Window 1 (top left):

```
. // 1. Load Cleaned DTA File
. use "390_data_latin1_cleaned.dta", clear

. // 2. Convert wide to long format
. reshape long Ref_, i(Microbe) j(Reference)
(note: j = 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
> 27 28 29 30 31 32 33 34 35 36 37 38 39)

Data                               wide   ->   long

Number of obs.                       15   ->   585
Number of variables                  40   ->   3
j variable (39 values)                    ->   Reference
xij variables:
                    Ref_1 Ref_2 ... Ref_39   ->   Ref_

. rename Ref_ Score

. // 3. Normalize scores within each Reference
. egen mean_score = mean(Score), by(Reference)

. egen sd_score = sd(Score), by(Reference)

. gen z_score = (Score - mean_score) / sd_score
(205 missing values generated)

. // 4. Summary statistics
. collapse (mean) z_score, by(Microbe)

. sort z_score
```

Window 2 (top right):

```
. // 5. Reload original data for visualizations
. use "390_data_latin1_cleaned.dta", clear

. reshape long Ref_, i(Microbe) j(Reference)
(note: j = 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
> 27 28 29 30 31 32 33 34 35 36 37 38 39)

Data                               wide   ->   long

Number of obs.                       15   ->   585
Number of variables                  40   ->   3
j variable (39 values)                    ->   Reference
xij variables:
                    Ref_1 Ref_2 ... Ref_39   ->   Ref_

. rename Ref_ Score

. egen mean_score = mean(Score), by(Reference)

. egen sd_score = sd(Score), by(Reference)

. gen z_score = (Score - mean_score) / sd_score
(205 missing values generated)

. // 6. Boxplot by Microbe
. graph box z_score, over(Microbe, label(angle(45))) ///
>     title("Normalized Microbe Scores") ///
>     ylabel(-2(0.5)2) scheme(s1color) ///
>     graphregion(color(white)) ///
>     ytitle("Z-score") ///
>     name(box_microbe, replace)
```

Window 3 (bottom left):

```
. // 4. Summary statistics
. collapse (mean) z_score, by(Microbe)

. sort z_score

. list

                      Microbe      z_score
  1.           P. aeruginosa    -1.748302
  2.      (Candida albicans)    -1.609224
  3.               S. aureus    -.7407857
  4.    (Ruminococcus gnavus)    -.7121019
  5.  (Enterococcus faecalis)    -.5469284

  6.         (Shigella spp.)     -.354531
  7.               B. longum    -.2536972
  8.  (Klebsiella pneumoniae)    -.2320847
  9. (Salmonella Typhimurium)     .1683169
 10.          A. muciniphila     .5013026

 11.   (Campylobacter jejuni)     .6418965
 12.                  (ETBF)      .780039
 13. (Clostridioides difficile)   .8062239
 14.                  (AIEC)     1.157871
 15.          F. prausnitzii     1.236219

. // 5. Reload original data for visualizations
. use "390_data_latin1_cleaned.dta", clear

. reshape long Ref_, i(Microbe) j(Reference)
```

