**פרוטוקול עיבוד 16S – בהדגמה על נתוני ניסוי כימותרפיה ID: 351316**חומר צפייה מומלץ: https://www.youtube.com/watch?v=cEbYCzTzQr8

פרוטוקול עבודה מוצלח: 

מיקום הקבצים ב-Dropbox:   
sequencing files/chemotherapy/Raw/16S\_Chemo\_351316

**כניסה לשרת:**שם השרת: matrix.lnx.biu.ac.il  
שם משתמש: nissan  
סיסמא: Nano2022@  
ישר כאשר השרת עולה:  
**bash**

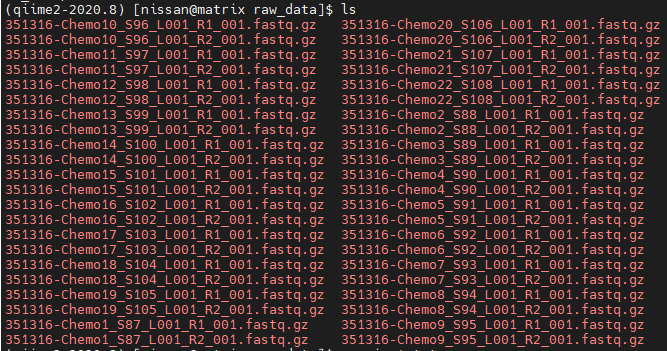
**source /home/private/software/packages/miniconda2/bin/activate qiime2-2020.8**

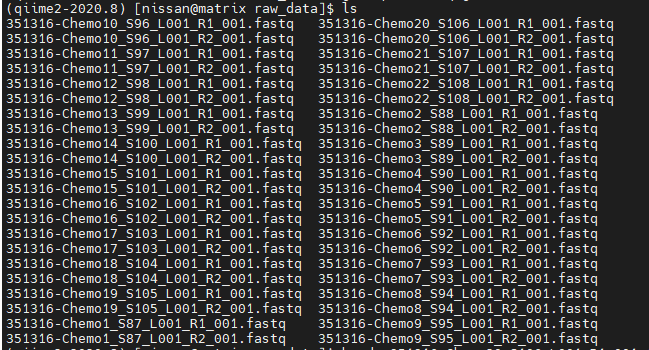
(הדומיין משתנה מ-nissan ל-qiime2-2020.8). מדובר בעצם בכלי qiime2, המותאם לעבודה עם ריצוף מיקרוביאלי.

\*הערה: אע"פ שכתוב כאן שהגרסה היא qiime2-2020.8, הגרסה יכולה להשתנות ולהיות עדכנית יותר, אבל בשרת היא תהיה שמורה עם אותו השם ע"מ שהפקודה תישאר זהה.

הקדמה כללית:

האזור שמרוצף במעבדה הוא בד"כ 4V, כי הוא האזור הקבוע היחיד שמסוגל לטרגט חיידקים מסוימים שאנו רוצים בהם (כמו b. adolescentis). אזור זה הוא באורך של כ-254 נוקלאוטידים, ואת הרצף הזה אנחנו מרצפים ב"חתיכות" של 250 נוקלאוטידים בכל ריד.

**שלב 1 – העברת הקבצים לשרת Matrix ולתצורת qza**קבצי ה-raw data המגיעים מהריצוף הינם בתצורת fastq.gz.  
ככלל, הקבצים יגובו קודם כל בdropbox, משם נוכל לטעון אותם בכל עבודה חדשה.  
1. נכנס לתיקיית העבודה הרצויה בשרת, נפתח תיקיית raw\_data (**mkdir raw\_data**) ולתוכה נעביר את הקבצים הדחוסים (גרירה מתיקייה ב-dropbox השולחני אל תוך תיקיית raw\_data)  
2. נוודא כי הקבצים עברו:  


3. כעת, נבצע פתיחה של הדחיסה עם הפקודה: **gunzip \*.fastq.gz, ונוודא לאחר סיומה כי כל הקבצים נפתחו:**

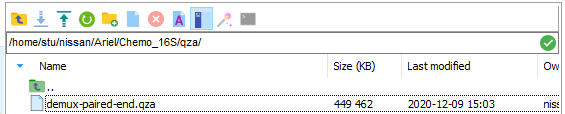
4. כתיבת קובץ manifest.tsv: אשר יגדיר לqiime את המיקום של כל אחד מהקבצים שלנו בשרת. כותבים אותו ידנית ומכניסים לתיקייה אחת "לפני" תיקיית הraw\_data:  


מיקום הקובץ: /home/stu/nissan/Ariel/Chemo\_16S/

5. העברת הקבצים לפורמט העבודה של qiime2 🡨 qza.  
נפתח תיקיית qza ע"י **mkdir qza**, תיקייה זו תהווה תיקיית יעד לקבצי הfastq לאחר שיעברו המרה לקבצי qza.  
מה שקורה זה בעצם נתינת הקובץ manifest.tsv כקובץ input, ממנו קורא ה-qiime לאיזה קבצים עליו לעשות את ההמרה ומאיפה לקחת אותם. הoutput שהגדרנו הוא בעצם בתוך תיקיית qza שזה עתה פתחנו, ובתוכה – תיכתב ההמרה בקובץ "demux-paired-end.qza".  
הפקודה:  
**qiime tools import --type 'SampleData[PairedEndSequencesWithQuality]' --input-path manifest.tsv --input-format PairedEndFastqManifestPhred33V2 --output-path qza/demux-paired-end.qza**

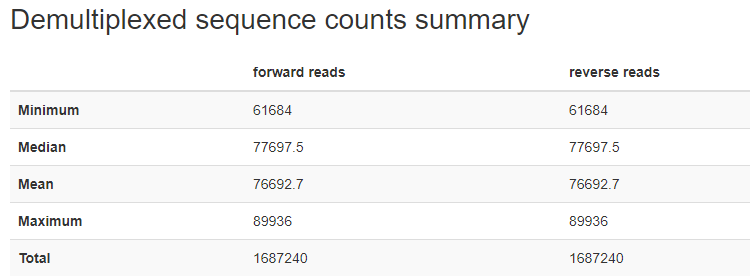
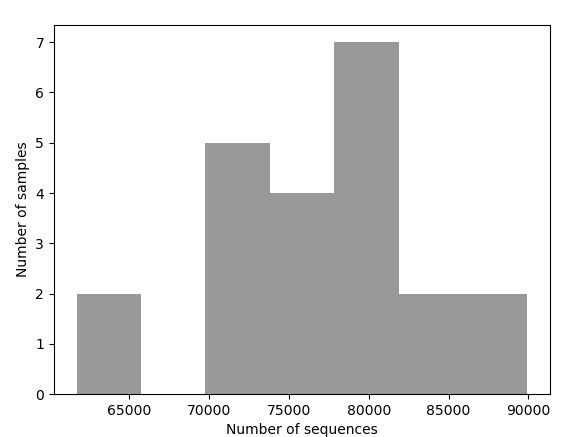
התהליך לוקח כ-15 דקות.

