

Detected biomarkers by LEfSe analyses

LEfSe was used for comparing characteristic taxonomy among Han to non-Han in all groups; Hans to another Hans in different location; non-Hans to another non-Hans in different location; Hans to non-Hans in same location. Finally, biomarkers were obtained by four steps as follow.

Step 1: the Kruskal-Wallis test analyzes all features, testing whether the values in different classes are differentially distributed. Features violating the null hypothesis are further analyzed in Step 2. The significant threshold of $pvalue \leq 0.05$ in Kruskal-Wallis test.

Step 2: the pairwise Wilcoxon test checks whether all pairwise comparisons between subclasses within different classes significantly agree with the class level trend. The significant threshold of $pvalue \leq 0.05$ in Wilcoxon test.

Step 3: the resulting subset of vectors is used to build a Linear Discriminant Analysis model from which the relative difference among classes is used to rank the features. The final output thus consists of a list of features that are discriminative with respect to the classes, consistent with the subclass grouping within classes, and ranked according to the effect size with which they differentiate classes. The significant threshold of $LDA\ SCORE(\log_{10}) > 2$.

Step 4: Based on Step 3, the representative biomarkers of Hans in a certain location must be the largest or smallest in all different locations, and there are significant differences compared with other locations. Similarly, the representative biomarkers of non-Hans in a certain location must be the largest or smallest in all different locations, and there are significant differences compared with other locations.

Results as follow:

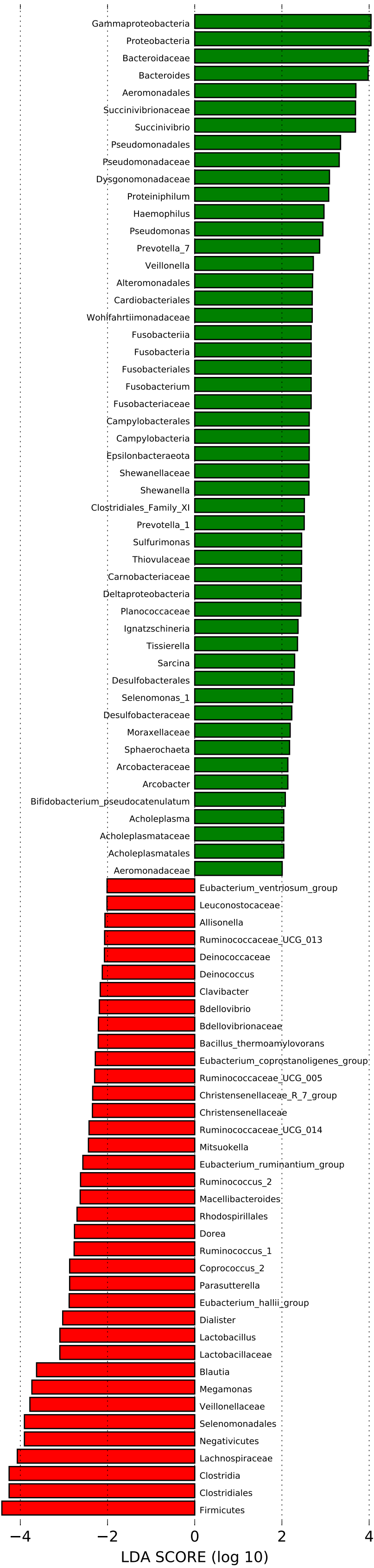
Differences are represented in the color of the most abundant class (Red and green represent different geographic location or ethnicity); LEfSe provides features that were differential among conditions of interest with statistical and biological significance, ranking them according to the effect size.

Hans to another Hans in different location: Pages 2-22

non-Hans to another non-Hans in different location: Pages 23-43

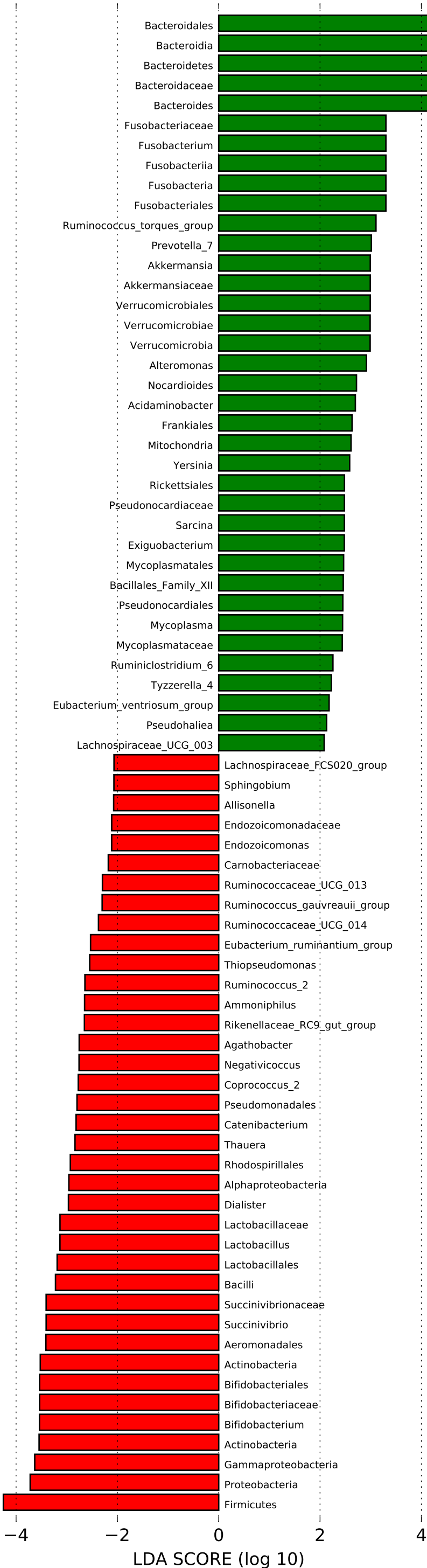
Hans to non-Hans in same location: Pages 44-50

Gansu Xinjiang



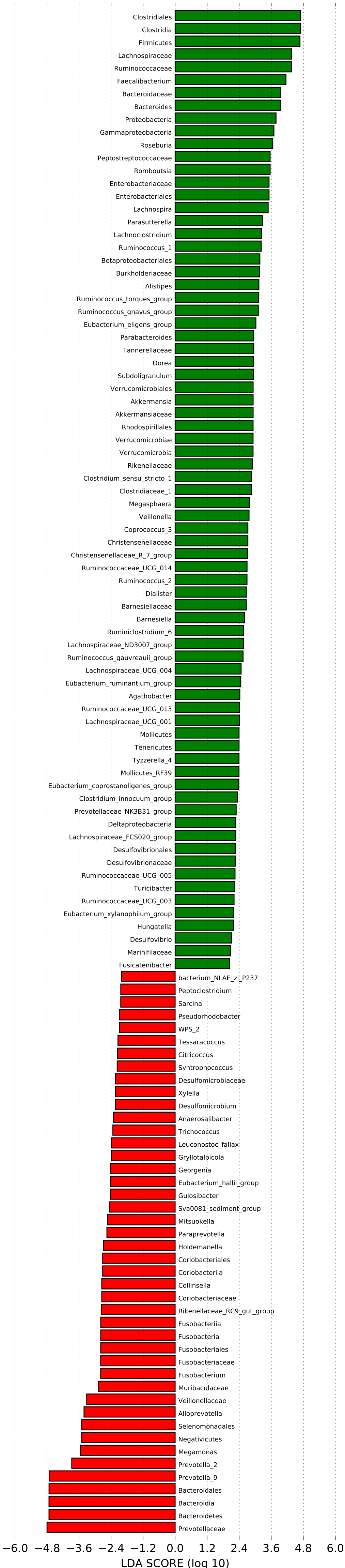
LDA SCORE (log 10)