Window 4 (bottom right):

```
. // 8. ANOVA
. encode Microbe, gen(microbe_id)

. anova z_score Reference
```

```
              Number of obs =        380    R-squared     =  0.0000
              Root MSE      =          1    Adj R-squared = -0.1082

      Source |  Partial SS         df         MS           F     Prob>F
      Model  |  1.819e-11          37   4.916e-13        0.00   1.0000

   Reference |  1.805e-11          37   4.880e-13        0.00   1.0000

    Residual |       342          342          1

       Total |       342          379   .90237467
```

```
. anova z_score microbe_id
```

```
              Number of obs =        380    R-squared     =  0.8825
              Root MSE      =    .331009    Adj R-squared =  0.8780

      Source |  Partial SS         df         MS           F     Prob>F
      Model  |  301.81444          14   21.558175      195.81   0.0000

  microbe_id |  301.81444          14   21.558175      195.81   0.0000

    Residual |  40.185556         365   .11009741

       Total |       342          379   .90237467
```
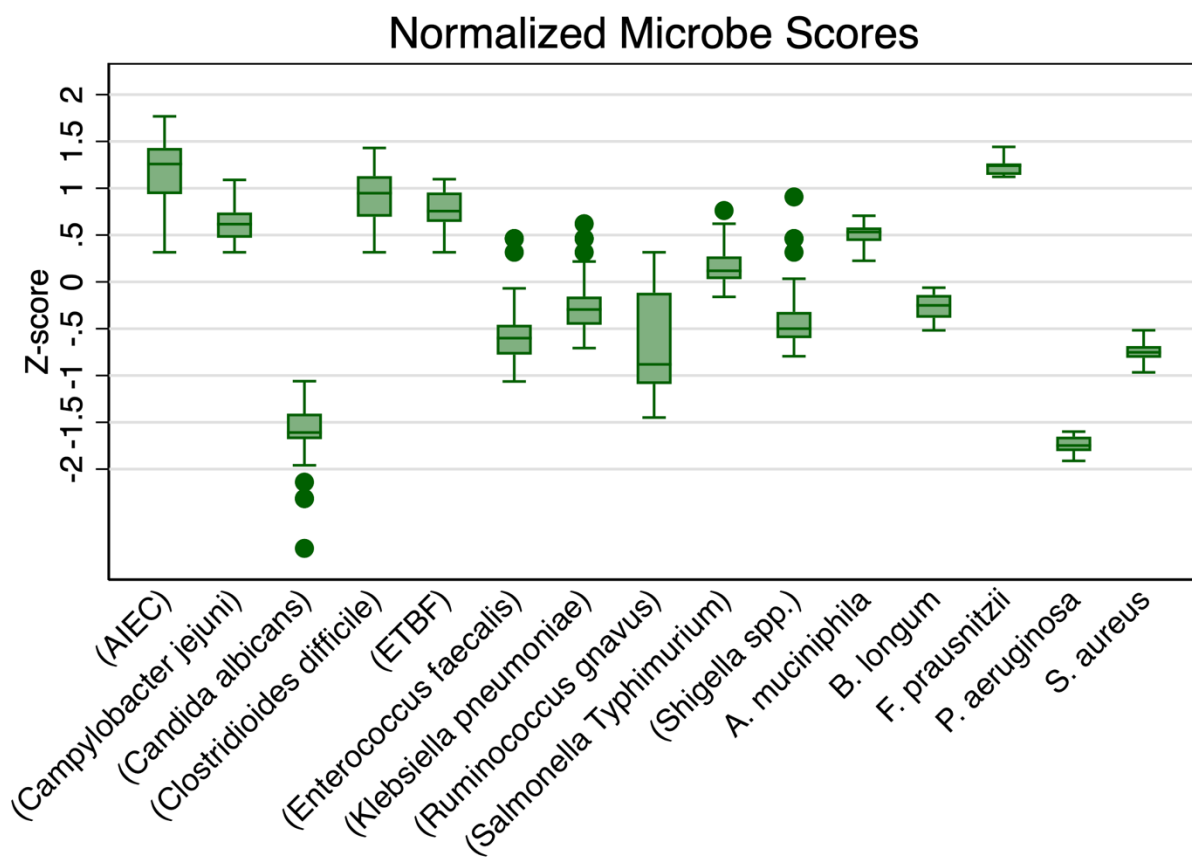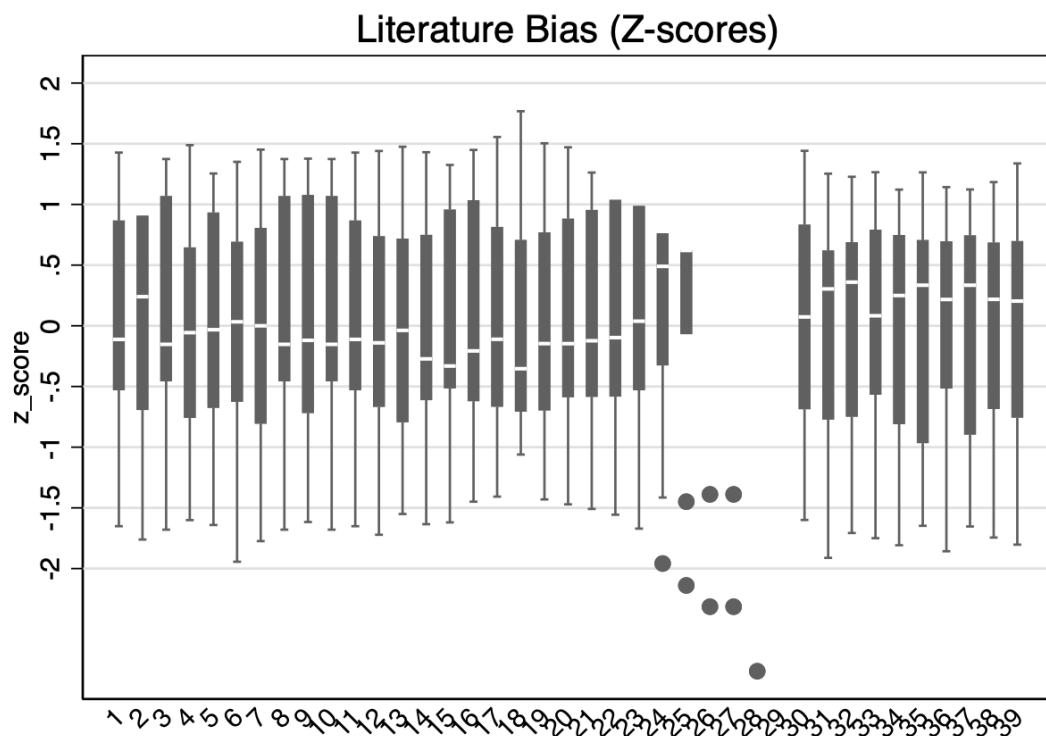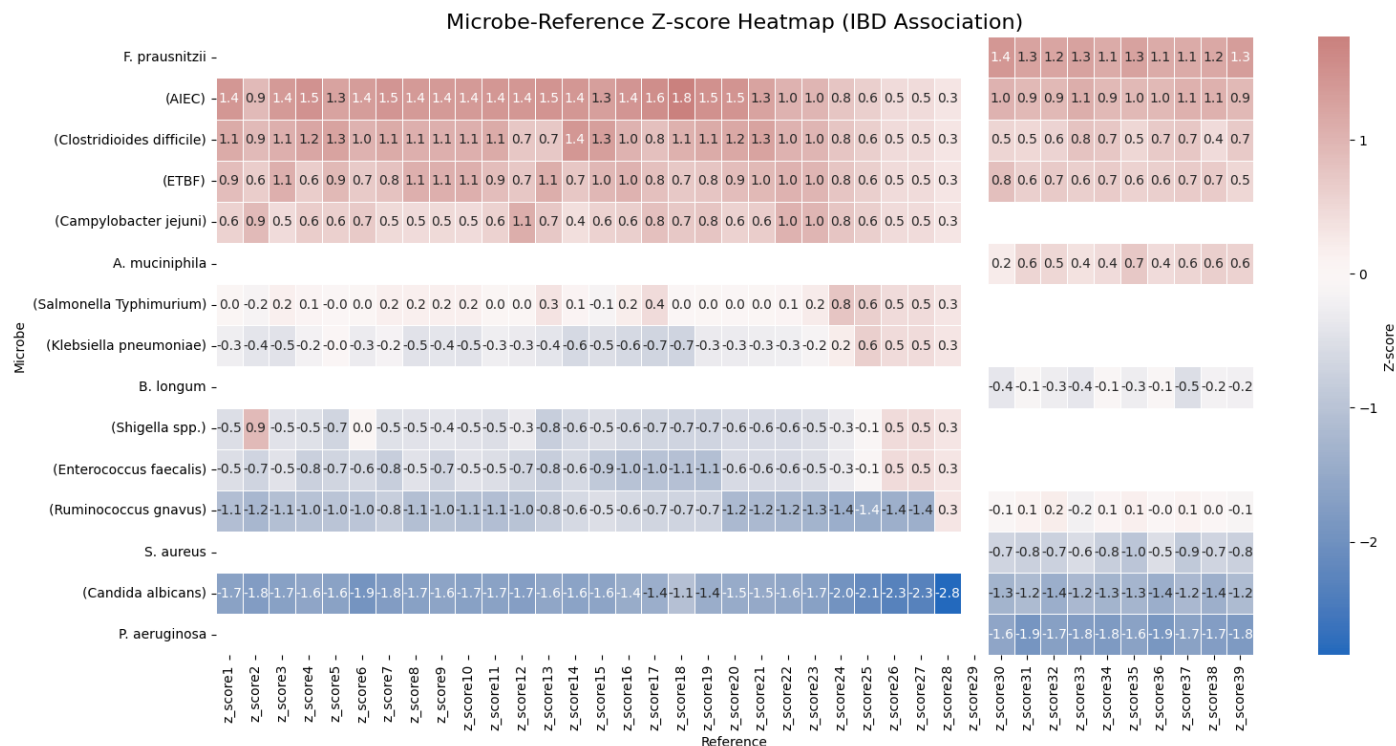
```
. // 9. Prepare wide format for clustering
```

