In the name of god

M.A.Shirazi

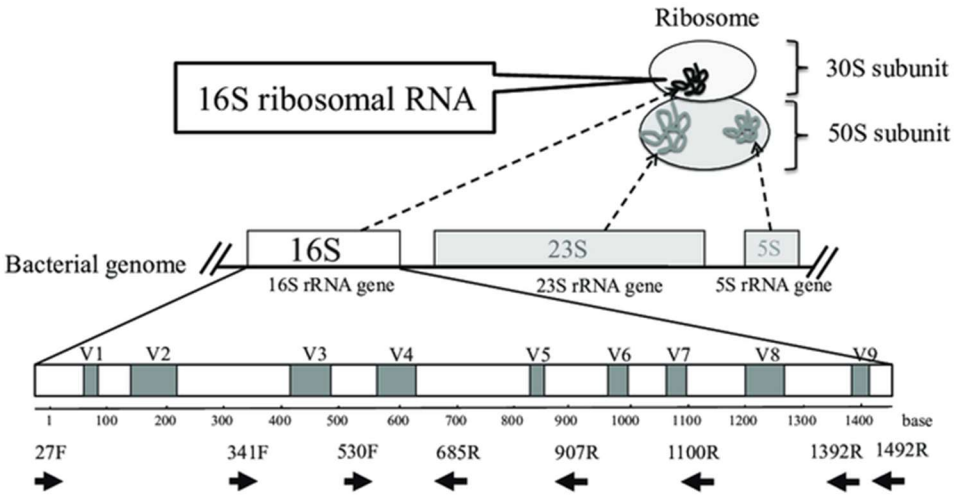
# 16S rRna analysis & metagenomic with QIIME2

Linkedin: Mohammad Amin Shirazi

Email: mashirazi96@ut.ac.ir

Mohammad Amin Shirazi

ژن 16S rRNA یک بخش مهم از ژن‌های ریبوزومی در باکتری‌ها و آرکی‌هاست و نقش کلیدی در شناسایی و طبقه‌بندی این موجودات دارد. این ژن در همه باکتری‌ها و آرکی‌ها وجود دارد و به دلیل داشتن بخش‌هایی پایدار و بخش‌هایی متغیر، یک ابزار قوی در مطالعات فیلوژنتیکی و شناسایی میکروبی است.




## ساختار ژن 16S rRNA

ژن 16S rRNA دارای دو ویژگی ساختاری مهم است :

بخش‌های پایدار(Conserved Regions) : این بخش‌ها در سراسر باکتری‌ها و آرکی‌ها بسیار پایدار هستند و تغییرات کمی در طول زمان تکامل دارند. این ویژگی باعث می‌شود که بتوانیم این بخش‌ها را در طیف گسترده‌ای از گونه‌های میکروبی شناسایی کنیم و به‌عنوان نقاطی برای هم‌تراز کردن توالی‌های مختلف استفاده کنیم. در تجزیه و تحلیل‌های فیلوژنتیکی، این بخش‌ها به محققان کمک می‌کنند تا موجودات را با دقت بیشتری دسته‌بندی کنند .

بخش‌های متغیر(Variable Regions) : این بخش‌ها که به‌صورت نواحی V1 تا V9 نام‌گذاری شده‌اند، بین گونه‌های مختلف تغییرات بیشتری دارند و به ما امکان شناسایی و تمایز گونه‌های نزدیک به هم را می‌دهند. هر کدام از نواحی V1 تا V9 می‌توانند اطلاعات خاصی در مورد گروه‌های خاصی از باکتری‌ها داشته باشند. به‌همین دلیل، معمولاً در مطالعات میکروبیوم یک یا چند تا از این نواحی متغیر برای شناسایی دقیق‌تر انتخاب می‌شوند.

Mohammad An

## کاربردهای ژن 16S rRNA در مطالعات میکروبی

ژن 16S rRNA به دلیل ساختار خاصش در تحقیقات میکروبی و زیست‌شناسی کاربردهای مهمی دارد:

شناسایی گونه‌های میکروبی: از ژن 16S rRNA به‌طور گسترده‌ای برای شناسایی گونه‌های میکروبی استفاده می‌شود. با تعیین توالی ژن 16S rRNA از نمونه‌های محیطی یا بیولوژیکی، می‌توان به‌طور مستقیم انواع مختلف باکتری‌ها را شناسایی کرد. این روش به‌ویژه در مطالعات میکروبیوم (مثلاً میکروبیوم روده، خاک، و آب) استفاده می‌شود .

طبقه‌بندی فیلوژنتیکی: از طریق مقایسه توالی 16S rRNA گونه‌های مختلف، می‌توان روابط تکاملی بین آن‌ها را تعیین کرد و درخت فیلوژنتیکی ساخت. این درخت‌ها به ما کمک می‌کنند تا روابط نزدیک و دور بین گونه‌های مختلف را مشاهده کنیم.

پژوهش‌های میکروبیوم: در مطالعاتی که بر روی جوامع میکروبی و تنوع آن‌ها تمرکز دارند، توالی‌یابی ژن 16S rRNAیکی از ابزارهای اصلی است. این ژن به دلیل ویژگی‌های پایدار و متغیرش می‌تواند تنوع و غنای میکروبی یک محیط را نشان دهد.

## انتخاب نواحی متغیر ژن 16S برای توالی‌یابی

از آنجا که توالی‌یابی کامل ژن 16S ممکن است هزینه‌بر باشد، محققان اغلب از نواحی خاصی از این ژن برای تحلیل استفاده می‌کنند. این نواحی به‌طور کلی با نام‌های زیر شناخته می‌شوند:

V3–V4 : یکی از پرکاربردترین نواحی در شناسایی باکتری‌های محیطی. این ناحیه برای تحلیل‌های مربوط به میکروبیوم روده مناسب است.

V4: این ناحیه به دلیل اندازه کوتاه و قابلیت شناسایی دقیق در مطالعات بسیاری استفاده می‌شود.

V1–V3: این نواحی در برخی از پروژه‌ها برای شناسایی گسترده‌تر باکتری‌ها استفاده می‌شوند.

Mohammad Amin Shirazi

## چرا از ژن 16S rRNA استفاده می‌شود؟

ژن 16S rRNA به دلیل پایداری بالای بخش‌های محافظت‌شده و تنوع در بخش‌های متغیر، یک انتخاب ایده‌آل برای شناسایی و طبقه‌بندی میکروب‌هاست. برخی دلایل استفاده گسترده از آن عبارتند از:

حفظ شـــده بودن در طی تکامل: به دلیل اینکه بخش‌هایی از این ژن در اکثر موجودات تغییر نکرده، امکان شناسایی گسترده و دقیق گونه‌ها وجود دارد.

وجود نواحی متغیر: نواحی متغیر امکان شناسایی گونه‌های نزدیک به هم را فراهم می‌کنند.

قابلیت اسـتفاده در محیط‌های مختلف: این ژن به‌طور گسـترده‌ای در محیط‌های مختلف مانند بدن انسـان، خاک، آب، و محیط‌های دیگر استفاده می‌شود.

## مدل‌های طبقه‌بندی بر اساس ناحیه‌های متغیر

مدل‌های طبقه‌بندی، مثل GreenGenes وSILVA ، برای نواحی خاص آموزش دیده‌اند. بنابراین، وقتی شـــما داده‌های خود را توالی‌یابی می‌کنید، باید از مدل طبقه‌بندی‌ای اســـتفاده کنید که با ناحیه‌ی هدف داده‌هایتان مطابقت دارد.

برای مثال:

V3–V4 : این ناحیه در بسیاری از مطالعات میکروبیوم استفاده می‌شود.

V4: ناحیه‌ای است که معمولاً دقت خوبی در شناسایی بسیاری از گونه‌های باکتریایی دارد.

V1–V3: برای برخی گروه‌های خاص از باکتری‌ها حساسیت بالاتری دارد.

پایگاه‌های SILVA و GreenGenes دو منبع اصـــلی برای تحلیل و طبقه‌بندی میکروب‌ها به‌ویژه باکتری‌ها و آرکی‌ها هســـتند و برای تحلیل‌های توالی‌یابی ژن 16S rRNA و طبقه‌بندی تاکســـونومیک اســـتفاده می‌شوند. در ادامه این پایگاه‌ها و کاربردهای آن‌ها را به‌طور کامل‌تر توضیح می‌دهم:

Mohammad Amin Shirazi

| ویژگی | SILVA | GreenGenes |
|---|---|---|
| آخرین به‌روزرسانی | به‌طور مرتب به‌روزرسانی می‌شود | آخرین نسخه در سال 2013 |
| شامل یوکاریوت‌ها | بله | خیر |
| دقت طبقه‌بندی | بالا و دقیق | معتبر، اما بدون به‌روزرسانی جدید |
| سازگاری با ابزارها | با ابزارهای جدید مانند QIIME 2 و Mothur | عمدتاً با نسخه‌های قدیمی QIIME |
| کاربردها | کاربرد گسترده در محیط‌های مختلف | بیشتر در محیط‌های مشخص و مطالعات قدیمی‌تر |

پایگاه داده SILVA

پایگاه داده SILVA یک منبع جامع و دقیق است که مجموعه‌ای از توالی‌های ریبوزومی RNA (شامل ژن‌های 16S، 18S، و 23S) برای باکتری‌ها، آرکی‌ها و یوکاریوت‌ها را شامل می‌شود. از SILVA به‌عنوان منبعی برای طبقه‌بندی و شناسایی انواع میکروبی استفاده می‌شود. این پایگاه به‌طور دوره‌ای به‌روز می‌شود تا جدیدترین توالی‌ها و اطلاعات طبقه‌بندی را شامل شود.

ویژگی‌ها و مزایای پایگاه داده SILVA :

جامعیت و به‌روزرسانی مداوم SILVA: به دلیل دارا بودن تعداد زیادی توالی‌های ژن‌های ریبوزومی ۱۶ S و ۱۸ S و پوشش گسترده‌تر از موجودات مختلف در محیط‌های گوناگون به‌روز شده است.

طبقه‌بندی دقیق: داده‌های SILVA از منابع مختلف گردآوری و با دقت طبقه‌بندی شده‌اند، که به محققان اجازه می‌دهد تا تاکسون‌های مختلف را با اطمینان بیشتری شناسایی کنند.

پشتیبانی از ابزارهای مختلف SILVA : از فایل‌های قالب‌بندی‌شده برای استفاده در ابزارهای مختلف مانند QIIME 2 و Mothur پشتیبانی می‌کند.

Mohammad Amin Shi

قابلیت انتخاب نسخه‌های مختلف: می‌توانید از نسخه‌های مختلف SILVA استفاده کنید که هر نسخه‌ای به‌روزرسانی‌های جدیدتر و تغییرات تاکسونومیک بیشتری دارد.

پایگاه دادهGreenGenes

پایگاه داده GreenGenes یکی دیگر از منابع محبوب برای تحلیل ژن 16S rRNA است که بیشتر بر روی باکتری‌ها و آرکی‌ها متمرکز است. برخلافSILVA ، به‌روزرسانی GreenGenes متوقف شده و آخرین نسخه آن در سال ۲۰۱۳ منتشر شده است، اما به دلیل سازگاری با ابزارهای تحلیل مانند QIIME 2 و گستردگی استفاده، همچنان کاربرد دارد

.ویژگی‌ها و محدودیت‌های پایگاه داده:GreenGenes

پایایی: گرچه GreenGenes دیگر به‌روزرسانی نمی‌شود، اما به‌طور گسترده‌ای در تحقیقات میکروبیوم و توالی‌یابی 16S rRNAاستفاده شده است و داده‌های آن به خوبی بررسی شده‌اند.

دقت در طبقه‌بندی: این پایگاه داده شامل طبقه‌بندی‌هایی معتبر برای تاکسون‌های میکروبی است و با اینکه به‌روزرسانی نشده، هنوز هم در بسیاری از مطالعات متداول استفاده می‌شود.

سازگاری با ابزارهای قدیمی‌تر: ابزارهایی مانند QIIME نسخه ۱ و برخی از نسخه‌های قبلی QIIME 2 به طور خاص با GreenGenes کار می‌کنند و از طبقه‌بندی آن پشتیبانی می‌کنند.

مدل‌های طبقه‌بندی بر اساس ناحیه‌های متغیر مدل‌های طبقه‌بندی، مثل GreenGenes وSILVA ، برای نواحی خاص...

چرا از SILVA یا GreenGenes استفاده می‌کنیم؟

پایگاه‌های SILVA و GreenGenes هر دو برای ساخت مدل‌های طبقه‌بندی میکروبی از روی ژن 16S rRNA استفاده می‌شوند. درQIIME 2 ، وقتی که توالی‌های 16S خود را طبقه‌بندی می‌کنید، نیاز به یک مدل طبقه‌بندی دارید که بر اساس یکی از این پایگاه‌ها آموزش دیده باشد. این مدل‌ها کمک می‌کنند تا توالی‌های شما به‌طور دقیق به گروه‌های میکروبی مختلف دسته‌بندی شوند .جمع‌بندی برای تحلیل‌های

Mohammad Amin Shirazi

جدید و جامع‌تر، SILVA توصیه می‌شود، چرا که به‌روزتر و دقیق‌تر است و نواحی متنوع‌تری از موجودات میکروبی را پوشش می‌دهد. اما اگر برایتان مهم است که پایگاه داده با ابزارهای قدیمی‌تر سازگار باشد یا از روش‌های تحلیل قدیمی استفاده کنید، GreenGenes گزینه مناسبی است.

**: QIIME 2**

مفاهیم اصلی این صفحه چند مفهوم اصلی در QIIME 2 را توضیح می‌دهد که قبل از شروع به استفاده از این نرم‌افزار، آشنایی با آنها اهمیت دارد. همچنین، ممکن است به لغت‌نامه موجود در سایت برای درک بهتر مطالب این صفحه و دیگر مستندات نیاز داشته باشید .فایل‌های داده: مصنوعات (artifacts) QIIME 2 داده‌های تولید شده توسط QIIME 2 به‌صورت مصنوعات QIIME 2 ذخیره می‌شوند. یک مصنوع QIIME 2شامل داده و متاداده است. متاداده توصیفاتی درباره داده دارد، از جمله نوع آن، فرمت، و نحوه تولید آن (اصالت داده). یک مصنوع QIIME 2 معمولاً پسوند فایل qza. دارد .از آنجایی که QIIME 2 با مصنوعات به جای فایل‌های داده کار می‌کند (e.g. FASTA files) ، باید داده‌ها را با وارد کردن آنها به شکل یک مصنوع QIIME 2 درآورید. وارد کردن داده‌ها در هر مرحله‌ای از تحلیل ممکن است، ولی معمولاً با وارد کردن داده‌های خام توالی شروع می‌کنید QIIME 2 .همچنین ابزارهایی برای خروج داده از یک مصنوع ارائه می‌دهد .با استفاده از مصنوعات، QIIME 2به‌طور خودکار نوع، فرمت، و اصالت داده را برای پژوهشگران پیگیری می‌کند. به این ترتیب، پژوهشگران می‌توانند بر روی تحلیل خود تمرکز کنند به جای این که نگران فرمت خاص داده برای تحلیل باشند .اصالت داده در مصنوعات مصنوعات این امکان را به QIIME 2 می‌دهند که علاوه بر خود داده، اطلاعات مربوط به نحوه ایجاد داده را هم پیگیری کند. با اصالت داده یک مصنوع، می‌توانید تمامی تحلیل‌های قبلی که منجر به تولید آن شده را ردیابی کنید، از جمله داده‌های ورودی که در هر مرحله استفاده شده‌اند. این پیگیری یکپارچه و خودکار اصالت داده‌ها، به پژوهشگر امکان می‌دهد تا مصنوعات را آرشیو کند یا برای مثال، آن را به همکاری ارسال کند، و همچنین می‌تولند نحوه دقیق ایجاد مصنوع را درک کند. این ویژگی باعث تکرارپذیری و بازتولیدپذیری تحلیل‌ها می‌شود و نمودارها و متونی را تولید می‌کند که می‌توانند در بخش روش‌های یک مقاله استفاده شوند .
توجه: اصطلاح "مصنوع" ممکن است برای برخی پژوهشگران گیج‌کننده باشد زیرا معمولاً برای اشاره به منبع خطاهای تجربی استفاده می‌شود. در اینجا، "مصنوع" به معنای یک شیء تولید شده توسط فرآیندی

Mohammad Amin Shirazi

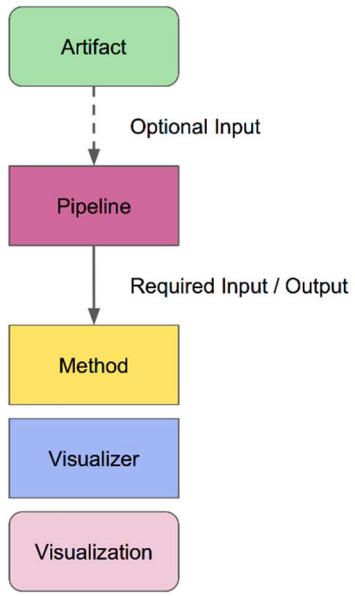
خاص است .فایل‌های داده: تصویری‌سازی‌ها تصویری‌سازی‌ها نیز نوعی از داده‌های تولید شده توسط QIIME 2هستند و معمولاً پسوند qzv. دارند. مشابه مصنوعات QIIME 2 ، تصویری‌سازی‌ها نیز متاداده و اصالت داده دارند و می‌توانند به صورت جداگانه بایگانی شده یا با همکاران به اشتراک گذاشته شوند. برخلاف مصنوعات، تصویری‌سازی‌ها به عنوان خروجی نهایی یک تحلیل محسوب می‌شوند و نمی‌توانند به عنوان ورودی تحلیل‌های دیگر در QIIME 2 استفاده شوند .نوع معنایی هر مصنوع تولید شده توسط QIIME 2دارای یک نوع معنایی است که به QIIME 2 امکان می‌دهد مصنوعات مناسب برای ورودی یک تحلیل را شناسایی کند. مثلاً اگر یک تحلیل به یک ماتریس فاصله نیاز داشته باشد، QIIME 2 می‌تواند مصنوعات دارای این نوع معنایی را پیدا کند و از استفاده مصنوعات ناسازگار جلوگیری کند . پلاگین‌ها تحلیل‌های میکروبیوم در QIIME 2 از طریق پلاگین‌ها در دسترس هستند. برای انجام تحلیل با QIIME 2 ، باید یک یا چند پلاگین نصب کنید که تحلیل‌های خاصی را ارائه می‌دهند. پلاگین‌ها بسته‌های نرم‌افزاری هستند که توسط هر کسی می‌توانند توسعه داده شوند .روش‌ها و بصری‌سازها پلاگین‌های QIIME 2 روش‌ها و بصری‌سازهایی را تعریف می‌کنند که برای انجام تحلیل‌ها به کار می‌روند. روش‌ها مصنوعات و پارامترهایی را به عنوان ورودی دریافت می‌کنند و یک یا چند مصنوع به عنوان خروجی تولید می‌کنند. بصری‌سازها فقط یک تصویری‌سازی به عنوان خروجی تولید می‌کنند و نمی‌توانند به عنوان ورودی به تحلیل‌های دیگر استفاده شوند .برای مشاهده مصنوعات و فایل‌های تصویری‌سازی ( QIIME 2معمولاً فایل‌های qza. و (qzv. بدون نیاز به نصب QIIME می‌توانید از qiime2 view استفاده کنید.

