse superare una percentuale stabilita della biomassa totale, anche a scapito dell'effetto di compensazione. Questa analisi con tetto di nodulazione è stata quindi ripetuta per cinque valori diversi del tetto di nodulazione (3%, 6%, 9%, 12%, 30%). Questa

grandezza, quindi, è stata trattata come una seconda variabile indipendente – a fianco dell'altra rappresentata dall'efficienza di azotofissazione – che influisce sul valore della variabile dipendente, cioè il tasso di crescita della pianta.

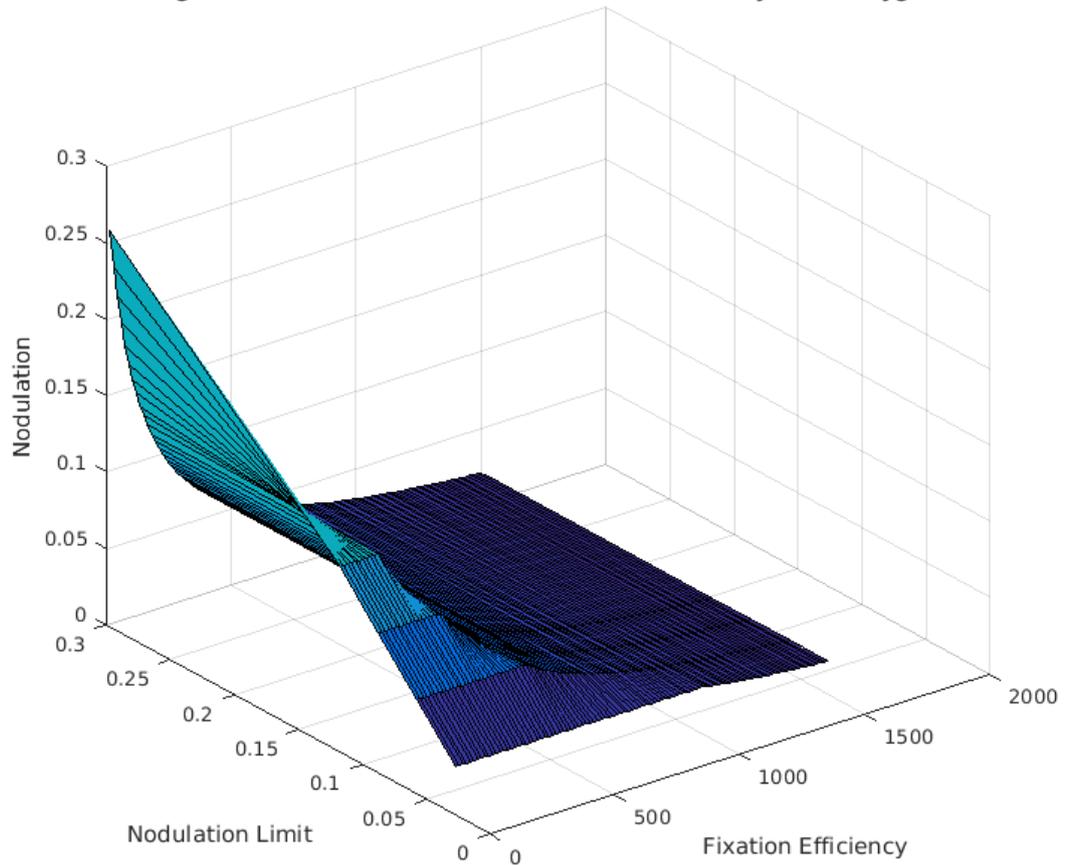
L'intero studio è stato ripetuto due volte, nel primo caso imponendo la limitazione dell'ossigeno per la zona nodulare III, e la seconda volta in condizione di ossigeno non limitante. I risultati di questo studio in due variabili sono stati rappresentati mediante un grafico 3D (Figura 28 in condizione di limitazione dell'ossigeno e Figura 31 in condizione di ossigeno non limitante); si riporta anche un grafico 3D esplicativo che riporta la percentuale di nodulazione in funzione ancora delle solite due variabili indipendenti (Figura 29 in condizione di limitazione dell'ossigeno e Figura 32 in condizione di ossigeno non limitante), con lo scopo di poter visualizzare immediatamente la percentuale di nodulazione per ogni combinazione di valori dell'efficienza sub-ottimale di azotofissazione e del tetto di nodulazione imposto. Inoltre, per ogni valore testato del tetto di nodulazione, si è tracciato un grafico del tasso di crescita della pianta in funzione dell'efficienza di azotofissazione (Figura 30 in condizione di limitazione dell'ossigeno e Figura 33 in condizione di ossigeno non limitante). Infine, anziché incollare l'intero dataset dello studio si preferisce riportare in Tabella 2, Tabella 3, Tabella 4 e Tabella 5 alcune informazioni fondamentali volte a riassumere i risultati numerici dello studio.

**Plant Growth Rate against Nodulation Limit and Fixation Efficiency (with oxygen limitation)**



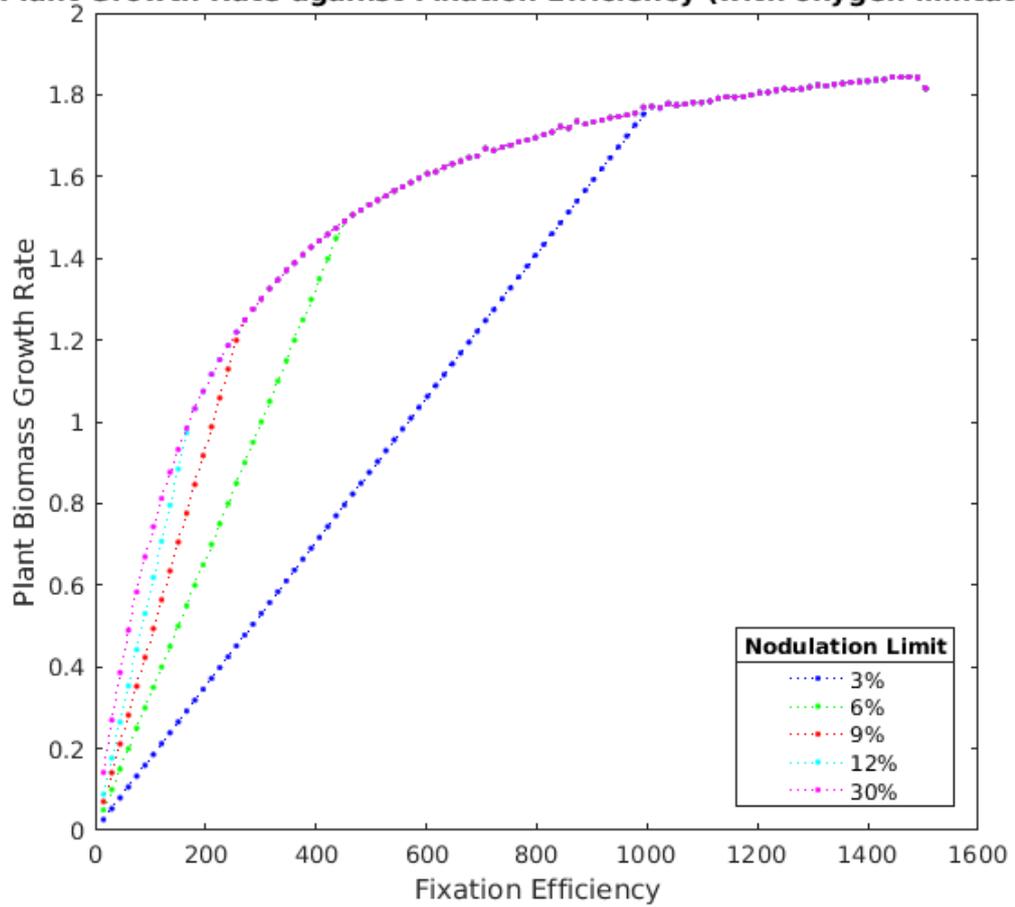
**Figura 28:** Tasso di crescita della biomassa della pianta in funzione dell'efficienza sub-ottimale di azotofissazione e del tetto di nodulazione imposto, in condizione di limitazione dell'ossigeno.

