

### **General comments:**

This manuscript addresses questions of great relevance for ecosystems under development. Although it does not answer all questions related to the potential of nitrogen fixation in these ecosystems, it certainly represents a starting point. The conclusions are to a certain extent valid, but I would argue that they would have had much more weight if, for instance, more samples had been analyzed.

Overall, the presentation and writing style are clear, but some sentences a bit too long, which makes the interpretation difficult (see technical corrections). I have a few general remarks that should be taken into account when discussing the results. First, it refers to the fact that the authors collected samples from a developmental stage where clover had already exceeded the highest density. Wouldn't be interesting to take samples from sites where highest density is found? In the same lines, I miss description about each developmental site in terms of plants species, specially other legume plants, as this could influence the relative abundance of nifH harboring bacteria in the soil. My second remark refers to the strong fluctuations in nifH copy numbers found per soil age or nodule size. Specifically, I would suggest performing the same analyses for more nodules per treatment. Third remark is about sampling, which was performed at one moment of the year. I can imagine that the size of the nodules will greatly vary according to the developmental stage of the plant. I advise the authors to add few sentences in the discussion explaining how they would expect the size of the nodules to vary at different stages of plant development.

### **Specific comments:**

Ln 298: could you specify what Invsimpson measures and why it was chosen?

Ln 346: please explain what indeterminate nodules are and their relevance.

Ln 369: according to lines 297-298, site 5a had highest diversity

### **Technical corrections:**

In 58: "To address these questions..." Which questions? Replace questions by issues

In 99: "due to that..." due to what? Not clear.

Ln 154: replace respectively by and

Ln292-295: this sentence is too long and very difficult to follow. Please divide in two parts

Ln 297-300 and In 360-364: same as above

Figure 3: Please add "of nifH gene sequences" after "...clone libraries".

Figure 4: please add "partial" before "...nifH gene sequences..."