Literature Bias (Z-scores)

Normalized Microbe Scores

Microbe-Reference Z-score Heatmap (IBD Association)

Microbe	Mean Z-score	Microbe Group
F. prausnitzii	1.236	Anti-inflammatory
(AIEC)	1.158	Pathobiont
(Clostridioides difficile)	0.886	Pathobiont
(ETBF)	0.78	Pathobiont
(Campylobacter jejuni)	0.642	Pathogen
A. muciniphila	0.501	Commensal
(Salmonella Typhimurium)	0.168	Pathogen
Klebsiella pneumoniae	-0.232	Opportunistic
B. longum	-0.254	Probiotic
(Shigella spp.)	-0.355	Pathogen
(Enterococcus faecalis)	-0.547	Opportunistic
(Ruminococcus gnavus)	-0.712	Pathobiont
S. aureus	-0.741	Opportunistic
(Candida albicans)	-1.609	Fungal Pathogen
P. aeruginosa	-1.748	Pathogen

- **F. prausnitzii**：Z = 1.236，功能分类：抗炎共生菌
- **(AIEC)**：Z = 1.158，功能分类：致病性条件菌（Pathobiont）
- **Clostridioides difficile**：Z = 0.886，功能分类：致病性条件菌

- **ETBF**：Z = 0.780，功能分类：致病性条件菌
- **Campylobacter jejuni**：Z = 0.642，功能分类：病原体
- **Akkermansia muciniphila**：Z = 0.501，功能分类：共生菌
- **Salmonella Typhimurium**：Z = 0.168，功能分类：病原体
- **Klebsiella pneumoniae**：Z = -0.232，功能分类：机会致病菌
- **Bifidobacterium longum**：Z = -0.254，功能分类：益生菌
- **Shigella spp.**：Z = -0.355，功能分类：病原体
- **Enterococcus faecalis**：Z = -0.547，功能分类：机会致病菌
- **Ruminococcus gnavus**：Z = -0.712，功能分类：致病性条件菌
- **Staphylococcus aureus**：Z = -0.741，功能分类：机会致病菌
- **Candida albicans**：Z = -1.609，功能分类：真菌病原体
- **Pseudomonas aeruginosa**：Z = -1.748，功能分类：病原体