בסופו מקבלים הודעת אישור:   
Imported manifest.tsv as PairedEndFastqManifestPhred33V2 to qza/demux-paired-end.qza  

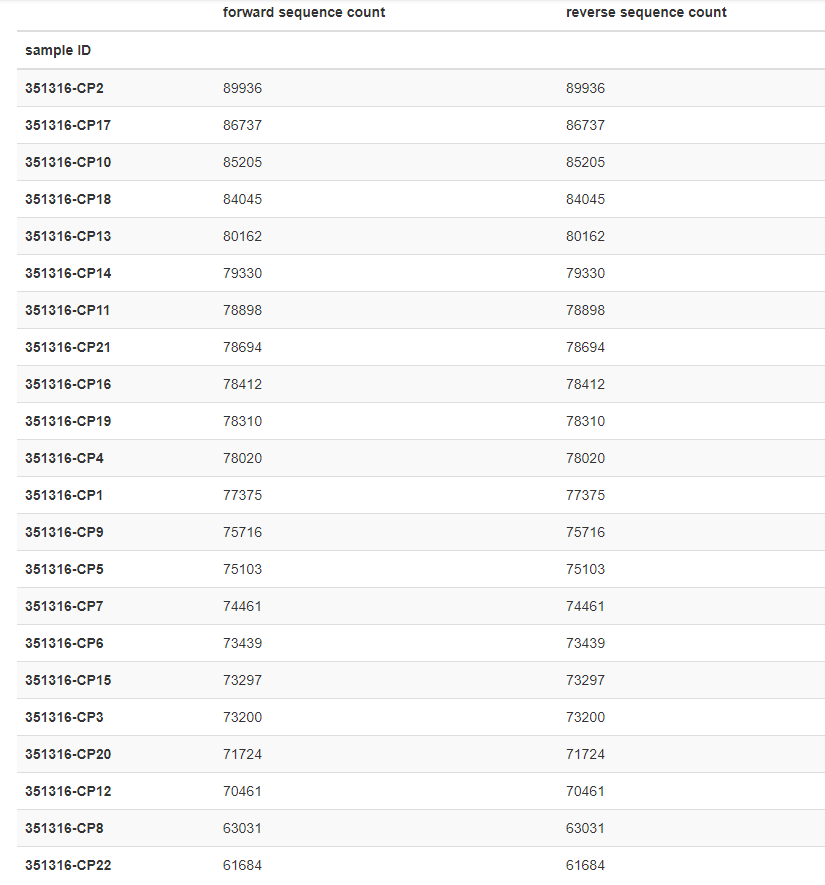

וניתן לראות שהקובץ אכן נוצר:  


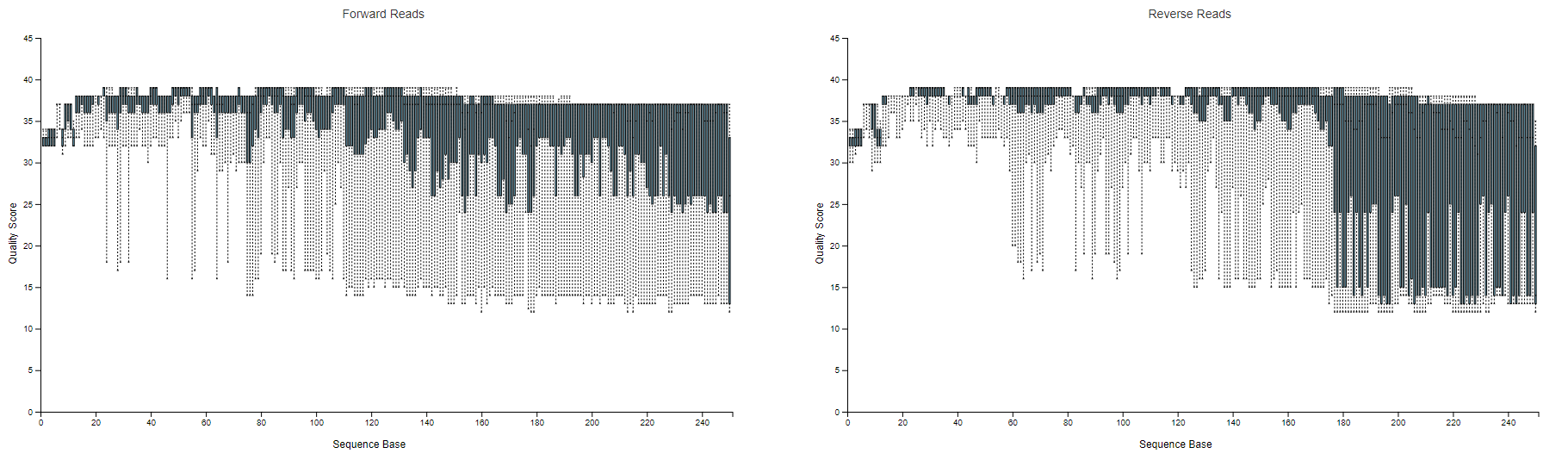
6. המרת קובץ qza לקובץ qzv  
פותחים תיקיית vis 🡨 **mkdir vis**  
כדי שנוכל לצפות בתוכן הקובץ (qza לא נגיש לצפייה), פותחים תיקיית vis ונותנים את הפקודה:  
**qiime demux summarize --i-data qza/demux-paired-end.qza --o-visualization vis/demux-paired-end.qzv**מהות הפקודה: --i-data מסמן את קובץ ה-input, הפעולה היא visualization, המיקום בסיום הפקודה הוא המיקום ושם הקובץ שנרצה לקבל (סיומת qzv).  
מיד לאחר מכן, מתקבלת הודעת אישור על שמירת הוויזואליזציה:  
 Saved Visualization to: vis/demux-paired-end.qzv

7. צפייה בנתונים – נכנסים לתיקייה vis על השרת, שומרים את הקובץ שקיבלנו (סיומת .qzv) במיקום נגיש על המחשב שלנו או על ה-dropbox. מעלים אותו אל הכתובת:  
<https://view.qiime2.org/> בדפדפן.

**שלב 2 – בדיקת איכות הרצפים והכנה לניקוי רעשים**בשלב זה "נפענח" את התוצאות שקיבלנו בדו"ח הוויזואלי של qiime2, אותו פתחנו בסעיף 7, שלב 1.  
1. נסתכל בלשונית overview 🡨 על הטבלה שהתקבלה תחת Demultiplexed sequence count summary:  
  
במקרה זה, אנחנו רואים כי סך כל הקריאות שלנו (reads) היה 1,687,240. אנו רואים כי המספר המינימלי של קריאות פר דגימה היה 61,684, ואילו המספר המקסימלי של קריאות פר דגימה היה 89,936. ערכי ה-Median, Maximum הינם ערכים אשר מחושבים באופן אוטומטי: ממוצע וחציון.  
את אותם ערכים בדיוק ניתן לראות באופן ויזואלי גם בגרף:  
  
משמאל לימין: אנו רואים כי 2 דגימות היו עם בערך 63,000 קריאות, 5 דגימות היו עם כ-72,500 קריאות, 4 דגימות היו עם כ-75,000 קריאות, 7 דגימות היו עם כ-80,000 דגימות, 2 דגימות היו עם כ-85,000 קריאות ו-2 דגימות עם כ-87,000 דגימות.

נמשיך לגלול את העמוד כלפי מטה, ונגיע אל ה-Per-sample sequence counts:  
מדובר בעצם בפירוט של כל הנתונים שראינו מסוכמים קודם לכן – כמה בדיוק קריאות היו בכל דגימה.

  
בשלב זה, אם נראה דגימה שהיו בה משמעותית פחות קריאות מאשר לכל האחרות (למשל אם היינו מקבלים פה דגימה עם 5,000 קריאות – זה מעט מאוד ביחס לכל השאר), נבין כי יתכן שהיא עברה זיהום או שהעיבוד שלה לא היה תקין, ונסיר אותה.  
\* **במידה ומסירים דגימה מסוימת**: יש לחזור לשלב 4 בסעיף הקודם, להסיר את הדגימה מקובץ ה-manifest.tsv, ליצור שוב אובייקט qzv המבוסס על ה-manifest.tsv המעודכן וכן הלאה.