مرور بر جریان‌های کاری پلاگین‌های QIIME 2 توجه تمام امید خود را رها کنید، اگر واژه‌نامه را نخوانده‌اید . توجه این راهنما برای کاربران تازه‌کار QIIME 2 است، به ویژه برای کسانی که به تحقیق در زمینه میکروبیوم تازه وارد هستند. برای کاربران با تجربه که در تحلیل میکروبیوم آشنا هستند، و کسانی که از استفاده بی‌ضابطه از اموجی‌ها پرهیز دارند، به راهنمایی برای کاربران حرفه‌ای مراجعه کنید .به تمامی تازه‌واردان خوش آمد می‌گوییم .این راهنما یک نمای کلی از بسیاری از پلاگین‌ها و اقدام‌های اصلی موجود در QIIME 2 به شما می‌دهد و شما را به سمت راهنماهای آموزشی مربوطه هدایت می‌کند تا

Mohammad Amin Shirazi

عمیق‌تر کاوش کنید. به عبارت دیگر، ممکن است سوال شما این باشد که "چطور از QIIME 2 استفاده کنم"، اما این راهنما شما را به مسیر درست هدایت می‌کند. این را به عنوان نقشه گنج خود در نظر بگیرید: اقدامات QIIME 2 سنگ‌فرش‌هایی هستند که شما را به سوی شکوه هدایت می‌کنند و نمودارهای جریان زیر به شما می‌گویند که همه چیزهای خوب کجا دفن شده‌اند .به یاد داشته باشید که مسیرهای زیادی از دامنه کوه به بالا می‌روند، اما در قله، همه ما به همان ماه نگاه می‌کنیم .بیایید راه را پیدا کنیم: نمودارهای جریان

قبل از اینکه درباره پلاگین‌ها و اقدامات خاص صحبت کنیم، ابتدا یک نمای کلی از جریان کاری استاندارد QIIME 2برای تحلیل داده‌های توالی آمپلیکون را بررسی خواهیم کرد. و قبل از اینکه به آن نمای کلی نگاه کنیم، باید به کلید نقشه گنج خود نگاه کنیم



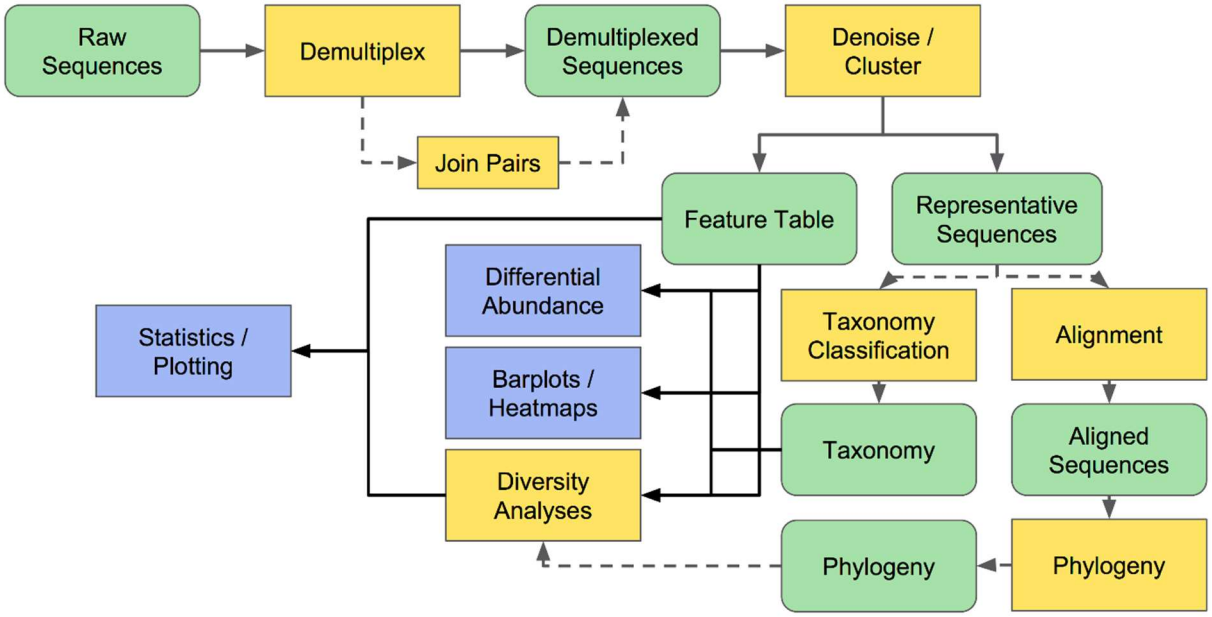

هر نوع داده (یعنی، آثار و تجسم‌ها) و اقدام (یعنی، روش‌ها، تجسم‌کننده‌ها و پایپلاین‌ها) با یک گره رنگی متفاوت نمایان شده است. لبه‌هایی که هر گره را به هم متصل می‌کنند یا از نوع ممتد هستند (که ورودی یا خروجی الزامی را نشان می‌دهند) یا از نوع خط‌چین (که ورودی اختیاری را نشان می‌دهند). نمی‌دانید

این اصـــطلاحات به چه معنی اســـت؟ اه! واژه‌نامه بالا را بخوانید .در نمودارهای جریان زیر :اقدامات با نام پلاگین و نام اقدام برچسب‌گذاری شده‌اند. برای استفاده از آن اقدام، "qiime"را تایپ کرده و سپس متن موجود در آن گره را وارد کنید. مثلا، .qiime demux emp-singleتوجه داشته باشید که پایپلاین‌ها نوع خاصـــی از اقدام‌ها هســـتند که چندین اقدام را در یک دســـتور اجرا می‌کنند. جالب اســـت! در (برخی از) نمودارهای جریان، پایپلاین‌ها به صـــورت جعبه‌هایی نمایش داده می‌شـــوند که اقدامات داخل آن را در بر دارند .آثار با نوع معنایی فایل برچســـب‌گذاری شـــده‌اند. نگران نباشـــید - شـــما نیازی به تایپ آن نام‌های طولانی ندارید مگر اینکه بخواهید فایلی را به آن فرمت وارد کنید .تجســـم‌ها به طور مختلف با برچســـب "تجسم"، نامی که نمایانگر اطلاعات داخل تجسم است، یا با تصویری که نشان‌دهنده برخی از اطلاعات خوشمزه‌ای است که ممکن است در آن تجسم پیدا کنید، جایگزین شده‌اند ...نکات مفید برای مبتدیان قبل از اینکه ادامه دهیم، چند نکته مهم دیگر :راهنمای زیر به هیچ وجه جامع نیســـت. این تنها برخی از اقدامات اصلی در بیشتر پلاگین‌های "هسته‌ای QIIME 2 "را پوشش می‌دهد. پلاگین‌ها و اقدامات بسیاری برای کشف وجود دارند. کنجکاوید بیشتر یاد بگیرید؟ برای دیدن تمام پلاگین‌های نصب شده، "qiime -- help"را در ترمینال خود تایپ کنید) و "Enter" را فشـــار دهید .(برای دیدن تمام اقدامات پلاگین demux، "qiime demux --help"را تایپ کنید. برای یادگیری درباره روش emp-single درdemux ، "qiime demux emp-single --help"را تایپ کنید. حالا می‌دانید چگونه به مستندات کمک برای سایر پلاگین‌ها و اقدامات دسترسی پیدا کنید .نمودارهای جریان زیر به‌گونه‌ای طراحی شده‌اند که ساده‌ترین حالت ممکن باشند، بنابراین بسیاری از ورودی‌ها (به ویژه ورودی‌های اختیاری و متا دیتا) و خروجی‌ها (به ویژه خلاصه‌های آماری و سایر خروجی‌های جزئی) و تمام پارامترهای ممکن برای بیشتر اقدامات حذف شده‌اند. بسیاری از اقدامات اضافی (مثلاً برای نمایش خلاصه‌های آماری یا دستکاری جداول ویژگی‌ها ) نیز حذف شده‌اند. حالا که همه چیز درباره مستندات کمک را می‌دانید، از آن برای یادگیری بیشتر درباره اقدامات فردی و ســـایر اقدامات موجود در یک پلاگین اســـتفاده کنید (پیشـــنهاد: اگر یک پلاگین اقدامات اضافی دارد که در اینجا توضیح داده نشده است، احتمالاً برای بررسی خروجی اقدامات دیگر در آن پلاگین استفاده می‌شود .(متا دیتا مفهومی مرکزی در QIIME 2 است. ما به‌طور گسترده‌ای در این راهنما به متا دیتا نمی‌پردازیم، زیرا کار با متا دیتا به‌طور کامل در اینجا توضـــیح داده شـــده اســـت. خوب بخوانید و دور شوید .فایل‌های آثار (qza.) و تجسم‌ها (qzv.) در واقع فقط آرشیوهای فشرده‌ای هستند که شامل یک یا

Mohammad Amin Shirazi

چند فایل داده و فایل‌های جانبی حاوی اطلاعات منابع هستند. شما می‌توانید در هر زمانی یک اثر/تجسم را از حالت فشرده خارج کنید تا نگاهی به داخل آن بیاندازید، اما روش بهتر برای انجام این کار استفاده از "qiime tools export" است که در آموزش‌های صادرات ما به تفصیل توضیح داده شده است. امتحان کنید و ببینید می‌توانید چه نوع فرمت‌های فایلی در آثار مختلف ذخیره شده‌اند! اگر واقعاً می‌خواهید بیشتر در مورد ساختار فایل‌های آثار/تجسم‌ها بخوانید، ادامه دهید .هیچ راهی برای انجام کارها در QIIME 2وجود ندارد. همچنین یک "رویکرد "QIIME 2 وجود ندارد. بیشتر پلاگین‌ها و اقدامات در QIIME 2نرم‌افزارهای مستقل یا روش‌های از پیش موجود هستند QIIME .چسبی است که جادوی آن را به هم می‌چسباند. بسیاری از مسیرها از پای کوه به بالا می‌روند ...فراموش نکنید که به‌درستی ارجاع دهید! مطمئن نیستید چه چیزی را ارجاع دهید؟ برای دیدن ارجاع‌ها برای یک اقدام یا پلاگین خاص، دستورهای کمکی که یاد گرفته‌اید را تایپ کنید، اما "--help" را با "--citations" جایگزین کنید .حتی بهتر: به https://view.qiime2.org/ بروید و هر اثر یا تجسم QIIME 2 را به پنجره بکشید و رها کنید. به شرطی که آن فایل در QIIME 2018.4+ ایجاد شده باشد، تب "citations" باید اطلاعاتی درباره تمام ارجاع‌های مرتبط با تولید آن فایل را شامل شود. فوق‌العاده است .مرور مفهومی QIIME 2 حالا که واژه‌نامه و کلید را خواندیم، بیایید یک نمای کلی از جریان‌های کاری مختلف برای بررسی داده‌های توالی آمپلیکون را بررسی کنیم:

Mohammad Amin Shirazi




لبه‌ها و گره‌ها در این نمای کلی نمایانگر اقدامات خاص یا انواع داده‌ها نیســـتند، بلکه نمایانگر دســـته‌بندی‌های مفهومی هســتند، مثلاً انواع اصلی داده‌ها یا اهداف تحلیلی که ممکن است در یک آزمایش داشته باشیم. تمام این مراحل و اصطلاحات در ادامه به تفصیل توضیح داده می‌شوند .تمام داده‌ها باید به‌عنوان یک اثر QIIME 2وارد شوند تا توسط یک اقدام QIIME 2 استفاده شوند (به‌استثنای برخی از متا دیتا). کاربران مختلف ممکن اســت وارد این جریان کاری در مراحل مختلف شــوند. بیشــتر آن‌ها نوعی داده توالی خام )مثلاً داده‌های FASTQ یا (FASTA خواهند داشت که باید طبق طرح وارد کردن توالی مناسب وارد شود. ســایر کاربران ممکن اســت با داده‌های توالی دیمولتی‌پلاکس شــده یا حتی یک جدول ویژگی که توســط همکاران به آن‌ها داده شــده شروع کنند. راهنمای وارد کردن داده‌ها انواع داده‌های رایجی که کاربران باید وارد QIIME 2 کنند را پوشش می‌دهد .حالا که می‌دانیم واقعاً می‌توانیم وارد این جریان کاری نمای کلی در هر یک از گره‌ها شویم، بیایید از بخش‌های فردی آن عبور کنیم.

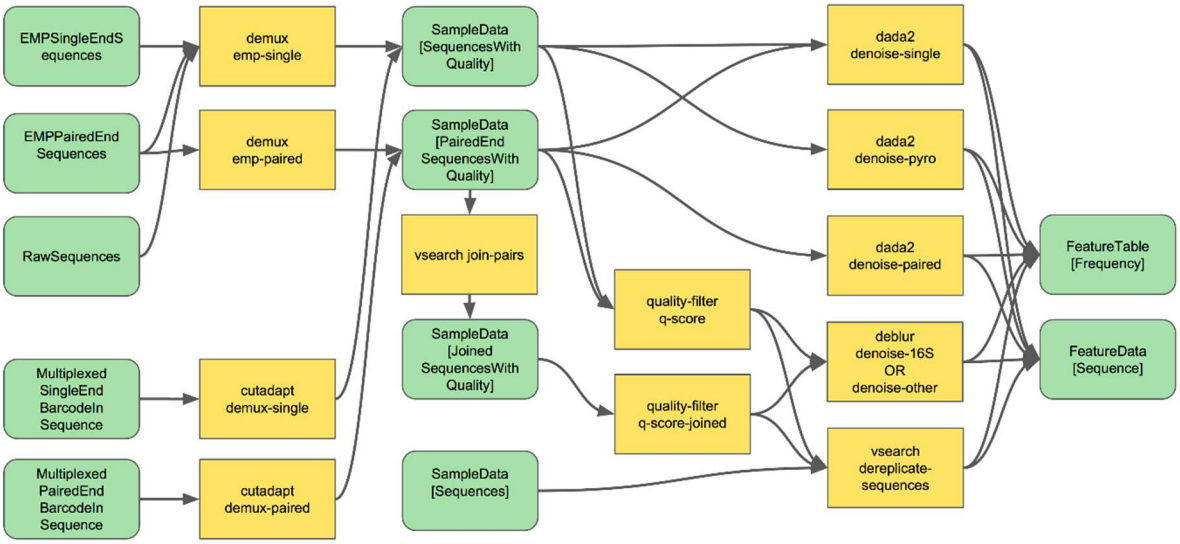
تمام آزمایش‌های توالی‌یابی آمپلیکون/متاژنتیک در نهایت از داده‌های توالی خام شـــروع می‌شـــوند. این داده‌ها معمولاً فرمت FASTQ هســتند که شــامل توالی‌های DNA و نمرات کیفیت برای هر باز هســتند .باید این

Mohammad An

توالی‌ها را دموپلکس کنیم تا مشخص کنیم هر توالی از کدام نمونه آمده است .توالی‌ها سپس باید به نسخه‌های آمپلیکون توالی‌های مختلف (ASVs) تبدیل شوند یا به واحدهای تاکسونومیک عملیاتی (OTUs)خوشه‌بندی شوند تا دو هدف زیر محقق شود :کاهش خطاهای توالی از بین بردن تکرار توالی‌ها جدول ویژگی‌ها و توالی‌های نمایشی که به دست می‌آیند، داده‌های کلیدی هستند. آن‌ها را از دست ندهید! جدول ویژگی‌ها اساساً یک ماتریس از نمونه‌ها در برابر مشاهدات است، یعنی تعداد دفعاتی که هر "ویژگی(OTU) "ها، ASVها و غیره (در هر نمونه از مجموعه داده‌ها مشاهده می‌شود .با این جدول ویژگی‌ها می‌توان کارهای زیادی انجام داد. تحلیل‌های رایج شامل موارد زیر است :طبقه‌بندی تاکسونومیکی توالی‌ها (یعنی، "چه گونه‌هایی وجود دارند؟") تحلیل‌های تنوع آلفا و بتا، یا اندازه‌گیری‌های تنوع درون و بین نمونه‌ها به ترتیب (یعنی، "نمونه‌های من چقدر شبیه به هم هستند؟") بسیاری از تحلیل‌های تنوع به شباهت فیلوژنتیک بین ویژگی‌ها بستگی دارند. اگر شما در حال توالی‌یابی نشانگرهای فیلوژنتیک) مثلاً ژن‌های (rRNA 16S باشید، می‌توانید این توالی‌ها را برای ارزیابی رابطه فیلوژنتیک بین ویژگی‌های مختلف خود هم‌راستا کنید .اندازه‌گیری‌های تفاوت در فراوانی تعیین می‌کنند که کدام ویژگی‌هاOTU) ها، ASVها، تاکسا و غیره (به‌طور معناداری در گروه‌های تجربی مختلف فراوان‌تر یا کمتر فراوان هستند .این تنها شروع کار است و بسیاری از تست‌های آماری و روش‌های ترسیم داده‌ها در اختیار شماست (QIIME 2) و فراتر از آن. دنیا در دستان شماست. بیایید شیرجه بزنیم .هشدار مراقب باشید! ما داریم وارد مفاهیم فنی پیچیده می‌شویم .آیا شما مفاهیم پایه و نوع‌های معنایی خود را مطالعه کرده‌اید؟ اگر هنوز این کار را نکرده‌اید، الان این کار را انجام دهید یا با خطر خود ادامه دهید .دموپلکسی خب! تصور کنید که ما به تازگی داده‌های FASTQ را از دستگاه توالی‌یابی دریافت کرده‌ایم. بیشتر دستگاه‌های توالی‌یابی نسل جدید توانایی آنالیز صدها یا حتی هزاران نمونه در یک باند/دستگاه را دارند. این کار با مولتی‌پلکسی این نمونه‌ها انجام می‌شود که به معنای مخلوط کردن تعداد زیادی نمونه است. چگونه می‌توانیم متوجه شویم هر توالی از کدام نمونه است؟ این معمولاً با افزودن یک بارکد منحصر به فرد (که به آن ایندکس یا تگ هم گفته می‌شود) به یکی یا هر دو انتهای هر توالی انجام می‌شود. شناسایی این توالی‌های بارکد و ارتباط آن‌ها با نمونه‌های مربوطه، به ما امکان دموپلکس کردن توالی‌ها را می‌دهد .اگر می‌خواهید شروع به دموپلکس کردن کنید، شما (یا فردی که نمونه‌ها را آماده کرده و توالی‌یابی کرده است) باید بدانید که هر بارکد به کدام نمونه تعلق دارد. اگر نمی‌دانید، با همکاران آزمایشگاهی یا مرکز

Mohammad Amin Shirazi

توالی‌یابی خود صحبت کنید. این اطلاعات بارکد باید در فایل متاداده نمونه‌ها شما گنجانده شود .فرایند دموپلکسی در QIIME 2 شبیه به این جریان کار است (برای حالا سمت راست این نمودار را نادیده بگیرید):




تولید می‌کند، یعنی تمام توالی‌های منحصر به فرد مشاهده شده در مجموعه داده) نیز در جریان کار دموپلکسی و دنوایزینگ نشان داده شده است و نقطه شروع ضروری برای تمام دیگر روش‌های خوشه‌بندی در QIIME 2است، همانطور که در اینجا نشان داده شده است:

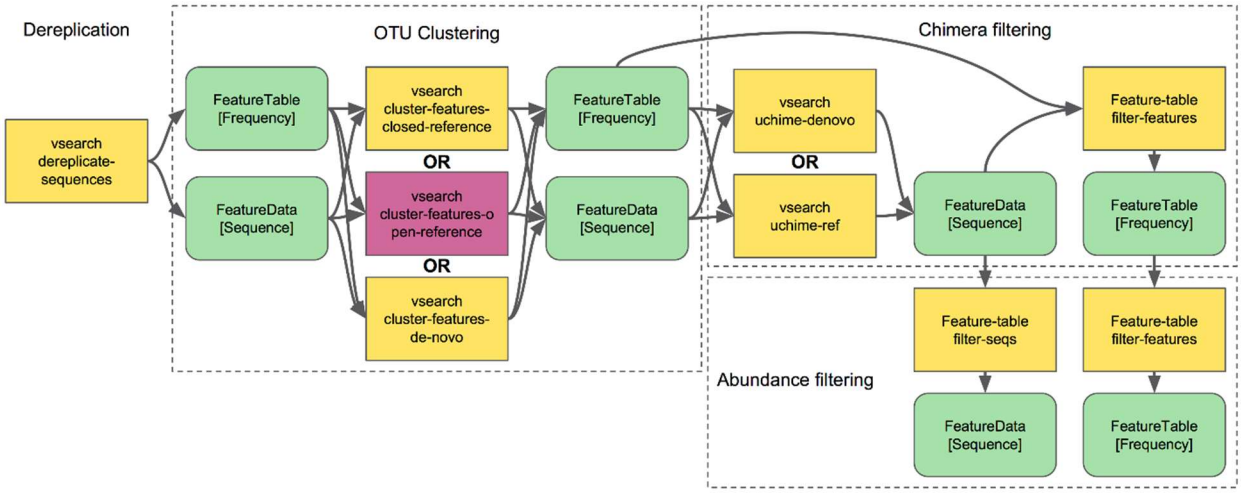
این نمودار جریان تمام مراحل دموپلکسی که در حال حاضر در QIIME 2 امکان‌پذیر است را شرح می‌دهد، که بستگی به نوع داده‌های خامی دارد که وارد کرده‌اید. معمولاً تنها یکی از اقدامات مختلف دموپلکسی که در q2-demux یا q2-cutadapt موجود است برای داده‌های شما کاربرد خواهد داشت و همان کافی خواهد بود .برای یادگیری بیشتر در مورد دموپلکسی و امتحان کردن آن، می‌توانید از آموزش‌های تصویری متحرک (برای داده‌های تک‌سر) و آموزش‌های خاک‌های آتاکاما (برای داده‌های جفتی) استفاده کنید. این آموزش‌ها شامل داده‌های فرمت EMP هستند (همانطور که در مستندات واردات توضیح داده شده است). آیا بارکدها و پرایمرها درون توالی‌هایتان وجود دارند؟ آموزش‌های cutadapt را برای استفاده از روش‌های دموپلکسی در q2-cutadapt ببینید. آیا توالی‌ها دو ایندکس دارند یا از فرمت‌های مختلط یا جهت‌گیری

Mohammad

متفاوت برخوردارند؟ به شدت دعا کنید .سپس به انجمن QIIME 2 مراجعه کنید تا ببینید آیا کسی راه حلی پیدا کرده است .توالی‌های جفتی در برخی مراحل تحلیل نیاز به پیوند دارند. اگر آموزش خاک‌های آتاکاما را دنبال کرده باشید، خواهید دید که این کار به طور خودکار هنگام دنوایزینگ با q2-dada2 انجام می‌شود. اما اگر می‌خواهید از q2-deblur یا یک روش خوشه‌بندی OTU (همانطور که در ادامه توضیح داده شده است) استفاده کنید، از q2-vsearch برای پیوند این توالی‌ها قبل از ادامه کار استفاده کنید، همانطور که در جریان کار دموپلکسی نشان داده شده است. برای یادگیری بیشتر در مورد پیوند توالی‌ها، آموزش پیوند توالی‌ها را ببینید .اگر شروع به کندن موهای خود کردید و از دهان شما کف می‌آید، ناامید نشوید QIIME 2 :معمولاً هرچه بیشتر پیش بروید، ساده‌تر می‌شود. وارد کردن و دموپلکسی داده‌های خام توالی‌یابی بخش بسیار پرتنش برای بیشتر کاربران جدید است .اما یک بار که به آن مسلط شوید، مثل تکه کیک خواهد بود .دنوایزینگ و خوشه‌بندی تبریک می‌گویم که به این مرحله رسیدید! مراحل دنوایزینگ و خوشه‌بندی کمی کمتر گیج‌کننده از وارد کردن و دموپلکسی داده‌ها هستند !نام‌های این مراحل بسیار توصیفی هستند :ما توالی‌های خود را دنوایز می‌کنیم تا توالی‌های پر سر و صدا را حذف و/یا اصلاح کنیم .ما توالی‌هایمان را از نظر تکرار کاهش می‌دهیم تا از تکرار و نیازهای اندازه فایل/حافظه در مراحل بعدی جلوگیری کنیم (نگران نباشید! شمارش هر تکرار را نگه می‌داریم) .ما توالی‌ها را خوشه‌بندی می‌کنیم تا توالی‌های مشابه (برای مثال، آن‌هایی که ≤ ۹۷٪ مشابه هم هستند) را به توالی‌های تکراری واحد تبدیل کنیم. این فرایند که به آن انتخاب OTU نیز گفته می‌شود، زمانی یک روش رایج بود که همزمان با حذف تکرارها، نوعی دنوایزینگ سریع و ابتدایی انجام می‌شد) برای گرفتن خطاهای تصادفی توالی‌یابی و PCR که باید نادر و مشابه به توالی‌های مرکزی فراوان‌تر باشند .(به جای آن از روش‌های دنوایزینگ استفاده کنید اگر می‌توانید. دوران تغییر کرده است. به آینده خوش آمدید .دنوایزینگ بیایید با دنوایزینگ شروع کنیم، که در سمت راست جریان کار دموپلکسی و دنوایزینگ نشان داده شده است . روش‌های دنوایزینگ موجود در QIIME 2 شامل DADA2 و Deblur هستند. برای یادگیری بیشتر در مورد این روش‌ها، می‌توانید مقالات اصلی هر یک را مطالعه کنید. نمونه‌هایی از DADA2 در آموزش‌های تصویری متحرک و آموزش مطالعه میکروبیوم مدفوعی (برای داده‌های تک‌سر) و آموزش خاک‌های آتاکاما (برای داده‌های جفتی) موجود هستند. نمونه‌هایی از Deblur نیز در آموزش‌های تصویری متحرک (برای داده‌های تک‌سر) و آموزش پیوند توالی‌ها (برای داده‌های جفتی) موجود است. توجه داشته باشید که

Mohammad Amin Shirazi

deblur ) همچنین (vsearch dereplicate-sequences باید با فیلترسازی ابتدایی بر اساس نمرات کیفیت انجام شود، اما این برای dada2 لازم نیست. هر دو روش Deblur و DADA2 شامل روش‌های بررسی شیمرا و فیلترسازی فراوانی داخلی هستند، بنابراین پس از این روش‌ها نیازی به فیلترسازی اضافی نیست .به طور ساده، این روش‌ها توالی‌های پر سر و صدا را فیلتر کرده، خطاهای توالی‌های حاشیه‌ای) در مورد (DADA2 را اصلاح کرده، توالی‌های شیمرا را حذف کرده، توالی‌های تک‌تایی را حذف کرده، توالی‌های جفتی دنوایز شده را) در مورد (DADA2 پیوند داده و سپس آن توالی‌ها را از نظر تکرار کاهش می‌دهند .ویژگی‌هایی که توسط روش‌های دنوایزینگ تولید می‌شوند، اسامی زیادی دارند، معمولاً برخی از انواع "نسخه توالی(SV) "، "نسخه آمپلیکون(ASV) "SV ، "نسخه واقعی"SV ، "دقیق ..."SV من معتقدم که ما قبلاً این‌ها را به عنوان ASV در این آموزش معرفی کرده‌ایم، پس بیایید نام‌گذاری خود را منظم نگه داریم .خوشه‌بندی در ادامه، روش‌های خوشه‌بندی را بررسی خواهیم کرد. دروپلیکیشن (ساده‌ترین روش خوشه‌بندی که عملاً OTU های ۱۰۰٪ منحصر به فرد را

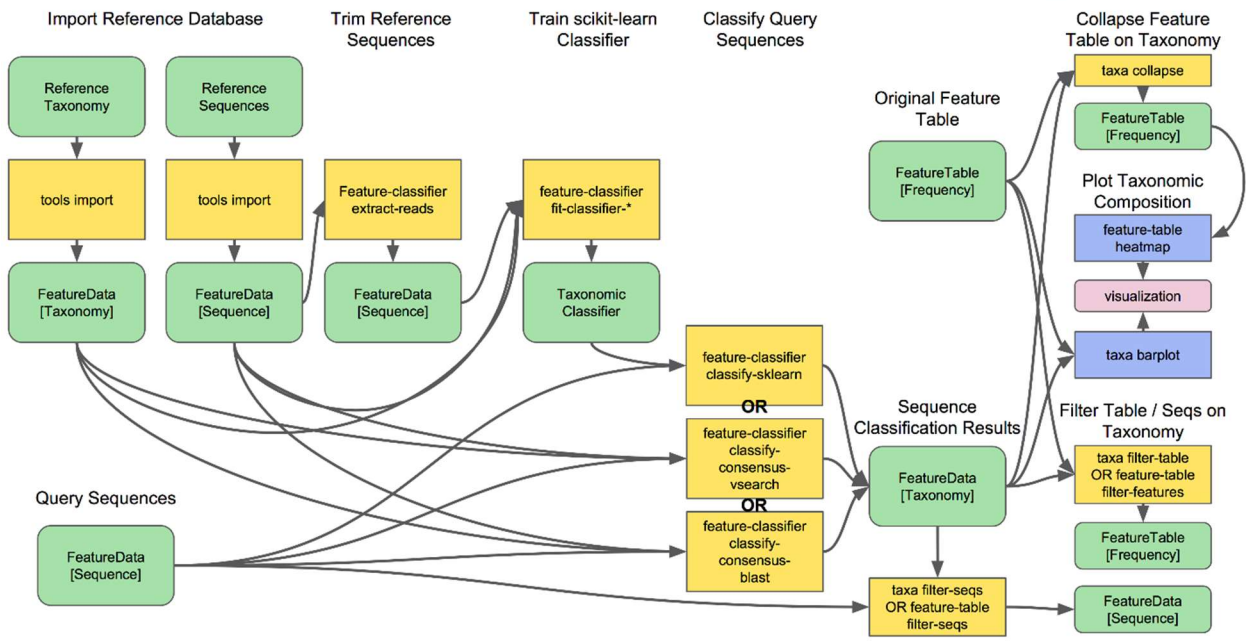



q2-vsearch استراتژی مختلف خوشه‌بندی OTU را پیاده‌سازی می‌کند: closed reference، de novo و open reference. همه این‌ها باید با فیلترسازی بر اساس نمرات کیفیت ابتدایی آغاز شده و با فیلترسازی شیمرا و فیلترسازی قوی OTU (که به آن روش بوکولیش نیز گفته می‌شود) دنبال شوند .

آموزش خوشه‌بندی OTU استفاده از چندین روش خوشه‌بندی q2-vsearch را نشان می‌دهد. فراموش نکنید که آموزش فیلترسازی شیمرا را هم بخوانید !ویژگی‌هایی که توسط روش‌های خوشه‌بندی تولید می‌شوند، به نام واحدهای تاکسونومیک عملیاتیOTU) ها (شناخته می‌شوند، که به زبان اسپرانتو به معنای زباله‌های نامناسب و نامشخص است .جدول ویژگی‌ها محصولات نهایی تمام روش‌ها/جریان‌های دنوایزینگ و خوشه‌بندی، دو نوع از مهم‌ترین آثار در یک جریان کار توالی‌یابی آمپلیکون هستند : FeatureTable[Frequency](جدول ویژگی) و FeatureData[Sequence] (توالی‌های نماینده). این‌ها دو تا از مهم‌ترین آثار در یک جریان کار توالی‌یابی آمپلیکون هستند و برای بسیاری از تحلیل‌های پایین‌دست استفاده می‌شوند، همانطور که در ادامه بحث خواهد شد. در واقع، جدول‌های ویژگی به عنوان سوابق مرکزی تمام مشاهدات هر نمونه، برای هر تحلیل QIIME 2 ضروری هستند. این آثار مهم باید دارای پلاگین قدرتمند خود باشند، یعنی .q2-feature-table ما در اینجا به همه عملیات این پلاگین نمی‌پردازیم (برخی از آن‌ها در زیر ذکر شده‌اند)، اما این پلاگین عملیات‌های مفیدی بر روی جدول‌های ویژگی انجام می‌دهد، بنابراین با مستندات آن آشنا شوید !تکرار می‌کنم: جدول‌های ویژگی در تحلیل‌های QIIME 2 مرکزی هستند. تقریباً تمام مراحل تحلیل (یعنی پس از دموپلکسی و دنوایزینگ/خوشه‌بندی) به نوعی به جدول‌های ویژگی مرتبط هستند. دقت کنید !توجه می‌خواهید ببینید کدام توالی‌ها با هر شناسه ویژگی مرتبط هستند؟ از دستور qiime metadata tabulate با استفاده از اثر FeatureData[Sequence] خود به عنوان ورودی استفاده کنید .تبریک می‌گویم !شما از مراحل وارد کردن، دموپلکسی و دنوایزینگ/خوشه‌بندی داده‌هایتان عبور کرده‌اید، که برای اکثر کاربران سخت‌ترین و پیچیده‌ترین مراحل هستند (فقط به این دلیل که روش‌های مختلفی برای انجام آن‌ها وجود دارد!). اگر به این مرحله رسیدید، بقیه آن آسان خواهد بود. حالا سرگرمی شروع می‌شود .طبقه‌بندی تاکسونومی و تحلیل‌های تاکسونومیک برای بسیاری از آزمایش‌ها، پژوهشگران می‌خواهند ارگانیسم‌های موجود در یک نمونه را شناسایی کنند. مثلاً، چه جنس‌ها یا گونه‌هایی در نمونه‌های من وجود دارند؟ آیا پاتوژن‌های انسانی در نمونه بیمار وجود دارند؟ چه چیزی در شراب من شنا می‌کند؟ ما می‌توانیم این کار را با مقایسه توالی‌های پرس‌وجو) یعنی ویژگی‌های ما، چهASV ها یاOTU ها باشند (با یک پایگاه داده مرجع از توالی‌هایی که ترکیب تاکسونومیک شناخته شده دارند انجام دهیم. پیدا کردن نزدیک‌ترین تطابق تنها کافی نیست — چون توالی‌های دیگری که تطابق‌های نزدیک یا تقریباً به اندازه آن دارند، ممکن است آنالیزهای تاکسونومیک

Mohammad Amin Shirazi

مختلفی داشــته باشــند. بنابراین ما از طبقه‌بندی‌کننده‌های تاکســونومی اســتفاده می‌کنیم تا نزدیک‌ترین ارتباط تاکسـونومیک را با درجه‌ای از اطمینان یا اجماع تعیین کنیم (که ممکن اسـت نام یک گونه نباشـد اگر نتوان آن را با قطعیت پیش‌بینی کرد!)، که این امر بر اسـاس تطابق، فرکانس‌های کِی-مر و غیره انجام می‌شـود. کسـانی که علاقه‌مند به یادگیری بیشـتر در مورد طبقه‌بندی تاکسـونومی در QIIME 2 هسـتند، می‌توانند تا زمان بازگشـت گاوها بخوانند .بیایید ببینیم که یک جریان کار طبقه‌بندی تاکسـونومی چگونه می‌تواند به نظر برسد:




q2-feature-classifierشامل سه روش مختلف طبقه‌بندی است classify-consensus-blast :و classify-consensus-vsearchهر دو روش‌های مبتنی بر تطابق هســتند که یک تخصــیص اجماعی را از بین N تطابق برتر پیدا می‌کنند. این روش‌ها فایل‌های FeatureData[Taxonomy] و FeatureData[Sequence] مرجع را به طور مستقیم می‌گیرند و نیازی به پیش‌آموزش ندارند classify-sklearn .روش‌های طبقه‌بندی مبتنی بر یادگیری ماشین را فراهم می‌کند و از نظر نظری می‌تواند هر یک از روش‌های طبقه‌بندی موجود در scikit-learn را اعمال کند. این طبقه‌بندها باید آموزش ببینند، برای مثال، برای یادگیری ویژگی‌هایی که بهترین تمایز را بین هر گروه تاکسـونومیک ایجاد می‌کنند. این گام اضـافی به فرآیند طبقه‌بندی افزوده