Analysis	Details
ANOVA: Microbe effect	$F(14, 365) = 195.81$, $p < 0.0001$, R¬≤ = 0.88
ANOVA: Reference effect	$F(37, 342) = 0.00$, $p = 1.0000$, not significant
Boxplot export	Saved to 'stata_figures/box_microbe_clear.png' and 'box_reference.png'
Top microbe (Z-score)	F. prausnitzii (Z = 1.236)——抗炎共生菌？
Bottom microbe (Z-score)	P. aeruginosa (Z = -1.748)
Z-score mean range	-1.748 to 1.236

Literature × Microbe Z-score Matrix

Microbe Z-score Radar Plot
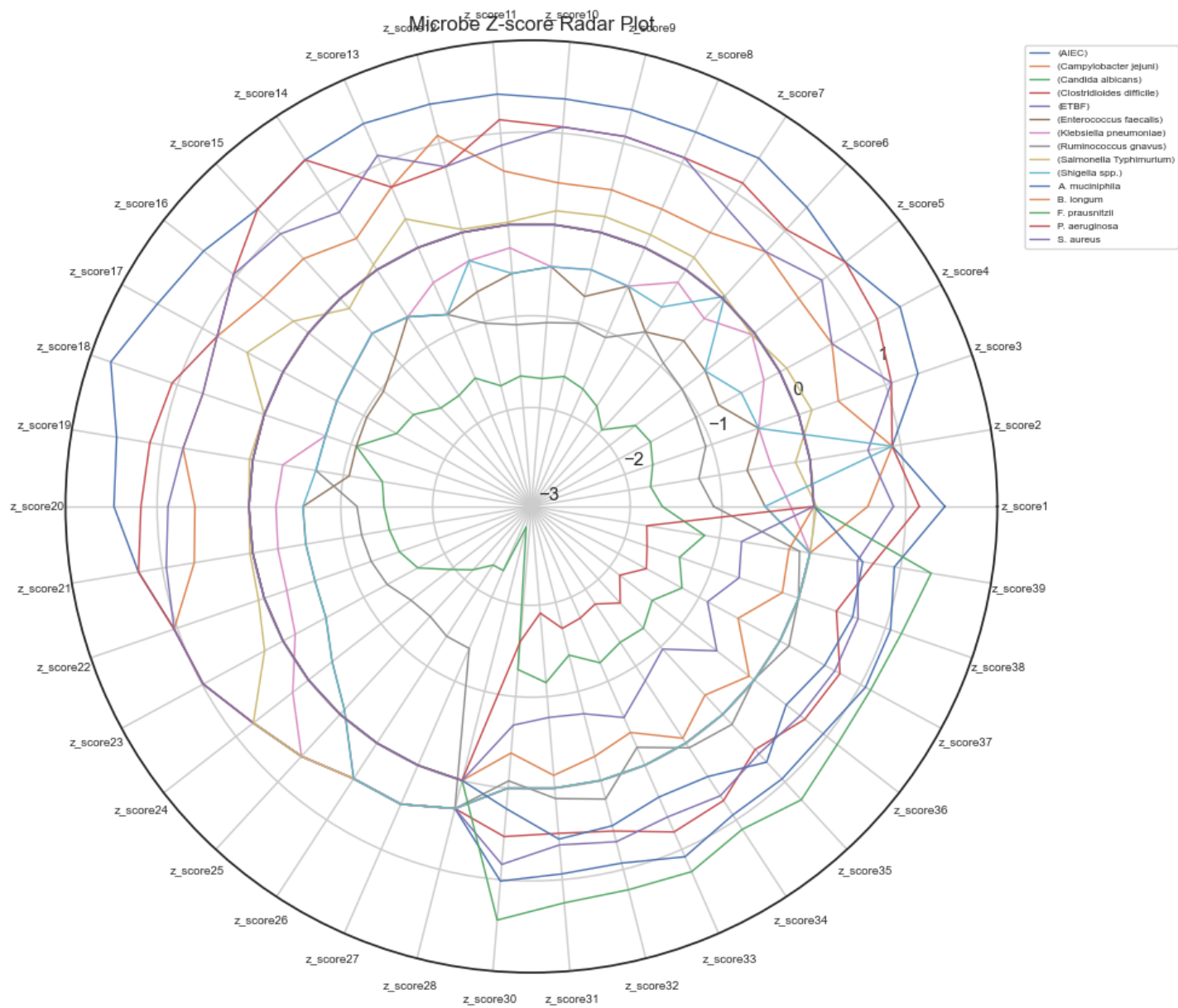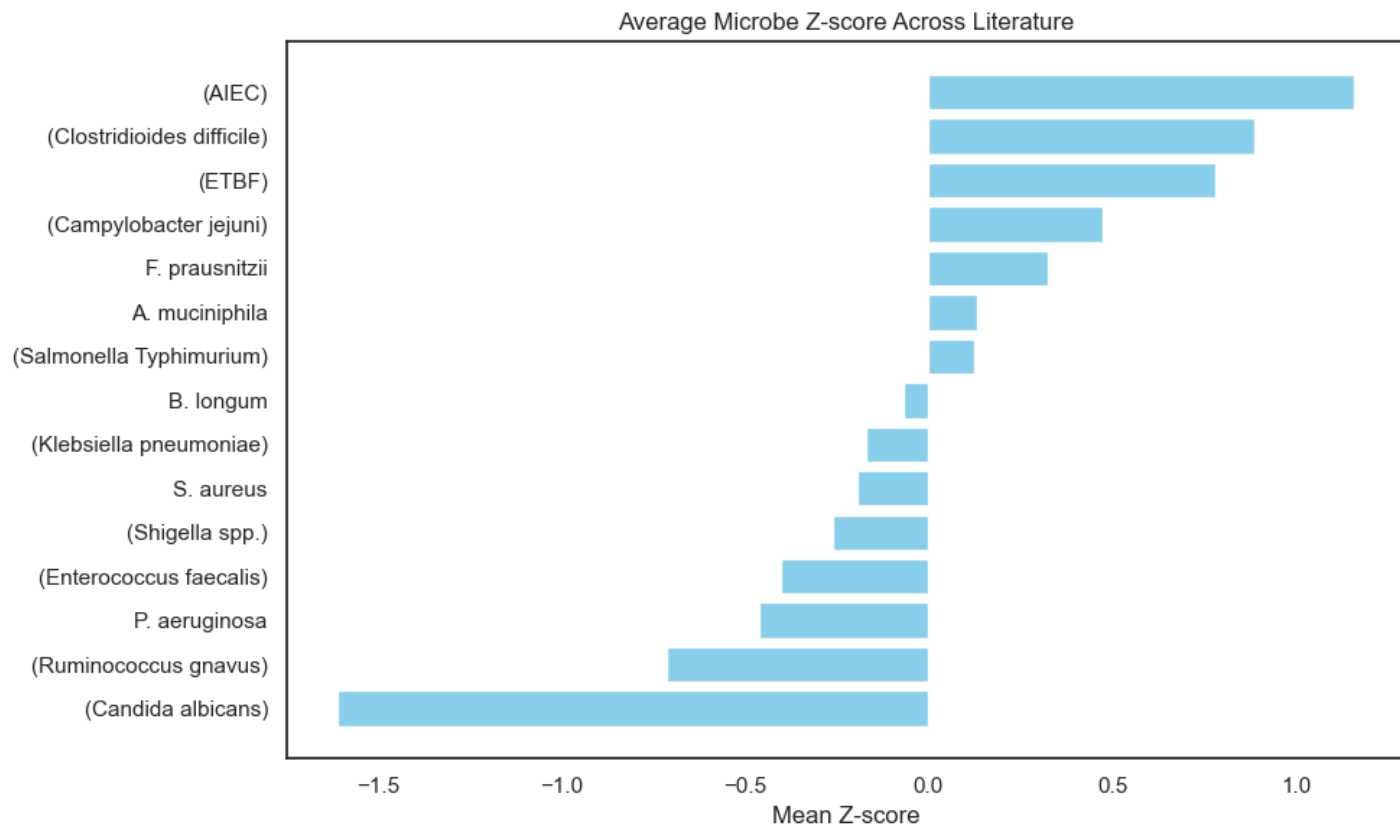
Legend:
- (AIEC)
- (Campylobacter jejuni)
- (Candida albicans)
- (Clostridioides difficile)
- (ETBF)
- (Enterococcus faecalis)
- (Klebsiella pneumoniae)
- (Ruminococcus gnavus)
- (Salmonella Typhimurium)
- (Shigella spp.)
- A. muciniphila
- B. longum
- F. prausnitzii
- P. aeruginosa
- S. aureus