**Nodulation against Nodulation Limit and Fixation Efficiency (with oxygen limitation)**



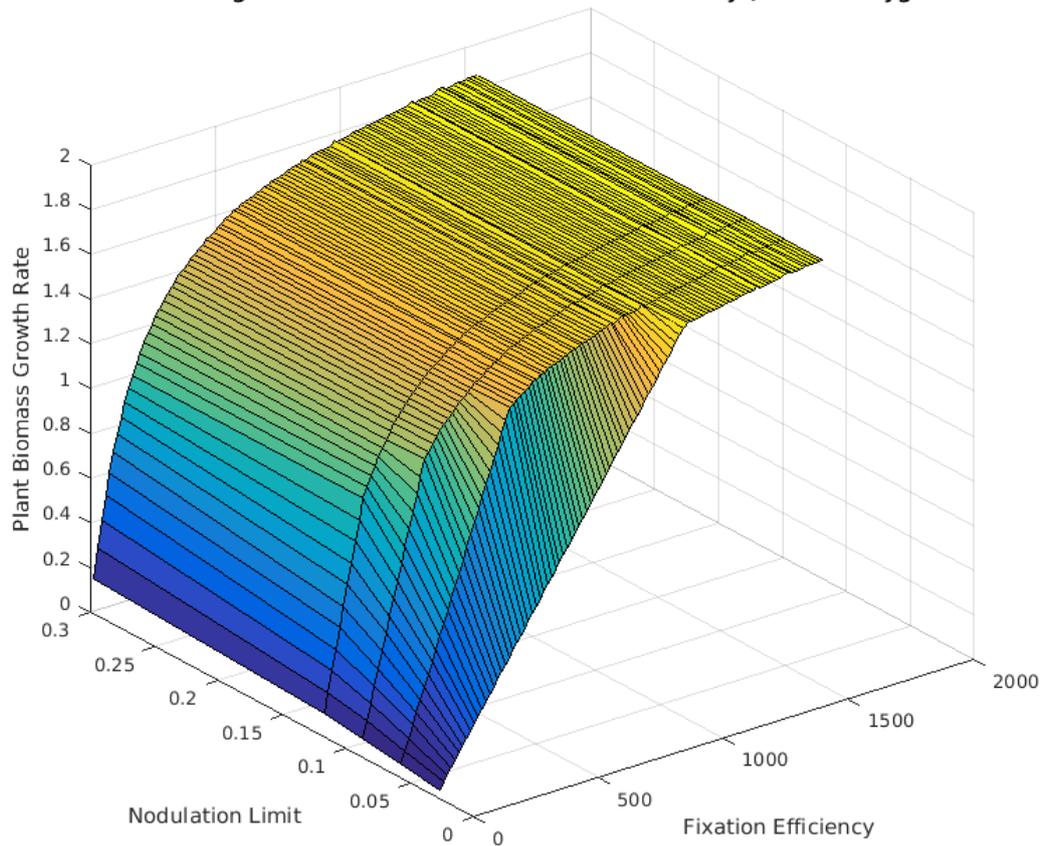
**Figura 29:** Frazione di nodulazione in funzione dell'efficienza sub-ottimale di azotofissazione e del tetto di nodulazione imposto, in condizione di limitazione dell'ossigeno.

**Plant Growth Rate against Fixation Efficiency (with oxygen limitation)**



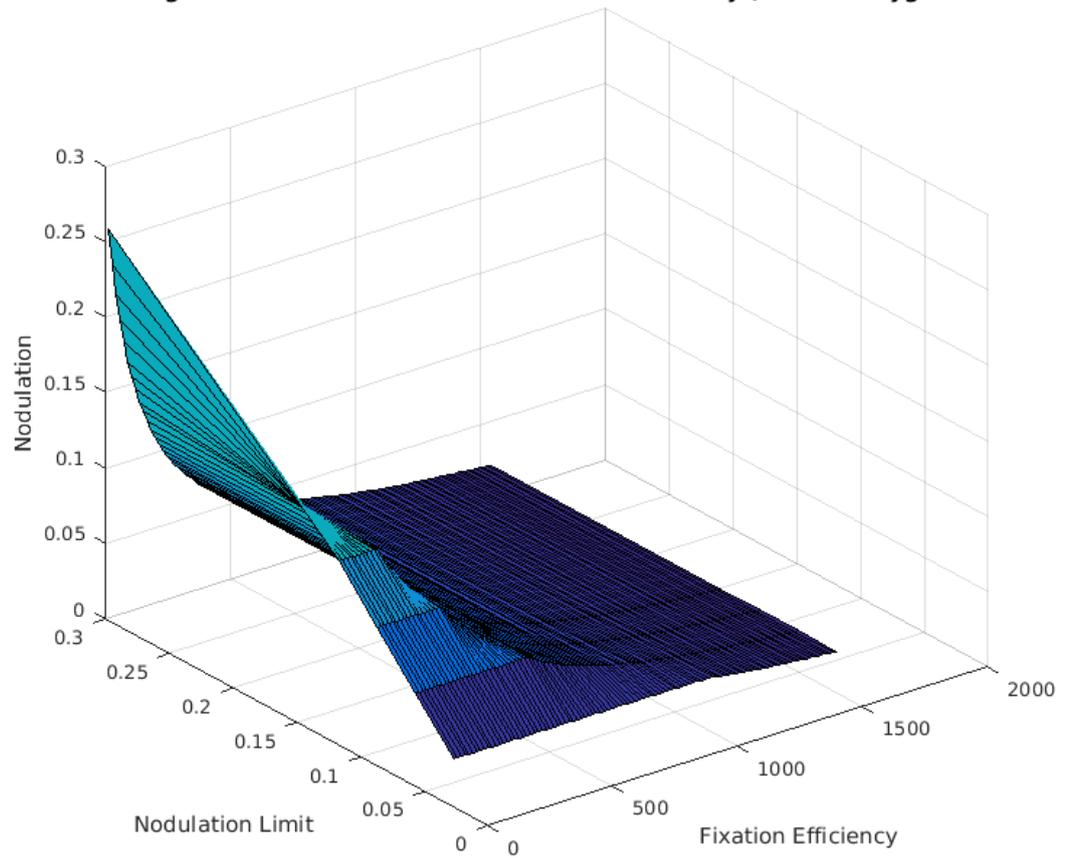
**Figura 30:** Tasso di crescita della biomassa della pianta in funzione dell'efficienza sub-ottimale di azotofissazione, mostrata per diversi valori imposti del tetto di nodulazione, in condizione di limitazione dell'ossigeno.

**Plant Growth Rate against Nodule Limit and Fixation Efficiency (without oxygen limitation)**



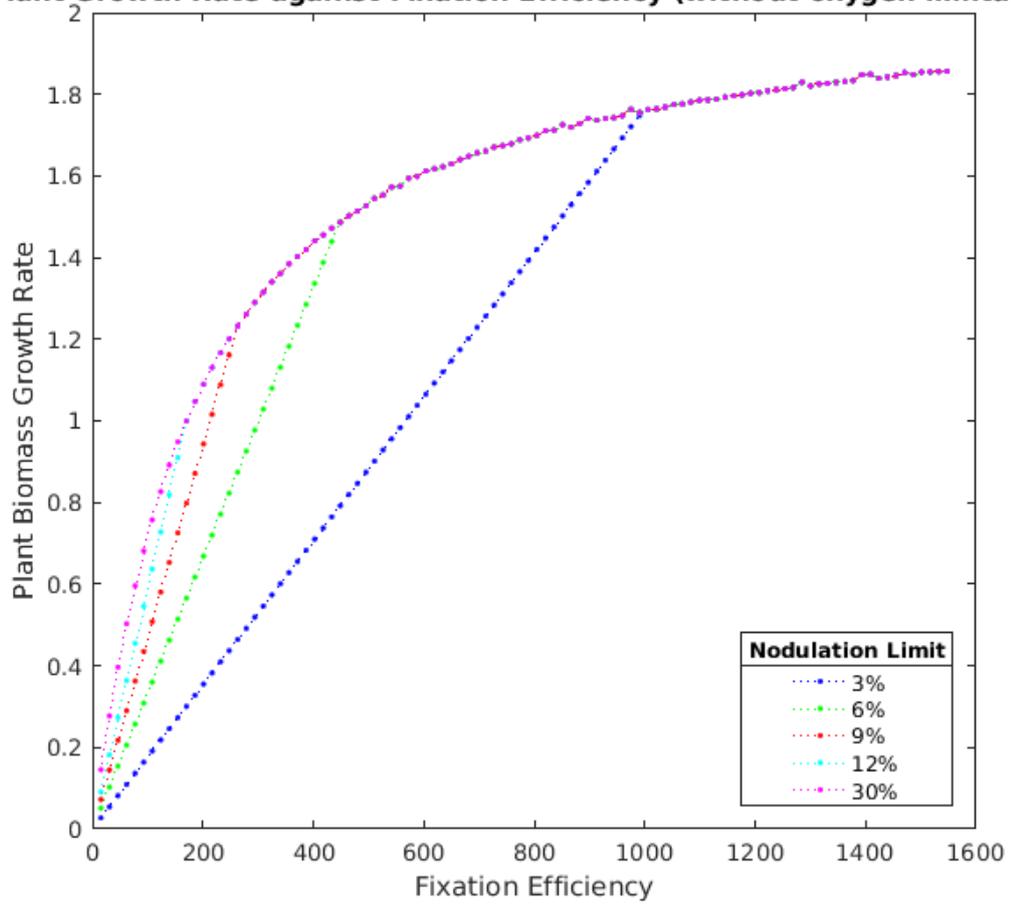
**Figura 31:** Tasso di crescita della biomassa della pianta in funzione dell'efficienza sub-ottimale di azotofissazione e del tetto di nodulazione imposto, in condizione di ossigeno non limitante.

**Nodulation against Nodulation Limit and Fixation Efficiency (without oxygen limitation)**



**Figura 32:** Frazione di nodulazione in funzione dell'efficienza sub-ottimale di azotofissazione e del tetto di nodulazione imposto, in condizione di ossigeno non limitante.

**Plant Growth Rate against Fixation Efficiency (without oxygen limitation)**



**Figura 33:** Tasso di crescita della biomassa della pianta in funzione dell'efficienza sub-ottimale di azotofissazione, mostrata per diversi valori imposti del tetto di nodulazione, in condizione di ossigeno non limitante.

### LIMITED OXYGEN

Fixation Efficiency (Percentage of the value predicted in standard conditions)	Plant Biomass Growth Rate				
	Nodule_MAX: 3_Percent	Nodule_MAX: 6_Percent	Nodule_MAX: 9_Percent	Nodule_MAX: 12_Percent	Nodule_MAX: 30_Percent
2	0,053	0,100	0,141	0,177	0,270
10	0,266	0,500	0,705	0,884	0,932
20	0,531	1,000	1,300	1,300	1,300
30	0,797	1,490	1,490	1,490	1,490
40	1,062	1,607	1,607	1,607	1,607
50	1,327	1,676	1,676	1,676	1,676
60	1,593	1,734	1,734	1,734	1,734
70	1,775	1,775	1,775	1,775	1,775
80	1,805	1,805	1,805	1,805	1,805
90	1,828	1,828	1,828	1,828	1,828
98	1,844	1,844	1,844	1,844	1,844

**Tabella 2:** Per ogni valore del tetto di nodulazione imposto, viene riportato il tasso di crescita della biomassa della pianta in funzione di alcuni valori esemplificativi dell'efficienza di azotofissazione (riportata come percentuale del valore predetto in condizioni standard), in condizione di ossigeno limitante.

### UNLIMITED OXYGEN

Fixation Efficiency (Percentage of the value predicted in standard conditions)	Plant Biomass Growth Rate				
	Nodule_MAX: 3_Percent	Nodule_MAX: 6_Percent	Nodule_MAX: 9_Percent	Nodule_MAX: 12_Percent	Nodule_MAX: 30_Percent
2	0,055	0,103	0,145	0,182	0,277
10	0,273	0,514	0,726	0,910	0,949
20	0,546	1,028	1,315	1,315	1,315
30	0,819	1,502	1,502	1,502	1,502
40	1,092	1,618	1,618	1,618	1,618
50	1,366	1,688	1,688	1,688	1,688
60	1,639	1,740	1,740	1,740	1,740
70	1,781	1,781	1,781	1,781	1,781
80	1,811	1,811	1,811	1,811	1,811
90	1,848	1,848	1,848	1,848	1,848
98	1,855	1,855	1,855	1,855	1,855

**Tabella 3:** Per ogni valore del tetto di nodulazione imposto, viene riportato il tasso di crescita della biomassa della pianta in funzione di alcuni valori esemplificativi dell'efficienza di azotofissazione (riportata come percentuale del valore predetto in condizioni standard), in condizione di ossigeno non limitante.

IN CONDIZIONE DI OSSIGENO LIMITANTE	Nodule_MAX: 3_Percent	Nodule_MAX: 6_Percent	Nodule_MAX: 9_Percent	Nodule_MAX: 12_Percent	Nodule_MAX: 30_Percent
Con quale valore di efficienza sub-ottimale di azotofissazione si raggiunge il tasso di crescita semi-massimale?	35%	19%	13%	11%	10%
Che percentuale del tasso di crescita massimale si ottiene quando l'efficienza di azotofissazione testata è il 50% dell'ottimo?	72%	91%	91%	91%	91%

**Tabella 4:** Per ciascun valore del tetto di nodulazione imposto, sono riportati due parametri che riassumono la diversità di rendimento che si osserva variando il tetto di nodulazione.

IN CONDIZIONE DI OSSIGENO NON LIMITANTE	Nodule_MAX: 3_Percent	Nodule_MAX: 6_Percent	Nodule_MAX: 9_Percent	Nodule_MAX: 12_Percent	Nodule_MAX: 30_Percent
Con quale valore di efficienza sub-ottimale di azotofissazione si raggiunge il tasso di crescita semi-massimale?	34%	18%	13%	10%	10%
Che percentuale del tasso di crescita massimale si ottiene quando l'efficienza di azotofissazione testata è il 50% dell'ottimo?	74%	91%	91%	91%	91%

**Tabella 5:** Per ciascun valore del tetto di nodulazione imposto, sono riportati due parametri che riassumono la diversità di rendimento che si osserva variando il tetto di nodulazione.

Per discutere questi risultati, iniziamo con l'osservare che questi sono molto simili – sia dal punto di vista qualitativo che quantitativo – nella condizione di ossigeno limitante per la zona nodulare III e nella condizione di ossigeno non limitante per la zona nodulare III. La differenza principale, per quanto piccola, consiste nel fatto che in generale il modello con ossigeno illimitato ha mostrato un tasso di crescita leggermente superiore, indipendentemente da quale fosse il valore testato di efficienza di azotofissazione. Con questa premessa, vengono discussi soltanto i risultati relativi al modello con ossigeno limitato.

Le simulazioni hanno confermato le aspettative teoriche mostrando che il tasso di crescita della pianta e l'efficienza di azotofissazione sono positivamente correlate: a valori di efficienza di azotofissazione che si avvicinano a quello massimale corrispondono tassi di crescita più elevati. Il risultato inatteso (ma non in contrasto con la teoria) è stato il tipo di rapporto numerico tra le due grandezze. Infatti, le simulazioni suggeriscono un profilo di rendimenti decrescenti associati all'aumento dell'efficienza di azotofissazione: il dimezzamento dell'efficienza ottimale di azotofissazione ha comportato una riduzione soltanto del 9% della crescita delle piante (ad eccezione del modello in cui il tetto di nodulazione era imposto al 3%, in cui comunque si è osservata una riduzione del tasso di crescita piuttosto piccola, cioè del 18%). Nel modello in cui il tetto di nodulazione era imposto al 30% (ovvero il modello era praticamente privo di un tetto di nodulazione), il tasso di crescita semi-massimale è stato raggiunto con un'efficienza di azotofissazione di appena il 10% del valore ottimale, sebbene ciò richiedesse che i noduli rappresentassero ben il 13% della biomassa totale. Nei modelli in cui, invece, il tetto di nodulazione era imposto a valori del 3%, 6%, 9% e 12%, i benefici di una ridotta efficienza di azotofissazione sono minori, ma comunque si osserva il profilo di rendimenti decrescenti. In questi casi, il tasso di crescita semi-massimale della pianta viene raggiunto rispettivamente al 35%, 19%, 13% e 11% dell'efficienza di azotofissazione ottimale. Complessivamente, queste simulazioni forniscono un'evidenza relativamente al fatto che anche una debole simbiosi potrebbe fornire un notevole beneficio alla pianta.

## 5. Conclusioni

Per entrambi gli studi effettuati in questo lavoro di tesi, è opportuno considerare i risultati sia in relazione allo scopo specifico di ciascuno dei due, sia in relazione alla validazione del modello ViNE. Per “validazione del modello” si intende un processo per cui si eseguono simulazioni di vario genere attraverso il modello, al fine di metterne alla prova l’aderenza con i dati sperimentali e con la conoscenza consolidata del sistema biologico modellizzato. Il modello ViNE è già stato sottoposto ad un’ardua validazione iniziale dai suoi autori, ma il fatto che abbia dimostrato un elevato grado di affidabilità non significa che raffiguri alla perfezione il sistema biologico modellizzato. Tornando al caso particolare dei due studi portati avanti in questo lavoro, il confronto con dati sperimentali sarebbe difficile: i dati andrebbero prodotti *de novo* attraverso un disegno sperimentale dedicato, e non sarebbe facile farlo, dato che in uno dei due studi si è indagato l’effetto di variazioni di flusso in singole reazioni sull’intero network, mentre nell’altro si è analizzato l’effetto di una ridotta efficienza di azotofissazione batterica sul tasso di crescita della pianta (e non sarebbe facile riprodurre queste situazioni in laboratorio). Comunque, come discusso approfonditamente nell’apposita sezione, da ogni simulazione eseguita mediante un modello metabolico ci si aspetta un certo tipo di risultato, sulla base della nostra conoscenza del sistema biologico (cioè, in questo caso, della simbiosi tra *M. truncatula* e *S. meliloti*), e pertanto un primo grado di validazione del modello consiste semplicemente nel valutare la verosimiglianza dei risultati. Nel contesto delle due analisi condotte in questo lavoro, il modello ViNE ha mostrato di descrivere accuratamente l’interazione metabolica tra i due partner della simbiosi, pur dimostrando qualche limite e un certo margine di perfettibilità.

Discussa l’attendibilità dei risultati, si può passare al confronto con lo scopo del lavoro. Per quanto riguarda lo studio di *Robustness Analysis*, questo ha fornito un *insight* su quale sia il comportamento generico delle reazioni batteriche relativamente al tasso di crescita della pianta, permettendo inoltre di tracciare una prima distinzione tra reazioni

essenziali (quelle sinergiche in senso stretto) e ridondanti (quelle che si sono indicate come “competitive di tipo beta”), e producendo così una base grezza di informazioni utile per un ulteriore approfondimento del rapporto tra il metabolismo del batterio e quello della pianta. Inoltre, lo studio ha evidenziato un limite (benché puntiforme) del modello nel descrivere il comportamento di alcune reazioni (quelle che si sono indicate come “competitive di tipo alfa”).

Quanto allo studio dell’effetto della ridotta efficienza di azotofissazione sulla crescita della pianta, questo ha permesso di approfondire il rapporto numerico che lega le due grandezze, e quindi di trarre la conclusione che anche una simbiosi debole potrebbe apportare un notevole beneficio alla pianta. Uno studio come questo potrebbe essere il punto di partenza per un piano di ricerca sperimentale guidata da ipotesi, ed è un buon esempio dell’utilità dei modelli metabolici.