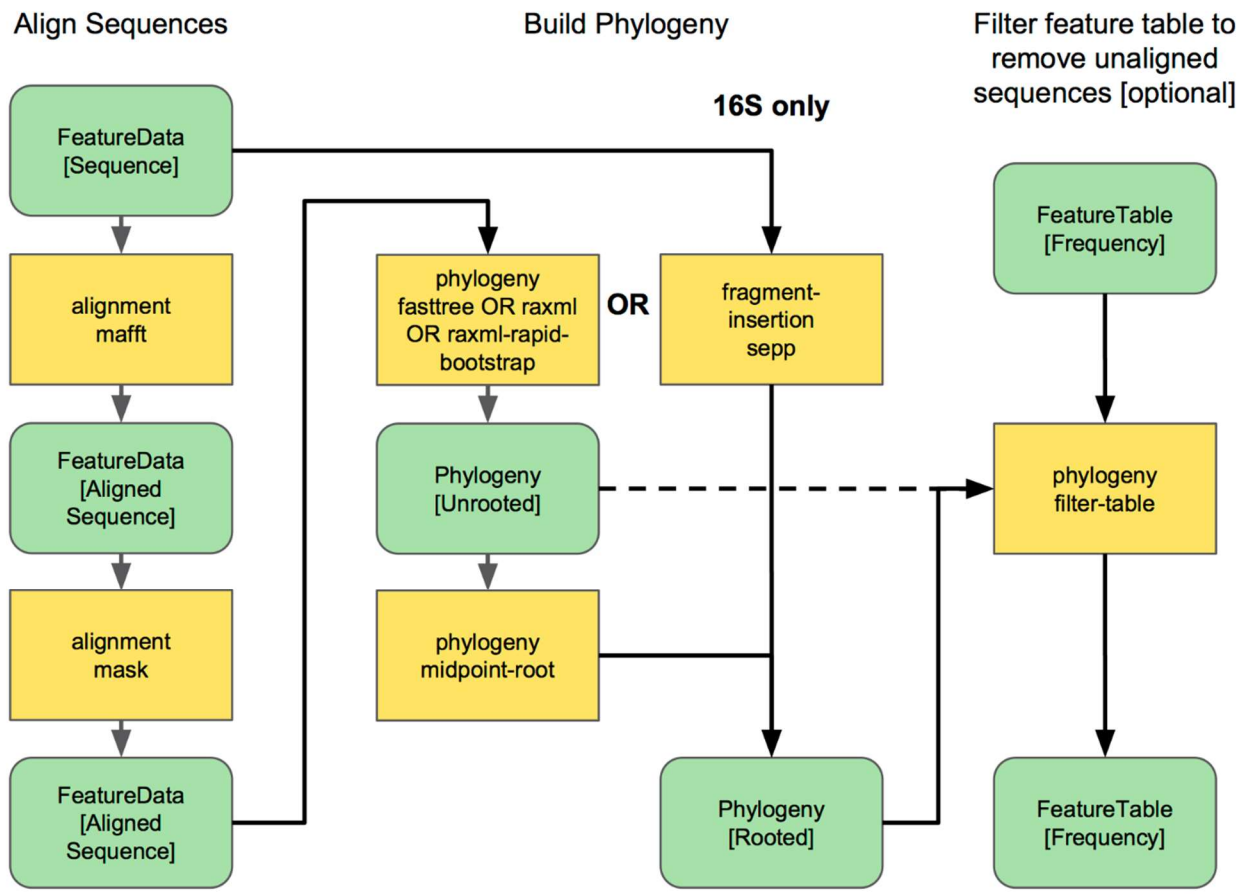
می‌شود. آموزش طبقه‌بندی خاص پایگاه داده مرجع و ژن‌های مارکر است و تنها یک‌بار برای هر ترکیب Marker-gene/reference database انجام می‌شود؛ سپس می‌توانید از آن طبقه‌بند برای استفاده‌های بعدی بدون نیاز به آموزش مجدد استفاده کنید !بیشتر کاربران نیازی به انجام گام آموزش نخواهند داشت، چون توسعه‌دهندگان خوب QIIME 2 چندین طبقه‌بند از پیش‌آموزش دیده برای استفاده عمومی فراهم کرده‌اند .کدام روش بهتر است؟ همه آن‌ها خوب هستند، در غیر این صورت ما وقت خود را برای معرفی آن‌ها نمی‌گذاشتیم .اما به طور کلی، classify-sklearn با استفاده از Naive Bayes classifier ممکن است کمی بهتر از سایر روش‌ها عمل کند، بر اساس معیارهای مختلف برای طبقه‌بندی توالی‌های ژن ۱۶ rRNA S و توالی‌های ITS قارچی. با این حال، ممکن است برای برخی کاربران دشوارتر و پر استرس‌تر باشد، چون نیاز به گام آموزش اضافی دارد. این گام آموزشی می‌تواند منابع حافظه زیادی مصرف کند و برای کاربرانی که نمی‌توانند از طبقه‌بندهای از پیش‌آموزش دیده استفاده کنند، یک مانع ایجاد کند. برخی کاربران همچنین ممکن است روش‌های مبتنی بر تطابق را ترجیح دهند، چون نحوه عملکرد آن‌ها بسیار شفاف‌تر است و پارامترهای آن‌ها راحت‌تر قابل تنظیم هستند (برای توضیحات بیشتر درباره این پارامترها و تنظیمات توصیه‌شده برای کاربردهای مختلف به لینک بالا مراجعه کنید) .زمان‌بری و مصرف حافظه در طبقه‌بندی ویژگی‌ها طبقه‌بندی ویژگی‌ها ممکن است زمان‌بر باشد. همه چیز بستگی به تعداد توالی‌ها و تعداد توالی‌های مرجع دارد. توالی‌های خوشه‌بندی شده OTU زمان بیشتری برای طبقه‌بندی نیاز دارند (چون اغلب تعداد بیشتری دارند). قبل از طبقه‌بندی، ویژگی‌های با فراوانی کم را از فایل توالی خود فیلتر کنید و از پایگاه‌های داده مرجع کوچکتر استفاده کنید، اگر نگرانی‌هایی در مورد زمان اجرا دارید. در عمل، در آزمایش‌های توالی‌یابی با اندازه "معمولی" (هرچند این یعنی چه؟ (ما شاهد تفاوت‌های زمانی از چند دقیقه (چند صد ویژگی) تا چند ساعت (صدها هزار ویژگی) برای تکمیل طبقه‌بندی هستیم. اگر می‌خواهید اعداد دقیق‌تری داشته باشید، به بنچمارک‌های عملکرد زمان اجرای طبقه‌بندها مراجعه کنید .طبقه‌بندی ویژگی‌ها ممکن است منابع حافظه زیادی مصرف کند. معمولاً حداقل ۴ گیگابایت RAM و حداکثر ۳۲+ گیگابایت نیاز است. این بستگی به اندازه توالی‌های مرجع، طول آن‌ها و تعداد توالی‌های پرس‌وجو دارد ...نتایج طبقه‌بندی ویژگی‌ها همه طبقه‌بندها یک FeatureData[Taxonomy] ایجاد می‌کنند که شامل فهرستی از طبقه‌بندی‌های تاکسونومیک برای هر توالی پرس‌وجو است .توجه می‌خواهید ببینید کدام توالی‌ها و تخصیص‌های تاکسونومیک با هر شناسه

Mohammad Amin Shirazi

ویژگی مرتبط هستند؟ از دستور qiime metadata tabulate با استفاده از آثار FeatureData[Taxonomy] و FeatureData[Sequence] خود به عنوان ورودی اســتفاده کنید .حالا که توالی‌های خود را طبقه‌بندی کرده‌ایم طبقه‌بندی تاکســونومیک دنیای جدیدی از امکانات را برای شــما باز می‌کند .در اینجا برخی از اقدامات اصلی هستند که با داشتن اثر FeatureData[Taxonomy] فعال می‌شوند :ادغام جدول ویژگی‌ها با taxa collapse! این کار تمام ویژگی‌هایی که تخصیص تاکسونومیک مشابه دارند را به یک ویژگی واحد ادغام می‌کند. این تخصیص تاکسونومیک به شناسه ویژگی در جدول ویژگی جدید تبدیل می‌شود. این جدول ویژگی جدید می‌تواند به همان شیوه‌های جدول اصلی استفاده شود. برخی کاربران ممکن است به‌طور خاص علاقه‌مند به انجام تحلیل‌های تنوع مبتنی بر تاکســونومی باشــند، اما حداقل کســانی که تاکسونومی را تخصیص می‌دهند، احتمالاً به تحلیل فراوانی متفاوت آن تاکسون‌ها علاقه‌مند خواهند بود. مقایسه تحلیل‌های فراوانی متفاوت با استفاده از تاکسون‌ها به عنوان ویژگی در مقابل استفاده از ASV ها یا OTU ها می‌تواند تشخیصی و اطلاعاتی برای تحلیل‌های مختلف باشد .ترسیم ترکیب تاکسونومیک برای مشاهده فراوانی انواع مختلف تاکسون‌ها در هر یک از نمونه‌های

شما. برای جزئیات بیشتر از taxa barplot و feature-table heatmap استفاده کنید .فیلتر کردن جدول ویژگی‌ها و توالی‌های نماینده (FeatureData[Sequence]) برای حذف برخی گروه‌های تاکسونومیک. این کار برای حذف آلاینده‌ها یا گروه‌های غیرهدف، مانند DNA میزبان شــامل توالی‌های میتوکندری یا کلروپلاســت مفید اســت. همچنین می‌تولند برای تمرکز بر گروه‌های خاص برای تحلیل عمیق‌تر مفید باشــد. برای اطلاعات بیشتر و مثال‌ها به آموزش فیلترسازی مراجعه کنید .هم‌راستاسازی توالی‌ها و ساخت درخت فیلوژنی بسیاری از تحلیل‌های تنوع به شباهت فیلوژنتیکی بین ویژگی‌های مختلف وابسته هستند. اگر شما در حال توالی‌یابی مارکرهای فیلوژنتیکی) مثلاً ژن‌های ۱۶ (S rRNA هستید، می‌توانید این توالی‌ها را هم‌راستا کنید تا رابطه فیلوژنتیکی بین هر یک از ویژگی‌هایتان را ارزیابی کنید. این درخت فیلوژنی سپس می‌تواند توسط سایر تحلیل‌های پایین‌دست، مانند تحلیل‌های فاصله UniFrac استفاده شود .گزینه‌های مختلف برای هم‌راستاسازی توالی‌ها و تولید یک درخت فیلوژنی در نمودار جریان زیر نشان داده شده است. برای توضیحات دقیق‌تر در مورد هم‌راستاسازی/ساخت درخت فیلوژنی، به آموزش q2-phylogeny و آموزش q2-fragment-insertion مراجعه کنید.

Mohammad Amin Shirazi



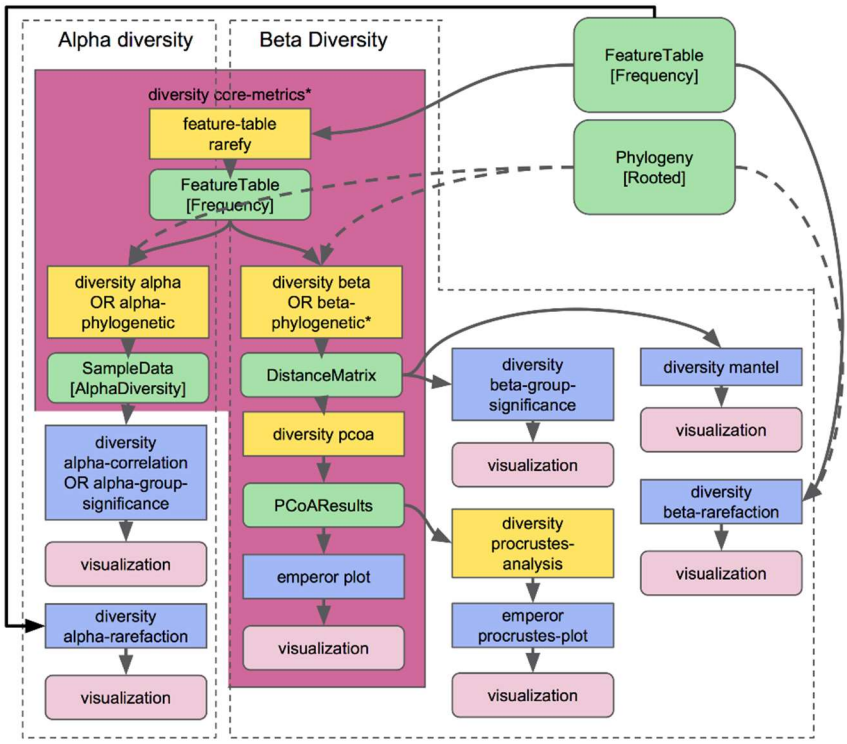

حالا که Phylogeny[Rooted] اثر هنری خود را داریم، توجه کنید که کجاها از آن اســـتفاده می‌شـــود . تحلیل تنوع در آزمایشـــات میکروبیوم، پژوهشـــگران معمولاً به ســـوالات زیر می‌پردازند :چه تعداد گونه OTU/ASV/مختلف در نمونه‌های من وجود دارد؟ چه میزان تنوع فیلوژنتیکی در هر نمونه وجود دارد؟ نمونه‌ها و گروه‌های مختلف نمونه‌ها چقدر مشابه یا متفاوت هستند؟ چه عواملی) مانندpH ، ارتفاع، فشار خون، محل بدن یا گونه میزبان (با تفاوت‌های ترکیب میکروبی و تنوع زیستی مرتبط هستند؟ این ســوالات می‌توانند از طریق تحلیل‌های آلفا-تنوع و بتا-تنوع پاسخ داده شـوند .آلفا-تنوع میزان تنوع در داخل یک نمونه خاص را اندازه‌گیری می‌کند و به ما نشـــان می‌دهد که چه تعداد گونه یا ویژگی) مثلاً ASVیا (OTU در آن نمونه وجود دارد .بتا-تنوع میزان تنوع یا تفاوت بین نمونه‌ها را اندازه‌گیری می‌کند

Mohammad

و به ما کمک می‌کند تا تفاوت‌های ترکیب میکروبی بین گروه‌های مختلف نمونه‌ها را بررسی کنیم. به‌عنوان‌مثال، اگر تنوع بتا بین گروه‌ها بیشتر باشد، این ممکن است نشان‌دهنده این باشد که نمونه‌های داخل یک گروه به یکدیگر شبیه‌تر از نمونه‌های گروه‌های دیگر هستند، که نشان می‌دهد عضویت در این گروه‌ها در شکل‌گیری ترکیب میکروبی آن‌ها تأثیر دارد .چگونگی انجام تحلیل‌های تنوع در QIIME 2 QIIME 2این امکان را به شما می‌دهد که تحلیل‌های آلفا و بتا-تنوع را انجام دهید و این تحلیل‌ها می‌توانند به روش‌های مختلفی انجام شوند. در اینجا یک نمای کلی از روند کار آورده شده است :تحلیل آلفا-تنوع: بعد از تولید جدول ویژگی و انتساب‌های تاکسونومی (از طریق طبقه‌بندی)، می‌توانید با استفاده از ابزار qiime diversity alpha شاخص‌های آلفا-تنوع را محاسبه کرده و نتایج را با ابزارهایی مانند qiime diversity alpha-rarefactionتجسم کنید .تحلیل بتا-تنوع: برای مقایسه تنوع بین چندین نمونه، معمولاً با استفاده از ابزار qiime diversity beta و انتخاب معیارهای مختلف فاصله) مثل UniFrac یا Bray-Curtis)تنوع بتا محاسبه می‌شود و نتایج آن با استفاده از تکنیک‌هایی مانند تحلیل مولفه‌های اصلی (PCoA) یا گام‌های غیرمتریک مقیاس‌بندی چندبعدی (NMDS) تجسم می‌شود .آزمون‌های آماری: پس از محاسبه تنوع آلفا و بتا، می‌توانید از آزمون‌های آماری مانند ANOVA یا PERMANOVA برای بررسی تفاوت‌های معنادار بین گروه‌ها استفاده کنید. این به شما کمک می‌کند تا فرضیه‌هایی مانند تفاوت در غنای گونه‌ای بین گروه‌ها را آزمایش کنید .تصویرسازی تنوع QIIME 2همچنین ابزارهای تصویری قدرتمندی برای نمایش نتایج تحلیل‌های تنوع فراهم می‌کند :آلفا-تنوع: می‌توانید تنوع درون‌نمونه‌ای را با استفاده از نمودارهای جعبه‌ای، نمودارهای نادرسازی یا نمودارهای ویولن برای مقایسه گروه‌های مختلف تجسم کنید .بتا-تنوع: می‌توانید تنوع بین‌نمونه‌ای را با استفاده از نمودارهای PCoA یا نمودارهای حرارتی مشاهده کنید و الگوهای خوشه‌بندی را بررسی کرده و تفاوت‌ها یا شباهت‌ها بین گروه‌های نمونه‌ها را شناسایی کنید .تحلیل‌های اضافی تنوع در QIIME 2 علاوه بر تحلیل‌های اصلی آلفا و بتا-تنوع، QIIME 2امکان انجام تحلیل‌های مختلف دیگری را نیز فراهم می‌آورد :آزمون تفاوت فراوانی: پس از تجمیع جدول ویژگی‌ها بر اساس گروه‌های تاکسونومی، می‌توانید فراوانی انواع مختلف تاکسون‌ها را در گروه‌های نمونه مقایسه کنید .تصویرسازی ترکیب تاکسونومی: با استفاده از ابزارهایی مانند نمودارهای میله‌ای تاکسون‌ها می‌توانید ترکیب نسبی تاکسون‌ها در نمونه‌های مختلف را مشاهده کرده و بینش‌هایی درباره ترکیب میکروبی نمونه‌ها به‌دست آورید .این تحلیل‌ها از طریق افزونه‌های مختلف QIIME 2 مانند

