Reviewer's report

Title: Ulcerative colitis and irritable bowel patients exhibit distinct abnormalities of the gut microbiota

Version: 1 Date: 1 March 2010

Reviewer: Grietje Holtrop

Reviewer's report:

STATISTICAL REVIEW

The alignment and scaling of DGGE gels has been done in a thorough manner. Unfortunately, subsequent statistical analyses are poorly explained and not well-justified, and the methodology used is inappropriate in places (such as t-tests instead of analysis of variance). More details are given below.

COMPULSORY REVISION

C1. How was the replication mentioned in comment E1 taken into account in the statistical analyses? Please clarify.

C2. Second page of Results, line 3 onwards and elsewhere: there are three treatment groups and these data should be analysed by Analysis of Variance (ANOVA). If the effect of Treatment is significant, post-hoc t-tests (i.e. based on the mean residual sum of squares and residual degrees of freedom from the ANOVA output) can be performed to investigate which of the treatments differ. Replicate samples obtained from the same subject should be taken into account appropriately, this to avoid overly significant findings.

C3. Were assumptions of normality and equality of variance (with the latter begin more important than the former) across the treatment groups investigated? Were transformations or alternative analyses considered when these two assumptions were not met?

C4. Second page of Results, 2nd paragraph 2nd line: how was P < 10E-6 obtained?

C5. (see also comment E4) If PLS is already based on class membership, then what additional information is provided by the CVA analysis?

C6. CVA: how were the data ‘condensed’ before implementing CVA (as there are many more RF variables than samples)? Was this done by principal component analysis? How many principal components were used in the subsequent CVA? Please explain.

C7. Fourth page of Materials & Methods, line 6-7: why were the DGGE bands treated as absence/presence data for calculating similarity indices, whilst for the Shannon diversity index the band intensity was taken into account? If band
intensity is regarded reliable then why was the similarity index calculated from absence/presence?

C8. Why were bands to be excised selected based on the first PLS component only? Why was this not based on the Euclidean length of each RF in the first four PLS dimensions, say (based on classification success rates in Fig 3a)? If CVA gives the optimal 2-dimensional separation of the three treatment groups, then why was band-selection not based on maximum peaks from the first CV axis (back-transformed to the original RF) instead?

C9. Why were four bands selected for sequencing? Fig 4a seems to suggest that there are a further three peaks (0.221, 0.644, 0.746) that may be influential in segregation of the treatment groups. Were the four top-ranking bands always the same in the cross-validation exercise?

C10. Third page of Results section, 2nd paragraph (similarity). How were the similarity data analysed statistically? How were P-values obtained? It should be noted that we cannot perform t-tests or ANOVA on similarities, as there will be many more similarities than data points (for example, 4 data points will give 6 similarities, but t-tests/anova can not be used as the similarities are not independent from one another).

MINOR ESSENTIAL REVISION

E1. Please specify how many faecal samples per subject were analysed. At the start of Materials & Methods we have 13+11+22=46 subjects, whilst the data matrix based on DGGE mentions 96 lanes. The number of data points in Figure 3b also suggests that there were multiple samples per volunteer. Please also specify whether the multiple DGGE lanes were replicates from the same faecal sample or consisted of faecal samples collected at different days but from the same volunteer.

E2. Please explain the statistical methodology in layman’s terms, as many readers with a microbiology background will not be familiar with metabolomics techniques such as PLS.

E3. Second page of Results, line 6 onwards: why is the median reported here instead of the mean?

E4. Third page of Materials & Methods, bottom line: Please mention what was used as the Y-variable. Was this the so-called ‘class membership’?

E5. Is the information presented in Figure 3a used at all? See also comment C8.

E6. For how many samples and/or volunteers were the four selected bands sequenced? Was this done for all samples or for a subset of samples? If done for a subset of samples, how were the samples selected? Please clarify

E7. Materials & Methods, Subjects and sample collection. Line 4 states ‘age 18-60’ whilst line 6 gives ranges up to 64 years. Please make this consistent.
E8. Fig 6: what do the error bars represent?

E9. Fig 3a: the permutation test results are not well-explained. What do the mean +- 2 SD represent? Why are the means always approximately 37%, what does this mean?

E10. What does the +- refer to in the text (e.g. bottom line of first page of Results and various other places)?

DISCRETIONARY REVISION

None

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests