

مروری جامع بر استفاده از رویکردهای بیوانفورماتیک در آنالیزهای میکروبیوم و میکروبیوتا

بهاره نوحی و منصور زاهدی*

ایران، تهران، دانشگاه شهید بهشتی، دانشکده علوم شیمی و نفت، گروه شیمی تجزیه و شیمی فیزیک

تاریخ دریافت: ۱۴۰۳/۰۱/۲۷ تاریخ پذیرش: ۱۴۰۳/۰۶/۰۳

چکیده

در دهه‌های اخیر، مطالعه میکروبیوم و میکروبیوتا به‌عنوان شاخه‌ای نوین در زیست‌شناسی، توجه بسیاری از پژوهشگران حوزه‌های پزشکی، کشاورزی و زیست‌محیطی را به خود جلب کرده است. پیشرفت چشمگیر فناوری‌های توالی‌یابی نسل جدید، امکان تحلیل عمیق و دقیق جوامع میکروبی را فراهم ساخته است؛ با این حال تفسیر و استخراج از داده‌های عظیم و پیچیده حاصل از این فناوری‌ها بدون بهره‌گیری از ابزارهای بیوانفورماتیکی ممکن نیست. این مقاله، به مرور جامع کاربردهای بیوانفورماتیک در آنالیزهای میکروبیومی پرداخته است و ابزارها و پایگاه‌های داده کلیدی مانند ENA، DADA2، QIIME2 و SILVA معرفی و مقایسه شده‌اند. همچنین، رویکردهای چند-اومیکس، به‌عنوان روش‌هایی نوآورانه در تحلیل عملکرد میکروبیوم بررسی شده‌اند. مقاله حاضر همچنین نشان می‌دهد که چگونه روش‌های یادگیری ماشین و مدل‌سازی، قادرند اطلاعات زیستی را به ابزارهایی قدرتمند برای تشخیص زودهنگام بیماری، طراحی درمان‌های شخصی‌سازی شده و کشف نشانگرهای زیستی تبدیل کنند. در نهایت، ضمن بیان چالش‌هایی مانند نیاز به منابع محاسباتی بالا، تفسیر پیچیده نتایج و نبود پایگاه‌های داده جامع، بر ضرورت توسعه زیرساخت‌های محاسباتی و آموزش بین‌رشته‌ای در این حوزه تأکید شده است و در واقع نقشه راهی علمی برای پژوهشگران فراهم می‌سازد تا با بهره‌گیری از بیوانفورماتیک، پیچیدگی‌های زیستی میکروبیوم را با دقت و عمق بیشتری کشف کنند.

واژه‌های کلیدی: میکروبیوم، میکروبیوتا، بیوانفورماتیک، پایگاه‌های داده و نرم افزارها، داده‌های چند-اومیکس.

* نویسنده مسئول، پست الکترونیکی: m-zahedi@sbu.ac.ir

مقدمه

کشاورزی و علوم زیست‌محیطی ضروری است (۵-۱). تأثیر میکروبیوم و میکروبیوتا بر فیزیولوژی میزبان، عملکرد سیستم ایمنی و سلامت عمومی کاملاً آشکار است. برای نمونه، میکروبیوتای روده انسان نقش مهمی در هضم غذا و سنتز ویتامین‌ها دارد و همچنین بدن را در برابر باکتری‌های بیماری‌زا محافظت می‌کند. تغییر در ترکیب میکروبیوتا، که به عنوان دیس‌بیوز (Dysbiosis) شناخته می‌شود، با طیف وسیعی از بیماری‌ها از جمله چاقی و دیابت، بیماری‌های پوستی همچون پسوریازیس، بیماری‌های التهابی روده و حتی اختلالات روانی نظیر افسردگی و اضطراب اجتماعی، آلزایمر و اوتیسم و ... در ارتباط است. این تعامل پیچیده و ظریف میان میکروبیوتا و میزبان، اهمیت حیاتی درک عملکرد باکتری‌ها را در شرایط سلامت و بیماری برجسته می‌کند و می‌تواند به توسعه روش‌های تشخیصی جدید،

گرچه دو اصطلاح «میکروبیوم» و «میکروبیوتا» اغلب به جای یکدیگر استفاده می‌شوند، از لحاظ علمی دو مفهوم مرتبط اما متمایز را نشان می‌دهند. واژه «میکروبیوم» به مجموع ژنوم تمامی میکروارگانیسم‌هایی اشاره دارد که در یک محیط مشخص مانند بدن انسان، خاک و اقیانوس حضور دارند. به عبارتی دیگر، میکروبیوم بیانگر تمام مواد ژنتیکی موجود در اجتماع میکروبی است. باکتری‌ها، آرکی‌ها، ویروس‌ها و یوکاریوت‌های میکروبی مانند قارچ‌ها و پروتوزوئرها از فراوان‌ترین موجودات در هر میزبان و اکوسیستم‌ها به‌شمار می‌روند که در عملکرد زیستی میزبان‌ها و حفظ تعادل اکوسیستم‌ها نقشی حیاتی دارند. از سوی دیگر، میکروبیوتا به اجتماع واقعی این میکروارگانیسم‌ها اشاره دارد که در این زیستگاه‌ها ساکن هستند. تمایز دقیق میان این دو مفهوم برای درک جامع از پژوهش‌های مرتبط در زمینه‌هایی همچون پزشکی،

رویکردها نسبت به روش‌های سنتی و همچنین محدودیت‌ها و چالش‌هایی که همچنان وجود دارند، است.