Mohammad Amin Shirazi

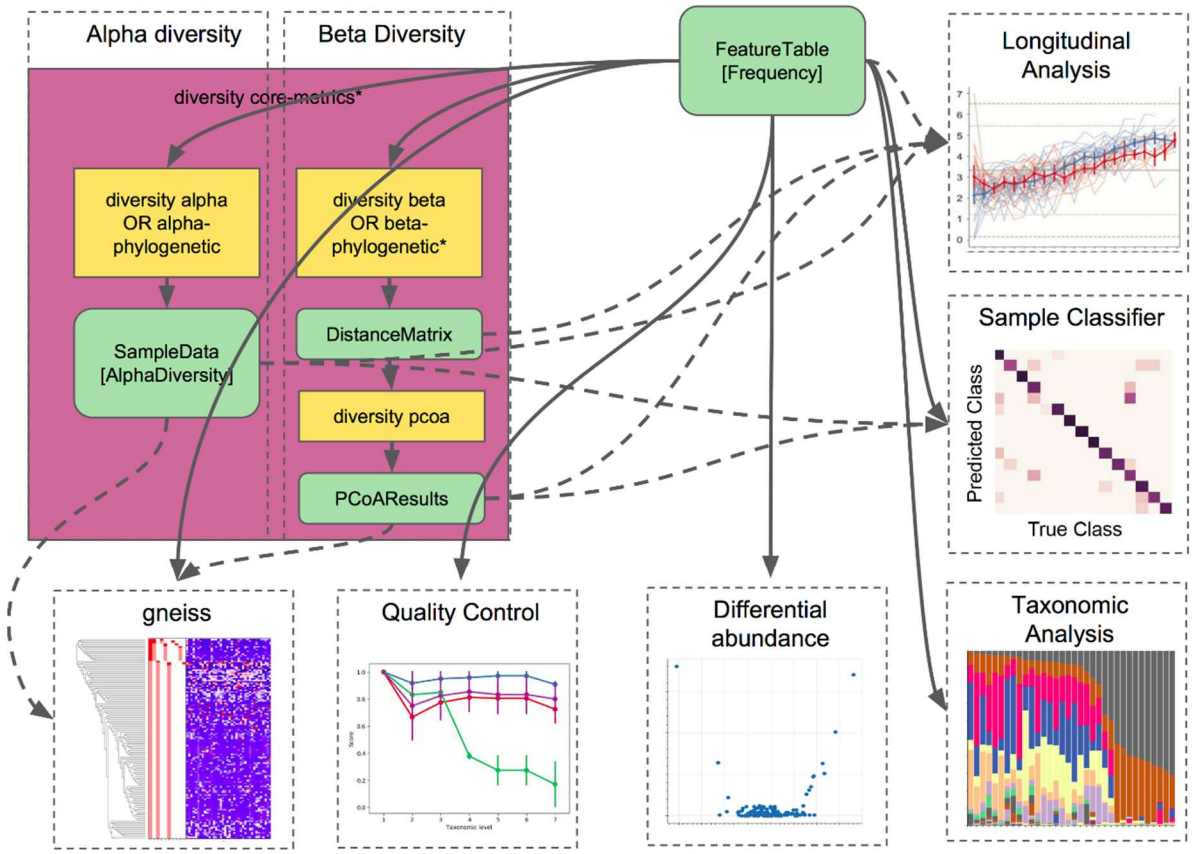
q2-diversityبرای شاخص‌های تنوع و q2-taxa برای تحلیل تاکسونومی انجام می‌شوند و به شما کمک می‌کنند تا داده‌های خود را به روش‌های مختلف بررسی کنید. با انجام این تحلیل‌ها می‌توانید الگوها و بینش‌های مهمی درباره نمونه‌های میکروبیوم خود به‌دست آورید و رابطه آن‌ها با عواملی مانند شرایط محیطی، وضعیت سلامتی یا اثرات درمانی را شبیه‌سازی کنید.




در پلاگین q2-diversity بسیاری از اعمال مفید و متنوع وجود دارند! برای یادگیری بیشتر، حتماً آن‌ها را بررسی کنید. همان‌طور که در نمودار جریان مشاهده می‌کنید، خطوط اصلی (diversity core-metrics* مثل core-metrics و core-metrics-phylogenetic) شامل دستورات متنوعی برای محاسبه شاخص‌های تنوع و تولید آثار اصلی مرتبط با تنوع هستند که می‌توانند در تحلیل‌های بعدی استفاده شوند. این آثار عبارتند از: SampleData[AlphaDiversity]: این آثار حاوی برآوردهای تنوع آلفا برای هر نمونه در جدول ویژگی شما هستند و اصلی‌ترین اثر برای تحلیل‌های تنوع آلفا است. DistanceMatrix: این آثار شامل ماتریس فاصله/تفاوت جفتی بین هر جفت نمونه در جدول ویژگی شما هستند و اصلی‌ترین اثر برای تحلیل‌های

تنوع بتا محسوب می‌شوند :PCoAResults .این آثار نتایج تحلیل مختصات اصلی (PCoA) را برای هر معیار فاصله/تفاوت شامل می‌شوند. تحلیل مختصات اصلی یک تکنیک کاهش ابعاد است که مقایسه‌های بصری تفاوت‌ها/شباهت‌ها بین نمونه‌ها را در فضای ۲ Dیا ۳ Dتسهیل می‌کند .این‌ها آثار اصلی مرتبط با تنوع هستند. آنها را به دقت نگهداری کنید! می‌توانیم این داده‌ها را در انواع تحلیل‌های بعدی یا در اعمال مختلف پلاگین q2-diversity که در نمودار جریان نشان داده شده‌اند، دوباره استفاده کنیم. بیشتر این اعمال در آموزش moving pictures توضیح داده شده‌اند، پس به آن مراجعه کنید تا اطلاعات بیشتری بدست آورید !توجه داشته باشید که در QIIME 2 شاخص‌های تنوع آلفا و بتا متنوعی وجود دارند. برای یادگیری بیشتر (و مشخص کردن اینکه باید به کدام مقاله ارجاع دهید!)، این منبع جالب را بررسی کنید که توسط یکی از کاربران دوست‌داشتنی QIIME 2 برای روشن‌سازی این موضوع به اشتراک گذاشته شده است. از استیفانی متشکریم !سرگرمی با جداول ویژگی در این مرحله شما جدول ویژگی، نتایج طبقه‌بندی تاکسونومی، نتایج تنوع آلفا و بتا را دارید. وای !تحلیل‌های تاکسونومی و تنوع، همان‌طور که در بالا توضیح داده شد، انواع تحلیل‌های اساسی هستند که اکثر کاربران QIIME 2 احتمالاً باید در نقطه‌ای از کار خود انجام دهند. اما این فقط آغاز کار است و تحلیل‌های پیشرفته زیادی در دسترس ما قرار دارند .

Mohammad Amin Shirazi




در اینجا یک مرور کلی از برخی تحلیل‌های پیشرفته در QIIME 2 داریم که هرکدام از این تحلیل‌ها آموزش‌های اختصاصی خود را دارند :تحلیل داده‌های طولی: پلاگین q2-longitudinal برای انجام تحلیل‌های آماری در آزمایش‌های طولی طراحی شده است، یعنی جایی که نمونه‌ها از بیماران/موضوعات/مکان‌ها به طور مکرر در طول زمان جمع‌آوری می‌شوند. این شامل مطالعات طولی بر روی تنوع آلفا و بتا و برخی نمودارهای تعاملی جالب است .پیش‌بینی آینده (یا گذشته) :پلاگین q2-sample-classifier برای انجام تحلیل‌های یادگیری ماشین بر روی داده‌های ویژگی است. مدل‌های طبقه‌بندی و رگرسیون هر دو پشتیبانی می‌شوند. این به شما امکان می‌دهد که کارهایی مانند موارد زیر را انجام دهید :پیش‌بینی متادیتای نمونه به عنوان تابعی از داده‌های ویژگی (مثلاً آیا می‌توانیم از نمونه مدفوع برای پیش‌بینی حساسیت به سرطان استفاده کنیم؟ یا پیش‌بینی کیفیت شراب بر اساس ترکیب میکروبی انگور قبل از تخمیر؟) شناسایی ویژگی‌هایی که پیش‌بینی‌کننده ویژگی‌های مختلف نمونه هستند .سنجش نرخ‌های رشد میکروبی (مثلاً برای پیگیری توسعه طبیعی میکروبیوم در روده نوزاد و تأثیرات سوء تغذیه مداوم یا

Mohammad

آنتی‌بیوتیک‌ها، رژیم غذایی، و روش زایمان) .پیش‌بینی نمونه‌های خارج از محدوده و نمونه‌های اشـــتباه برچســـب‌گذاری شـــده .آزمون تفاوت در فراوانی: برای تعیین اینکه کدام ویژگی‌ها در گروه‌های مختلف نمونه‌ها به طور معناداری بیشتر یا کمتر فراوان هستند QIIME 2 .در حال حاضر چندین رویکرد مختلف برای آزمایش تفاوت در فراوانی پشـتیبانی می‌کند، از جمله ancom و ancom-bc اعمال در پلاگین q2-composition).ارزیابی و کنترل کیفیت داده‌ها: پلاگین q2-quality-control برای ارزیابی و کنترل کیفیت داده‌های توالی طراحی شـــده اســـت. این شـــامل اعمالی اســـت که :دقت روش‌های مختلف بیوانفورماتیکی یا مولکولی یا تغییرات کیفیت بین اجرای آزمایش‌ها را بررســـی می‌کند. این اعمال معمولاً زمانی استفاده می‌شوند که نمونه‌هایی با ترکیب شناخته‌شده وجود داشته باشد، مانند جوامع ساختگی، زیرا دقت به عنوان شباهت بین ترکیب‌های مشاهده‌شده و پیش‌بینی‌شده محاسبه می‌شود. اما کاربردهای خلاقانه‌تری نیز ممکن است امکان‌پذیر باشد ...توالی‌ها را بر اساس تطابق با یک پایگاه داده مرجع فیلتر می‌کند یا توالی‌هایی که حاوی بخش‌های خاصی از DNA (مثلاً توالی‌های پرایمر) هستند. این برای حذف توالی‌هایی که با گروه خاصی از ارگانیسم‌ها، DNAغیرهدف، یا سایر مواد بی‌ربط تطابق دارند مفید است . و این فقط یک مرور کلی بود QIIME 2 !همچنان در حال رشد است، پس منتظر پلاگین‌های جدید در نسخه‌های آینده باشید ، و چشم‌های خود را به روی پلاگین‌های شخص ثالث که به توسعه قابلیت‌های QIIME 2ادامه خواهند داد باز نگه دارید .حالا بروید و لذت ببرید !

در QIIME 2 ، تفاوت بین فرمت‌های EMP و CASAVA مربوط به نحوه‌ی ســـاخت و ذخیره‌ســـازی داده‌های توالی‌های میکروبی است. هر کدام از این فرمت‌ها به روش خاصی اطلاعات را سازماندهی می‌کنند . EMP format (Earth Microbiome Project): فرمت EMP یک ســـاختار اســـتاندارد برای داده‌های میکروبیوم اســـت که معمولاً برای داده‌های 16S rRNA اســـتفاده می‌شـــود. این فرمت شـــامل یک فایل داده توالی (FASTQ)و یک فایل متادیتا است که اطلاعات مربوط به نمونه‌ها را شـرح می‌دهد.معمولاً در پروژه‌هایی که بر اســـاس 16S rRNA و داده‌های میکروبیوم از مجموعه‌های بزرگ نمونه‌ها کار می‌کنند، اســـتفاده می‌شـــود CASAVA format: .فرمت CASAVA یک فرمت قدیمی‌تر و رایج اســـت که توســـط نرم‌افزار CASAVA )که توسط Illumina برای پردازش داده‌های توالی‌دهی خود طراحی شده (استفاده می‌شود.این فرمت شامل سه فایل است R1, R2, :و index که نشان‌دهنده‌ی توالی‌های پیشرفته (paired-end reads)

Mohammad Amin Shirazi

و اطلاعات مربوط به بارکدها است.این فرمت معمولاً برای داده‌های توالی‌دهی Illumina و برای نمونه‌گیری بارکدی اســـتفاده می‌شـــود.تفاوت‌ها EMP format:معمولاً برای داده‌های 16S rRNA و میکروبیوم‌های بزرگتر استفاده می‌شود و با اطلاعات متادیتا به راحتی قابل استفاده است CASAVA format.به‌طور خاص برای داده‌های توالی‌دهی Illumina با فرمت paired-end و بارکدهای نمونه طراحی شـــده اســـت .در QIIME 2، هنگام وارد کردن داده‌ها، باید دقت کنید که فرمت درست را انتخاب کنید تا ابزارها بتوانند آن را به‌درستی پردازش کنند.

۱- we need install mini conda:

check Linux version by command [uname -m]

If it outputs x86_64, go with Miniconda3 Linux 64-bit.

If it outputs aarch64, choose Miniconda3 Linux-aarch64 64-bit.

If it outputs s390x, then choose Miniconda3 Linux-s390x 64-bit.

now we can select from this link to download:

https://docs.anaconda.com/miniconda/#miniconda-latest-installer-links

go to terminal:

```
wget https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
bash Miniconda3-latest-Linux-x86_64.sh
```

Then, follow the prompts:

Mohammad Amin Shirazi

Press Enter to review the license.Type yes to accept the license.Press Enter to confirm the installation location or specify a custom path.Type yes to add Miniconda to your PATH for easy access.

conda –version

conda update conda

QIIME 2 توزیع‌های مختلفی برای کاربردهای خاص دارد و شـــما می‌توانید بر اســـاس نیاز خود یکی از آن‌ها را انتخاب کنید:

Amplicon Distribution: مناســـب برای تحلیل داده‌های توالی‌یابی ژن‌های آمپلیکون (مثل 16S rRNA ). اگر قصـد تحلیل داده‌های آمپلیکون را دارید (که در آن از ژن‌های خاصـی مانند 16S اسـتفاده می‌شـود)، این توزیع مناسب است.

Metagenome Distribution: شـــامل ابزارهایی برای تحلیل متاژنوم‌ها اســـت که بر توالی‌یابی تمام ژنوم‌های میکروبی تمرکز دارد. اگر کار شـما روی کل ژنوم‌های میکروبی یا پروژه‌های متاژنومیکس اسـت، این توزیع را انتخاب کنید.

Pathogenome Distribution: مناســب برای تحلیل داده‌های پاتوژنومیک که بر پاتوژن‌ها تمرکز دارد. این توزیع بیشتر برای پروژه‌های خاص مرتبط با پاتوژن‌ها مناسب است.

Tiny Distribution: یک نسـخه کوچک و سـبک از QIIME 2 اسـت که فقط ابزارهای پایه‌ای را شـامل می‌شـود. برای یادگیری و تسـت کردن QIIME 2 بدون نیاز به فضـای زیادی اسـتفاده می‌شـود، اما برای تحلیل‌های کامل مناسب نیست.

Mohammad Amin Shirazi

اگر شما برای تحلیل داده‌های آمپلیکون (مانند 16S rRNA) از QIIME 2 استفاده می‌کنید، فقط دستور زیر را اجرا کنید:

```
conda env create -n qiime2-amplicon-2024.10 --file https://data.qiime2.org/distro/amplicon/qiime2-amplicon-2024.10-py310-linux-conda.yml
```

activate conda:

```
conda activate qiime2-amplicon-2024.10
```

```
qiime –help
```

START WITH DATA:

```
git clone https://github.com/gibbons-lab/isb_course_2020 materials
```

for check all type:

```
qiime tools list-types
```

for manifest making:

https://docs.qiime2.org/2024.10/tutorials/importing/

output:

Mohammad Amin Shirazi

https://docs.qiime2.org/2024.10/tutorials/exporting/

```
qiime tools import\
-- type 'SampleData[SequencesWithQuality]\ '
-- input-path manifest.tsv\
-- output-path cdiff.qza\
-- input-format SingleEndFastqManifestPhred33V2
```

برای وارد کردن داده‌ها به QIIME 2 و تبدیل فایل‌های خام به فرمت QIIME 2، باید از ابزارهای مختلفی استفاده کنید که به شما اجازه می‌دهند داده‌های خود را از فرمت‌های مختلف (مثل FASTQ یا دیگر فرمت‌های خام) به فرمت‌های خاص QIIME 2 (مثل FeatureData[Sequence] و FeatureData[Taxonomy]) تبدیل کنید. این فرایند معمولاً شامل مراحل زیر است:

مراحل وارد کردن داده‌ها به QIIME 2:

بررسی داده‌های خام:

قبل از شروع، مطمئن شوید که فایل‌های خام شما در فرمت مناسب برای پردازش قرار دارند. برای داده‌های Illumina، معمولاً فایل‌های FASTQ (شامل ریدهای Forward و Reverse) و فایل‌های بارکد وجود دارند.

استفاده از qiime tools import برای وارد کردن داده‌ها:

Mohammad Amin Shirazi

برای تبدیل داده‌های خام به فرمت‌های QIIME 2، از دستور qiime tools import استفاده می‌کنیم. این دستور به شما اجازه می‌دهد که داده‌ها را به انواع مختلف فرمت‌های QIIME 2 مانند FeatureData[Sequence] یا FeatureData[Taxonomy] وارد کنید.

وارد کردن داده‌های توالی (مثل داده‌های FASTQ): برای وارد کردن فایل‌های FASTQ (که شامل توالی‌های خام است)، دستور زیر را استفاده می‌کنید:

```
qiime tools import\
-- type 'SampleData[PairedEndSequencesWithQuality]\ '
-- input-path /path/to/your/raw_data\ /
-- input-format CasavaOneEightLanelessPerSampleDirFmt\
-- output-path demux-paired-end.qza
```

در اینجا:

--type 'SampleData[PairedEndSequencesWithQuality]': نوع داده‌ای که وارد می‌کنید (داده‌های توالی با کیفیت، در اینجا نمونه‌های paired-end).

--input-path: مسیر فایل‌های ورودی (فایل‌های خام).

Mohammad Amin Shirazi

--input-format: فرمت ورودی (در این مثال، CasavaOneEightLanelessPerSampleDirFmt برای داده‌های Illumina است).

--output-path: مسیر و نام فایل خروجی که فرمت QIIME 2 را دریافت خواهد کرد (در اینجا، demux-paired-end.qza).