Gansu Yunnan



■ Guizhou

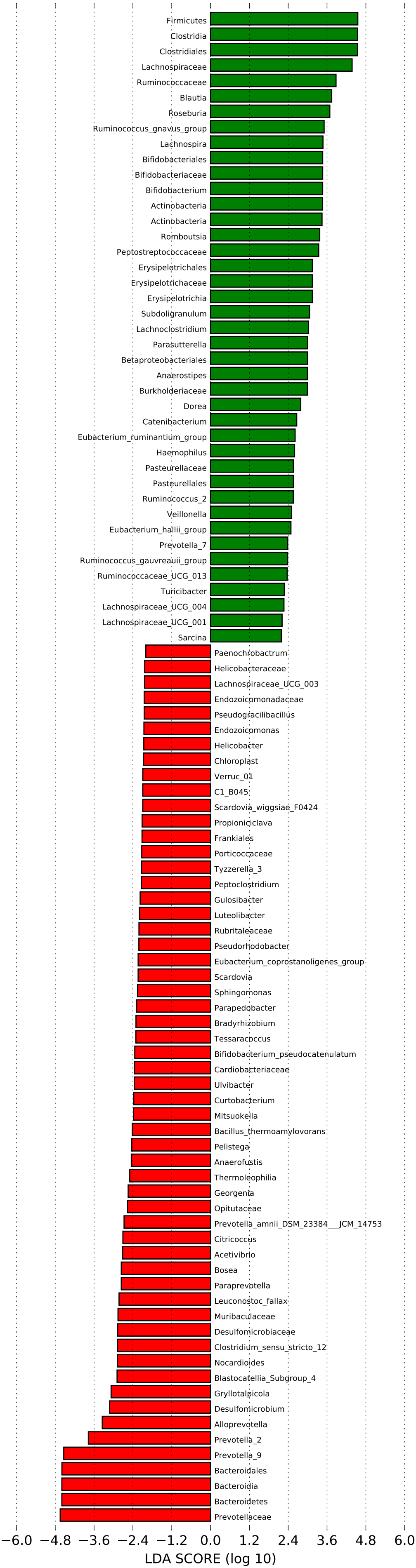
■ InnerMongolia



Guizhou Jilin

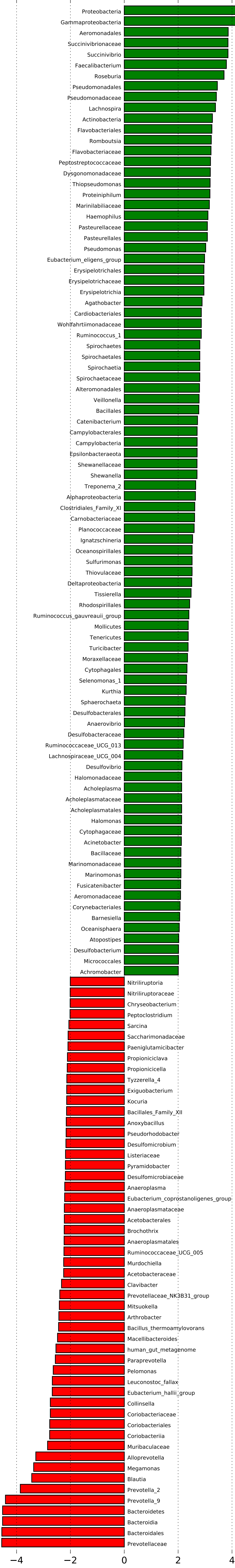


Guizhou Tibet

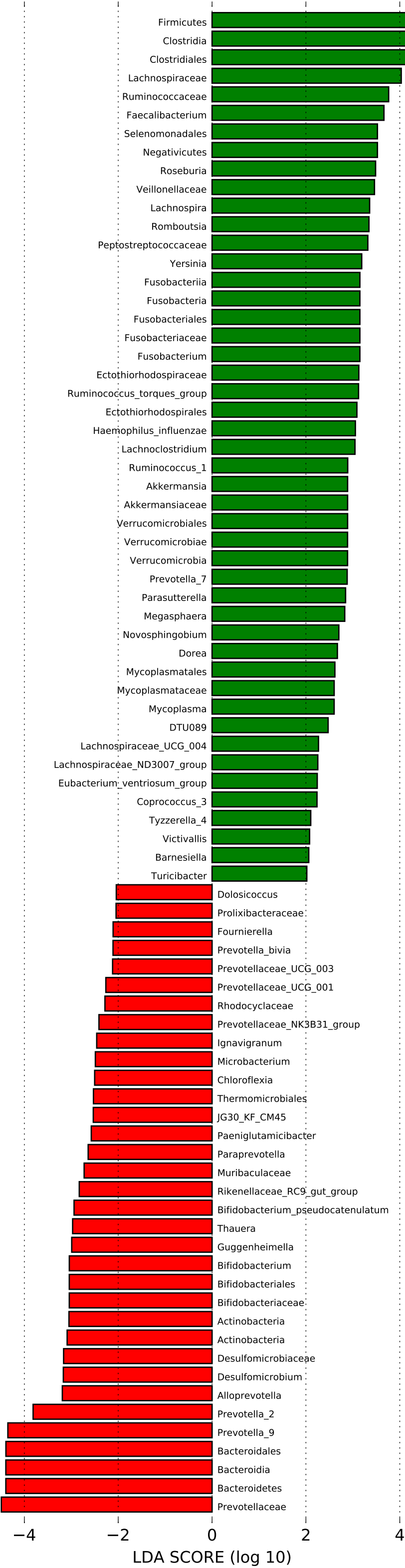


Guizhou

Xinjiang

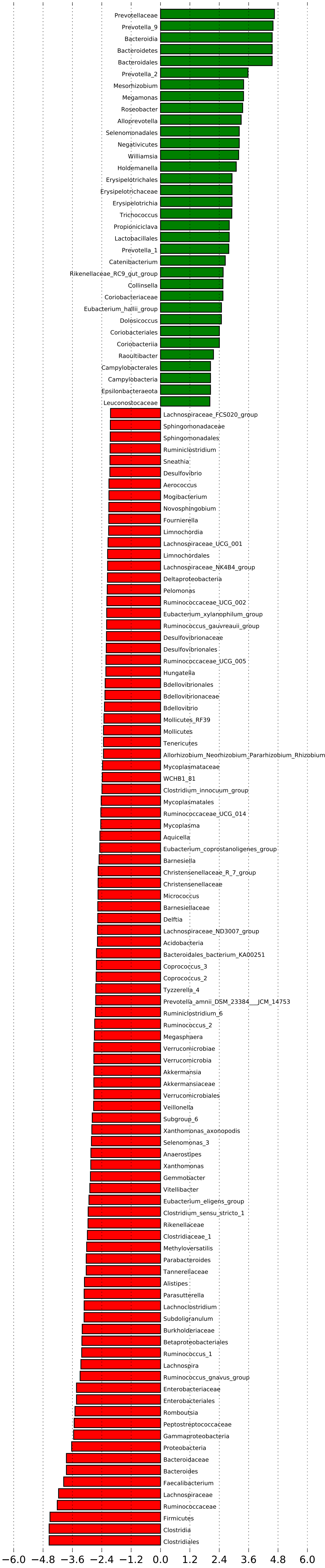


Guizhou Yunnan

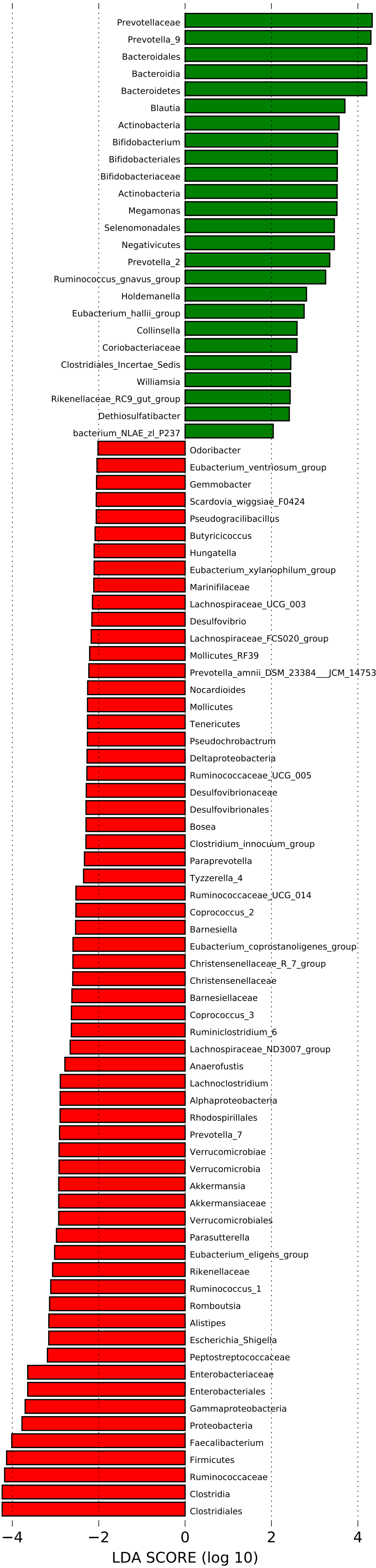


InnerMongolia

Jilin

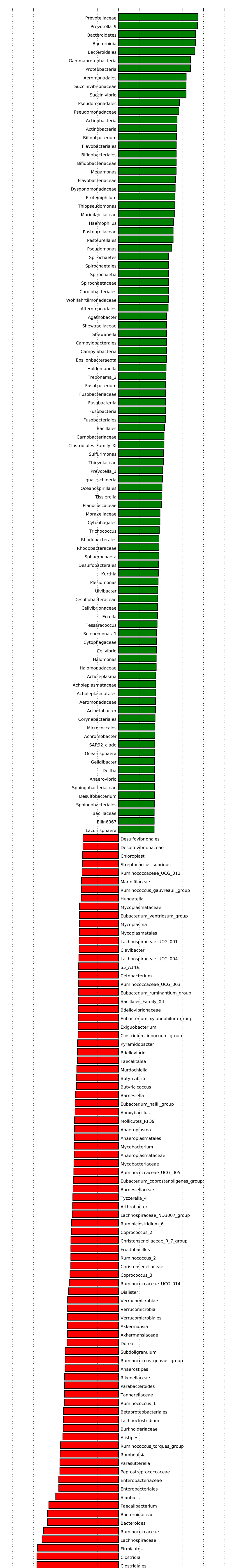


InnerMongolia Tibet

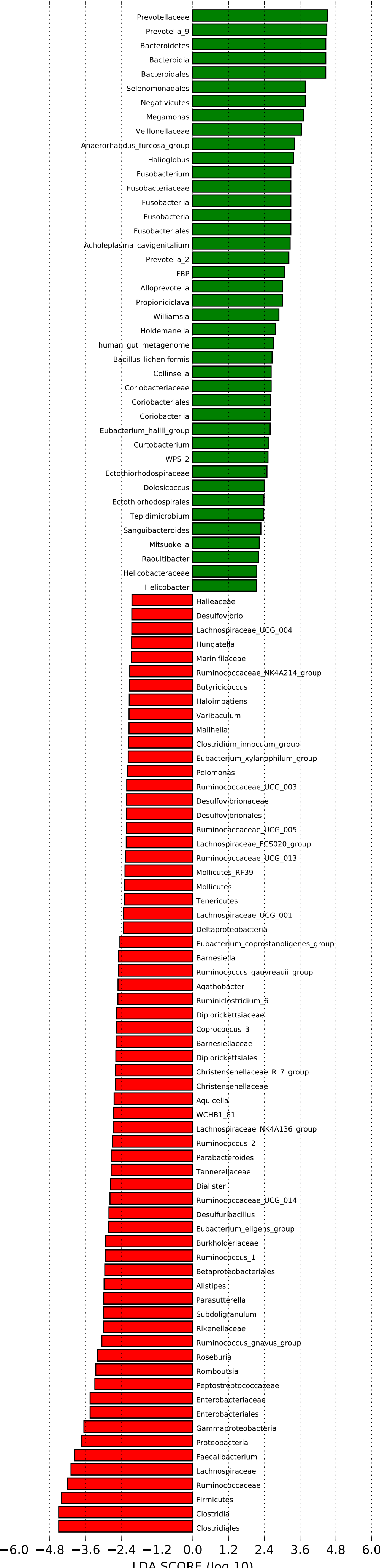


LDA SCORE (log 10)

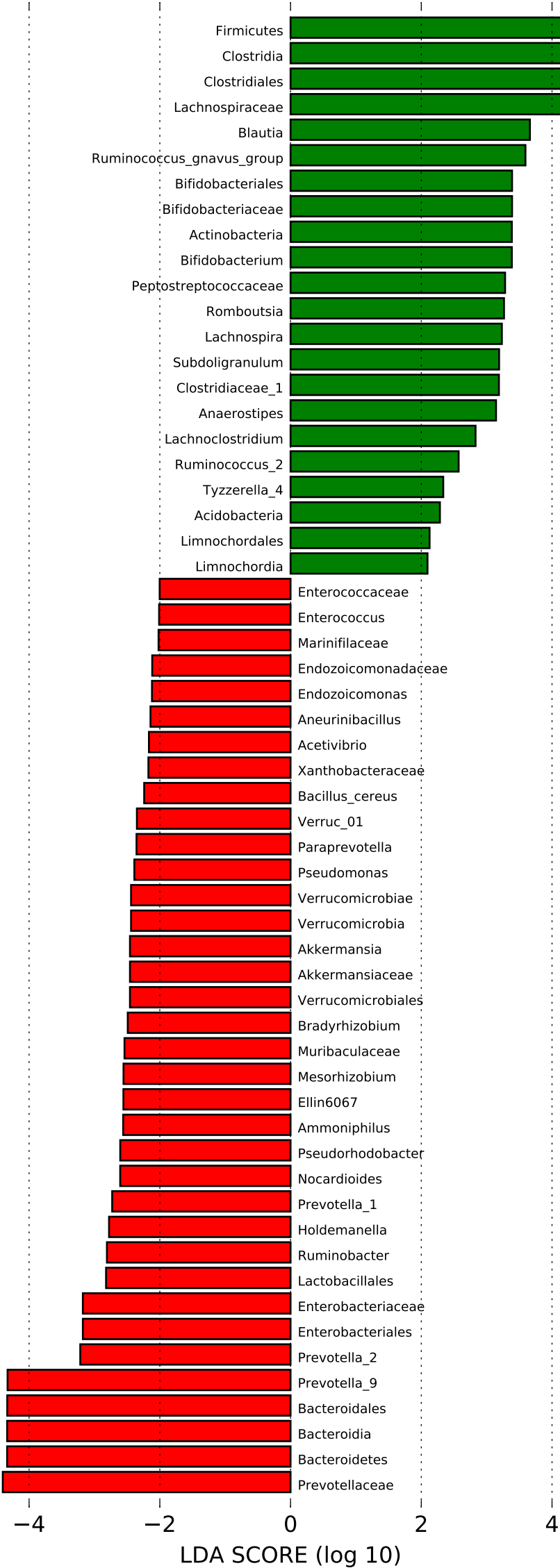
InnerMongolia Xinjiang



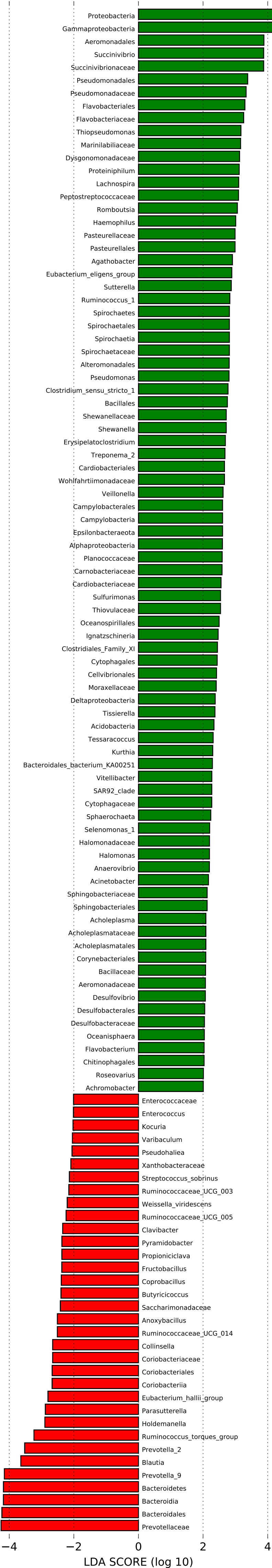
InnerMongolia Yunnan



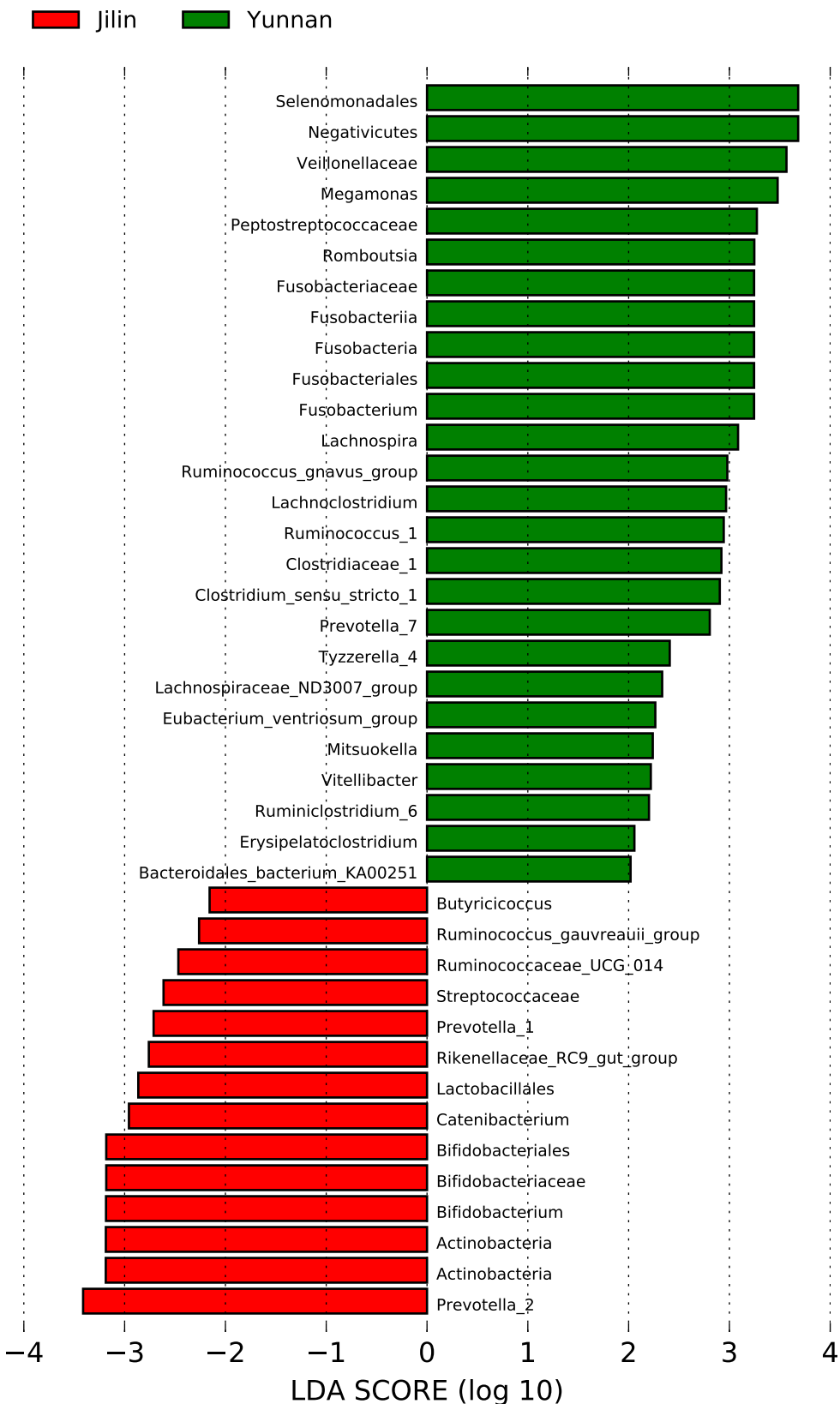
Jilin Tibet



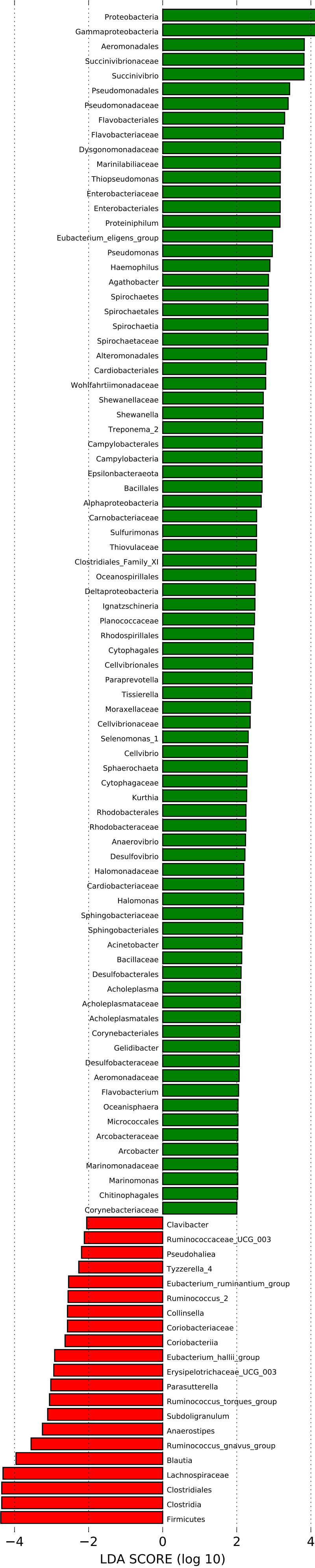
■ Jilin ■ Xinjiang



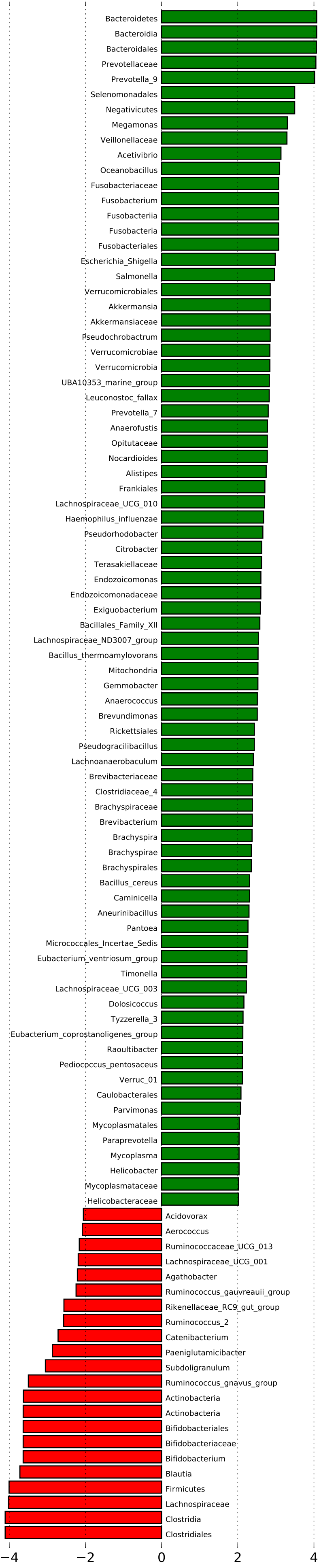
LDA SCORE (log 10)



Tibet Xinjiang

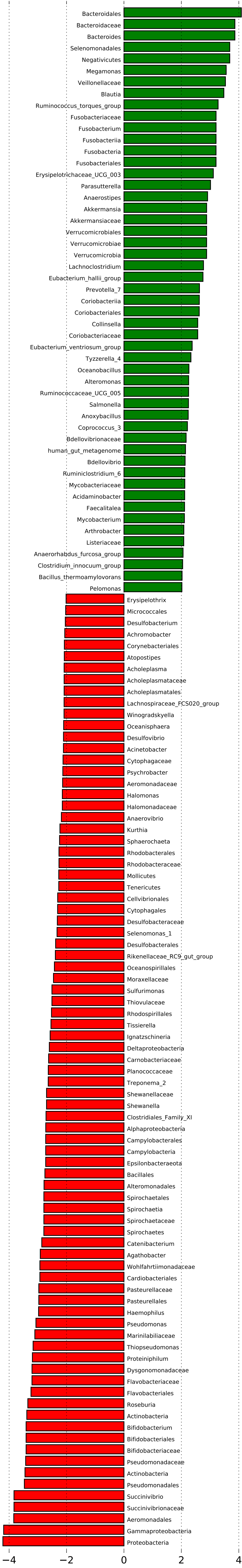


Tibet Yunnan

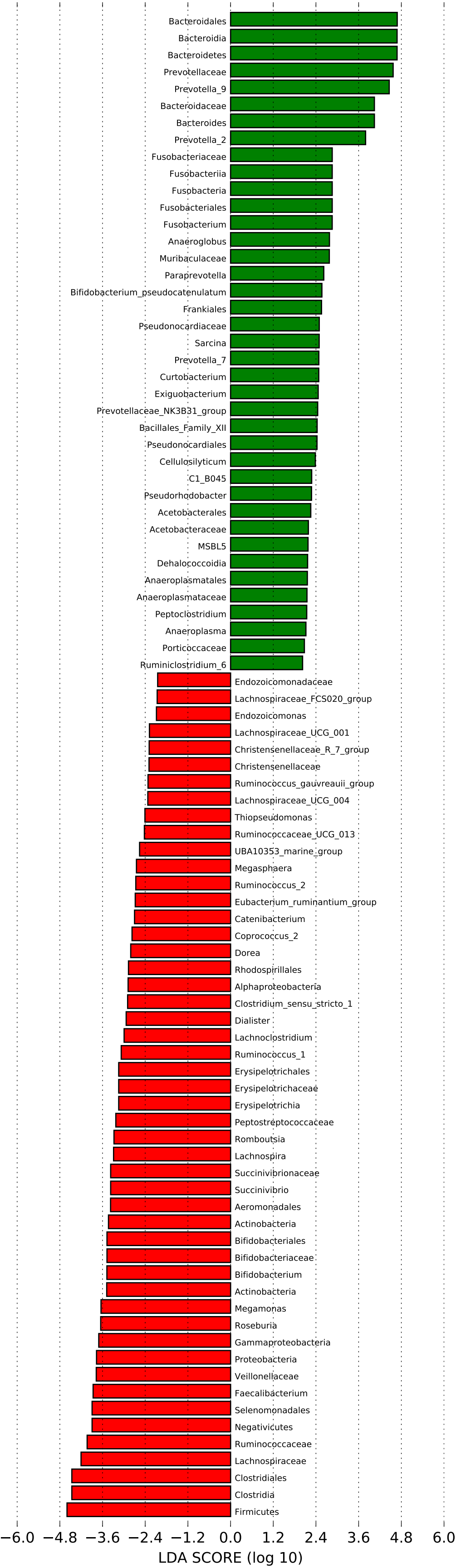


Xinjiang

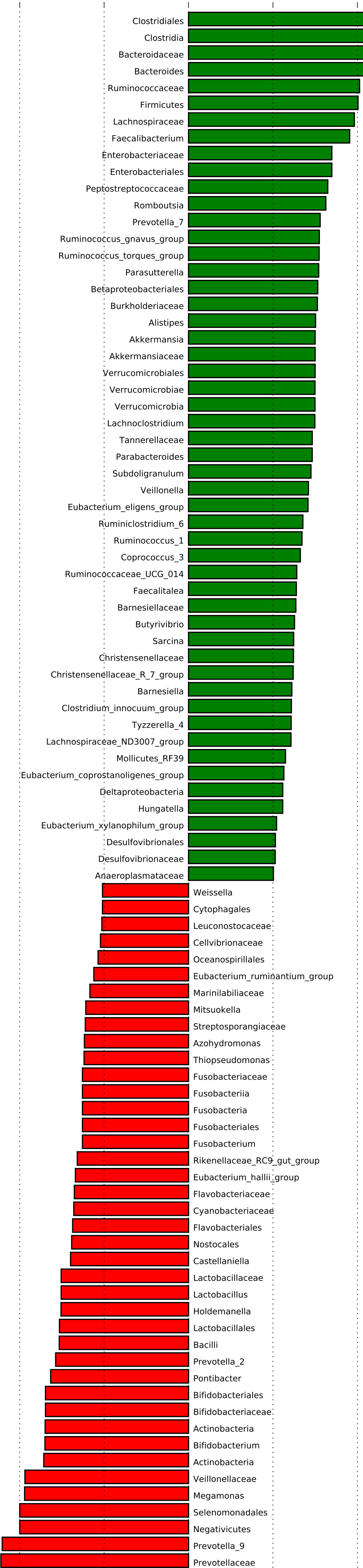
Yunnan



Gansu Guizhou

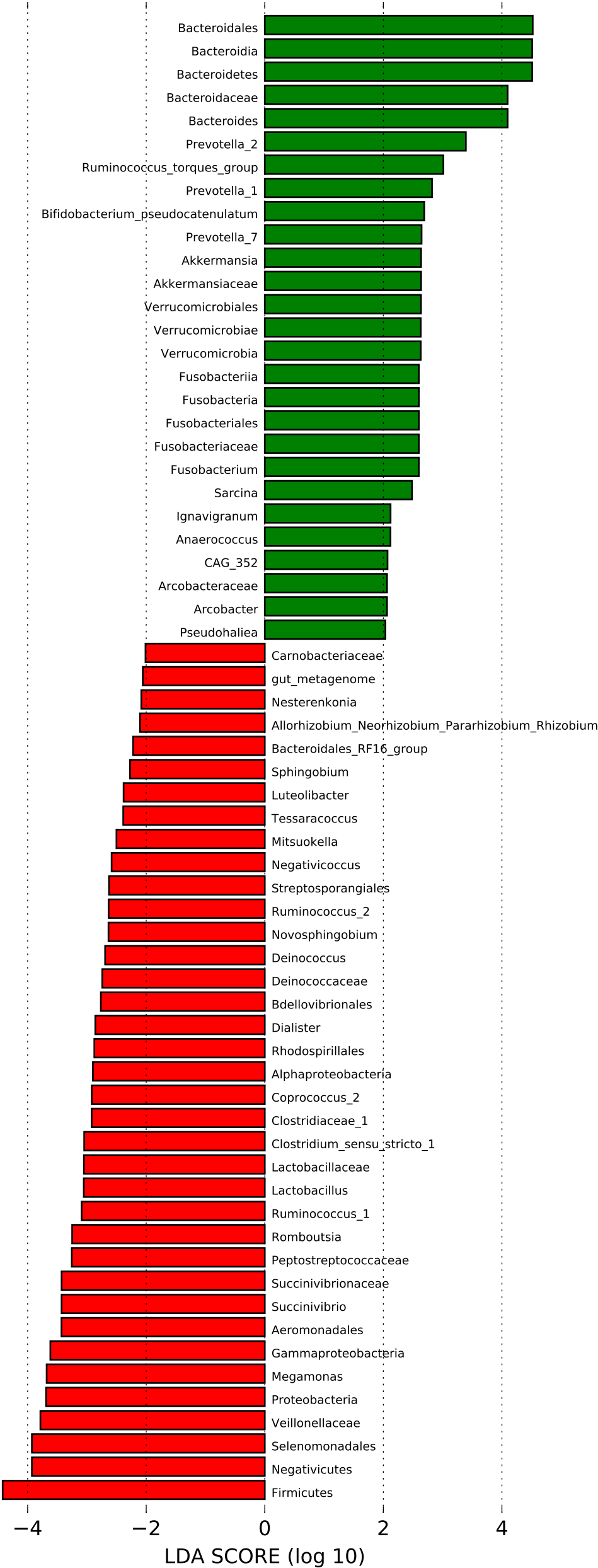


Gansu InnerMongolia

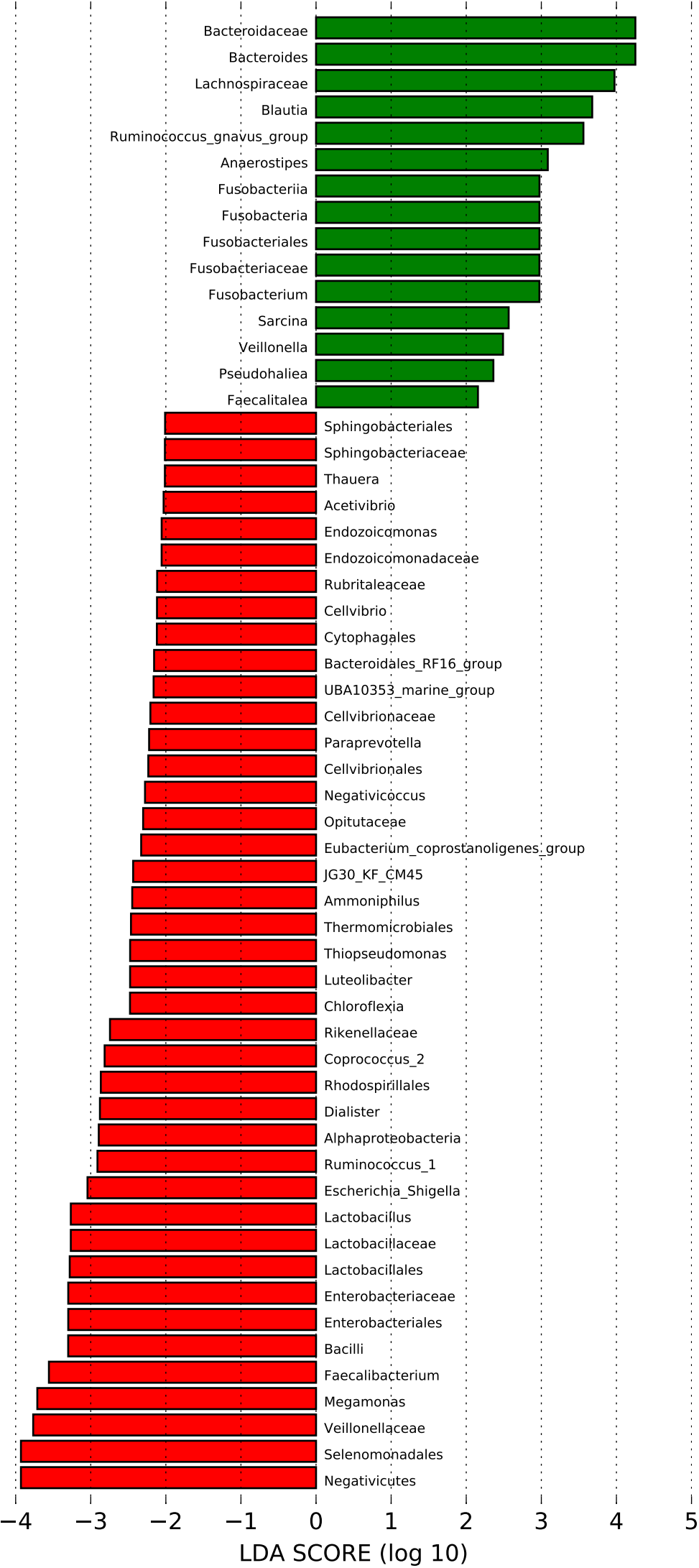


LDA SCORE (log 10)

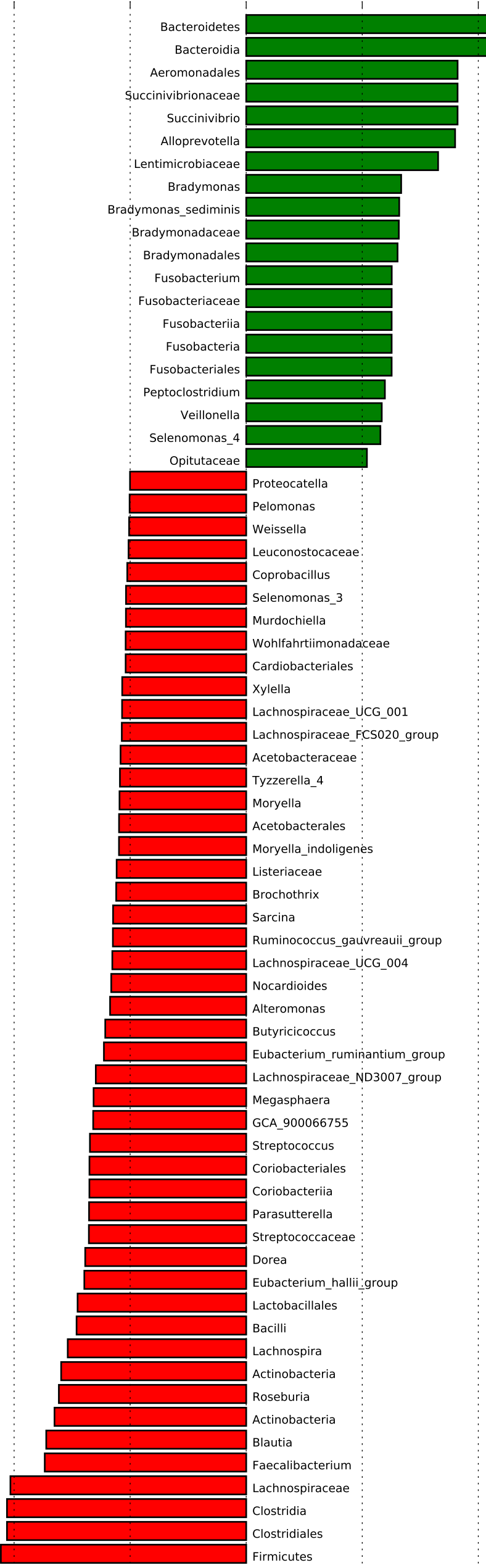
Gansu Jilin



■ Gansu ■ Tibet

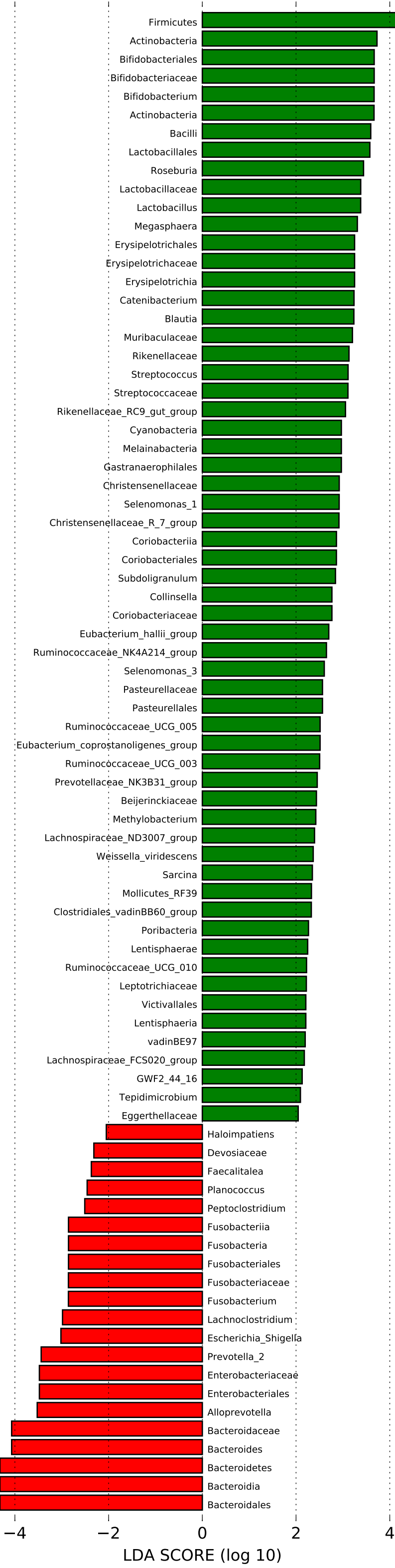


■ Korean ■ Miao



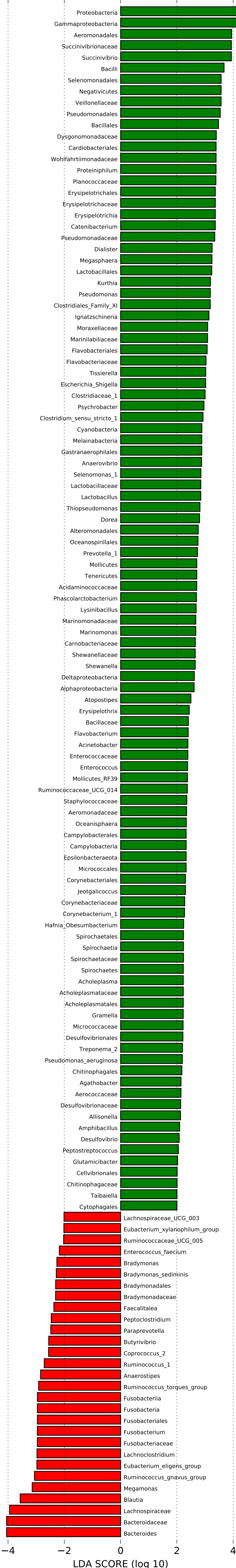
LDA SCORE (log 10)

Miao Tibetan



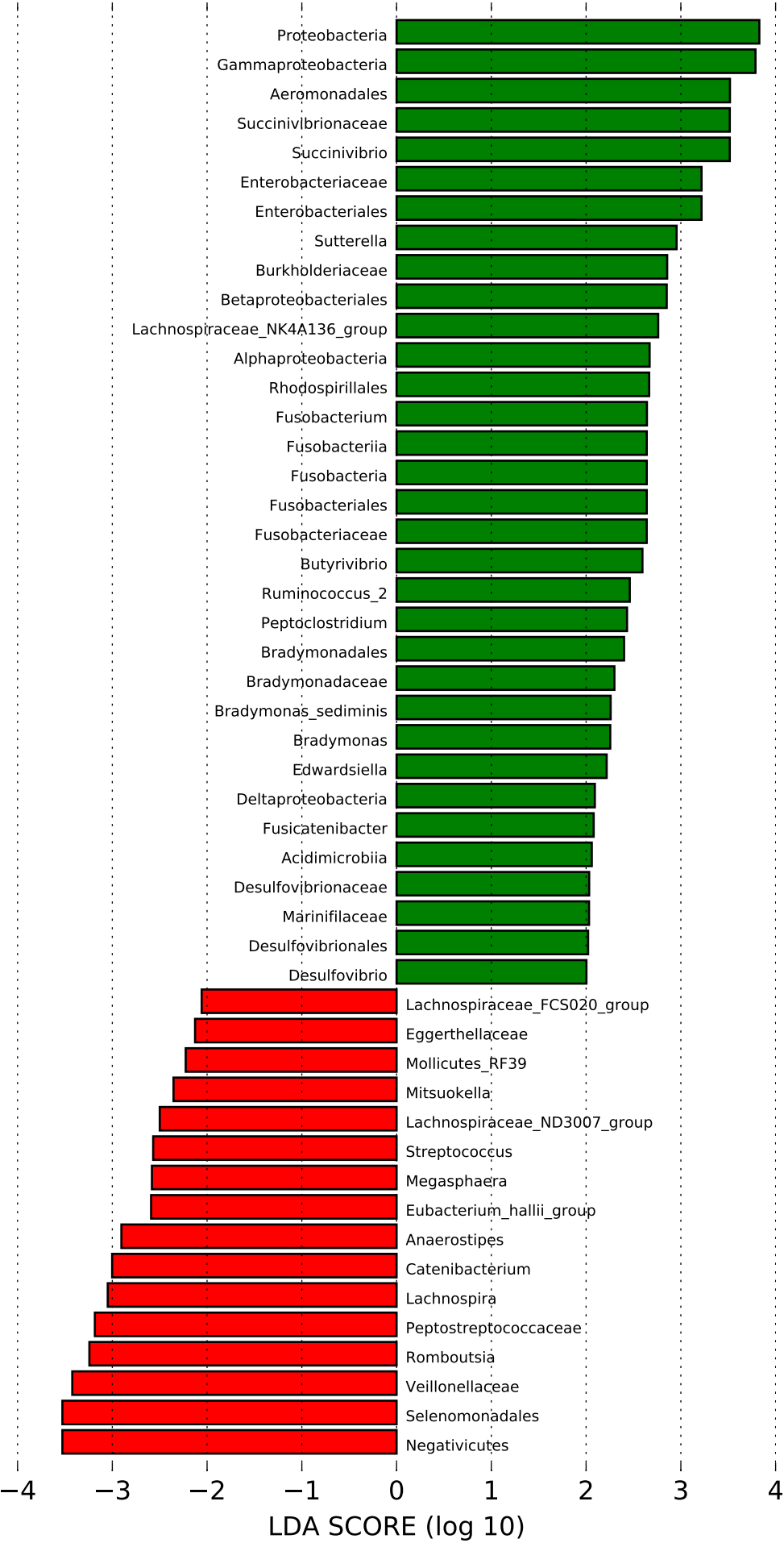
Miao

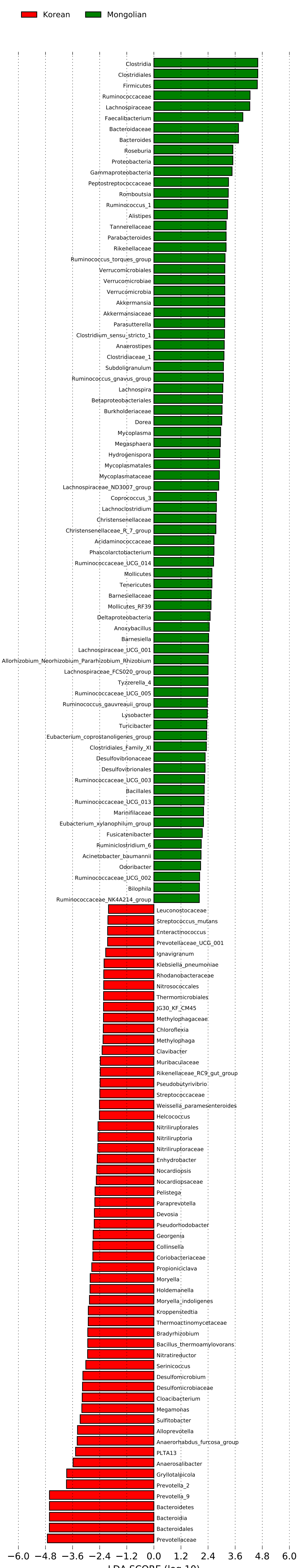
Uygur



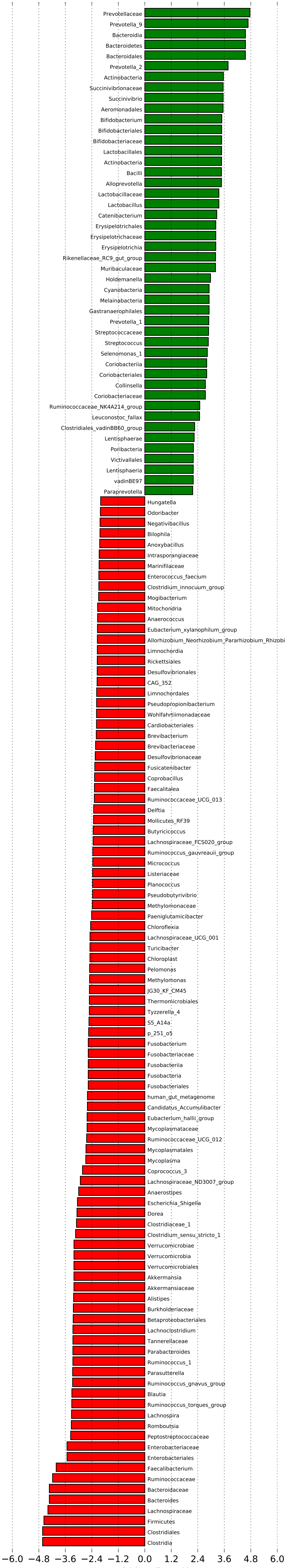
LDA SCORE (log 10)

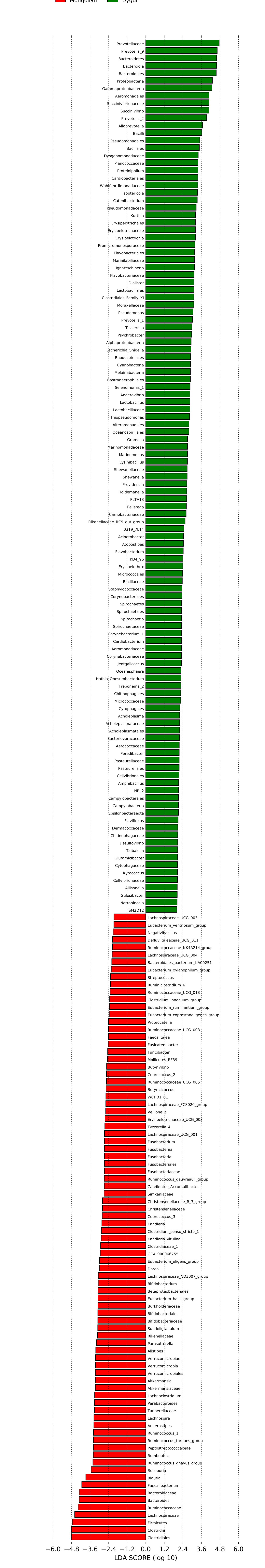
■ Bai ■ Miao

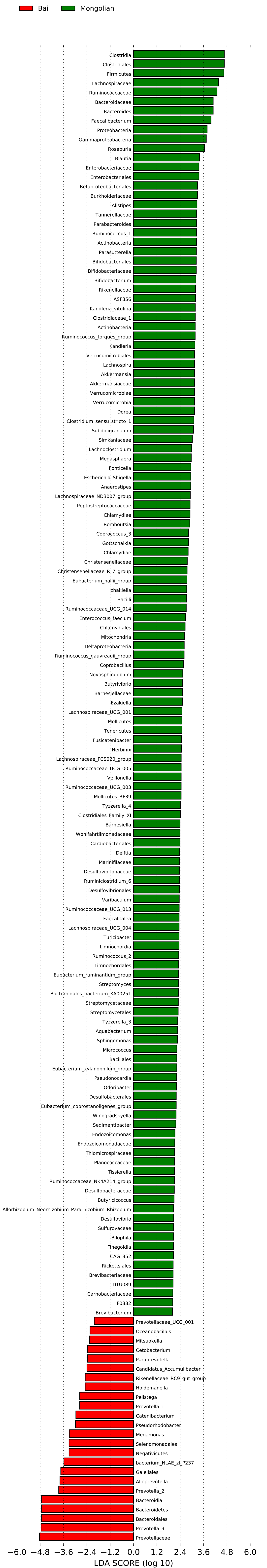




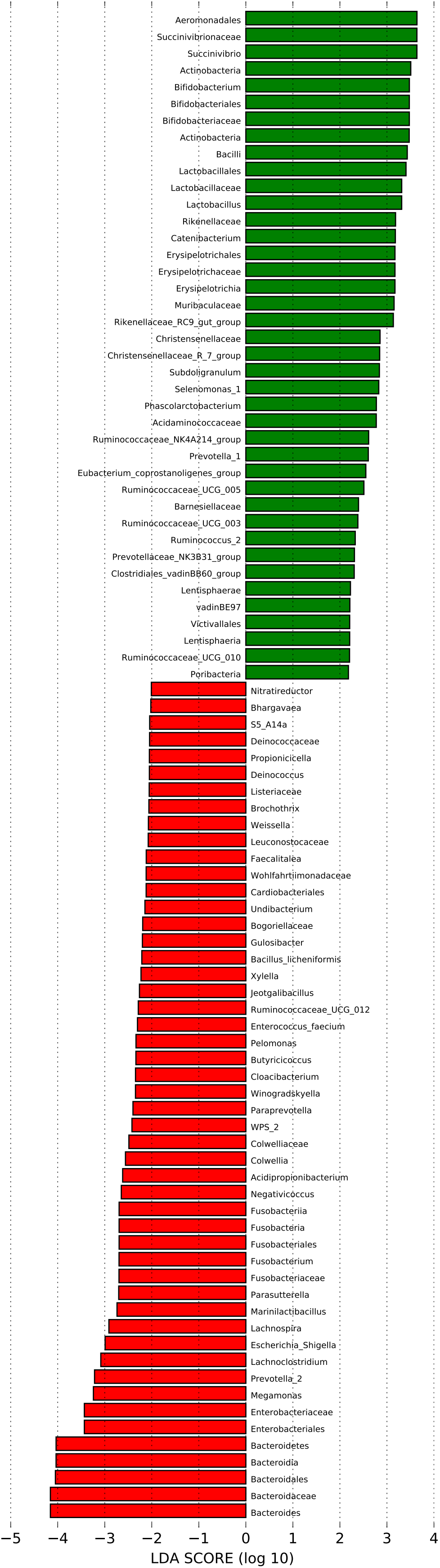
Mongolian Tibetan





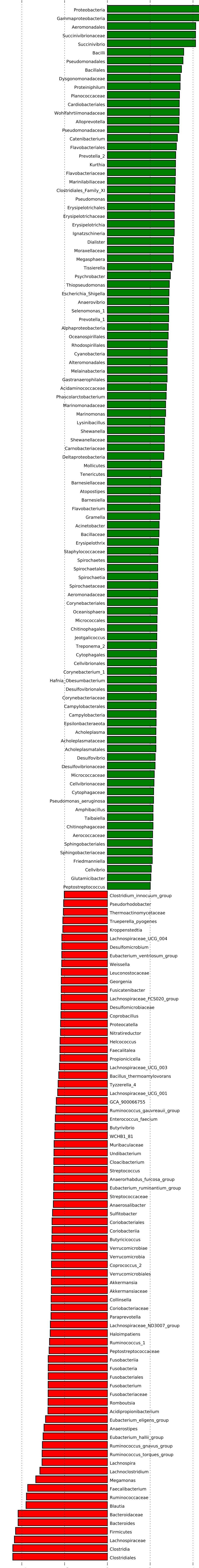


■ Korean■ Tibetan



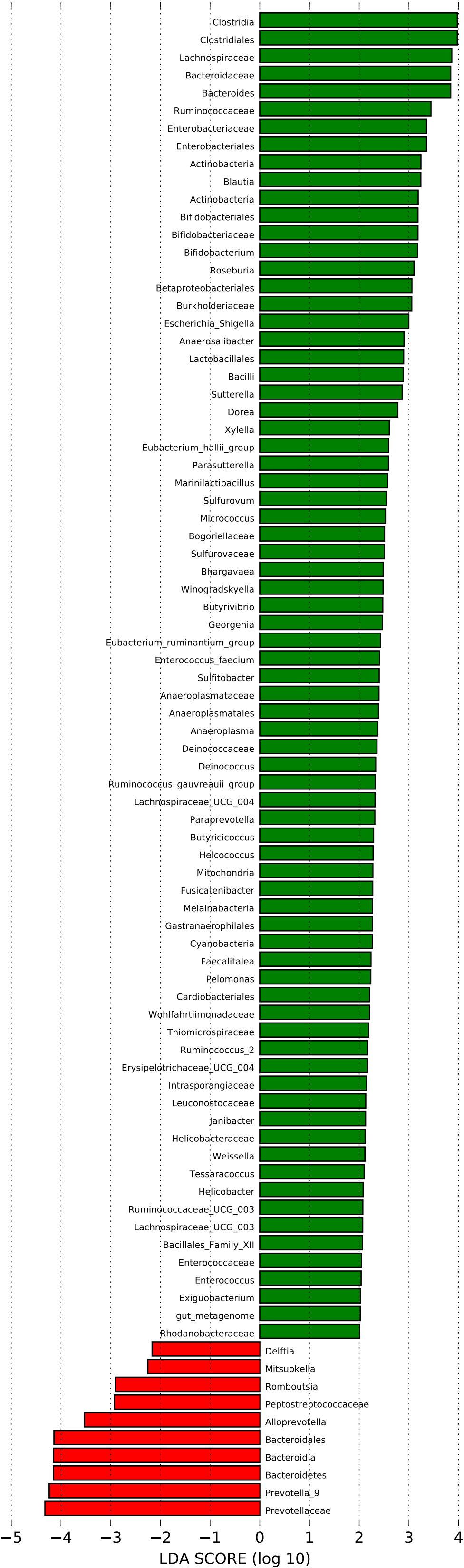
Korean

Uygur

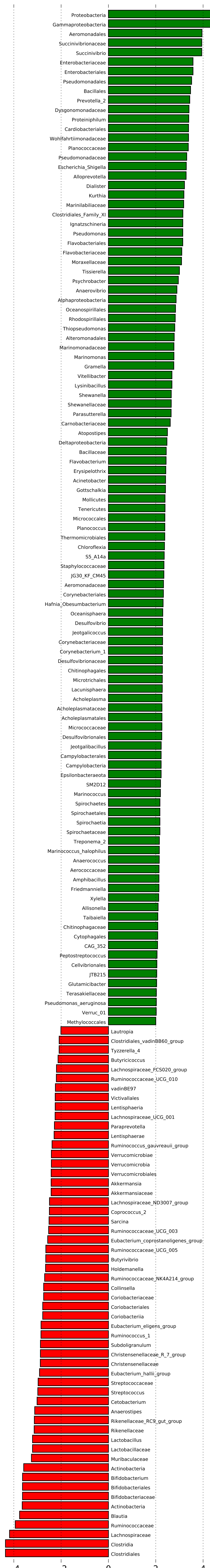


LDA SCORE (log 10)

Bai Korean

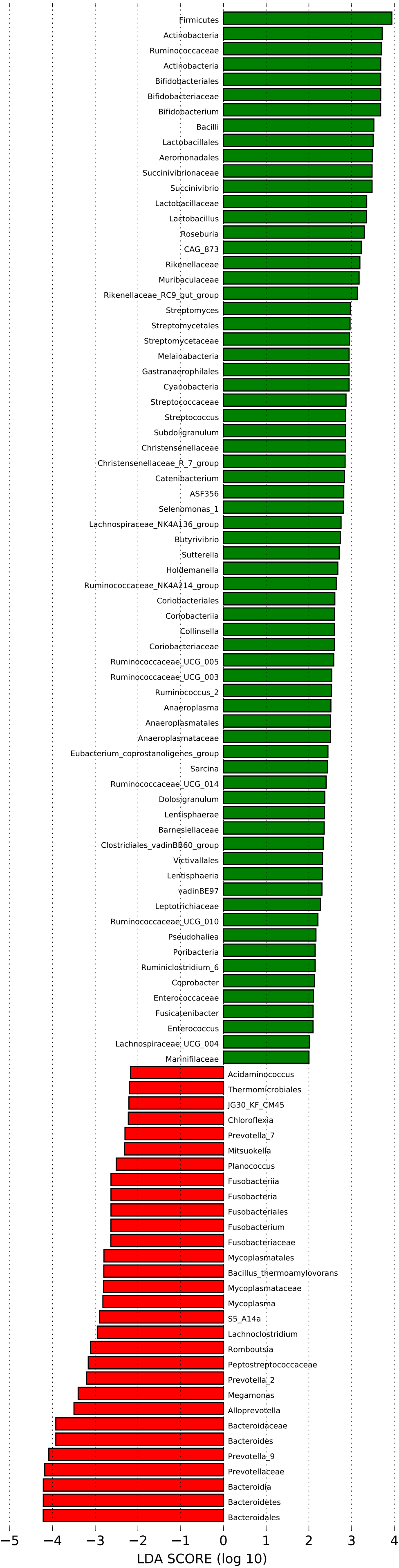


Tibetan Uygur



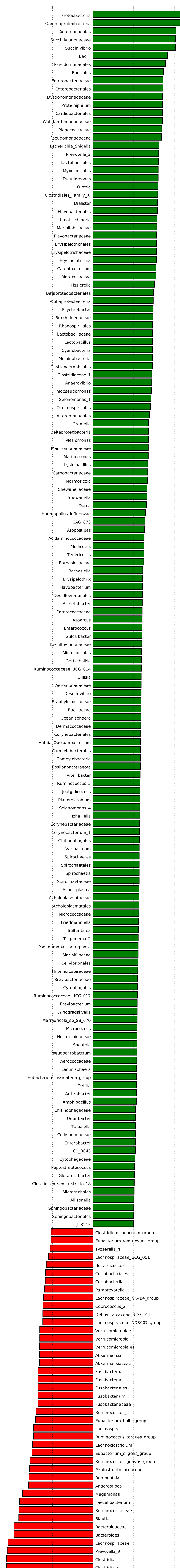
LDA SCORE (log 10)

Bai Tibetan



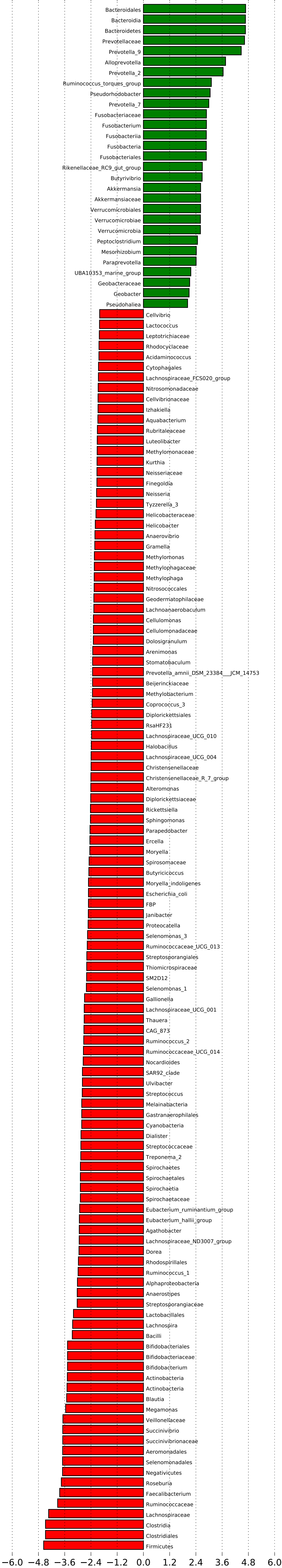
Bai

Uygur



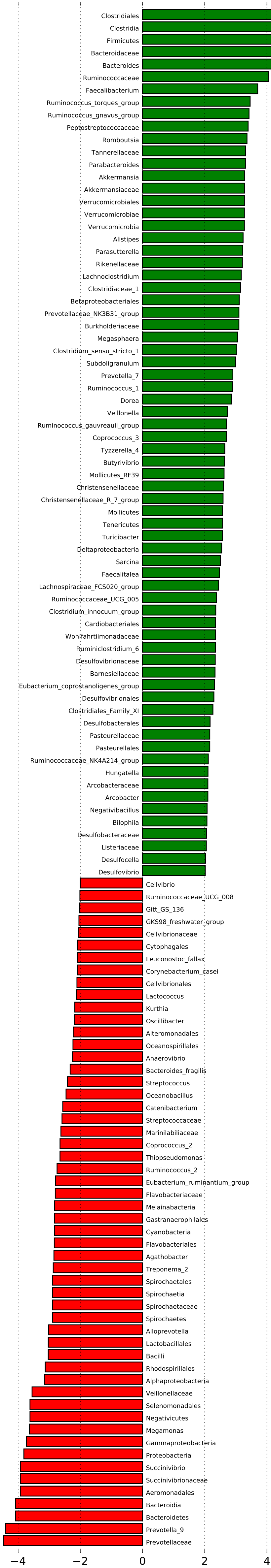
Hui

Miao

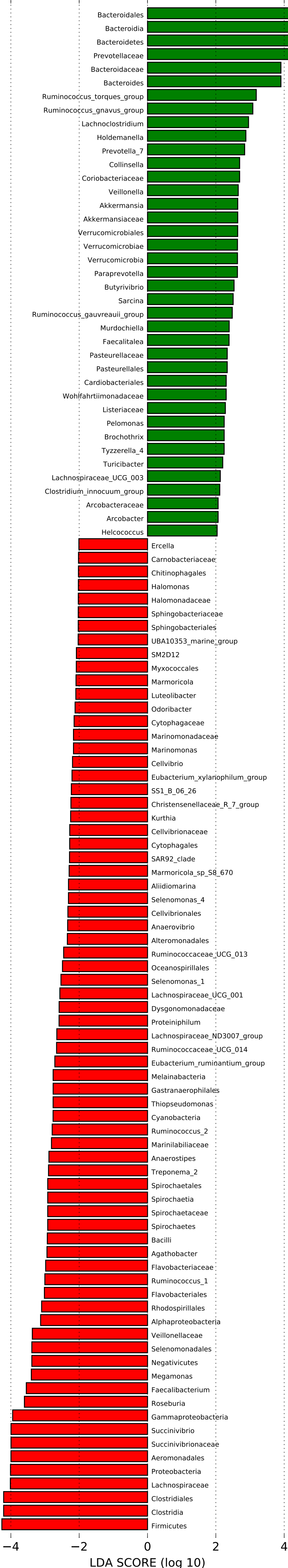




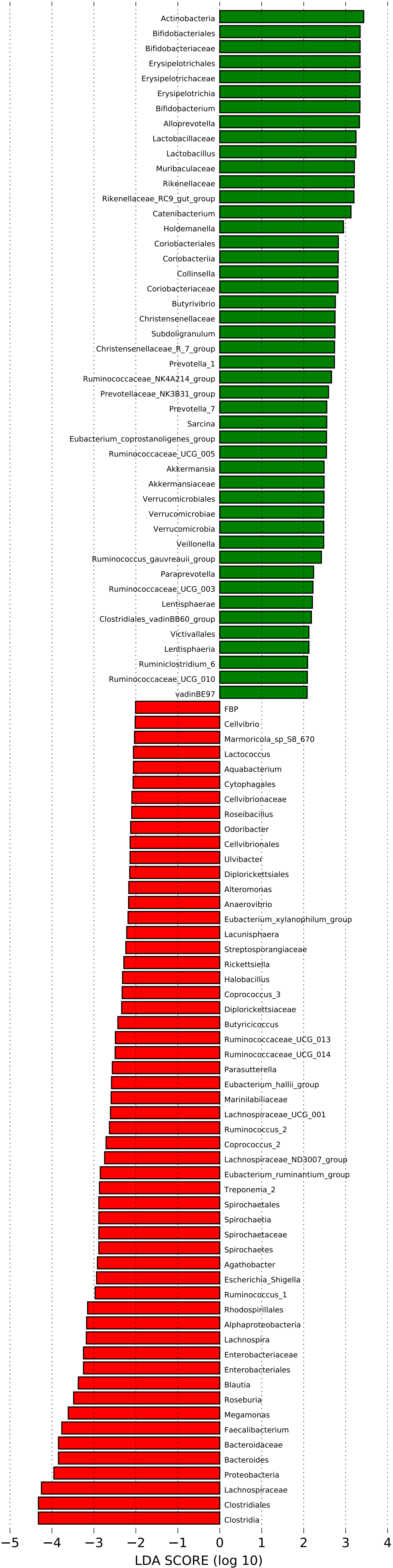
Hui Mongolian



Hui Korean

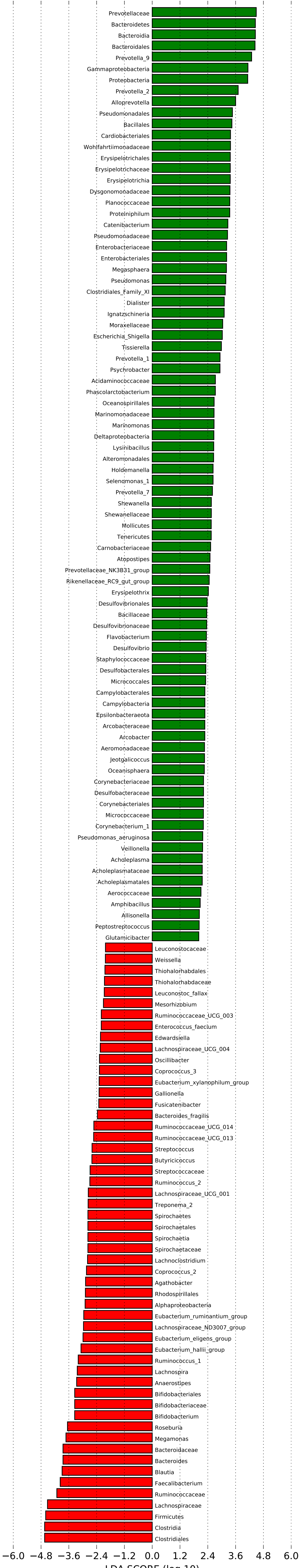


Hui Tibetan

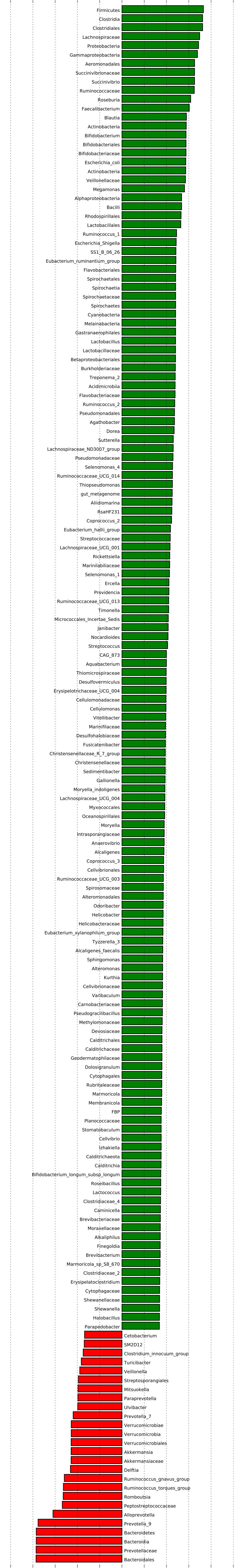


Hui

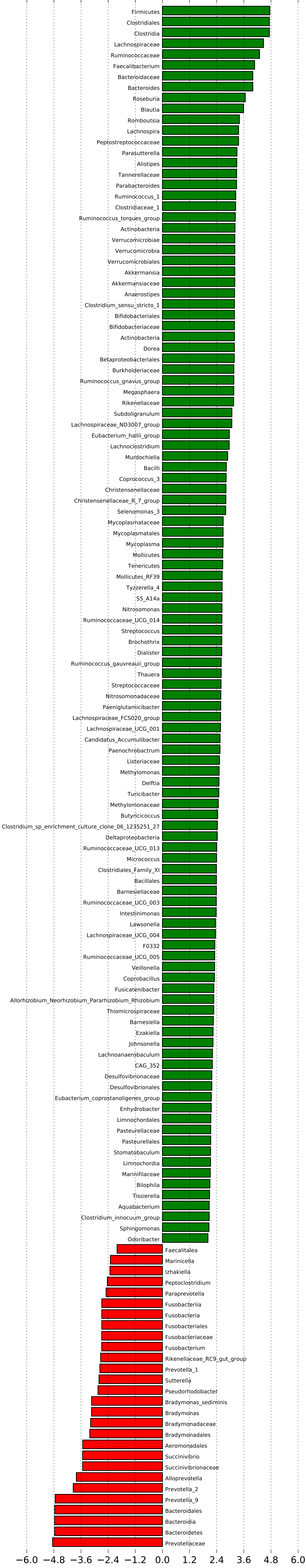
Uygur



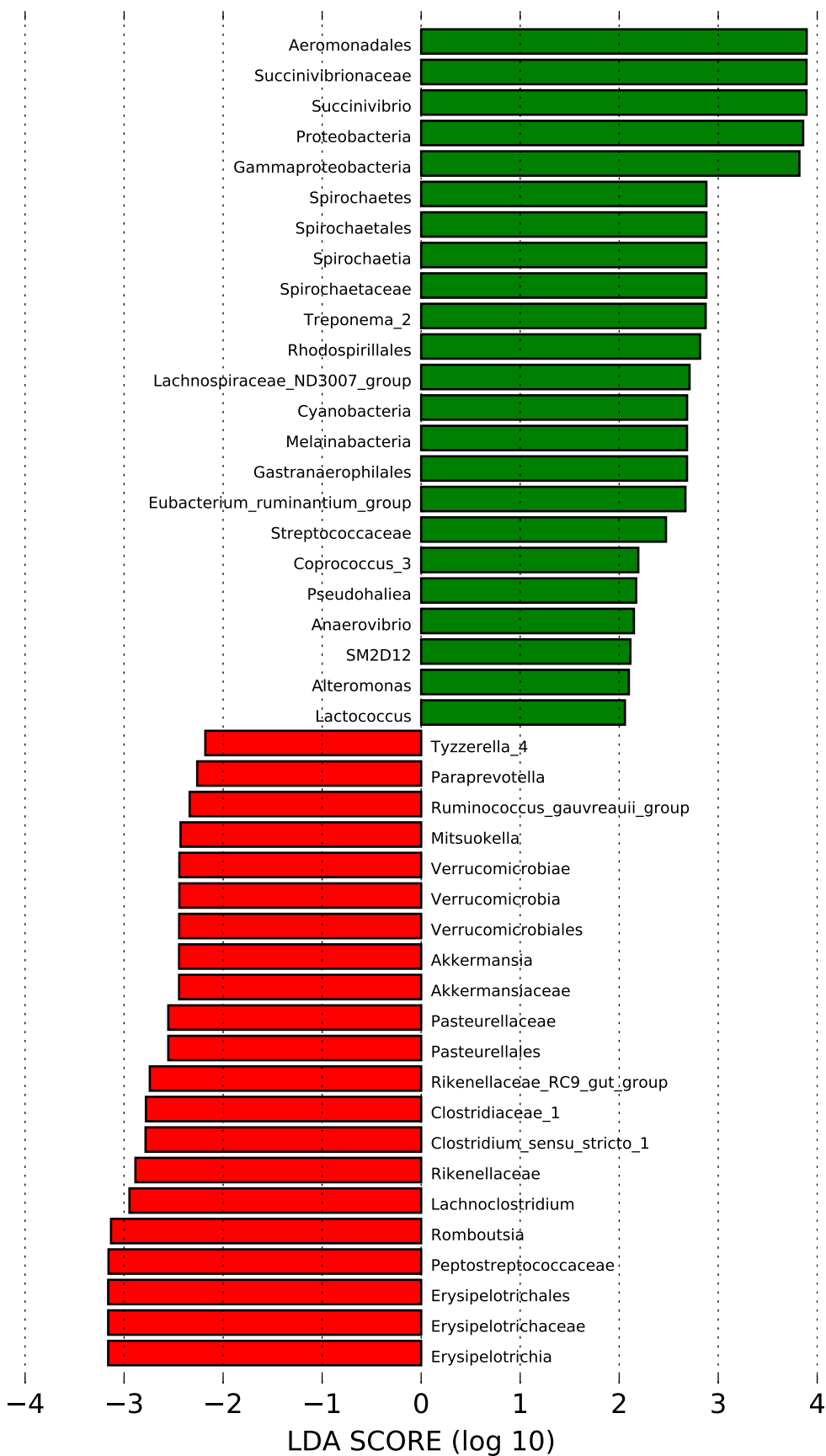
Bai Hui

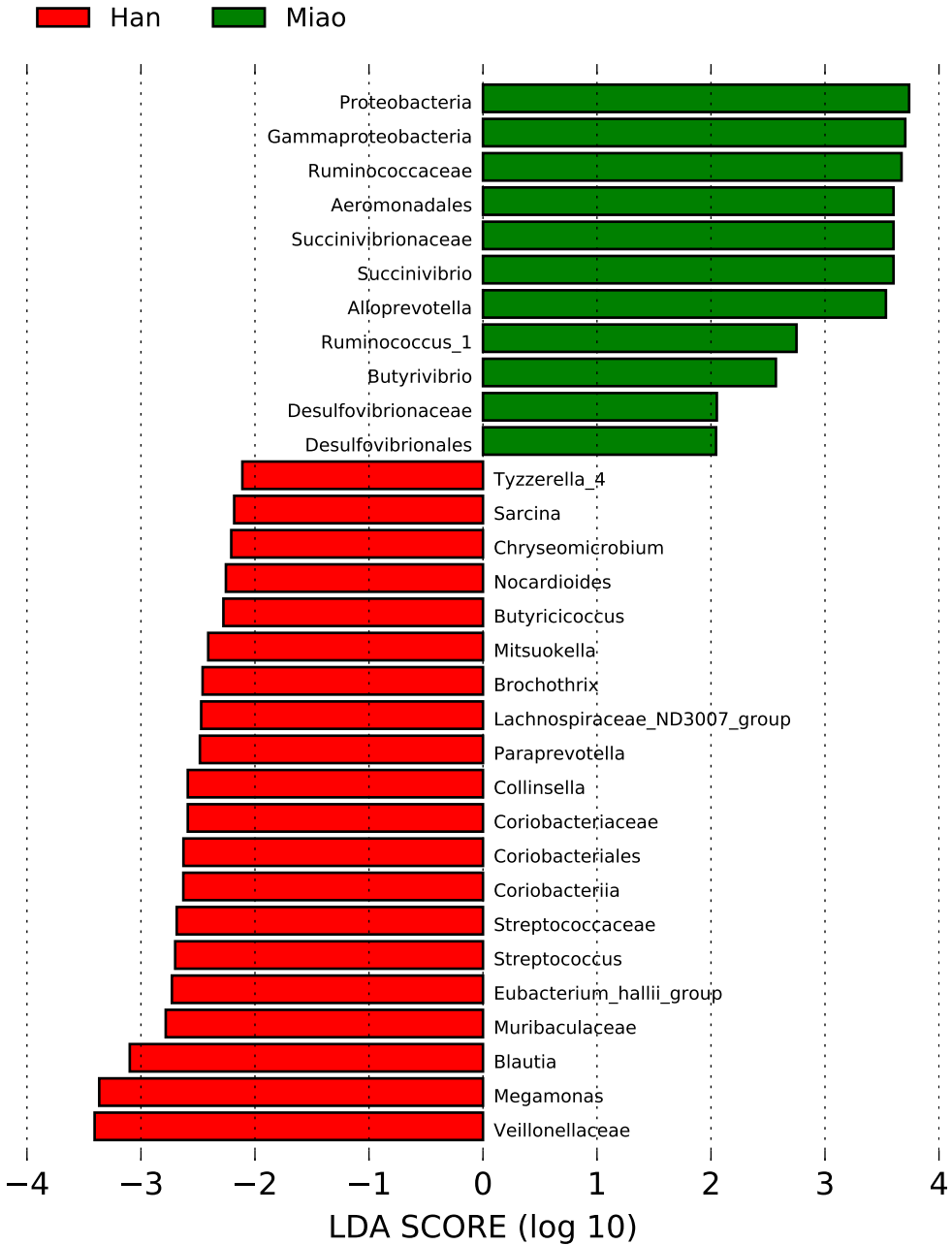


Miao Mongolian

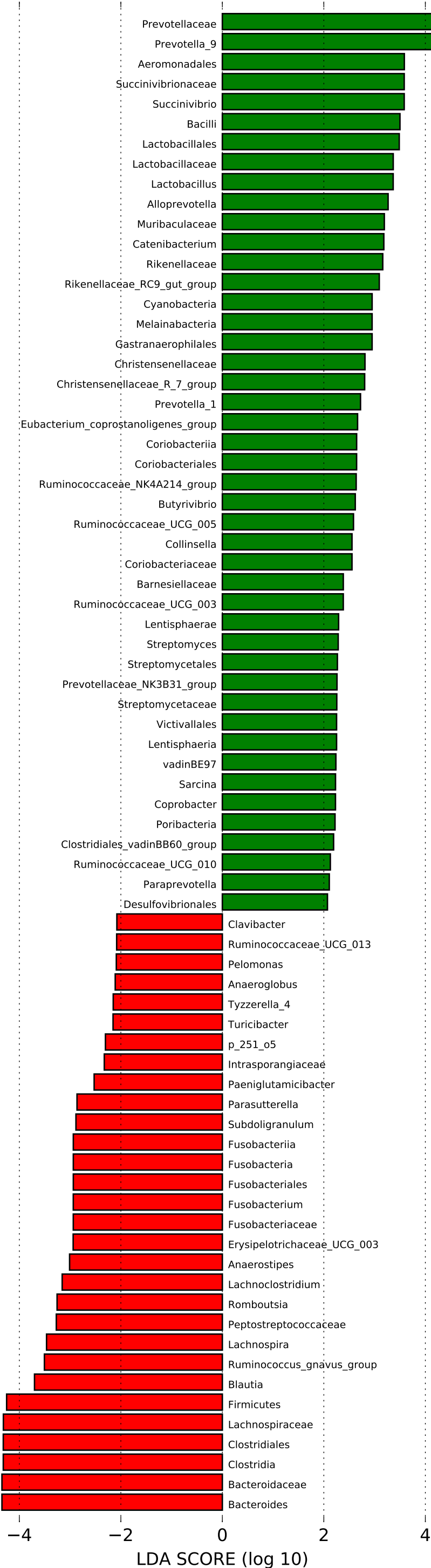


■ Han ■ Hui

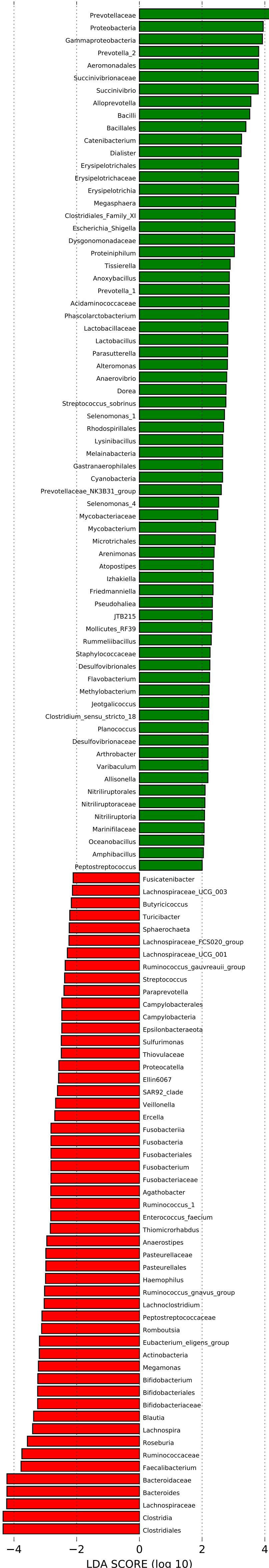




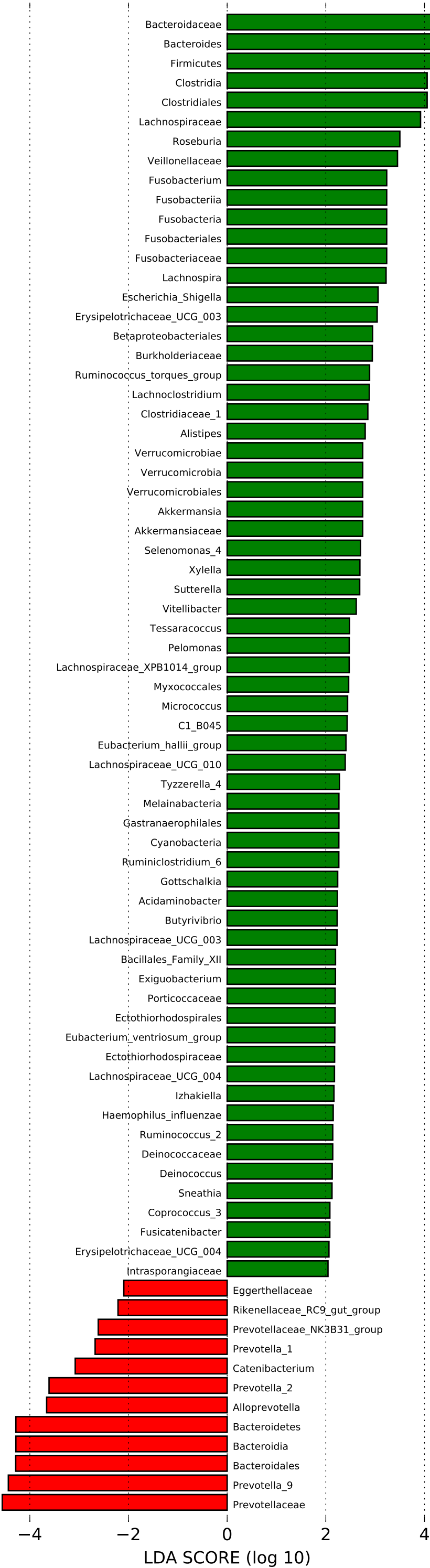
■ Han
■ Tibetan



■ Han ■ Uygur



Bai Han



LDA SCORE (log 10)