## 6. Bibliografia

1. Daims, H., Lückner, S. & Wagner, M. A New Perspective on Microbes Formerly Known as Nitrite-Oxidizing Bacteria. *Trends Microbiol.* **24**, 699–712 (2016).
2. Biavati, B. & Sorlini, C. *Microbiologia agroambientale*. (2007).
3. Masoni, A. & Ercoli, L. Azoto nel terreno. in *Riduzione dell'inquinamento delle acque dai nitrati provenienti dall'agricoltura* 211–241 (Felici editore, 2010).
4. MarÃ³ti, G. & Kondorosi, Ã. Nitrogen-fixing Rhizobium-legume symbiosis: are polyploidy and host peptide-governed symbiont differentiation general principles of endosymbiosis? *Front. Microbiol.* **5**, 326 (2014).
5. Suzaki, T., Yoro, E. & Kawaguchi, M. Leguminous Plants: Inventors of Root Nodules to Accommodate Symbiotic Bacteria. *Int. Rev. Cell Mol. Biol.* **316**, 111–158 (2015).
6. Ané, J.-M., Zhu, H. & Frugoli, J. Recent Advances in *Medicago truncatula* Genomics. *Int. J. Plant Genomics* **2008**, 256597 (2008).
7. Kang, Y., Li, M., Sinharoy, S. & Verdier, J. A Snapshot of Functional Genetic Studies in *Medicago truncatula*. *Front. Plant Sci.* **7**, 1175 (2016).
8. documentation on *Medicago truncatula*. Available at: [http://dryades.units.it/FVG/index.php?procedure=taxon\\_page&id=2584&num=4809](http://dryades.units.it/FVG/index.php?procedure=taxon_page&id=2584&num=4809).
9. Jiang, Q. & Gresshoff, P. M. Classical and Molecular Genetics of the Model Legume *Lotus japonicus*. *Mol. Plant-Microbe Interact.* **10**, 59–68 (1997).
10. Documentation on *Glycine max*. Available at: <https://knowpulse.usask.ca/node/7>.
11. Rose, R. J. *Medicago truncatula* as a model for understanding plant interactions with other organisms, plant development and stress biology: past, present and future. *Funct. Plant Biol.* **35**, 253 (2008).
12. Documentation on *Medicago sativa*. Available at: [http://dryades.units.it/FVG/index.php?procedure=taxon\\_page&id=2568&num=606](http://dryades.units.it/FVG/index.php?procedure=taxon_page&id=2568&num=606).
13. Covert, M. W. *et al.* Metabolic modeling of microbial strains in silico. *Trends Biochem. Sci.* **26**, 179–186 (2001).

14. Edwards, J. S. & Palsson, B. O. Systems properties of the Haemophilus influenzae Rd metabolic genotype. *J. Biol. Chem.* **274**, 17410–6 (1999).
15. Edwards, J. S. & Palsson, B. O. The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5528–33 (2000).
16. Förster, J., Famili, I., Fu, P., Palsson, B. Ø. & Nielsen, J. Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. *Genome Res.* **13**, 244–53 (2003).
17. Zhang, C. & Hua, Q. Applications of Genome-Scale Metabolic Models in Biotechnology and Systems Medicine. *Front. Physiol.* **6**, 413 (2016).
18. Oberhardt, M. A., Palsson, B. Ø. & Papin, J. A. Applications of genome scale metabolic reconstructions. *Mol. Syst. Biol.* **5**, 320 (2009).
19. Thiele, I. & Palsson, B. Ø. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* **5**, 93–121 (2010).
20. Terzer, M., Maynard, N. D., Covert, M. W. & Stelling, J. Genome scale metabolic networks. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **1**, 285–297 (2009).
21. Thiele, I., Jamshidi, N., Fleming, R. M. T. & Palsson, B. Ø. Genome-Scale Reconstruction of Escherichia coli's Transcriptional and Translational Machinery: A Knowledge Base, Its Mathematical Formulation, and Its Functional Characterization. *PLoS Comput. Biol.* **5**, e1000312 (2009).
22. Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–8 (2010).
23. diCenzo, G. C., Tesi, M., Pfau, T., Mengoni, A. & Fondi, M. A Virtual Nodule Environment (ViNE) for modelling the inter-kingdom metabolic integration during symbiotic nitrogen fixation. *bioRxiv* 765271 (2019). doi:10.1101/765271
24. Roux, B. *et al.* An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant J.* **77**, 817–837 (2014).
25. MATLAB. *Wikipedia*
26. Schellenberger, J. *et al.* Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.* **6**, 1290–307 (2011).

## 7. Appendice

Si è scelto di riportare l'intero codice (chiaramente solo quello definitivo) usato su MATLAB per:

1. Eseguire la *Robustness Analysis* su ognuna delle reazioni dei batteroidi della zona nodulare III e classificare ogni reazione sulla base del risultato della *Robustness Analysis* (i codici usati per la zona II prossimale, per la zona II distale e per l'interzona II-III sono del tutto simili);
2. Testare tutti i 100 valori di efficienza di azotofissazione compresi tra l'1% e il 100% del valore ottimale (predetto in condizioni standard), compensati da una modulazione adeguatamente accresciuta, e per ciascuno di essi ottimizzare il modello massimizzando il tasso di crescita della pianta (è riportato soltanto il codice usato per simulare la condizione di ossigeno limitante).

### 7.1 Codice per la *Robustness Analysis*

```
%% Start Cobra and set the 'ibm_cplex' Cobra solver
initCobraToolbox;
changeCobraSolver('ibm_cplex', 'all');

%% Bacteroid III reactions Classification based on Robustness Analysis

clearvars
modelFileName = load ('finalNodulatedPlant.mat');
model = modelFileName.finalNodulatedPlant; % load the model

BacIIIRxnsPos = strfind(model.rxns, 'BacteroidIII'); % find reactions whose name contains the string
'BacteroidIII'
BacIIIRxnsPos(cellfun('isempty',BacIIIRxnsPos))={0}; % replace every empty cell with a zero
BacIIIRxnsPos = cell2mat(BacIIIRxnsPos); % transform cell array into double array
BacIIIRxns = model.rxns(BacIIIRxnsPos==1); % list Bacteroid III reactions
nRxns = length(BacIIIRxns); % number of BacteroidIII reactions
sol = optimizeCbModel(model); % optimize model maximizing growth
nPoints = 41; % number of points for robustness analysis

% make a vector for every class
SteadyRxns = cell(nRxns,1);
SynergicRxns = cell(nRxns,1);
CompetingRxns = cell(nRxns,1);

for i=1:nRxns
    RxnStr = BacIIIRxns{i}; % reaction specified by the i-th position of BacIIIRxns
    disp(RxnStr);
    optFlux = sol.x(findRxnIDs(model, RxnStr)); % store optimal flux
    disp(strcat('optimal flux =',num2str(optFlux)))
end
```

```

% store the original bounds
minFlux = model.lb(findRxnIDs(model,RxnStr));
maxFlux = model.ub(findRxnIDs(model,RxnStr));
disp(strcat('lower bound =',num2str(minFlux),'...','upper bound =',num2str(maxFlux)))

% walking away from the optimal flux, find those flux values which
% nullify growth

% start with finding the zeroing-value standing at the right of the
% optimal flux
increment = 10^-5;
z=1;
while z==1
    forcedFlux = optFlux + increment;
    modelTest = changeRxnBounds(model,RxnStr,forcedFlux,'b');
    solTest = optimizeCbModel(modelTest);
    if forcedFlux >= maxFlux
        z=0;
        rightLimit=maxFlux;
    else
        if solTest.f>0
            z=1;
            increment=increment*10;
        else
            z=0;
            rightLimit=forcedFlux;
        end
    end
end
end

% then find the zeroing-value standing at the left of the optimal flux
modelTest = model;
solTest = optimizeCbModel(model);
increment = 10^-5;
z=1;
while z==1
    forcedFlux = optFlux - increment;
    modelTest = changeRxnBounds(model,RxnStr,forcedFlux,'b');
    solTest = optimizeCbModel(modelTest);
    if forcedFlux <= minFlux
        z=0;
        leftLimit=minFlux;
    else
        if solTest.f>0
            z=1;
            increment=increment*10;
        else
            z=0;
            leftLimit=forcedFlux;
        end
    end
end
end
clearvars modelTest solTest
disp(strcat('left limit =',num2str(leftLimit),'...','right limit =',num2str(rightLimit)))

% perform Robustness Analysis between leftLimit and RightLimit
RangeWidth = rightLimit - leftLimit;
Step = RangeWidth / (nPoints-1);
RobuArray = zeros(nPoints,2);
for j=1:nPoints
    testedFlux = leftLimit + (j-1)*Step;
    model = changeRxnBounds(model,RxnStr,testedFlux,'b');
    solNew = optimizeCbModel(model);
    RobuArray(j,1) = testedFlux;
    RobuArray(j,2) = solNew.f;
    clearvars testedFlux solNew
end
end
ctrlFlux = RobuArray(:,1); % vector of control flux values
objFlux = RobuArray(:,2); % vector of objective flux values

% find first relevant point of the Robustness Analysis array