وارد کردن داده‌های بارکد (اگر دارید): اگر فایل‌های بارکد دارید، می‌توانید از دستور مشابه برای وارد کردن آن‌ها استفاده کنید:

```
qiime tools import\
-- type 'SampleData[SequencesWithQuality]\ '
-- input-path /path/to/barcodes\ /
-- input-format SingleEndFastqManifestPhred33V2\
-- output-path barcodes.qza
```

وارد کردن داده‌های متادیتا (اختیاری): شما می‌توانید یک فایل متادیتا شامل اطلاعات مربوط به نمونه‌ها (مثل وضعیت نمونه، نوع آزمایش، شرایط محیطی و غیره) وارد کنید. این اطلاعات در تجزیه و تحلیل‌های بعدی مفید خواهند بود:

```
qiime tools import\
-- type 'SampleData[Metadata]\ '
```

Mohammad Amin Shirazi

```
-- input-path /path/to/metadata.tsv\
-- output-path metadata.qza
```

بررسی وارد شدن صحیح داده‌ها: پس از وارد کردن داده‌ها، می‌توانید از دستور qiime tools view برای مشاهده و تأیید صحت داده‌ها استفاده کنید. مثلاً:

```
qiime tools view demux-paired-end.qza
```

انجام فرایندهای اولیه (مثل دموکس یا فیلتر کردن داده‌ها): پس از وارد کردن داده‌ها، معمولاً باید داده‌ها را فیلتر کنید یا توالی‌های نامعتبر را حذف کنید. برای این کار، می‌توانید از دســـتورات QIIME 2 مانند demux summarize برای بررسی کیفیت توالی‌ها استفاده کنید:

```
qiime demux summarize\
-- i-data demux-paired-end.qza\
-- o-visualization demux-summary.qzv
```

انواع داده‌های QIIME 2:

Mohammad Amin Shirazi

پس از وارد کردن داده‌ها به QIIME 2، شـــما می‌توانید از انواع مختلف داده‌ها برای تجزیه و تحلیل‌های بیشـــتر استفاده کنید:

FeatureData[Sequence]: شامل توالی‌های ویژگی‌ها (برای نمونه، ASVs یا OTUs).

FeatureData[Taxonomy]: شامل طبقه‌بندی‌های تاکسونومی برای ویژگی‌ها.

SampleData[AlphaDiversity]: شامل نتایج تنوع آلفا.

DistanceMatrix: شامل ماتریس فاصله/عدم شباهت بین نمونه‌ها.

PCoAResults: شامل نتایج تحلیل مؤلفه‌های اصلی.

نکات:

فرمت‌های مختلف داده‌ها (مثل FASTQ، CASAVA، و EMP) ممکن اســـت نیاز به دســـتور واردات متفاوت داشته باشند.

برخی از فرمت‌ها نیاز به فایل‌های متادیتا دارند که اطلاعات اضافی را در مورد نمونه‌ها فراهم می‌کنند.

با انجام این مراحل، داده‌های شـــما آماده‌ی اســـتفاده در QIIME 2 برای انجام تحلیل‌های مختلف میکروبیوم و تنوع زیستی خواهند بود.

Mohammad Amin Shirazi

برای وارد کردن فایل‌های FASTQ به QIIME 2، در واقع لازم است یک فایل به نام manifest file ایجاد کنیم که به QIIME 2 نشان دهد فایل‌های FASTQ کجا قرار دارند و اطلاعاتی مثل نام نمونه‌ها و مسیر فایل‌ها را در خود داشته باشد.

ساخت فایل Manifest

این فایل معمولاً در قالب TSV یا CSV (که همان فایل‌های متنی جدولی هستند) ساخته می‌شود و به QIIME 2 کمک می‌کند مسیر و اطلاعات فایل‌های FASTQ را شناسایی کند. ساختار فایل به شکل زیر است:

هر ردیف، شامل اطلاعات یک نمونه و مسیر مربوط به فایل FASTQ آن است.

یک فایل معمولی manifest به صورت زیر ساخته می‌شود:

نمونه‌ای از فایل manifest.tsv (Single-End Data)

| sample-id | absolute-filepath | direction |
|---|---|---|
| sample1 | path/to/sample1.fastq.gz/ | forward |
| sample2 | path/to/sample2.fastq.gz/ | forward |
| sample3 | path/to/sample3.fastq.gz/ | forward |

sample-id: شناسه‌ای که برای هر نمونه در نظر می‌گیری.

absolute-filepath: مسیر کامل فایل FASTQ برای هر نمونه.

Mohammad Amin Shirazi

direction: برای داده‌های single-end همیشـــه مقدار forward دارد؛ برای paired-end می‌تواند forward یا reverse باشد.

quality control:

/https://docs.qiime2.org/2024.10/tutorials/moving-pictures-usage

qiime demux summarize --i-data cdiff.qza --o-visualization qualities.qzv

/https://view.qiime2.org

برای بررسی کیفیت توالی‌ها در QIIME 2، از دستور qiime demux summarize استفاده می‌کنیم که توالی‌های وارد شـــده را به صـــورت گرافیکی و آماری بررســـی می‌کند و نتیجه را در قالب یک فایل qzv . (Visualizations) ذخیره می‌کند. این فایل .qzv را می‌توانیم بعداً در QIIME 2 بررســـی کنیم تا کیفیت توالی‌ها و هرگونه مشکل احتمالی را شناسایی کنیم.

مراحل کیفیت‌سنجی توالی‌ها:

وارد کردن داده‌های توالی (اگر هنوز وارد نکرده‌اید): ابتدا، اگر داده‌های توالی خود را وارد نکرده‌اید، باید آن‌ها را به فرمت QIIME 2 تبدیل کنید. برای مثال، اگر داده‌های شما به صورت paired-end هستند، دستور واردات به شکل زیر خواهد بود:

qiime tools import\

Mohammad Amin Shirazi

```
-- type 'SampleData[PairedEndSequencesWithQuality]\ '
-- input-path /path/to/raw_data\ /
-- input-format CasavaOneEightLanelessPerSampleDirFmt\
-- output-path demux-paired-end.qza
```

استفاده از دستور qiime demux summarize: پس از وارد کردن داده‌ها، می‌توانید از دستور زیر برای بررسی کیفیت توالی‌ها استفاده کنید:

```
qiime demux summarize\
-- i-data demux-paired-end.qza\
-- o-visualization demux-summary.qzv
```

در اینجا:

--i-data: ورودی داده‌های توالی وارد شده (در اینجا، فایل demux-paired-end.qza که توالی‌های paired-end را شامل می‌شود).

--o-visualization: خروجی یک فایل .qzv است که نتایج کیفیت‌سنجی را در قالب تصویری ذخیره می‌کند. در این مثال، فایل خروجی demux-summary.qzv است.

Mohammad Amin Shirazi

مشاهده فایل خروجی: پس از اجرای دستور بالا، فایل خروجی demux-summary.qzv ایجاد خواهد شد. برای مشاهده و تجزیه و تحلیل کیفیت توالی‌ها، باید این فایل را در QIIME 2 باز کنید:

```
qiime tools view demux-summary.qzv
```

این فایل شامل نمودارهایی است که نشان می‌دهند که آیا توالی‌ها در هر موقعیت از توالی دارای کیفیت مناسب هستند یا خیر. معمولاً این گراف‌ها شامل اطلاعاتی در مورد:

کیفیت کلی توالی‌ها

نوسانات کیفیت در طول توالی‌ها

تعداد توالی‌ها برای هر نمونه

تحلیل کیفیت:

نمودار کیفیت توالی: نشان می‌دهد که کیفیت توالی‌ها در موقعیت‌های مختلف (در ابتدا، وسط یا انتهای توالی) چگونه است. معمولاً در انتهای توالی‌ها کیفیت کاهش می‌یابد.

تعداد توالی‌ها: تعداد توالی‌های موجود در هر نمونه و اینکه آیا توالی‌های کافی برای هر نمونه وجود دارد یا خیر.

نکات:

Mohammad Amin Shirazi

اگر کیفیت توالی‌ها پایین است (به ویژه در انتهای توالی‌ها)، ممکن است نیاز به فیلتر کردن یا کوتاه کردن توالی‌ها (trimming) داشته باشید.

برای رفع مشکلات کیفیتی، ممکن است از فیلتر کردن توالی‌های با کیفیت پایین یا استفاده از توالی‌های کوتاه‌تر بهره ببرید.

با استفاده از این روش، می‌توانید مطمئن شوید که توالی‌های شما برای تجزیه و تحلیل‌های بعدی در QIIME 2 آماده هستند.

trimming:

```
qiime dada2 denoise-single\
--  i-demultiplexed-seqs cdiff.qza\
--  p-trunc-len 150\
--  output-dir dada2 –verbose
```

این بخش از دستور به QIIME 2 می‌گوید که از پلاگین DADA2 برای denoising یا فیلتر کردن نویز در داده‌های توالی‌یابی تک‌انتهایی (single-end) استفاده کند. این مرحله به حذف نویزهای رایج در داده‌های توالی‌یابی کمک می‌کند.

این آرگومان نشان می‌دهد که فایل ورودی داده‌های توالی‌یابی demultiplex شده با نام cdiff.qza است. این فایل باید قبلاً در مرحله import یا demultiplexing ایجاد شده باشد.

رفع نویز توالی‌ها با DADA2 در QIIME 2

Mohammad Amin Shirazi

در این مرحله، افزونه DADA2 در QIIME 2 برای انجام تصفیه توالی‌ها (denoising)، فیلتر کردن (trimming)، حذف کیمراها (chimera removal) و شناسایی ASV (Amplicon Sequence Variants) استفاده می‌شود. هدف اصلی این فرآیند این است که کیفیت داده‌های توالی‌برداری شده بهبود یابد و تنها توالی‌های واقعی (بدون نویز) برای تحلیل‌های بیشتر در دسترس قرار گیرند.

مراحل رفع نویز توالی‌ها با DADA2:

وارد کردن داده‌ها (Importing Data): ابتدا، داده‌های توالی باید وارد QIIME 2 شوند. فرض می‌کنیم که داده‌ها به صورت paired-end (دو توالی مرتبط با هر نمونه) هستند. این داده‌ها باید با فرمت مناسب وارد شوند، برای مثال:

```
qiime tools import\
-- type 'SampleData[PairedEndSequencesWithQuality]\ '
-- input-path /path/to/raw_data\ /
-- input-format CasavaOneEightLanelessPerSampleDirFmt\
-- output-path demux-paired-end.qza
```

در این دستور:

Mohammad Amin Shirazi

--input-path مسیر داده‌های خام شما است.

--output-path خروجی‌ای است که داده‌های وارد شده را در فرمت .qza ذخیره می‌کند.

استفاده از DADA2 برای رفع نویز: در این مرحله، از دستور qiime dada2 denoise-paired برای پردازش توالی‌های وارد شده و یافتن ASVها (Amplicon Sequence Variants) استفاده می‌کنیم. این دستور به طور خودکار فیلترینگ، تریمینگ و حذف کیمراها را انجام می‌دهد.

برای داده‌های paired-end (که دو فایل توالی برای هر نمونه دارند) به شکل زیر عمل می‌کنیم:

```
qiime dada2 denoise-paired\
-- i-demultiplexed-seqs demux-paired-end.qza\
-- p-trim-left-f 0\
-- p-trim-left-r 0\
-- p-trunc-len-f 240\
-- p-trunc-len-r 160\
-- o-table table.qza\
-- o-representative-sequences rep-seqs.qza\
-- o-denoising-stats denoising-stats.qza
```

Mohammad Amin Shirazi

توضیحات این دستور:

--i-demultiplexed-seqs: ورودی داده‌های توالی که به صورت demultiplexed (تقسیم شده بر اساس نمونه‌ها) وارد شده‌اند.

--p-trim-left-f و --p-trim-left-r: تعیین موقعیت‌های برش از سمت چپ توالی‌ها برای فایل‌های forward (f) و reverse (r). این اعداد می‌توانند بر اساس کیفیت توالی‌ها تنظیم شوند.

--p-trunc-len-f و --p-trunc-len-r: طول توالی‌های برش داده شده (truncated) برای فایل‌های forward و reverse. معمولاً در اینجا به میزان ۲۴۰ برای forward و ۱۶۰ برای reverse استفاده می‌شود.

--o-table: خروجی شامل feature table است که شامل تعداد ASVها برای هر نمونه می‌باشد.

--o-representative-sequences: خروجی شامل نماینده‌های توالی‌ها است که به عنوان ASV شناخته می‌شوند.

--o-denoising-stats: فایل خروجی شامل آماری از فرآیند denoising است که می‌تواند برای بررسی کیفیت و کارایی فرآیند استفاده شود.

مشاهده خروجی: پس از انجام فرآیند denoising، چندین فایل خروجی ایجاد می‌شود:

table.qza: جدول ویژگی‌ها (feature table) که شامل فراوانی هر ASV برای هر نمونه است.

rep-seqs.qza: فایل توالی‌های نماینده یا ASVها.

denoising-stats.qza: فایل آماری که اطلاعات مربوط به فرآیند DADA2 را نشان می‌دهد (مثلاً تعداد توالی‌های اولیه، توالی‌های باقی‌مانده پس از فیلتر کردن، و تعداد ASVهای شناسایی شده).

Mohammad Amin Shirazi

برای مشاهده آماری که در denoising-stats.qza ذخیره شده است، می‌توانید از دستور زیر استفاده کنید:

```
qiime metadata tabulate\
-- m-input-file denoising-stats.qza\
-- o-visualization denoising-stats.qzv
```

این دستور، فایل آماری را به صورت یک visualization (فایل .qzv) درآورده که می‌توانید آن را مشاهده کنید.

بررسی کیفیت نتایج: پس از انجام فرآیند DADA2، می‌توانید بررسی کنید که چه تعداد ASV از داده‌های اولیه استخراج شده است و میزان بهبود کیفیت توالی‌ها به وضوح قابل مشاهده خواهد بود.

گام‌های بعدی: پس از اینکه ASVها شناسایی شدند، می‌توانید:

این ASVها را به کمک یک پایگاه داده برای تعیین taxonomic classification (تشخیص تاکسونومیک) تحلیل کنید.

از جدول ویژگی‌ها (feature table) برای انجام تحلیل‌های تنوع زیستی (diversity analysis) مانند alpha و beta diversity استفاده کنید.

Mohammad Amin Shirazi

```
qiime phylogeny align-to-tree-mafft-fasttree\
--    i-sequences dada2/representative_sequences.qza\
--    output-dir tree
```

ساخت درخت فیلوژنتیکی در QIIME 2

در این مرحله، هدف ساخت یک درخت فیلوژنتیکی است که می‌توان از آن در تحلیل‌های بعدی مانند تحلیل‌های تنوع فیلوژنتیکی (phylogenetic diversity) استفاده کرد. برای ساخت این درخت فیلوژنتیکی از توالی‌های نماینده (ASVها) استفاده می‌شود که در مرحله قبلی با استفاده از افزونه DADA2 به دست آمده‌اند.

درخت فیلوژنتیکی معمولاً برای محاسبه شاخص‌های تنوع فیلوژنتیکی و بررسی ارتباط تکاملی میان نمونه‌ها و ویژگی‌ها استفاده می‌شود.

مراحل ساخت درخت فیلوژنتیکی:

استفاده از دستور align-to-tree-mafft-fasttree: برای ساخت درخت فیلوژنتیکی، باید توالی‌های نماینده (ASVها) را ابتدا با استفاده از الگوریتم MAFFT به یکدیگر هم‌راستا (alignment) کنیم و سپس با استفاده از FastTree یک درخت فیلوژنتیکی بسازیم.

دستور زیر برای ساخت درخت فیلوژنتیکی از توالی‌های نماینده استفاده می‌شود:

Mohammad Amin Shirazi

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

توضیحات این دستور:

--i-sequences rep-seqs.qza: ورودی توالی‌های نماینده (ASVها) که از مرحله DADA2 به دست آمده است.

--o-alignment aligned-rep-seqs.qza: خروجی شامل توالی‌های هم‌راسته شده است.

--o-masked-alignment masked-aligned-rep-seqs.qza: خروجی شامل توالی‌های هم‌راسته شده که قسمت‌های کم‌اطلاعات از آن‌ها حذف شده است (برای کاهش نویز).

--o-tree unrooted-tree.qza: خروجی شامل درخت فیلوژنتیکی بدون ریشه است.

--o-rooted-tree rooted-tree.qza: خروجی شامل درخت فیلوژنتیکی ریشه‌دار است.

Mohammad Amin Shirazi

مشاهده درخت فیلوژنتیکی: پس از ساخت درخت، می‌توانید درخت فیلوژنتیکی ریشه‌دار را به صورت تصویری مشاهده کنید:

\ qiime phylogeny visualize

\ i-tree rooted-tree.qza--

o-visualization rooted-tree.qzv--

در این دستور:

--i-tree rooted-tree.qza: درخت فیلوژنتیکی ریشه‌دار که از دستور قبلی تولید شده است.

--o-visualization rooted-tree.qzv: خروجی تصویری در قالب فایل .qzv که می‌توانید آن را در QIIME 2 View مشاهده کنید.

استفاده از درخت فیلوژنتیکی در تحلیل‌های بعدی: درخت فیلوژنتیکی که ساخته شده می‌تواند در تحلیل‌های مختلفی مانند:

Phylogenetic Diversity (تنوع فیلوژنتیکی)

Beta Diversity (تنوع بین نمونه‌ها)

Alpha Diversity (تنوع درون نمونه‌ها) استفاده شود.

Mohammad Amin Shirazi

```
qiime diversity core-metrics-phylogenetic\
--  i-table dada2/table.qza\
--  i-phylogeny tree/rooted_tree.qza\
--  p-sampling-depth 8000\
--  m-metadata-file metadata.tsv\
--  output-dir diversity
```

تنوع آلفا و بتا دو نوع معیار برای اندازه‌گیری تنوع میکروبی در محیط‌های زیستی مختلف هستند، که برای تحلیل و مقایسه جامعه‌های میکروبی استفاده می‌شوند. در مطالعات میکروبیوم، این معیارها به ما کمک می‌کنند تا ببینیم چطور تنوع میکروبی در یک نمونه یا بین نمونه‌های مختلف تغییر می‌کند.

۱. تنوع آلفا (Alpha Diversity):

تنوع آلفا میزان تنوع گونه‌ها (Species Richness) یا تعداد گونه‌ها در یک نمونه خاص یا یک محیط خاص را نشان می‌دهد. این معیار تنها به هر نمونه به صورت جداگانه نگاه می‌کند و به ما می‌گوید که در هر نمونه، چند نوع گونه میکروبی وجود دارد و فراوانی نسبی آن‌ها چگونه است. روش‌های مختلفی برای محاسبه تنوع آلفا وجود دارد، مانند:

Richness (غنای گونه‌ای): تعداد کل گونه‌ها یا OTU (Operational Taxonomic Units) ها در هر نمونه.Shannon Diversity Index: شاخصی که هم تعداد گونه‌ها و هم فراوانی نسبی آن‌ها را در نظر می‌گیرد. مقدار بالاتر این شاخص نشان‌دهنده توزیع یکنواخت‌تر گونه‌ها و تنوع بالاتر است. Simpson Diversity Index: مشابه شاخص شانن است اما حساسیت بیشتری به گونه‌های غالب دارد و برای محیط‌هایی که گونه‌های غالب بیشتری دارند مناسب‌تر است.

Mohammad Amin Shirazi

در QIIME 2، از طریق دستور qiime diversity alpha می‌توانیم تنوع آلفا را محاسبه کنیم و با دستور core-metrics-phylogenetic نیز می‌توان به طور هم‌زمان چندین شاخص تنوع آلفا و بتا را محاسبه کرد.

۲. تنوع بتا (Beta Diversity):

تنوع بتا برای مقایسه تنوع میکروبی بین نمونه‌های مختلف یا محیط‌های مختلف به کار می‌رود و به ما کمک می‌کند بفهمیم تفاوت در ترکیب گونه‌ها بین دو یا چند نمونه چگونه است. این معیار تنوع بین نمونه‌ها را محاسبه می‌کند و نشان می‌دهد چقدر نمونه‌ها از نظر ترکیب میکروبی شبیه یا متفاوت هستند. روش‌های مختلفی برای محاسبه تنوع بتا وجود دارد، که معروف‌ترین آن‌ها عبارتند از:

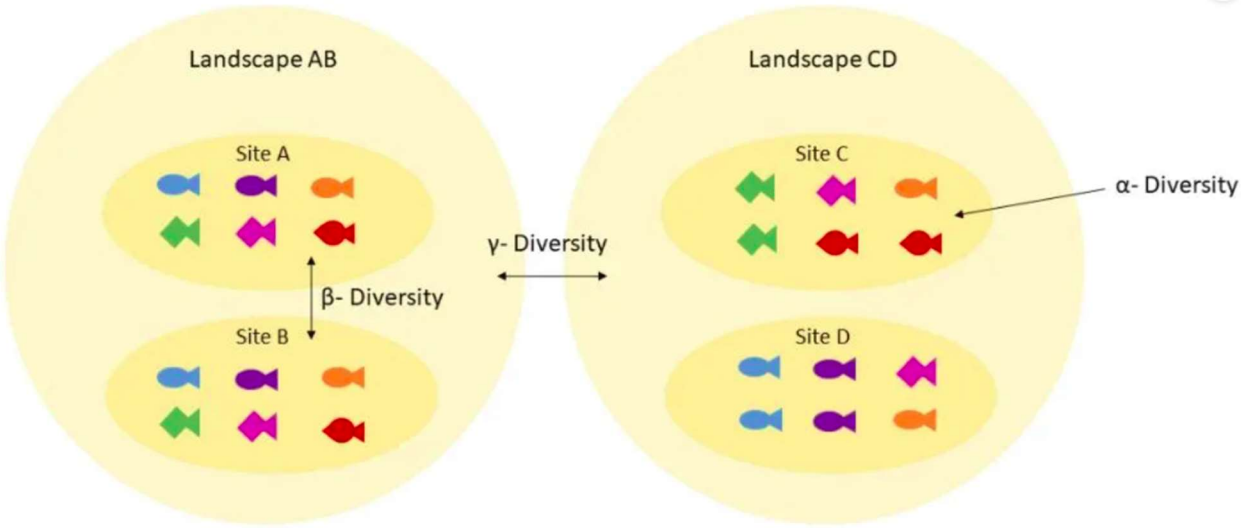
Unweighted UniFrac: تفاوت بین نمونه‌ها را بر اساس حضور یا عدم حضور گونه‌ها بدون در نظر گرفتن فراوانی‌شان محاسبه می‌کند.Weighted UniFrac: علاوه بر حضور یا عدم حضور گونه‌ها، فراوانی نسبی گونه‌ها را نیز در محاسبات در نظر می‌گیرد.Bray-Curtis Dissimilarity: شباهت یا تفاوت بین نمونه‌ها را بر اساس فراوانی نسبی گونه‌ها محاسبه می‌کند.

در QIIME 2، می‌توانیم با دستور qiime diversity beta یا qiime diversity core-metrics-phylogenetic تنوع بتا را محاسبه کنیم. به طور خاص، دستور adonis نیز به ما امکان می‌دهد تفاوت‌های تنوع بتا بین گروه‌های مختلف را از نظر آماری بررسی کنیم و ببینیم که آیا تفاوت‌های قابل توجهی وجود دارد یا خیر.

به طور خلاصه:تنوع آلفا: تنوع و فراوانی گونه‌ها را در یک نمونه نشان می‌دهد.تنوع بتا: تفاوت تنوع بین نمونه‌ها را مقایسه می‌کند.

در نهایت، این شاخص‌ها به تحلیل‌های زیست‌محیطی و میکروبی کمک می‌کنند تا تغییرات میکروبی را در شرایط مختلف یا در طول زمان بررسی کنیم.

Mohammad Amin Shirazi




```
qiime diversity adonis\
--   i-distance-matrix diversity/weighted_unifrac_distance_matrix.qza\
--   m-metadata-file metadata.tsv\
--   p-formula "disease_stat\ "
--   p-n-jobs 2\
--   o-visualization permanova.qzv
```

آزمون آماری PERMANOVA (Permutational Multivariate Analysis of Variance) یک روش آماری برای مقایسه تنوع بتا بین گروه‌های مختلف است. در تحلیل‌های میکروبیوم، PERMANOVA به ما کمک می‌کند تا تفاوت در ترکیب میکروبی بین گروه‌های مختلف را از نظر آماری بررسی کنیم. برای مثال، اگر دو گروه مختلف از نمونه‌ها (مانند بیمار و سالم) داشته باشیم، می‌توانیم با PERMANOVA بررسی کنیم که آیا این گروه‌ها به طور معناداری از نظر ترکیب میکروبی با هم تفاوت دارند یا خیر.

چگونه PERMANOVA کار می‌کند؟

Mohammad Amin

PERMANOVA بر اســاس فاصــله‌های موجود بین نمونه‌ها عمل می‌کند. این فاصــله‌ها می‌توانند بر اســاس معیارهای مختلفی مانند Bray-Curtis، Weighted UniFrac یا Unweighted UniFrac محاســبه شــده باشــند. در این روش، ترکیب میکروبی هر نمونه به صــورت یک بردار عددی در فضــای چند بعدی در نظر گرفته می‌شــود، و فاصــله بین این بردارها نشــان‌دهنده میزان شــباهت یا تفاوت ترکیب میکروبی بین نمونه‌هاست.

مراحل انجام PERMANOVA:

محاســبه ماتریس فاصــله: ابتدا، یک ماتریس فاصــله برای داده‌ها ایجاد می‌شــود که نشــان‌دهنده فاصــله بین همه جفت‌های نمونه‌هاست. این ماتریس فاصله می‌تواند بر اساس معیارهایی مانند UniFrac یا Bray-Curtis محاسبه شود.

تعیین فرضیات:

فرض صفر (Null Hypothesis): ترکیب میکروبی بین گروه‌ها تفاوت معناداری ندارد، و تفاوت‌های مشاهده‌شده به دلیل تصــادف هســتند.فرض مخالف (Alternative Hypothesis): ترکیب میکروبی بین گروه‌ها تفاوت معناداری دارد.

محاســبه آماره آزمون: آماره آزمون PERMANOVA میزان پراکندگی بین و درون گروه‌ها را محاســبه می‌کند. این آماره با استفاده از مربع میانگین فاصله‌ها بین نمونه‌های هر گروه و فاصله بین گروه‌ها به دست می‌آید.

انجام پرموتاســیون: برای آزمون آماری، PERMANOVA داده‌ها را به تعداد زیادی (مثلاً ۹۹۹ یا ۹۹۹۹ بار) بازترتیب یا پرموت می‌کند. با هر بار پرموت، برچســب‌های گروهی بین نمونه‌ها جابجا می‌شــوند، و هر بار آماره آزمون جدید محاسبه می‌شود. سپس، تعداد دفعاتی که آماره آزمون حاصل از پرموتاسیون‌ها از آماره آزمون اصلی بزرگتر یا مساوی است محاسبه می‌شود. این تعداد تقسیم بر تعداد کل پرموتاسیون‌ها مقدار p-value آزمون را می‌دهد.

تفسیر نتیجه:

Mohammad Amin Shirazi

اگر مقدار p-value کمتر از سطح خطای آماری (مثلاً ۰.۰۵) باشد، فرض صفر رد می‌شود و نتیجه می‌گیریم که ترکیب میکروبی بین گروه‌ها به طور معناداری متفاوت است.اگر p-value بزرگتر از سطح خطای آماری باشد، فرض صفر رد نمی‌شود، یعنی تفاوت معناداری بین گروه‌ها وجود ندارد.مزایا و معایب PERMANOVA:

مزایا:

نیازی به فرض توزیع خاصی برای داده‌ها ندارد، چون یک آزمون ناپارامتریک است.می‌تواند با داده‌های چندبعدی و ماتریس‌های فاصله کار کند.

معایب:

نسبت به تفاوت‌های موجود در پراکندگی داخل گروه‌ها حساس است؛ یعنی اگر پراکندگی درون گروه‌ها خیلی متفاوت باشد، نتایج PERMANOVA می‌تواند به اشتباه معنادار شود.

این آزمون به شما می‌گوید که آیا تفاوت معناداری بین گروه‌ها از نظر ترکیب میکروبی وجود دارد یا خیر.

```
wget https://data.qiime2.org/2021.4/common/gg-13-8-99-515-806-nb-classifier.qza
```

```
qiime feature-classifier classify-sklearn\
--  i-reads dada2/representative_sequences.qza\
--  i-classifier gg-13-8-99-515-806-nb-classifier.qza\
--  o-classification taxa.qza
```

طبقه‌بندی تاکسونومیک در QIIME 2

Mohammad Amin Shirazi

در این مرحله، هدف طبقه‌بندی توالی‌ها به تاکسون‌ها (مانند جنس و گونه) است. برای این کار از یک مدل طبقه‌بندی تاکسونومیک استفاده می‌شود. در QIIME 2، می‌توان از مدل‌های GreenGenes یا Silva برای طبقه‌بندی استفاده کرد. این مدل‌ها بر اساس داده‌های مرجع، به تعیین تاکسونومیک توالی‌های DNA می‌پردازند.

مراحل طبقه‌بندی تاکسونومیک:

دانلود مدل طبقه‌بندی (GreenGenes یا Silva): ابتدا باید مدل طبقه‌بندی را از منابع معتبر دانلود کنید. برای مثال، مدل GreenGenes را می‌توانید از طریق دستور زیر دانلود کنید:

```
wget https://data.qiime2.org/2021.4/common/gg-13-8-99-515-806-nb-classifier.qza
```

این دستور مدل GreenGenes را که قبلاً آموزش دیده است و برای طبقه‌بندی توالی‌های 16S rRNA طراحی شده است، دانلود می‌کند.

طبقه‌بندی توالی‌ها با استفاده از مدل: پس از دانلود مدل طبقه‌بندی، از دستور classify-sklearn برای طبقه‌بندی توالی‌های نماینده (ASVها) به تاکسون‌ها استفاده می‌کنیم. در این مرحله، مدل به توالی‌های DNA ورودی نگاه کرده و آن‌ها را به تاکسون‌های مناسب نسبت می‌دهد.

دستور زیر برای طبقه‌بندی توالی‌ها استفاده می‌شود:

Mohammad Amin Shirazi

```
\ qiime feature-classifier classify-sklearn
\ i-reads dada2/representative_sequences.qza--
\ i-classifier gg-13-8-99-515-806-nb-classifier.qza--
o-classification taxa.qza--
```

توضیحات دستور:

--i-reads dada2/representative_sequences.qza: ورودی توالی‌های نماینده که از مرحله قبلی (DADA2) تولید شده‌اند.

--i-classifier gg-13-8-99-515-806-nb-classifier.qza: مدل طبقه‌بندی (در اینجا مدل GreenGenes) که برای طبقه‌بندی توالی‌ها استفاده می‌شود.

--o-classification taxa.qza: خروجی که شامل طبقه‌بندی تاکسونومیک توالی‌ها است و به صورت فایل taxa.qza ذخیره می‌شود.

مشاهده و تحلیل نتایج: پس از اجرای دستور بالا، طبقه‌بندی تاکسونومیک توالی‌ها در فایل taxa.qza ذخیره می‌شود. این فایل را می‌توان برای تحلیل‌های بیشتر یا مشاهده نتایج طبقه‌بندی استفاده کرد. برای مشاهده نتایج به صورت گرافیکی می‌توان از دستور qiime metadata tabulate استفاده کرد:

Mohammad Amin Shirazi

\ qiime metadata tabulate

\ m-input-file taxa.qza--

o-visualization taxa.qzv--

این دستور فایل taxa.qza را به یک فایل تصویری taxa.qzv تبدیل می‌کند که می‌توانید آن را در QIIME 2 View مشاهده کنید.

طبقه‌بندی تاکسونومیک در QIIME 2

طبقه‌بندی تاکسونومیک فرآیندی است که در آن توالی‌های DNA به گروه‌های مختلف (تاکسون‌ها) نسبت داده می‌شوند. این گروه‌ها می‌توانند در سطوح مختلف تاکسونومیک مانند شاه‌تبار، تبار، خانواده، جنس و گونه باشند. در QIIME 2، طبقه‌بندی تاکسونومیک به وسیله افزونه‌هایی مانند feature-classifier انجام می‌شود. این افزونه از مدل‌های مختلف آموزش دیده مانند GreenGenes یا Silva برای طبقه‌بندی توالی‌ها استفاده می‌کند.

مراحل طبقه‌بندی تاکسونومیک در QIIME 2:

دانلود مدل طبقه‌بندی: قبل از شروع طبقه‌بندی، باید یک مدل طبقه‌بندی معتبر را دانلود کنید. مدل‌هایی مانند GreenGenes و Silva برای این کار مناسب هستند. مدل‌های GreenGenes و Silva معمولاً برای داده‌های S rRNA۱۶ به‌طور گسترده استفاده می‌شوند. این مدل‌ها از داده‌های مرجع برای نسبت دادن توالی‌ها به تاکسون‌های مختلف استفاده می‌کنند.

Mohammad Amin Shirazi

برای مثال، برای دانلود مدل GreenGenes به دستور زیر نیاز دارید:

```
wget https://data.qiime2.org/2021.4/common/gg-13-8-99-515-806-nb-classifier.qza
```

این دستور مدل GreenGenes را از سرور QIIME 2 دانلود می‌کند. شما می‌توانید از مدل Silva نیز برای طبقه‌بندی استفاده کنید که مشابه همین روند است.

اجرای دستور classify-sklearn برای طبقه‌بندی توالی‌ها: پس از دانلود مدل طبقه‌بندی، از دستور classify-sklearn برای طبقه‌بندی توالی‌ها استفاده می‌شود. این دستور مدل را به داده‌های توالی شما اعمال کرده و آنها را به تاکسون‌های مختلف نسبت می‌دهد.

دستور زیر برای طبقه‌بندی توالی‌ها به کار می‌رود:

```
\ qiime feature-classifier classify-sklearn
\ i-reads dada2/representative_sequences.qza--
\ i-classifier gg-13-8-99-515-806-nb-classifier.qza--
o-classification taxa.qza--
```

Mohammad Amin Shirazi

توضیحات دستور:

--i-reads dada2/representative_sequences.qza: ورودی توالی‌های نماینده (ASVها) که از مرحله قبلی (DADA2) تولید شده‌اند.

--i-classifier gg-13-8-99-515-806-nb-classifier.qza: مدل طبقه‌بندی (در اینجا مدل GreenGenes)، که برای تعیین تاکسون‌ها استفاده می‌شود.

--o-classification taxa.qza: خروجی که شامل اطلاعات طبقه‌بندی است و در قالب فایل taxa.qza ذخیره می‌شود.

مشاهده نتایج طبقه‌بندی: پس از اتمام مرحله طبقه‌بندی، نتیجه در فایل taxa.qza ذخیره می‌شود. شما می‌توانید از این فایل برای تجزیه و تحلیل‌های بیشتر استفاده کنید. برای مشاهده نتایج طبقه‌بندی به‌صورت گرافیکی، از دستور qiime metadata tabulate استفاده می‌کنیم تا فایل taxa.qza را به یک فایل تصویری تبدیل کنیم که قابل مشاهده در QIIME 2 View باشد.

دستور زیر را برای تبدیل فایل به فایل تصویری و مشاهده نتایج استفاده کنید:

```
\ qiime metadata tabulate
\ m-input-file taxa.qza--
o-visualization taxa.qzv--
```

Mohammad Amin Shirazi

این دســـتور، فایل taxa.qza را به یک فایل taxa.qzv تبدیل می‌کند که شـــما می‌توانید آن را در محیط گرافیکی QIIME 2 مشاهده کنید.

کاربردهای طبقه‌بندی تاکسونومیک:

درک تنوع میکروبی: طبقه‌بندی تاکسـونومیک به شـما این امکان را می‌دهد که تنوع میکروبی نمونه‌های خود را در سطح‌های مختلف (از جمله گونه‌ها و جنس‌ها) بررسی کنید.

بررســی ارتباطات تکاملی: این فرآیند به شــما کمک می‌کند تا ببینید که کدام تاکســون‌ها به یکدیگر مرتبط هستند و چگونه انواع مختلف میکروب‌ها در طول زمان تکامل پیدا کرده‌اند.

بررســی تفاوت‌های محیطی: با طبقه‌بندی توالی‌ها به تاکســون‌های مختلف، می‌توان بررســی کرد که چگونه محیط‌های مختلف (مانند محیط‌های زیســـتی مختلف یا تغییرات فیزیولوژیکی در میزبان‌ها) می‌تولند بر ترکیب میکروبی اثر بگذارد.