2. נעבור ללשונית Interactive Quality Plot:  
ראשית כל, נראה איך נראים הריצופים שלנו באופן כללי – כאן למשל ניתן לראות כי ריצוף ה-forward היה יחסית יציב והתחיל "לזייף" בהדרגה (תהליך טבעי שקורה ככל שהריצוף מתקדם), ואילו גדיל הרוורס התחיל "לזייף" באופן חד יותר ומשמעותי יותר, החל ממיקום מסוים.  
  
נרצה לשמור על בסיסים כמה שיותר איכותיים, אולם לא "לקצור" יותר מידי מידע על הדרך.

נתקרב באמצעות הממשק האינטראקטיבי אל הקצוות ונבחן איפה נרצה לחתוך משמאל ומימין, כל אחד מהגדילים.  
תזכורת:  
הנוסחה היא -10log(p) = score. כאשר אנו שואפים לp נמוך ככל הניתן.  
לכן, אם ה-score הוא בערך 30, זה אומר שp = 0.001 וזה מצוין. נרצה להישאר תמיד מעל לציון 20, כי שם הטעות היא כבר p=0.01, וטעות אחת ל100 נוקליאוטידים זה די הרבה.

בכל מקרה, צריך לכל הפחות חפיפה של 40 נוקליאוטידים בין גדיל הreverse לגדיל הforward כדי להצליח בשלבים הבאים.

במקרה הזה – אני החלטתי להשאיר בגדיל הforward את הריצוף במקומות 23-249,  
ואילו בגדיל הReverse – להשאיר את מקומות 23-177.   
הסיבה בגללה החלטתי לקצור גם את ההתחלה היא שקיים שם אזור בו הרצף אחיד לחלוטין, כלומר כנראה החלק שם הוא עוד החלק הקבוע, ואותו נוריד. בד"כ נתחיל את הקצירה עד לנוקלאוטיד 23.

איך נעשה את החישוב:

בגדיל הforward נכתוב את הבסיס הראשון ואת הבסיס האחרון שאנחנו רוצים.  
בגדיל הרוורס נכתוב את אורך הגדיל (בדרך כלל 250 בסיסים) פחות הבסיס האחרון והבסיס הראשון, וזה ייתן לנו את הטווח שלהם לפי הגדיל הקדמי.  
ניצור סקאלה של שני הגדילים ונבדוק מה גודל החפיפה ביניהם, וכן מה אורך הגדיל שנישאר איתו בסוף לאחר החיתוך והאיחוד.

לדוגמה:

גדיל רוורס: מיקום התחלה: 73=250-177, מיקום סוף: 227=250-23, ולכן:  
גדיל רוורס: 227 |--------------------------------| 73  
גדיל forward: 249 |-----------------------------------------------------| 23   
  
כך נראה שגודל החפיפה בין הגדילים הוא 154bp, וגודל הגדיל לאחר החיתוכים+איחוד הגדילים הוא 226.

**שלב 3 – ביצוע החיתוך – Trimming**כלי עבודה: dada2התהליך לוקח כ-3 שעות  
הפקודה: **qiime dada2 denoise-paired --i-demultiplexed-seqs qza/demux-paired-end.qza --p-trim-left-f 23 --p-trim-left-r 23 --p-trunc-len-f 249 --p-trunc-len-r 177 --o-table qza/dada2\_table.qza --p-chimera-method consensus --verbose --o-representative-sequences qza/dada2\_rep-seqs.qza --o-denoising-stats qza/dada2\_denoising-stats.qza > qza/denoise.log**

הסבר: dada2 הינו כלי לביצוע trimming במיקומים אותם הגדרנו.   
אנו מבקשים חיתוך עבור ריצוף paired שנמצא בקובץ הqza שיצרנו בשלב 1, סעיף 5, כלומר המיקום של קובץ ה-input הוא qza/demux-paired-end.qza.  
מסומנים בירוק – החיתוכים במיקומים המבוקשים. left זה משמאל, trunc-len זה מימין, והסימונים f או r הם forward או reverse בהתאמה.  
בנוסף לחיתוך, dada2 גם מסיר לנו רצפי כימרה, וכן הוא מסיר מעט "רעש"- אם יש רצף מסוים שחוזר על עצמו פעמים רבות הוא מוגדר קונצנזוס, ואם קיים רצף יחיד ששונה ממנו בנוקלאוטיד בודד- הוא מוגדר טעות בריצוף ומתווסף לספירה של רצף הקונזצנוס.  
מסומנים בכחול – הפלטים שנקבל **(עליהם נעבור בהמשך).**

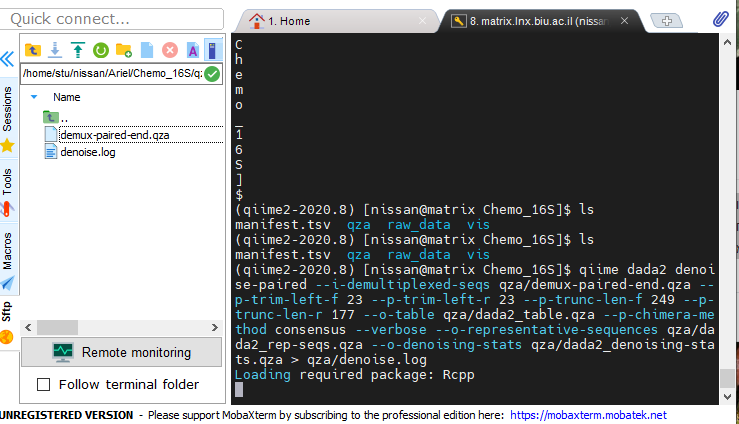
**הסבר לפלטים שקיבלנו:**

dada2\_table.qza: קובץ של FeatureTable[Frequency] שמכיל countx (תדירויות) של כל רצף ייחודי בכל דגימה בdataset.

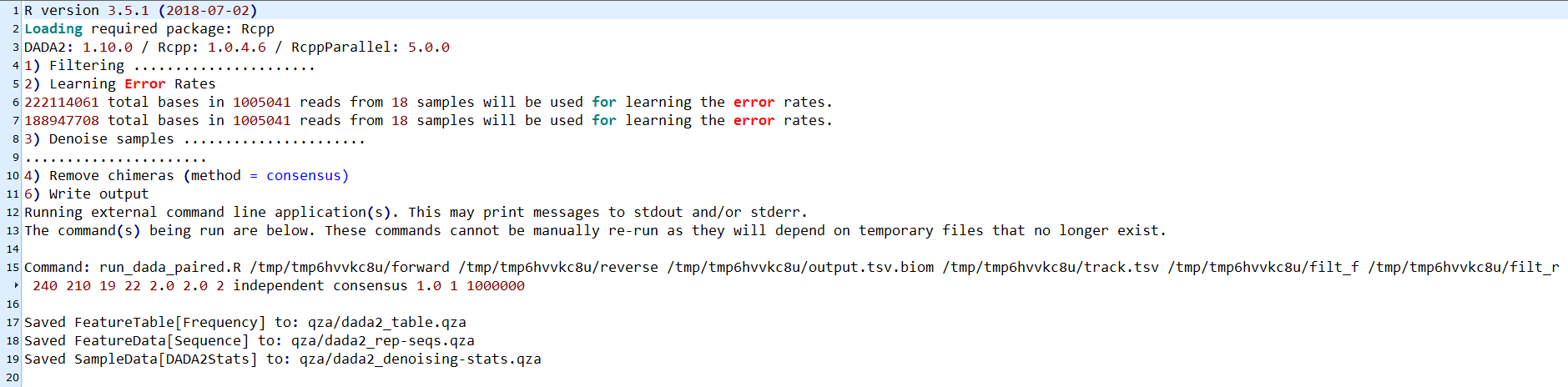
dada2\_rep-seqs.qza: קובץ של FeatureData[Sequence] (רצפים מייצגים) שממפה מזהים לכל פיצ'ר מהקובץ העליון לרצף שהם מייצגים.