```

```

firstPos = find(objFlux,1,'first');
if firstPos>1
    firstPos=firstPos-1;
end

%find last relevant point of the Robustness Analysis array
lastPos = find(objFlux,1,'last');
if lastPos<length(objFlux)
    lastPos = (lastPos+1);
end

% resize RobuArray
RobuArray((lastPos+1):(length(RobuArray)),:)=[]; % delete last useless rows from RobuArray
RobuArray(1:(firstPos-1),:)=[]; % delete first useless rows from RobuArray
ctrlFlux = RobuArray(:,1);
objFlux = RobuArray(:,2);

% add the optimal flux point to the array (if it isn't already there)
if isempty(find(ctrlFlux==optFlux))
    model = changeRxnBounds(model,RxnStr,optFlux,'b');
    solNew = optimizeCbModel(model);
    RobuArray(length(RobuArray)+1,1) = optFlux;
    RobuArray(length(RobuArray),2) = solNew.f;
    clearvars testedFlux solNew
end

% add the zero-point to the array, but only if it is within the bounds
if minFlux<0 && maxFlux>0 && isempty(find(ctrlFlux==0))
    model = changeRxnBounds(model,RxnStr,0,'b');
    solNew = optimizeCbModel(model);
    RobuArray(length(RobuArray)+1,1) = 0;
    RobuArray(length(RobuArray),2) = solNew.f;
    clearvars testedFlux solNew
end

% sort RobuArray's rows based on the ascending order of the elements of
% the ctrlFlux vector
ctrlFlux = RobuArray(:,1);
[ctrlFluxSorted, Indexes]=sort(ctrlFlux);
RobuArray=RobuArray(Indexes,:);
clearvars ctrlFluxSorted Indexes
disp(RobuArray)
ctrlFlux = RobuArray(:,1);
objFlux = RobuArray(:,2);

% find the objective value corresponding to the smallest ctrlFlux value
alphaPos = find(abs(ctrlFlux) == min(abs(ctrlFlux)));
alphaCtrl = ctrlFlux(alphaPos);
alphaObj = objFlux(alphaPos);

% find maximum and minimum of the objFlux vector
maxObj = max(objFlux);
minObj = min(objFlux);

% get the index of the first empty cell, for each class vector
SteadyRxnsIPos = find(cellfun('isempty',SteadyRxns),1);
SynergicRxnsIPos = find(cellfun('isempty',SynergicRxns),1);
CompetingRxnsIPos = find(cellfun('isempty',CompetingRxns),1);

% LET'S CLASSIFY!
if maxObj-minObj <= 0.2*maxObj
    SteadyRxns{SteadyRxnsIPos}=RxnStr;
    disp('STEADY')
else
    if alphaObj >= 0.95*maxObj
        CompetingRxns{CompetingRxnsIPos}=RxnStr;
        disp('COMPETING')
    else
        SynergicRxns{SynergicRxnsIPos}=RxnStr;
        disp('SYNERGIC')
    end
end
end

```

```

% restore original bounds and clear variables
model.lb(findRxnIDs(model,RxnStr)) = minFlux;
model.ub(findRxnIDs(model,RxnStr)) = maxFlux;
clearvars -except model* Bac* sol nPoints *Rxns
end
clearvars -except model* *Rxns nPoints

% concatenate the class vectors and delete empty rows
ListOfRxns = [SteadyRxns SynergicRxns CompetingRxns];
SteadyRxnsIPos = find(cellfun('isempty',SteadyRxns),1);
SynergicRxnsIPos = find(cellfun('isempty',SynergicRxns),1);
CompetingRxnsIPos = find(cellfun('isempty',CompetingRxns),1);
firstEmptyRow = max([SteadyRxnsIPos SynergicRxnsIPos CompetingRxnsIPos]);
ListOfRxns(firstEmptyRow:nRxns,:) = [];

% delete empty rows in each class vector, too
SteadyRxns(SteadyRxnsIPos:nRxns)=[];
SynergicRxns(SynergicRxnsIPos:nRxns)=[];
CompetingRxns(CompetingRxnsIPos:nRxns)=[];

% count reactions for every class vector
nSteadyRxns = length(SteadyRxns);
nSynergicRxns = length(SynergicRxns);
nCompetingRxns = length(CompetingRxns);

% visualize the list of sorted reactions
headers = {'Steady Reactions', 'Synergic Reactions', 'Competing Reactions'};
ListOfRxns = vertcat(headers, ListOfRxns);
disp(ListOfRxns);
ListOfRxns(1,:)=[];

% visualize the count of reactions for each class vector
CountOfRxns = [nSteadyRxns nSynergicRxns nCompetingRxns];
CountOfRxns = num2cell(CountOfRxns);
headers_bis = {'N2 Steady Reactions', 'N2 Synergic Reactions', 'N2 Competing Reactions'};
CountOfRxns = vertcat(headers_bis, CountOfRxns);
disp(CountOfRxns);

% convert absolute frequencies to relative frequencies
CountOfRxns = [nSteadyRxns/nRxns nSynergicRxns/nRxns nCompetingRxns/nRxns];

% make a bar plot, and, for each class, also print the ratio between number
% of reactions for the class and total number of reactions
barLabels = {strcat(num2str(nSteadyRxns), '/', num2str(nRxns), '=', num2str(nSteadyRxns/nRxns)), ...
    strcat(num2str(nSynergicRxns), '/', num2str(nRxns), '=', num2str(nSynergicRxns/nRxns)), ...
    strcat(num2str(nCompetingRxns), '/', num2str(nRxns), '=', num2str(nCompetingRxns/nRxns))};
figure
bar(CountOfRxns)
title('Frequencies of Reaction Classes in Bacteroids of Nodule Zone III')
axis = gca;
axis.XTickLabel = headers;
xPoints = axis.XTick;
text(xPoints, CountOfRxns, barLabels, 'HorizontalAlignment','center', 'VerticalAlignment','bottom')
clearvars -except model* *Rxns nPoints

```

## 7.2 Codice per lo studio sull'efficienza di azotofissazione

```

%% Start Cobra and set the 'ibm_cplex' Cobra solver
initCobraToolbox;
changeCobraSolver('ibm_cplex', 'all');

%% Test effect of changing N2-fixation efficiency - limited O2, varying nodule max
% PLOT 'PLANT_BIOMASS' AGAINST 'NODULE_LIMITATION' AND 'FIXATION_EFFICIENCY'

