



לפי ההסברים של ה-authors של החבילה הזו ב-R – זה מחשב p-value לכל cluster. יש 2 סוגים של pvalues – AU ו-BP. שניהם משתמשים ב-Bootstrapping. BP נחשב ל- biased – ו-AU נחשב ל-unbiased, כך שבסופו של דבר צריך להתייחס ל-AU כי הוא יותר מדויק.

בשביל להבין איך ה-bootstrapping עובד – עושים resampling ( במקרה שלך עשיתי 10000 איטרציות)וכל פעם עושים clustering. BP עבור cluster מסוים (מופיע **מעל** ה-cluster) מציין את אחוז האיטרציות שבהם כל הדגימות שמופיעות ב-cluster הופיע ביחד מתוך כל האיטרציות. החישוב של AU יותר מורכב ולא לגמרי הבנתי, אבל הרעיון דומה.

במקרה שלך – ב-2 הגרפים רואים שיש אשכול עיקרי אחד, ו-DH44. ה-AU באשכול העיקרי הוא 100, כלומר בכל האיטרציות כל הדגימות הופיעו ביחד באותו אשכול.

Methods:

Hierarchical clustering dendrogram with p-values was done  using the R package *pvclust* [REF - <https://academic.oup.com/bioinformatics/article/22/12/1540/207339> ], with multiscale bootstrap resampling of 10,000 iterations to assess statistical significance, represented by a 1–100 score (Supp. Fig X). Hierarchical clustering was performed using average linkage method with Euclidian distance as the distance measure.

Legend:

Hierarchical clustering dendrogram with p-values. Values on the edges of the clustering are p-values (%). Red values are AU (Approximately Unbiased) p-values, and green values are BP (Bootstrap Probability) values. Clusters with AU > = 95% are indicated by the rectangles and are considered to be strongly supported by data.