- בשורה התחתונה, קובץ של הרצפים המייצגים וקובץ שמכיל את מספר הפעמים שכל רצף כזה מופיע.

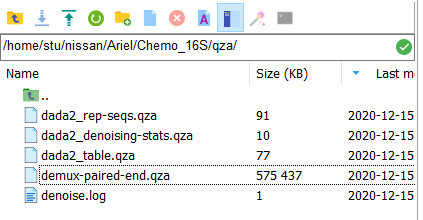
ביצוע denoising עם dada2 נותן לנו כפלט בעצם טבלת ASV- amplicon sequence variants.

קובץ ה-log יופיע בתיקייה ובאמצעותו נוכל לעקוב אחר התהליך.

**\*\* --p-chimera-method consensus –verbose 🡨 כימרות הן תוצר של ריצוף דנ"א לא ספציפי, משהו שנוצר בתהליך ה-PCR. נרצה להסיר מקטעים גנטיים אלה.**  
\* במידה ומרצפים צד אחד ולא paired – הפקודה תהיה qiime dada2 denoise-single ואז את ההמשך.

לאחר כ-3 שעות, נבדוק אם קיבלנו את 3 הפלטים הרצויים, ונבדוק שקובץ הלוג נראה תקין:  


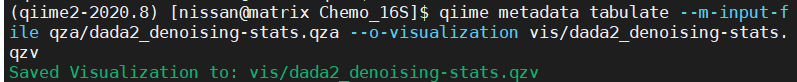
\*אם לא התקבל פלט, כנראה שעשינו trimming אגרסיבי מידי, ולא היו מספיק התאמות של חפיפה בין הקצוות. נצטרך לחזור לשלב הקודם ולחשוב מחדש על ביצוע הtrimming.

מצב התיקייה בסיום שלב זה:  


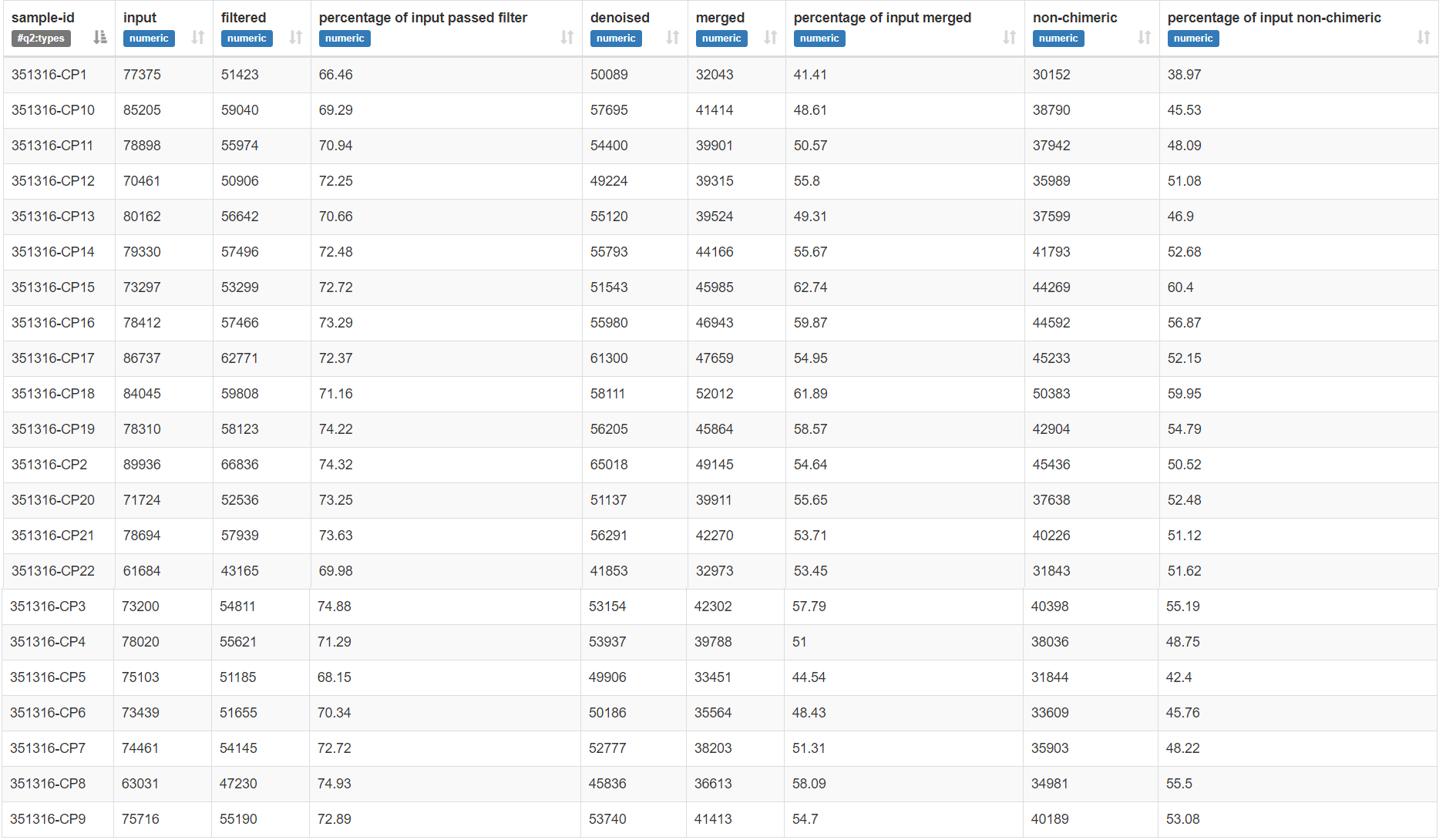
**שלב 4 – ניתוח תוצאות החיתוך בקובץ denoising\_stats**

נצא לתיקייה אחת "מעל" תיקיית qza, ונכתוב את הפקודה:  
**qiime metadata tabulate --m-input-file qza/dada2\_denoising-stats.qza --o-visualization vis/dada2\_denoising-stats.qzv**

נקבל הודעות אישור כי הוויזואליזציה נשמרה בקובץ נוסף, בסיומת **qzv**.

צפייה בנתונים – נכנסים לתיקייה vis על השרת, שומרים את הקובץ שקיבלנו (סיומת .qzv) במיקום נגיש על המחשב שלנו או על ה-dropbox. מעלים אותו אל הכתובת:  
<https://view.qiime2.org/> בדפדפן.

נבחן את התוצאות:



**שדה input** – כמה reads נכנסו לפילטור.  **שדה filtered** – כמה reads עברו את הפילטור בהצלחה.   
**Percentage of input passed filter** – בעצם חלוקה של filtered/input, כלומר כמה אחוזים עברו את הסינון.  
**denoised** – אמור להיות ערך קרוב יחסית לfiltered. במצב אידיאלי – יהיה בהם את אותו המספר, אבל בריצופים גדולים, נכנסים ASV (Amplicon Sequence Variant) בתוך הfiltered, ואח"כ הם יורדים בשלב הdenoised. כל עוד אנחנו באותו סדר גודל והמספרים קרובים זה לזה – הכול טוב, זה פשוט אומר שאחרי הפילטור, ירדו עוד כמה רצפים שלא היו שייכים לדגימות.  
**merged** – כמה מתוך הreads הצליחו להתמזג יחד כ-paired read.  
**percentage of input merged** – חלוקה של merged/input, כלומר כמה מתוך סך כל הInput, עברו את השלב הזה בהצלחה. בד"כ יהיה סביב ה-50%.  
**non-chimeric** – reads שלא שייכים לכימרה 🡨 כאן יש בד"כ "איבוד" מידע די גדול, כי הרבה מהריצוף הוא של כימרות. כלומר, זה לא מידע על חיידקים שנאבד לנו, אלא מידע שחשבנו שהוא של חיידקים – אולם הוא של כימרות.  
**percentage of input non-chimeric** – חלוקה של non-chimeric/input, כלומר כמה מתוך סך כל הInput, עברו את השלב הזה בהצלחה והם לא שייכים לכימרות.

\*\*\*\*\*\*\*\*

תוספת של תמר: לא בטוחה שהחלק הבא נחוץ בגלל שנעשה סינון דה-נובו.

ייתכן שעדיף שנעשה את הסינון הזה ידנית ונעבור ישר לשלב הקלסיפיקציה (ואז מדלגים על החלק של OTU).

**סינון דגימות בעל מספר רידים קטן:**

אם יש דגימות שמספר הרידים הסופי שלהן קטן במיוחד, נסנן אותן מיתר הדגימות שלנו:

בדוגמה הזאת נסנן דגימות עם מספר רידים שקטן מ-14,000. (יש ליצור קודם את התיקייה לרידים לאחר פילטור- filtered-sequences).

qiime feature-table filter-samples --i-table qza/dada2\_table.qza --p-min-frequency 14000 --o-filtered-table filtered-sequences/filtered-table.qza

**סינון רצפים בעל מספר חזרות קטן:**

נרצה לסנן רצפים שכמעט ולא חוזרים על עצמם, כמו סינגלטונים:

qiime feature-table filter-features --i-table filtered-sequences/filtered-table.qza --p-min-frequency 10 --o-filtered-table filtered-sequences/feature-frequency-filtered-table.qza

בדוגמה לעיל סיננו רצפים שחוזרים על עצמם פחות מ-10 פעמים.

אם נרצה לעבור מפה מיד לקלסיפיקציה:

qiime feature-classifier **classify-sklearn** --i-reads filtered-sequences/feature-frequency-filtered-table.qza --i-classifier **~/Documents/gg-13-8-99-nb-classifier.qza** --o-classification **qza/sk-classified.qza** --verbose

אם לא, נמשיך פשוט ל-OTU.

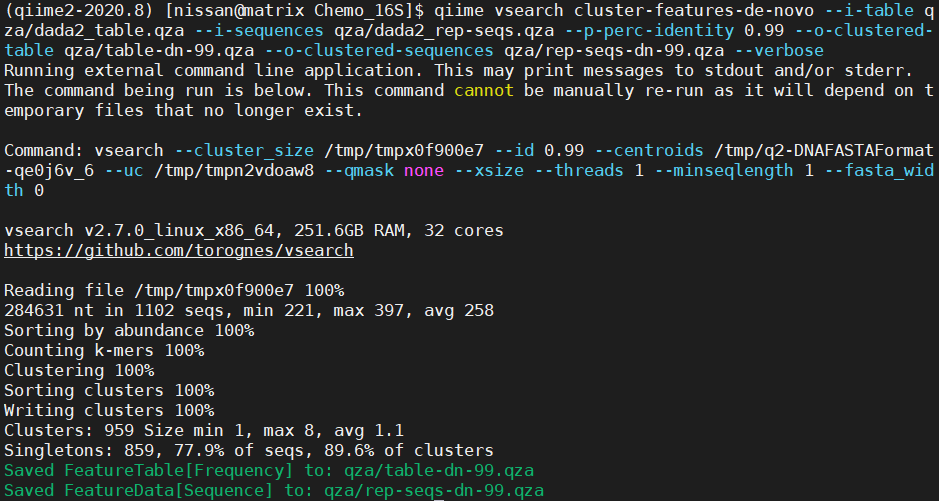
\*\*\*\*\*\*\*\*\*

**שלב 5 –OTU Clustering**

בשלב הזה של האנליזה יש לנו טבלת ASV, שהיא תוצר של פקודת הdenoise שעשינו לפני כן. בטבלה זו משתמשים ברוב אנליזות ה-S16 שעושים כיום, במקום בטבלת OTU. בד"כ אפשר לוותר על מעבר ל-OTU כי הוא יכול לאבד לנו מידע בדרך, ונלך ישר לשלב קבלת הזיהוי הטקסונומי. נעשה המרה ל-OTU בכל זאת אם נרצה את היכולת לחבר בין הפיצ'רים לבין מקטעי הריצוף עצמם, למקרה שנרצה להריץ רצף ב-Blast או משהו דומה.

פקודה: **qiime vsearch cluster-features-de-novo --i-table qza/dada2\_table.qza --i-sequences qza/dada2\_rep-seqs.qza --p-perc-identity 0.99 --o-clustered-table qza/table-dn-99.qza --o-clustered-sequences qza/rep-seqs-dn-99.qza --verbose**

כאן בעצם הגדרנו לבצע קליסטור פילוגנטי לפי התאמה של 99% לפחות (הכנה לעץ פילוגנטי).

לקרוא על קליסטור דה-נובו של OTU. בגדול, מצאנו את הרצפים הדומיננטיים (אנחנו מניחים שאלה רצפי המקור, שלא עברו מוטציות בשעתוק) ומהם נמשיך הלאה לביצוע מיפוי לdatabase כדי לזהות באילו חיידקים מדובר.  
 נקבל את הפלט הבא:

שורה נוספת:  
נמיר בעצם את הרצפים שקיבלנו לוויזואליזציה שאפשר לצפות בה:  
**qiime feature-table tabulate-seqs --i-data qza/rep-seqs-dn-99.qza --o-visualization vis/rep-seqs-dn-99.qzv**



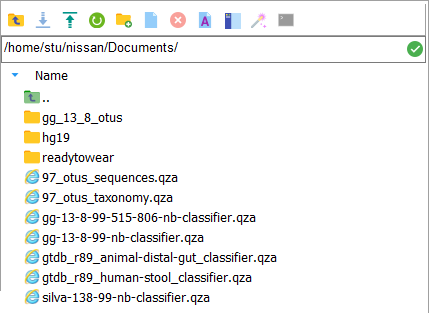
אני אוהבת להסתכל באתר של Qiime ולשמור לי כקובץ את כל הרצפים שקיבלנו כדי שאוכל לחזור אליהם מאוחר יותר. זה אמור להיראות ככה:

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

כעת נבדוק התאמה – כלומר נשווה לDatabases השונים כדי לזהות למי שייכים הרצפים שקיבלנו (כי הFeature ID לא באמת אומר לנו משהו).

**שלב 6 – קבלת זיהוי טקסונומי**

נרצה להשוות למאגרי המעבדה ששמורים אצלנו במיקום הבא בשרת:  
/home/stu/nissan/Documents/ זה נראה ככה:  


**נבנה את השאילתא באופן הבא (בערך – לא תבנית קבועה):**qiime feature-classifier **CLASIFFICATION\_METHOD** qza/rep-seqs-dn-99.qza  
 --i-reference-reads **OTU\_DATABASE\_PATH** --i-reference-taxonomy **TAXONOMY\_DATABASE\_PATH**   
--o-classification **RESULTS\_FULL\_PATH** --verbose

הדגמה – שיטת קלסיפיקציה vsearch עבור דאטאבייס:  
qiime feature-classifier **classify-consensus-vsearch** --i-query qza/rep-seqs-dn-99.qza  
 --i-reference-reads **~/Documents/97\_otus\_sequences.qza** --i-reference-taxonomy **~/Documents/97\_otus\_taxonomy.qza**  
 --o-classification **qza/vs-classifier.qza** --verbose

שיטה שנייה sklearn (עבור DB gg. יש אפשרות להשתמש בgg, gtdb או silva).

qiime feature-classifier **classify-sklearn** --i-reads **qza/rep-seqs-dn-99.qza** --i-classifier **~/Documents/gg-13-8-99-nb-classifier.qza** --o-classification **qza/sk-classified.qza** --verbose

במעבדה משתמשים בד"כ בשיטה של sklearn, ואכן ניתן לראות שהיא מקבלת ASV בלבד ואין צורך ב-OTU. נשתמש בדרך כלל בבסיס הנתונים gtdb, אלא אם כן אנחנו חושדים שקיים דנ"א אאוקריוטי בדגימה (לדוגמה- עכברים), ואז נשתמש ב-GG או silva.

path לסילבה:

"/home/stu/nissan/Documents/silva-138-99-nb-classifier.qza"

Path ל-gtdb:

"/home/stu/nissan/Documents/gtdb\_r89\_animal-distal-gut\_classifier.qza"

"/home/stu/nissan/Documents/gtdb\_r89\_human-stool\_classifier.qza"

לא משנה באיזה מהשיטות בחרנו (כמו שניסן אומר – למה צריך לבחור אם אפשר גם וגם), עלינו לנקות את המידע שם.

מעבר לקובץ לצפייה ויזואלית:

qiime metadata tabulate --m-input-file qza/sk-classified.qza --o-visualization vis/taxonomy.qzv

**שלב 7 – ניקוי המידע:  
  
ניקוי ראשון** – הסרת רצפי DNA מיטוכונדריאליים או כלורופלסטים:  
qiime taxa filter-table --i-table qza/table-dn-99.qza --i-taxonomy qza/sk-classified.qza --p-exclude mitochondria,chloroplast --o-filtered-table qza/table-clean.qza

בירוק – הטבלה שיצרנו בשלב 5  
בצהוב – הקובץ שיצרנו בשלב 6  
בכחול – הרצפים שנרצה להסיר  
בוורוד – שם הטבלה החדשה לאחר ההסרה וה-Path שלה

\*\*\*\*\*

תוספת תמר:

הסרת רצפים של ארכיאה ושל אאוקריוטים:

ארכיאה:

qiime taxa filter-table --i-table qza/table-clean.qza --i-taxonomy qza/sk-classified.qza --p-exclude "k\_\_Archaea" --o-filtered-table qza/table-archaea-clean.qza

אאוקריוטים:

qiime taxa filter-table --i-table qza/table-archaea-clean.qza --i-taxonomy qza/sk-classified.qza --p-exclude "k\_\_Eukaryota" --o-filtered-table qza/table-archaea-eukaryota-clean.qza

עכשיו נעשה את אותם סינונים מהרצפים: (לדברי אריאל אין צורך בשלב הזה, לבדוק עם/בלי)

מיטוכונדריה + כלורופלסט:

qiime taxa filter-seqs --i-sequences qza/rep-seqs-dn-99.qza --i-taxonomy qza/sk-classified.qza --p-exclude mitochondria,chloroplast --o-filtered-sequences qza/rep-seq-clean.qza

ארכיאה:

qiime taxa filter-seqs --i-sequences qza/rep-seq-clean.qza --i-taxonomy qza/sk-classified.qza --p-exclude "k\_\_Archaea" --o-filtered-sequences qza/rep-seq-archaea-clean.qza

אאוקריוטים:

qiime taxa filter-seqs --i-sequences qza/rep-seq-archaea-clean.qza --i-taxonomy qza/sk-classified.qza --p-exclude "k\_\_Eukaryota" --o-filtered-sequences qza/rep-seq-archaea-eukaryota-clean.qza

**צפייה בתוצאות:**

qiime taxa barplot --i-table qza/table-archaea-eukaryota-clean.qza --i-taxonomy qza/sk-classified.qza --m-metadata-file meta-data.tsv --o-visualization vis/taxa-bar-plots.qzv

נראה שהניקויים 2+3 לא משפיעים הרבה, אבל אם נראה שלא קיבלנו מספיק תוצאות לאחר שלושתם אפשר לחזור לראשון בלבד.

בכל מקרה נוודא שבסוף יש לקובץ הסופי שם אחיד:

mv qza/table-archaea-eukaryota-clean.qza qza/filtered-table.qza

mv qza/rep-seq-archaea-eukaryota-clean.qza qza/filtered-rep-seq.qza

שם הקובץ שאיתו בחרנו להמשיך הלאה השם החדש של הקובץ

\*\*\*\*\*\*\*

**ניקוי שני ושלישי:**  
אני אישית מדלגת על השלב הזה כדי לא לאבד data שגם ככה עבר "קיצוץ רציני" בשלבים של הtrimming והכימרות. אבל זה מה שעושים אם רוצים לסנן את המידע עוד קצת:  
**ניקוי שני** – סינון החיידקים (הטקסונומיות) שלא נמצאו לפחות ב-X דגימות. דוגמה עבור x=3:

qiime feature-table filter-features –i-table qza/vs-classifier-clean.qza –p-min-samples 3 –o-filtered-table qza/vs-mincount-table.qza

**ניקוי שלישי – להוריד חיידקים שהכמות שלהם הייתה פחות מy. דוגמה עבור y = 38:**

qiime feature-table filter-features –i-table qza/vs-mincount-table.qza –p-min-frequency 38 –o-filtered-table qza/vs-frequency-clean.qza  
  
**שלב 8 – הכנת עץ לחישוב a-diversity**  
**a-diversity-** רמת השונות של החיידקים בתוך אותה דוגמה

**β-diversity-** רמת השונות של החיידקים בין קבוצות שונות (מאפשר להשוות בין לפני ואחרי טיפול).

**ביצוע alignment de novo באמצעות mafft ליצירת עץ:**  
  
qiime alignment mafft --i-sequences qza/rep-seqs-dn-99.qza --o-alignment qza/aligned-rep-seqs.qza

**mask/ filter the alignments (remove highly variable positions):**  
qiime alignment mask --i-alignment qza/aligned-rep-seqs.qza --o-masked-alignment qza/masked-aligned-rep-seqs.qza

**בניית העץ:**qiime phylogeny fasttree --i-alignment qza/masked-aligned-rep-seqs.qza --o-tree qza/fasttree-tree.qza --verbose

**מציאת שורש העץ:   
qiime phylogeny midpoint-root --i-tree qza/fasttree-tree.qza --o-rooted-tree qza/fastree-rooted.qza**

ייצוא העץ:

qiime tools export --input-path qza/fastree-rooted.qza --output-path exports

הפקודה פותחת תיקייה חדשה exports ושמה שם את העץ המוכן.

\*\*\*\*\*\*\*\*\*\*\*

תוספת תמר:

שיטה מהירה וישירה יותר:

qiime phylogeny align-to-tree-mafft-fasttree --i-sequences qza/filtered-rep-seq.qza --o-alignment qza/aligned-rep-seqs.qza --o-masked-alignment qza/masked-aligned-rep-seqs.qza --o-tree qza/unrooted-tree.qza --o-rooted-tree qza/rooted-tree.qza

אבל: העץ שיוצא מושרש לא לפי mid-point, אז אולי עדיף את השיטה של אריאל.

קבצי הפלט:



קובץ מיפוי, קובץ מיפוי לאחר סינון, עץ לא מושרש ועץ מושרש.

ייצוא העץ:

qiime tools export --input-path qza/rooted-tree.qza --output-path exports

הפקודה פותחת תיקייה חדשה exports ושמה שם את העץ המוכן.

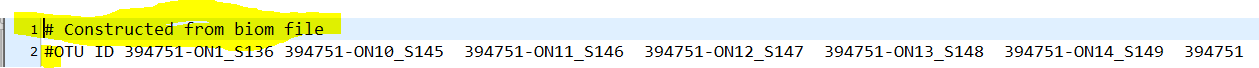
\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

שלב 9 – ייצוא התוצאות 😊:  
בעצם מכניסים את הטבלה הנקייה שלנו מהשלב הקודם:

qiime tools export --input-path qza/table-clean.qza --output-path exports

הפקודה מייצאת לתוך תיקיית exports אובייקט בשם feature-table.biom.

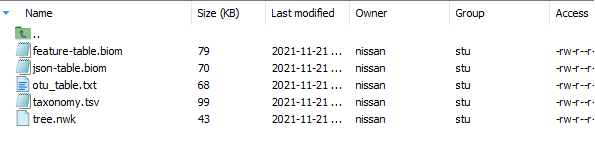
כעת, נכתוב את הפקודה הבאה כדי לייצר אובייקט בשם otu\_table.txt:  
biom convert -i exports/feature-table.biom -o exports/otu\_table.txt --to-tsv

כעת נשנה ידנית את הכותרת בתוך קובץ הטקסט: נמחק את הכותרת ואת ה-#:  
  
  
כך זה ייראה:  
  
   
נריץ את הפקודה:

biom convert -i exports/feature-table.biom -o exports/json-table.biom --table-type="OTU table" --to-json

נקבל אובייקט json-table.biom.

נרצה ליצור קובץ Taxonomy.tsv ולכן נעשה (על הטבלה לפני הניקוי):  
  
qiime tools export --input-path qza/sk-classified.qza --output-path exports

בסוף השלב הזה, אמורים להיות לנו 5 תוצרים בתיקיית exports:  


אותם נייצא ונעבור לR 😊

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

השלב הזה לא אמור להיות נחוץ, ועם זאת הוא היה בניסיון הראשון שלי לאנליזה.

לבדוק גם באנליזות המשך- לנסות לסיים פה ואם לא עובד להמשיך גם לשלב הזה.

שלב נוסף, רק אם צריכים את המידע הזה:

##קודם כל: לשנות את ההתחלה של הקובץ טקסונומיה מ-Feature ID ל- #OTU ID

biom add-metadata -i exports/feature-table.biom -o exports/feature-table-taxa.biom --observation-metadata-fp exports/taxonomy.tsv --sc-separated taxonomy

biom summarize-table -i exports/feature-table-taxa.biom

biom convert -i exports/feature-table-taxa.biom -o exports/feature-table-taxa-json.biom --table-type="OTU table" --to-json

לשנות את ההתחלה של הקובץ טקסונומיה ל- OTU ID (להוריד את ה-#)

qiime tools export --input-path qza/unrooted-tree.qza --output-path exports

\*\*\*\*\*\*\*\*\*\*\*\*

תוספת תמר:

ניצור ויזואליזציה:

qiime feature-table summarize --i-table qza/filtered-table.qza --o-visualization qza/filtered-table.qzv --m-sample-metadata-file meta-data.tsv

נבדוק מהי התדירות המינימלית שיש לנו בקובץ, ולפיה נבצע את הפקודה הבאה: (בדוגמה שלנו- 37199)

qiime diversity core-metrics-phylogenetic --i-phylogeny qza/rooted-tree.qza --i-table qza/filtered-table.qza --p-sampling-depth 37199 --m-metadata-file meta-data.tsv --output-dir diversity-metrics-results

נמצא אלפא דייברסיטי לפי faith (a measure of community richness):

qiime diversity alpha-group-significance --i-alpha-diversity diversity-metrics-results/faith\_pd\_vector.qza --m-metadata-file meta-data.tsv --o-visualization diversity-metrics-results/faith-pd-group-significance.qzv

ולפי שאנון:

qiime diversity alpha-group-significance --i-alpha-diversity diversity-metrics-results/shannon\_vector.qza --m-metadata-file meta-data.tsv --o-visualization diversity-metrics-results/shannon-group-significance.qzv

בטא דייברסיטי בשיטות שונות:

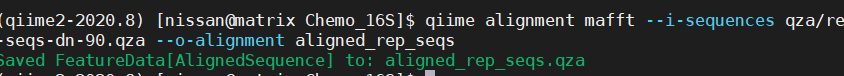
qiime diversity beta-group-significance --i-distance-matrix diversity-metrics-results/weighted\_unifrac\_distance\_matrix.qza --m-metadata-file meta-data.tsv --m-metadata-column day --o-visualization diversity-metrics-results/weighted-unifrac-day-significance.qzv --p-pairwise

qiime diversity beta-group-significance --i-distance-matrix diversity-metrics-results/unweighted\_unifrac\_distance\_matrix.qza --m-metadata-file meta-data.tsv --m-metadata-column day --o-visualization diversity-metrics-results/unweighted-unifrac-subject-group-significance.qzv --p-pairwise

PCoA:

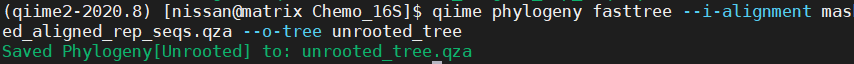
qiime emperor plot --i-pcoa diversity-metrics-results/unweighted\_unifrac\_pcoa\_results.qza --m-metadata-file meta-data.tsv --p-custom-axes mouseID --o-visualization diversity-metrics-results/unweighted-unifrac-emperor-day.qzv

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

עוד דברים שעשיתי (מהסרטון):  
**qiime alignment mafft --i-sequences qza/rep-seqs-dn-90.qza --o-alignment aligned\_rep\_seqs**

**qiime alignment mask --i-alignment aligned\_rep\_seqs.qza --o-masked-alignment masked\_aligned\_rep\_seqs.qza**  

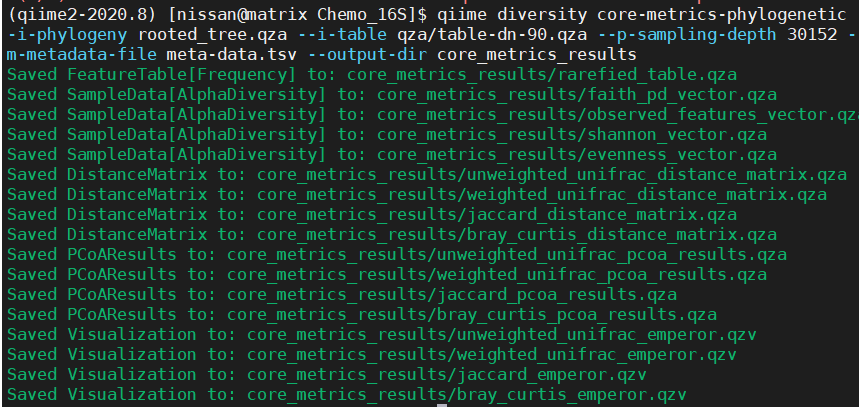

**qiime phylogeny fasttree --i-alignment masked\_aligned\_rep\_seqs.qza --o-tree unrooted\_tree**



**qiime phylogeny midpoint-root --i-tree unrooted\_tree.qza --o-rooted-tree rooted\_tree**

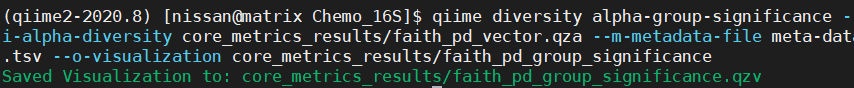


**qiime diversity core-metrics-phylogenetic --i-phylogeny rooted\_tree.qza --i-table qza/table-dn-90.qza --p-sampling-depth 30152 --m-metadata-file meta-data.tsv --output-dir core\_metrics\_results**

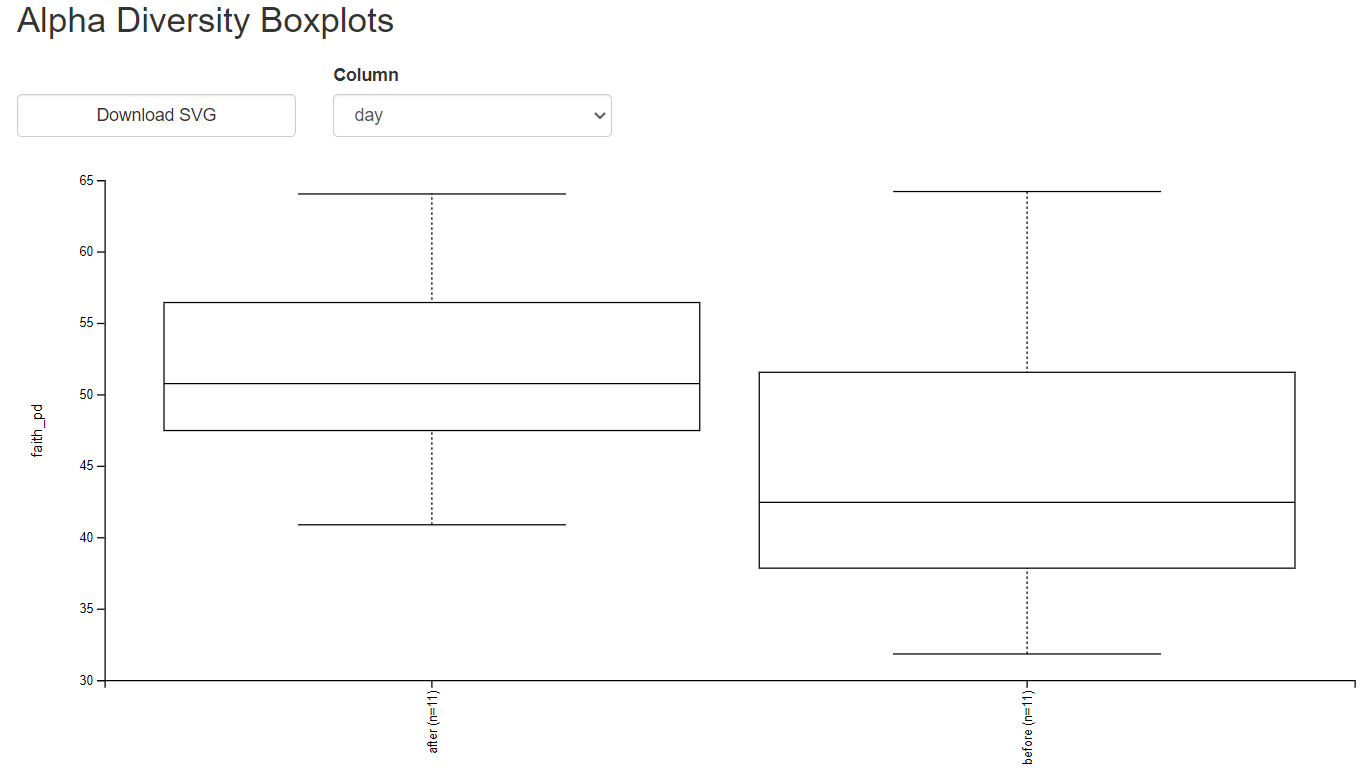


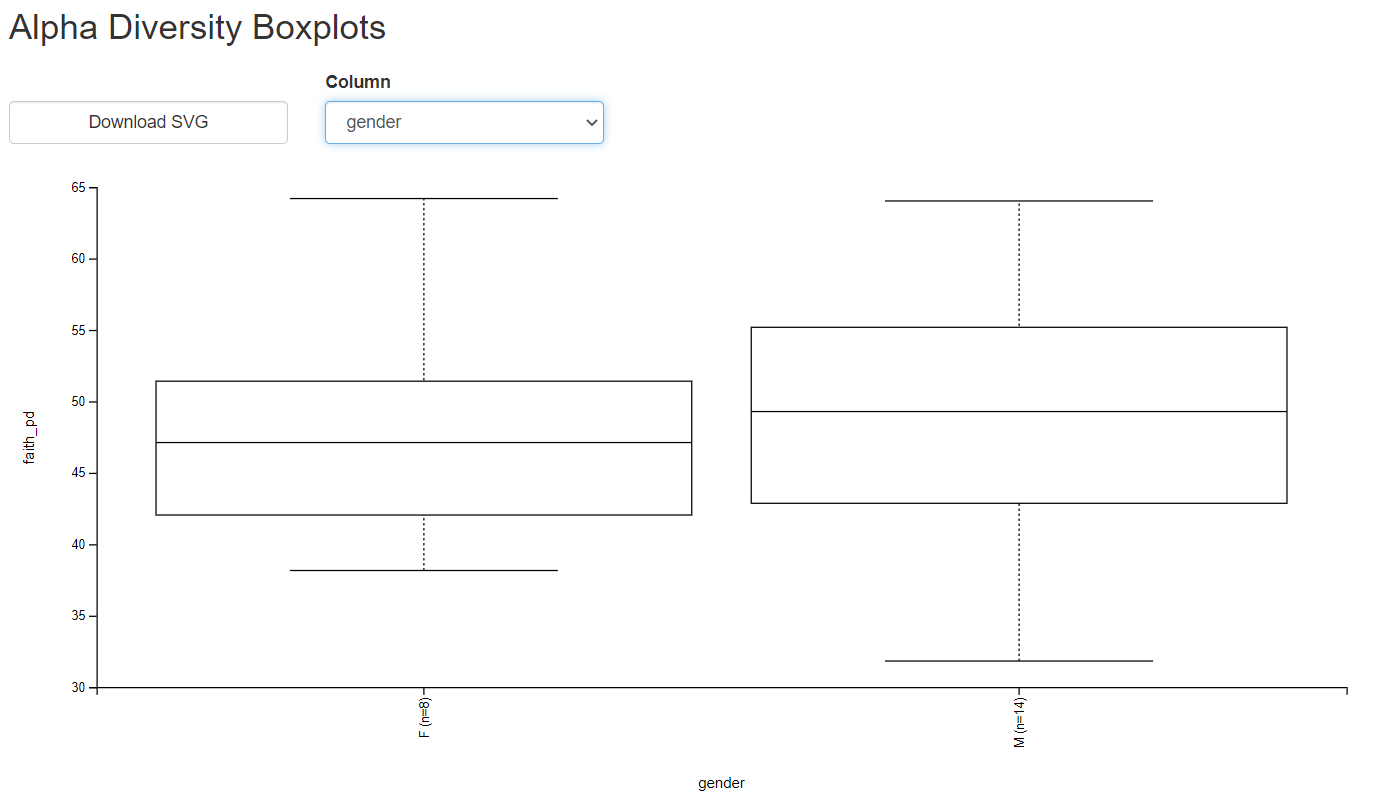
<https://forum.qiime2.org/t/alpha-and-beta-diversity-explanations-and-commands/2282>

**qiime diversity alpha-group-significance --i-alpha-diversity core\_metrics\_results/faith\_pd\_vector.qza --m-metadata-file meta-data.tsv --o-visualization core\_metrics\_results/faith\_pd\_group\_significance**



מבדיקה באתר של qiime:   
לפי day:



לפי gender:  


בטא דייברסיטי:

