

Reviewer 1:

In this paper, the authors reviewed studies on the molecular markers for identification of adherent-invasive *Escherichia coli* (AIEC) strains that are associated with Crohn's disease (CD). This is an interesting topic. Overall, the paper is clearly written. However, some revisions are required, please see my comments below:

1. Line 23. Please change "Adherent-invasive *Escherichia coli* (AIEC) have been extensively" to "Adherent-invasive *Escherichia coli* (AIEC) strains have been extensively"

The change suggested has been applied in **line 23**.

2. Line 52. Please also add "strains".

We have added "strains" in **line 53** as indicated.

3. Lines 60-62. "Additionally, induction of 61 high levels of cytokine secretion and exacerbation of intestinal inflammation in susceptible 62 hosts due to AIEC presence has been reported (6-8)." Please revise this. The authors need to clearly state that these studies are animal studies.

The sentence in **line 61-63** has been clarified with the information indicated.

4. Lines 67-71: Please revise this. These sentences do not make sense in their current form. "To further decipher the role of AIEC in CD (such as disease specificity or association with 68 active disease), in addition to uncovering the host range or AIEC reservoirs and 69 transmission paths is of paramount importance for eventually defining measures of 70 contamination risk and prevention and/or to provide personalized treatments for AIEC 71 carriers."

Thank you, we have modified these sentences in **Line 68-71**.

5. Line 84: "The first studies" sounds strange. Please revise this.

We have modified it (**line 85**).

6. Lines 91-98: The authors mentioned so many genes, what are they? Please add a brief explanation.

Based on the reviewer's comment, we have considered classifying the genes mentioned according to its function (**line 95-99**).

7. Table 1: Phylogroup: is this phylogroup based on house-keeping genes? This information should be there.

Information about the method used to determine strain phylogroup has been added (**line 831-832**).

8. Tables 2 and 3: please move the explanation sentences from the title to footnotes.

Change done as indicated.

9. Lines 368-369: How do you measure the abundance of AIEC? Given that no molecular marker is available. Please explain this further.

Since a molecular marker for AIEC identification is still not available, AIEC identification is currently determined by isolating *E. coli* strains from biological samples and then assessing whether they present adhesion and invasion abilities in intestinal epithelial cells and

replication capacity inside macrophages. Then, abundance can be calculated as percentage of *E.coli* isolates with AIEC-phenotype in relation to the total *E.coli* isolates. Since this requires a lot of work most studies only report AIEC prevalence and we believe that the discovery of an AIEC biomarker will facilitate the study of AIEC abundance and thus, epidemiological data will improve (line 373-377).

Reviewer 2:

The review is very well written and sums up the state of knowledge on the genetics of AIEC bacteria. Before publication, it would be important to clarify or comment on the 2 following points:

- Do the authors think, as indicated in the abstract, that AIEC can be used as a "diagnostic tool", or should it rather be considered as a therapeutic target? This point deserves to be clarified.

We appreciate the reviewer's comment. Our point of view nowadays is that AIEC could be used as a CD therapeutic target since is involved in CD pathogenesis. In the present, AIEC detection as a diagnostic tool would present many false positives (healthy subjects with AIEC) and many false negatives (CD patients without AIEC). Nonetheless, we think that in the future AIEC abundance might be useful as a diagnostic approach but much research needs to be done before stating that, so we prefer to focus on the identification of AIEC-carriers. Therefore we have deleted the part related with diagnosis from the abstract (line 34-35).

- Is the difficulty in identifying molecular markers due to the absence of a standardized method for characterizing AIEC strains, or to the fact that these are pathobiont bacteria which have adapted/evolved to the digestive environment in inflammatory condition, and that several genetic mechanisms can lead to the same phenotype? In the second case, even if the characterization methods are standardized, the identification of molecular markers will not improve. In fact, within the same collection characterized in a standardized way, the genetics of the strains are diverse. This point deserves to be clarified and discussed.

We agree on the fact that even in a given collection characterized in a particular way, the genetics of AIEC strains are diverse and, so far, no AIEC specific biomarker has been found. Thus we envisage two scenarios (I) such molecular marker does not exist because different genetic mechanisms can lead to the same phenotype or (II) it has still not been found with the current methodologies so new approaches are needed; for instance, transcriptomic, DNA methylation and epigenetic studies directed at examining these elements under particular conditions in which AIEC isolates behave differently from other strains.

In any case, to validate any purposed marker, it is important to use strains that are characterized equally. Then, more precise accuracy values will be obtained. If we look for a biomarker using strains that have been classified as AIEC or non-AIEC using different protocols, and that in a determined lab these could be grouped in a different phenotype, it could lead to incorrect data and conclusions.

We strongly believe that it will not be possible to determine the exact reason why the identification of an AIEC biomarker is so difficult until AIEC strains will be identified using the same method worldwide. Once this occurs, strains will be classified equally, so one AIEC will be categorized as AIEC everywhere and thus we will avoid incorrect associations between genetic and phenotypic features which in turn it will increase the probability to identify which is the exact reason.

In order to clarify this issue we have modified the sentence in **Line 229-233.**