Average Microbe Z-score Across Literature
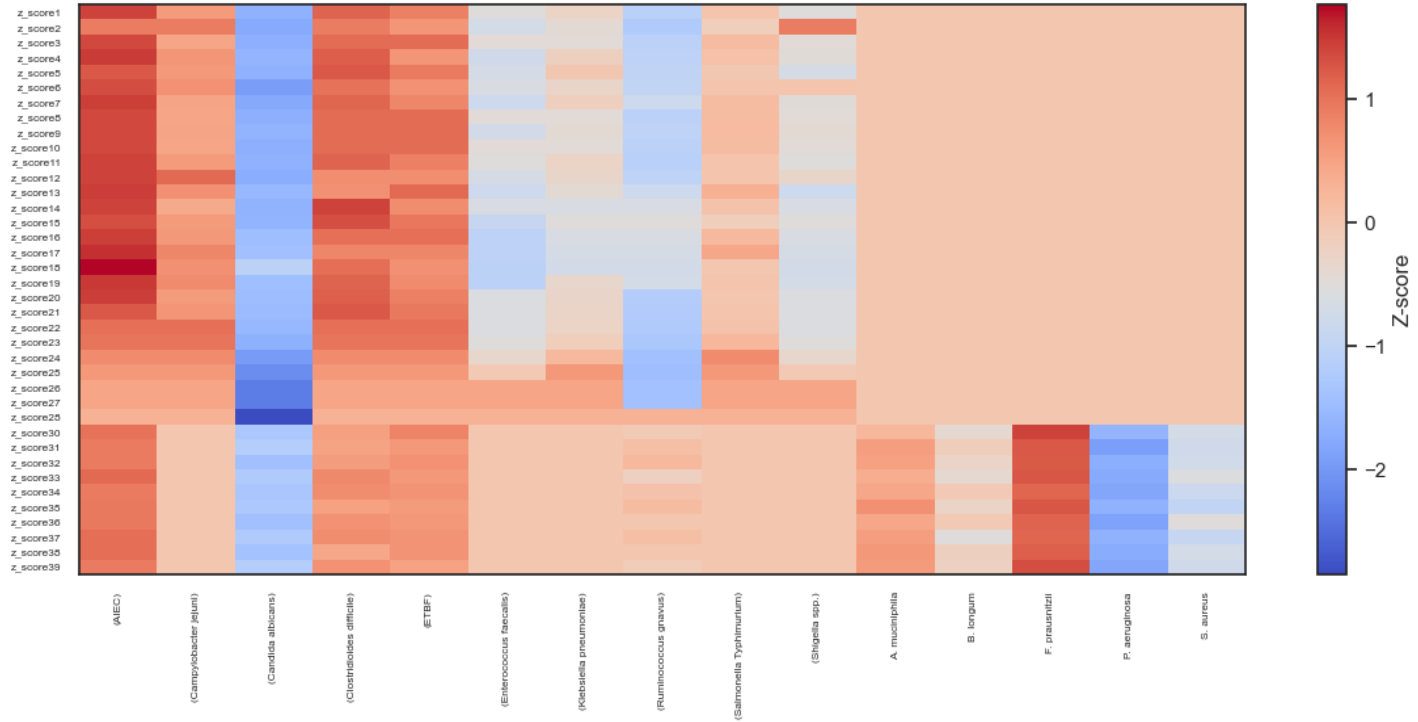
PCA of Literature-Derived Microbiome–IBD Relevance Scores