```

```

% load the model
modelFileName = load ('finalNodulatedPlant.mat');
model = modelFileName.finalNodulatedPlant;

% Add ammonium sinks for nodule zone III and for root
model = addReaction(model, 'NoduleIII_SINK_MNXM15', {'NoduleIII_MNXM15[C]'}, [-1], 0, 0, 1000000, 0);
model = addReaction(model, 'Root_SINK_MNXM15', {'Root_MNXM15[C]'}, [-1], 0, 0, 1000000, 0);

% Check standard flux through nodule biomass reaction
sol = optimizeCbModel(model);
noduleFlux = sol.x(findRxnIDs(model, 'NoduleBiomass'));

% Remove nodule biomass from overall biomass by zeroing the respective stoichiometric coefficient
model.S(findMetIDs(model, 'BiomassNodule[c]'),findRxnIDs(model, 'Biomass')) = 0;

% Add nodule biomass sink
model = addExchangeRxn(model, {'BiomassNodule[c]'}, 0, 1000000);

% Force nodule biomass through the sink
model.lb(findRxnIDs(model, 'EX_BiomassNodule[c]')) = noduleFlux;

% Determine rate of N2-fixation per nodule mass unit
sol = optimizeCbModel(model);
n2fix = 0.01 * sol.x(findRxnIDs(model, 'BacteroidIII_rxnConvFixed')) / ...
(sol.x(findRxnIDs(model, 'NoduleBiomass')) / (sol.x(findRxnIDs(model, 'PlantBiomass')) +
sol.x(findRxnIDs(model, 'NoduleBiomass'))));

% Find position of nitrogen fixation reaction
pos = findRxnIDs(model, 'BacteroidIII_rxnConvFixed');

% Get the original solution of the model
solOrig = optimizeCbModel(model);

% Determine the standard rate of N2-fixation
n2fixMax = solOrig.x(pos);

% Choose nodule limitation values
nodMaxValues=[0.03 0.06 0.09 0.12 0.30];
nodMaxStrings=cell(1,length(nodMaxValues));

% Prepare output variables
output_fixationEfficiency_limO2 = cell(100, 5);
PlantBiomass_limO2 = cell(length(nodMaxValues),100);
Nodulation_limO2 = cell(length(nodMaxValues),100);
figure

% Vary the rate of N2-fixation
for i=1:length(nodMaxValues)
    nodMax=nodMaxValues(i);
    nodMaxStrings{i}=num2str(100*nodMaxValues(i));

    for n = 1:100

        % Create a copy of the model that will be used to carry changes
        modelTemp = model;

        % Set efficiency
        n2fixNew = n2fix * n / 100;

        % Set nodule amount
        nodule = 0.01 * n2fixMax / n2fixNew;

        % Set initial nodule amount to be no more than 10% max
        if nodule > 0.1
            nodule = 0.1;
        end

        % Get original solution
        sol = optimizeCbModel(modelTemp);

        % Define the variable 'x'
        x = 1;
    end
end

```

```

% Get the initial local maximum
while x == 1

    % Record old solution
    solOld = sol;

    % Adjust oxygen availability and nitrogen fixation
    modelTemp.ub(findRxnIDs(modelTemp, 'TNIII_OXYGEN-MOLECULE')) = model.ub(findRxnIDs(model,
'TNIII_OXYGEN-MOLECULE')) * (nodule / 0.02);
    modelTemp.ub(pos) = n2fixNew * (nodule / 0.01);

    % Get new growth
    sol = optimizeCbModel(modelTemp);

    % if sol.x = [], then set to zero every flux
    if isempty(sol.x)
        sol.x = zeros(length(model.rxns),1);
    end

    % Adjust minimum flux through nodule biomass sink
    modelTemp.lb(findRxnIDs(modelTemp, 'EX_BiomassNodule[c]')) = (nodule * sol.x(findRxnIDs(model,
'PlantBiomass'))) / (1 - nodule);

    % Adjust minimum flux through NGAM reactions of nodule compartments
    modelTemp.lb(findRxnIDs(modelTemp, 'BacteroidIId_rxnNGAM')) = model.lb(findRxnIDs(model,
'BacteroidIId_rxnNGAM')) * (nodule / 0.02);
    modelTemp.lb(findRxnIDs(modelTemp, 'BacteroidIIp_rxnNGAM')) = model.lb(findRxnIDs(model,
'BacteroidIIp_rxnNGAM')) * (nodule / 0.02);
    modelTemp.lb(findRxnIDs(modelTemp, 'BacteroidIZ_rxnNGAM')) = model.lb(findRxnIDs(model,
'BacteroidIZ_rxnNGAM')) * (nodule / 0.02);
    modelTemp.lb(findRxnIDs(modelTemp, 'BacteroidIII_rxnNGAM')) = model.lb(findRxnIDs(model,
'BacteroidIII_rxnNGAM')) * (nodule / 0.02);
    modelTemp.lb(findRxnIDs(modelTemp, 'NoduleI_MNXR96136_C')) = model.lb(findRxnIDs(model,
'NoduleI_MNXR96136_C')) * (nodule / 0.02);
    modelTemp.lb(findRxnIDs(modelTemp, 'NoduleII_MNXR96136_C')) = model.lb(findRxnIDs(model,
'NoduleII_MNXR96136_C')) * (nodule / 0.02);
    modelTemp.lb(findRxnIDs(modelTemp, 'NoduleIIP_MNXR96136_C')) = model.lb(findRxnIDs(model,
'NoduleIIP_MNXR96136_C')) * (nodule / 0.02);
    modelTemp.lb(findRxnIDs(modelTemp, 'NoduleIZ_MNXR96136_C')) = model.lb(findRxnIDs(model,
'NoduleIZ_MNXR96136_C')) * (nodule / 0.02);
    modelTemp.lb(findRxnIDs(modelTemp, 'NoduleIII_MNXR96136_C')) = model.lb(findRxnIDs(model,
'NoduleIII_MNXR96136_C')) * (nodule / 0.02);

    % Adjust maximum flux through the reaction transferring oxygen to nodule zone III
    modelTemp.ub(findRxnIDs(modelTemp, 'TNIII_OXYGEN-MOLECULE')) = model.ub(findRxnIDs(model,
'TNIII_OXYGEN-MOLECULE')) * (nodule / 0.02);

    % Update nodule definition
    nodule = 0.01 * sol.x(pos) / n2fixNew;

    % Adjust maximum flux through fixed ammonium conversion
    modelTemp.ub(pos) = n2fixNew * (nodule / 0.01);

    % Get new growth
    sol = optimizeCbModel(modelTemp);

    % Stop the loop when local maximum obtained
    HHH = sol.x(pos) / ((sol.x(findRxnIDs(modelTemp, 'EX_BiomassNodule[c]')) /
(sol.x(findRxnIDs(modelTemp, 'EX_BiomassNodule[c]')) + sol.x(findRxnIDs(modelTemp, 'PlantBiomass'))))
* 100);
    if isnan(HHH) || (HHH > 0.995 * n2fixNew && HHH < 1.005 * n2fixNew)
        x = 0;
        nodule = 0.01 * sol.x(pos) / n2fixNew;
        if nodule > nodMax
            nodule = nodMax;
        end
    else
        nodule = 0.01 * sol.x(pos) / n2fixNew;
        if nodule > nodMax
            nodule = nodMax;
        end
    end
end

```

```

end
end

% Define 'y'
y = 1;

% See if increasing nodule amount can increase growth rate, iteratively
while y == 1

    % Record old values
    noduleOld = nodule;
    solOld2 = sol;

    % Increase the amount of nodule by 1%
    nodule = nodule + 0.001;
    if nodule > nodMax
        nodule = nodMax;
    end

    x = 1;
    % Get the local maximum as done above
    while x == 1
        solOld = sol;
        modelTemp.ub(findRxnIDs(modelTemp, 'TNIII_OXYGEN-MOLECULE')) =
model.ub(findRxnIDs(model, 'TNIII_OXYGEN-MOLECULE')) * (nodule / 0.02);
        modelTemp.ub(pos) = n2fixNew * (nodule / 0.01);
        sol = optimizeCbModel(modelTemp);
        if isempty(sol.x)
            sol.x = zeros(length(model.rxns),1);
        end
        modelTemp.lb(findRxnIDs(modelTemp, 'EX_BiomassNodule[c]')) = (nodule *
sol.x(findRxnIDs(model, 'PlantBiomass')) / (1 - nodule);
        modelTemp.lb(findRxnIDs(modelTemp, 'BacteroidIld_rxnNGAM')) = model.lb(findRxnIDs(model,
'BacteroidIld_rxnNGAM')) * (nodule / 0.02);
        modelTemp.lb(findRxnIDs(modelTemp, 'BacteroidIlp_rxnNGAM')) = model.lb(findRxnIDs(model,
'BacteroidIlp_rxnNGAM')) * (nodule / 0.02);
        modelTemp.lb(findRxnIDs(modelTemp, 'BacteroidIZ_rxnNGAM')) = model.lb(findRxnIDs(model,
'BacteroidIZ_rxnNGAM')) * (nodule / 0.02);
        modelTemp.lb(findRxnIDs(modelTemp, 'BacteroidIII_rxnNGAM')) = model.lb(findRxnIDs(model,
'BacteroidIII_rxnNGAM')) * (nodule / 0.02);
        modelTemp.lb(findRxnIDs(modelTemp, 'NoduleI_MNXR96136_C')) = model.lb(findRxnIDs(model,
'NoduleI_MNXR96136_C')) * (nodule / 0.02);
        modelTemp.lb(findRxnIDs(modelTemp, 'NoduleIld_MNXR96136_C')) = model.lb(findRxnIDs(model,
'NoduleIld_MNXR96136_C')) * (nodule / 0.02);
        modelTemp.lb(findRxnIDs(modelTemp, 'NoduleIlp_MNXR96136_C')) = model.lb(findRxnIDs(model,
'NoduleIlp_MNXR96136_C')) * (nodule / 0.02);
        modelTemp.lb(findRxnIDs(modelTemp, 'NoduleIZ_MNXR96136_C')) = model.lb(findRxnIDs(model,
'NoduleIZ_MNXR96136_C')) * (nodule / 0.02);
        modelTemp.lb(findRxnIDs(modelTemp, 'NoduleIII_MNXR96136_C')) = model.lb(findRxnIDs(model,
'NoduleIII_MNXR96136_C')) * (nodule / 0.02);
        modelTemp.ub(findRxnIDs(modelTemp, 'TNIII_OXYGEN-MOLECULE')) =
model.ub(findRxnIDs(model, 'TNIII_OXYGEN-MOLECULE')) * (nodule / 0.02);
        nodule = 0.01 * sol.x(pos) / n2fixNew;
        modelTemp.ub(pos) = n2fixNew * (nodule / 0.01);
        sol = optimizeCbModel(modelTemp);
        HHH = sol.x(pos) / ((sol.x(findRxnIDs(modelTemp, 'EX_BiomassNodule[c]')) /
(sol.x(findRxnIDs(modelTemp, 'EX_BiomassNodule[c]')) + sol.x(findRxnIDs(modelTemp, 'PlantBiomass'))))
* 100);

        % if there is no growth, then stop the loop
        if sol.f == 0
            x = 0;
        % if HHH is NaN or if it is included in '0.995*n2fixNew'~'1.005*n2fixNew' range, then update the
nodule and stop the loop
        elseif isnan(HHH) || (HHH > 0.995 * n2fixNew && HHH < 1.005 * n2fixNew)
            x = 0;
            nodule = 0.01 * sol.x(pos) / n2fixNew;
            if nodule > nodMax
                nodule = nodMax;
            end
        % in any other condition, update the nodule and repeat the loop
        else
            nodule = 0.01 * sol.x(pos) / n2fixNew;

```

```

        if nodule > nodMax
            nodule = nodMax;
        end
    end
end

% If the local maximum is lower than the previous local maximum,
% then the global maximum has already been found and we can stop
if sol.f <= solOld2.f
    y = 0;
end
end

% Save the output
nRiga = n;
output_fixationEfficiency_limO2{nRiga,1} = n2fixNew;
output_fixationEfficiency_limO2{nRiga,2} = solOld2.x(findRxnIDs(model, 'PlantBiomass'));
output_fixationEfficiency_limO2{nRiga,3} = solOld2.x(findRxnIDs(model, 'NoduleBiomass'));
output_fixationEfficiency_limO2{nRiga,4} = output_fixationEfficiency_limO2{nRiga,3} / ...
    (output_fixationEfficiency_limO2{nRiga,2} + output_fixationEfficiency_limO2{nRiga,3});
output_fixationEfficiency_limO2{nRiga,5} = solOld2.x(pos);
end

% Add headers
% headers = {'Fixation_efficiency', 'Plant_biomass', 'Nodule_biomass', ...
%   'Percent_nodule', 'Fixation_total'};
% output_fixationEfficiency_limO2 = vertcat(headers, output_fixationEfficiency_limO2);

% Plot 2D
xData = cell2mat(output_fixationEfficiency_limO2(:,1));
yData = cell2mat(output_fixationEfficiency_limO2(:,2));
plotMode = {'b.-'; 'g.-'; 'r.-'; 'c.-'; 'm.-'; 'b.-'; 'g.-'; 'r.-'; 'c.-'; 'm.-'; 'b.-'; 'g.-'; 'r.-'; 'c.-'; 'm.-'};
plot(xData, yData, plotMode{i});
hold on

PlantBiomass_limO2(i,:) = output_fixationEfficiency_limO2(:,2);
Nodulation_limO2(i,:) = output_fixationEfficiency_limO2(:,4);

end
nodMaxStrings=strcat(nodMaxStrings,'%');
lgd = legend (nodMaxStrings, 'Location','southeast');
title(lgd,'Nodulation Limit')
title('Plant Growth Rate against Fixation Efficiency (with oxygen limitation)')
xlabel('Fixation Efficiency')
ylabel('Plant Biomass Growth Rate')
hold off

% Plot 3D
xAxis = cell2mat(output_fixationEfficiency_limO2(:,1));
i=1:length(nodMaxValues);
yAxis = nodMaxValues(i);
zData = cell2mat(PlantBiomass_limO2);
figure
surf(xAxis, yAxis, zData);
title('Plant Growth Rate against Nodulation Limit and Fixation Efficiency (with oxygen limitation)')
xlabel('Fixation Efficiency')
ylabel('Nodulation Limit')
zlabel('Plant Biomass Growth Rate')

zDataBis = cell2mat(Nodulation_limO2);
figure
surf(xAxis, yAxis, zDataBis);
title('Nodulation against Nodulation Limit and Fixation Efficiency (with oxygen limitation)')
xlabel('Fixation Efficiency')
ylabel('Nodulation Limit')
zlabel('Nodulation')

% Clear workspace
clearvars -except PlantBiomass* Nodulation*

```