نتیجه‌گیری:

طبقه‌بندی تاکســـونومیک در QIIME 2 یکی از مراحل اســـاســـی برای تحلیل داده‌های میکروبی و درک تنوع میکروبی است. با استفاده از مدل‌های معتبر و ابزارهای موجود در QIIME 2، می‌توانیم توالی‌های خود را به تاکسون‌های مختلف نسبت دهیم و اطلاعات مفیدی از داده‌های میکروبی به دست آوریم

qiime taxa barplot\

Mohammad Amin Shirazi

```
--   i-table dada2/table.qza\
--   i-taxonomy taxa.qza\
--   m-metadata-file metadata.tsv\
--   o-visualization taxa_barplot.qzv
```

نمودار فراوانی نسبی تاکسون‌ها در QIIME 2

نمودار فراوانی نسبی تاکسون‌ها یکی از روش‌های مفید برای تجزیه و تحلیل ترکیب میکروبی نمونه‌ها است. این نمودار به شما کمک می‌کند تا ببینید هر تاکسون (مانند گونه‌ها یا جنس‌ها) در هر نمونه چه میزان از ترکیب میکروبی را تشکیل می‌دهد. این نمودار معمولاً برای مقایسه فراوانی نسبی تاکسون‌ها در نمونه‌های مختلف و شناسایی الگوهای غالب در هر نمونه استفاده می‌شود.

مراحل ایجاد نمودار فراوانی نسبی تاکسون‌ها:

ایجاد نمودار فراوانی نسبی با استفاده از داده‌های طبقه‌بندی شده: برای ایجاد نمودار فراوانی نسبی، باید ابتدا از داده‌های طبقه‌بندی شده استفاده کنید که قبلاً در مرحله قبل به‌دست آورده‌اید (یعنی داده‌های taxa.qza که شامل طبقه‌بندی تاکسون‌ها هستند).

دستور qiime taxa barplot: در QIIME 2 از دستور qiime taxa barplot برای ایجاد نمودار فراوانی نسبی استفاده می‌شود. این دستور به‌طور خودکار فراوانی نسبی هر تاکسون را برای هر نمونه محاسبه کرده و آن را به صورت یک نمودار میله‌ای نمایش می‌دهد.

Mohammad Amin Shirazi

دستور زیر برای ایجاد نمودار فراوانی نسبی به کار می‌رود:

\ qiime taxa barplot

\ i-table feature-table.qza--

\ i-taxonomy taxa.qza--

o-visualization taxa-barplot.qzv--

توضیحات دستور:

--i-table feature-table.qza: ورودی جدول ویژگی‌ها (Feature Table)، که داده‌های توالی‌ها و نمونه‌ها را شامل می‌شود.

--i-taxonomy taxa.qza: ورودی داده‌های طبقه‌بندی تاکسون‌ها که در مرحله قبلی ایجاد شده‌اند.

--o-visualization taxa-barplot.qzv: خروجی که شامل نمودار فراوانی نسبی تاکسون‌ها است و در قالب فایل taxa-barplot.qzv ذخیره می‌شود.

مشاهده نمودار فراوانی نسبی: پس از اجرای دستور بالا، فایل taxa-barplot.qzv تولید می‌شود. این فایل را می‌توان در محیط گرافیکی QIIME 2 مشاهده کرد. برای مشاهده نمودار، دستور زیر را برای باز کردن فایل در QIIME 2 استفاده کنید:

Mohammad Amin Shirazi

```
qiime tools view taxa-barplot.qzv
```

این دستور نمودار را در مرورگر شما باز می‌کند تا بتوانید به‌راحتی پراکندگی فراوانی نسبی تاکسون‌ها را در نمونه‌ها مشاهده کنید.

کاربردهای نمودار فراوانی نسبی تاکسون‌ها:

مقایسه تنوع میکروبی: با استفاده از این نمودار، می‌توانید تفاوت‌ها و شباهت‌های ترکیب میکروبی نمونه‌های مختلف را به‌صورت گرافیکی مقایسه کنید.

شناسایی تاکسون‌های غالب: این نمودار به شما کمک می‌کند تا تاکسون‌های غالب در هر نمونه را شناسایی کنید. این اطلاعات می‌تواند برای بررسی تفاوت‌های میکروبی میان گروه‌های مختلف مفید باشد.

بررسی تغییرات محیطی: با مشاهده الگوهای فراوانی نسبی در نمونه‌ها، می‌توان تاثیرات شرایط محیطی، فیزیولوژیکی یا درمانی را بر ترکیب میکروبی بررسی کرد.

نتیجه‌گیری:

نمودار فراوانی نسبی تاکسون‌ها ابزاری بسیار مفید در تجزیه و تحلیل داده‌های میکروبی است که به شما کمک می‌کند ترکیب میکروبی نمونه‌ها را در سطوح مختلف تاکسونومی بررسی کنید. این نمودار می‌تواند به شناسایی الگوهای غالب میکروبی و مقایسه تفاوت‌های میان گروه‌های مختلف کمک کند.

Mohammad Amin Shirazi

```
qiime tools export\
--    input-path taxa.qza\
--    output-path exported

biom convert -i exported/feature-table.biom -o taxa_table.tsv --to-tsv
```

استخراج داده‌ها به فرمت TSV در QIIME 2

اگر نیاز دارید که داده‌ها را به فرمت TSV (Tab-Separated Values) اســـتخراج کنید تا از آن‌ها در تحلیل‌های دیگر اسـتفاده کنید، می‌توانید از دسـتوراتی در QIIME 2 برای تبدیل داده‌ها به این فرمت اسـتفاده کنید. این کار معمولاً برای تحلیل‌های بعدی در نرم‌افزارهای مختلف یا برای به اشـــتراک‌گذاری داده‌ها با دیگران مفید است.

مراحل استخراج داده‌ها به فرمت TSV:

استخراج جدول ویژگی‌ها (Feature Table) به فرمت TSV: برای استخراج جدول ویژگی‌ها که شامل اطلاعات مربوط به حضور ویژگی‌ها (مانند ASVها یا OTUها) در هر نمونه است، از دستور زیر استفاده کنید:

```
\ qiime tools export
\ input-path feature-table.qza--
output-path feature-table--
```

Mohammad Amin Shirazi

این دســـتور داده‌های feature-table.qza را به فرمت TSV تبدیل کرده و در پوشـــه‌ی feature-table/ ذخیره می‌کند. پس از اجرای این دستور، فایل‌های خروجی در مسیر مشخص شده قرار می‌گیرند.

درون پوشـــه‌ی feature-table/، شـــما فایلی به نام feature-table.biom پیدا خواهید کرد. این فایل در فرمت BIOM است که برای کار با داده‌های میکروبی معمولاً استفاده می‌شود.

تبدیل فایل BIOM به فرمت TSV: برای تبدیل فایل BIOM به فرمت TSV، می‌توانید از ابزار biom اســـتفاده کنید. دستور زیر برای تبدیل این فایل به فرمت TSV به کار می‌رود:

```
biom convert -i feature-table.biom -o feature-table.tsv --to-tsv
```

این دســـتور فایل feature-table.biom را به فایل feature-table.tsv تبدیل کرده و در همان مســـیر ذخیره می‌کند.

استخراج داده‌های طبقه‌بندی تاکسون‌ها به فرمت TSV: برای استخراج داده‌های طبقه‌بندی تاکسون‌ها به فرمت TSV، از دستور زیر استفاده کنید:

```
\ qiime tools export

\ input-path taxa.qza--
```

Mohammad Amin Shirazi

output-path taxa--

پس از اجرا، داده‌های طبقه‌بندی‌شـــده به فرمت متنی صـــادر می‌شـــود که می‌توانید آن‌ها را به فرمت TSV تبدیل کنید.

مشـــاهده فایل‌های خروجی: بعد از انجام این مراحل، شـــما باید فایل‌های feature-table.tsv و taxa.tsv را در مسیر خروجی مشاهده کنید. این فایل‌ها آماده هستند تا در دیگر ابزارهای تحلیلی یا به‌عنوان ورودی برای تحلیل‌های بعدی استفاده شوند.

کاربردهای فایل‌های TSV:

انتقال داده‌ها: این فایل‌ها می‌توانند برای انتقال داده‌ها به نرم‌افزارهای دیگر یا برای به اشـــتراک‌گذاری با همکاران استفاده شوند.

تحلیل‌های بیشـــتر: شـــما می‌توانید این فایل‌ها را در نرم‌افزارهایی مانند Excel، R یا Python برای تحلیل‌های بیشتر بارگذاری کنید.

خروجی برای گزارش‌ها: فایل‌های TSV به راحتی برای گزارش‌های علمی یا مستندسازی داده‌ها قابل استفاده هستند.

نتیجه‌گیری:

Mohammad Amin Shirazi

استخراج داده‌ها به فرمت TSV یکی از روش‌های ساده و کارآمد برای ذخیره‌سازی داده‌ها در یک فرمت متنی است که می‌تواند برای تحلیل‌های بعدی در ابزارهای دیگر استفاده شود. با استفاده از دستورات QIIME 2 و ابزار biom، می‌توانید به راحتی داده‌های خود را به فرمت‌های مختلف تبدیل کرده و از آن‌ها در تحلیل‌های پیشرفته‌تر بهره ببرید.

```
import numpy as np
import pandas as pd
import seaborn as sns

abundances = pd.read_table("genus.tsv", skiprows=1, index_col=0)
abundances.index = abundances.index.str.split(";").str فقط نام جنس # [۵]
abundances = abundances[~abundances.index.isin(["g__", "__"])] # حذف جنس‌های نامشخص

#Transform برای ایجاد heatmap
transformed = abundances.apply(
    lambda xs: np.log(xs + 0.5) - np.log(xs.mean() + 0.5),
    axis=1(

sns.clustermap(transformed.T, cmap="magma", xticklabels=True, figsize=(16, 5))
```

برای اجرای این کد در محیط Linux، ابتدا باید اطمینان حاصل کنید که بسته‌های مورد نیاز (مثل numpy, pandas, seaborn) نصب شده باشند. سپس می‌توانید اسکریپت Python را اجرا کنید.

مراحل اجرا:

Mohammad Amin Shirazi

نصب بسته‌های مورد نیاز: ابتدا باید بسته‌های numpy, pandas, و seaborn را نصب کنید. برای این کار، در ترمینال دستور زیر را وارد کنید:

```
pip install numpy pandas seaborn
```

ایجاد اسکریپت Python: سپس کد زیر را در یک فایل Python ذخیره کنید، به عنوان مثال heatmap.py:

```
import numpy as np
import pandas as pd
import seaborn as sns

# خواندن داده‌ها از فایل TSV
abundances = pd.read_table("genus.tsv", skiprows=1, index_col=0)
abundances.index = abundances.index.str.split(";").str[۵]  # فقط نام جنس
abundances = abundances[~abundances.index.isin(["g__", "__"])]  # حذف جنس‌های نامشخص

# Transform برای ایجاد heatmap
transformed = abundances.apply(
```

Mohammad Amin Shirazi

```
,lambda xs: np.log(xs + 0.5) - np.log(xs.mean() + 0.5)
(axis=1

# ایجاد heatmap
sns.clustermap(transformed.T, cmap="magma", xticklabels=True, figsize=(16, 5))
```

اجرای اسکریپت: برای اجرای اسکریپت Python در ترمینال، دستور زیر را وارد کنید:

```
python heatmap.py
```

این کد فایل genus.tsv را که باید در همان مسیر اسکریپت قرار داشته باشد، بارگذاری کرده و پس از اعمال تبدیل‌های مورد نیاز، یک heatmap با استفاده از seaborn.clustermap ایجاد خواهد کرد.

نکات:

اطمینان حاصل کنید که فایل genus.tsv در همان مسیر اسکریپت قرار دارد.

Mohammad Amin Shirazi

این اسکریپت یک تصویر heatmap را تولید می‌کند که ممکن است در محیط‌های گرافیکی قابل مشاهده باشد. در صورتی که در محیطی بدون رابط گرافیکی هستید، می‌توانید از matplotlib برای ذخیره کردن تصویر استفاده کنید، به این صورت که در انتهای اسکریپت کد زیر را اضافه کنید:

```
import matplotlib.pyplot as plt
plt.savefig("heatmap.png")
```

این تصویر به نام heatmap.png ذخیره خواهد شد.

در کار با QIIME 2، در کنار مراحلی که قبلاً اشاره کردیم، چند نکته و قابلیت مهم دیگر وجود دارد که ممکن است برای شما مفید باشند:

۱. استفاده از Metadata:

Metadata اطلاعات مربوط به نمونه‌ها (مثل ویژگی‌های محیطی، فیزیولوژیکی یا بالینی) هستند که می‌توانند در تحلیل‌ها استفاده شوند. اطلاعات metadata معمولاً در یک فایل .tsv ذخیره می‌شود و هنگام انجام تحلیل‌ها می‌توانند به عنوان گروه‌بندی برای مقایسه‌ها، تست‌های آماری و تحلیل‌های تنوع استفاده شوند.

در QIIME 2، باید metadata خود را با دقت وارد کنید و به آن اشاره کنید، تا به درستی بتوانید از آن در آنالیزهای بعدی استفاده کنید.

۲. تعریف و استفاده از متغیرهای گروه‌بندی در تحلیل‌ها:

Mohammad Amin Shirazi

زمانی که می‌خواهید بین گروه‌های مختلف نمونه‌ها (مثلاً گروه‌های مختلف بیماران یا محیط‌ها) مقایسه انجام دهید، باید متغیرهایی را که می‌خواهید در تحلیل‌ها از آنها استفاده کنید به عنوان جنس/ویژگی معرفی کنید.

این کار معمولاً در مرحله اعمال alpha و beta diversity و دیگر تحلیل‌ها مهم است.

۳. افزونه‌ها و Plugins در QIIME 2:

افزونه‌ها (Plugins) امکانات و قابلیت‌های اضافی را به QIIME 2 اضافه می‌کنند. افزونه‌هایی مانند q2-diversity, q2-longitudinal, q2-composition, q2-taxa و دیگر افزونه‌ها ابزارهای خاصی برای تحلیل‌های پیچیده‌تر مانند تحلیل‌های طولی، پیش‌بینی ویژگی‌ها، و تحلیل غنی‌شدگی (abundance) و ترکیب گونه‌ها فراهم می‌آورند.

افزونه‌های مختلف می‌توانند نتایج به دست آمده از QIIME 2 را گسترش داده و تحلیل‌های تخصصی بیشتری را ممکن کنند.

۴. تست‌های آماری و تحلیل تفاوت‌های میان گروه‌ها:

یکی از بخش‌های مهم در تحلیل داده‌های میکروبیوم، انجام تست‌های آماری برای بررسی تفاوت‌ها بین گروه‌های مختلف نمونه‌ها است. این کار می‌تواند از طریق استفاده از تحلیل‌های differential abundance (بررسی تفاوت‌های غنی‌شدگی گونه‌ها) مانند ANCOM و ANCOM-BC انجام شود.

Mohammad Amin Shirazi

برای مثال، شما ممکن است بخواهید بررسی کنید که کدام میکروارگانیسم‌ها در بین گروه‌های مختلف (مثلاً بیماران مبتلا به بیماری‌های مختلف) تفاوت‌های معناداری دارند.

۵. PCA و Ordination:

تحلیل‌های PCA (تحلیل مؤلفه‌های اصلی) و دیگر تکنیک‌های کاهش ابعاد (مانند PCoA) برای بررسی رابطه بین نمونه‌ها و شبیه‌سازی فاصله‌ها بین آن‌ها بسیار مفید هستند. این تکنیک‌ها برای تجسم و تفسیر داده‌های پیچیده میکروبیوم کاربرد دارند.

این تکنیک‌ها می‌توانند برای شبیه‌سازی تنوع میکروبی و ارتباطات بین نمونه‌ها در فضاهای دو یا سه‌بعدی به کار گرفته شوند.

۶. استفاده از QIIME 2 Visualizer برای تحلیل‌های تصویری:

QIIME 2 Visualizer به شما این امکان را می‌دهد که نتایج تحلیل‌های مختلف را به صورت گرافیکی و بصری مشاهده کنید.

فایل‌های خروجی که به فرمت .qzv ذخیره می‌شوند، شامل گراف‌ها و نمودارهای مختلفی هستند که می‌توانید به راحتی آن‌ها را باز کرده و تحلیل‌های خود را به صورت تصویری مشاهده کنید.

۷. مدیریت داده‌های بزرگ (Big Data):

Mohammad Amin Shirazi

QIIME 2 برای کار با داده‌های بسیار بزرگ (مانند آنالیز داده‌های NGS) بهینه شده است. اما برای کار با داده‌های بزرگ و پیچیده، مدیریت حافظه و منابع سیستم مهم است. در این صورت، استفاده از سرورها و ماشین‌های محاسباتی با قدرت پردازش بالا می‌تواند به شما کمک کند تا کارایی بالاتری داشته باشید.

۸. ادغام با نرم‌افزارهای دیگر:

QIIME 2 امکان ادغام با نرم‌افزارهای دیگر برای تجزیه و تحلیل دقیق‌تر داده‌ها را دارد. برای مثال، می‌توانید از داده‌های حاصل از QIIME 2 برای انجام تحلیل‌های آماری پیچیده‌تر با نرم‌افزارهایی مانند R استفاده کنید.

۹. آموزش و مستندات بیشتر:

QIIME 2 مستندات بسیار کاملی دارد که شامل ویدیوهای آموزشی، دستورالعمل‌های گام به گام، و مثال‌های کاربردی است. اگر در هر مرحله‌ای با مشکل روبه‌رو شدید، مراجعه به مستندات QIIME 2 می‌تواند کمک بزرگی باشد.

۱۰. به‌روزرسانی و نگهداری سیستم:

به‌روزرسانی منظم QIIME 2 برای دریافت افزونه‌ها و اصلاحات جدید الزامی است. برای به‌روزرسانی QIIME 2، کافی است از دستور زیر در ترمینال استفاده کنید:

Mohammad Amin Shirazi

```
conda update qiime2
```

این نکات و ویژگی‌ها فقط بخشی از قابلیت‌های گسترده QIIME 2 هستند که می‌توانند به شما کمک کنند تا تحلیل‌های جامع‌تری در میکروبیوم‌ها و سایر داده‌های مشابه انجام دهید.

Mohammad Amin Shirazi

منابع:

https://bioinformatics.ccr.cancer.gov/docs/qiime2/Lesson2/

https://gibbons-lab.github.io/isb_course_2020/16S/