ادبیات تحقیق

بیوانفورماتیک به مؤلفه‌ای ضروری در پژوهش‌های مرتبط با میکروبیوم تبدیل شده است و امکان تحلیل جامع و دقیق اجتماعات میکروبی را فراهم آورده است. در این بخش، برخی از مهم‌ترین پژوهش‌هایی که از روش‌های بیوانفورماتیک برای مطالعه میکروبیوم و میکروبیوتا استفاده کرده‌اند، مرور می‌شوند. این بررسی‌ها نقش مهم بیوانفورماتیک را در فهم تعاملات زیستی میکروبی، سلامت و بیماری نشان می‌دهد.

تحلیل متاژنومیکس میکروبیوم روده انسان (۲۰۰۶)

گیل و همکاران، یکی از اولین مطالعات متاژنومیکس را بر روی میکروبیوم روده انسان انجام دادند که دیدگاه بنیادینی درباره تنوع و ظرفیت عملکردی میکروارگانیسم‌های روده ارائه کرد. این پژوهش با استفاده از روش توالی‌یابی شاتگان، اجتماعات میکروبی موجود در بخش پایینی روده دو فرد سالم را بررسی کرد. نتایج مطالعه نشان داد که اجتماعات میکروبی روده بسیار پیچیده و متنوع هستند و دارای مجموعه وسیعی از ژن‌ها هستند که در فرآیندهای متابولیکی همچون متابولیسم کربوهیدرات‌ها و اسیدهای آمینه نقش دارند. این مطالعه با نشان دادن این موضوع که میکروبیوم روده یک مجموعه منفعل نیست، بلکه فعالانه در متابولیسم میزبان نقش دارد، اهمیت زیادی یافت. گیل و همکاران با نقشه‌برداری از ژن‌های میکروبی و ارتباط آن‌ها با مسیرهای متابولیکی، نخستین گام مهم در تبیین نقش فعال میکروبیوم در متابولیسم میزبان را برداشتند و پایه‌ای برای مطالعات بعدی در این حوزه فراهم ساختند (۴).

پروژه میکروبیوم انسانی (۲۰۱۲)

کنسرسیوم پروژه میکروبیوم انسانی (Human Microbiome Project Consortium (HMP)، یک طرح گسترده برای توصیف تنوع میکروبیوم انسان در بخش‌های مختلف بدن بود. این پروژه، با استفاده هم‌زمان از روش‌های توالی‌یابی ژن 16S rRNA و متاژنومیکس، کاتالوگی مرجع از اجتماعات میکروبی مرتبط با افراد سالم را تهیه کرد که مبنایی برای پژوهش‌های آتی پیرامون نقش میکروبیوم در سلامت و بیماری ایجاد نمود. پروژه HMP با ارائه

مداخلات درمانی هدفمند و نشانگرهای زیستی برای پزشکی شخصی منجر شود (۶-۱۰).

فناوری‌های توالی‌یابی، چشم‌انداز مطالعات مرتبط با میکروبیوم و میکروبیوتا را متحول کرده‌اند. یکی از مزایای منحصربه‌فرد روش‌های توالی‌یابی پربازده مانند توالی‌یابی ژن 16S rRNA و متاژنومیکس شاتگان (Shotgun metagenomics)، آن است که این فناوری‌ها امکان بررسی‌های طولی و پایش زمانی را فراهم می‌کنند. این تکنیک‌ها به همراه روش‌های محاسباتی قدرتمند، شناسایی دقیق میکروبیوم‌های خاص و ژن‌ها یا مسیرهای عملکردی را میسر ساخته‌اند؛ این امر به فهم عمیق‌تری از تعاملات پیچیده میان اعضای این اجتماعات میکروبی منجر شده است. همچنین، تحلیل دقیق میکروبیوم به شناسایی گونه‌های میکروبی ناشناخته و نقش آن‌ها در فرایندهای زیستی گوناگون کمک کرده است (۱۱-۱۵).

پيامدهای پژوهش در حوزه میکروبیوم بسیار فراتر از سلامت انسان است. در کشاورزی، درک میکروبیوم خاک برای توسعه روش‌های پایدار کشت و بهبود بهره‌وری محصولات کشاورزی ضروری است. در علوم محیطی نیز پژوهش‌های مرتبط با میکروبیوم به حفظ اکوسیستم‌ها و کاهش اثرات تغییرات اقلیمی کمک می‌کند، زیرا میکروبیوم نقش اساسی در چرخه کربن و سایر فرآیندهای بیوشیمیایی ایفا می‌کنند. با تکامل مستمر تحقیقات در این حوزه، انتظار می‌رود کاربردهای بالقوه علوم میکروبیوم افزایش یافته و حتی به حوزه‌هایی مانند اکتشافات فضایی نیز گسترش یابد؛ در این زمینه شناخت جوامع میکروبی می‌تواند برای حفظ حیات در محیط‌های فرازمینی اهمیت حیاتی داشته باشد (۱۶-۲۰).

پژوهش در زمینه میکروبیوم و میکروبیوتا نیازمند تحلیل داده‌هایی حجیم و پیچیده است که از فناوری‌های توالی‌یابی پرظرفیت، مانند متاژنومیکس، متاترانسکریپتومیکس و توالی‌یابی ژن 16S rRNA به دست می‌آیند. به همین دلیل، استفاده از ابزارهای پیشرفته بیوانفورماتیک برای پردازش و تحلیل حجم عظیم این داده‌ها ضروری است. از اینرو، هدف از این پژوهش مروری معرفی و بیان ضرورت‌های استفاده از روش‌های بیوانفورماتیک در پژوهش‌های میکروبیومی، مزایای این

در پزشکی شخصی تأیید می‌کنند، به طوری که داده‌های میکروبیوم می‌تواند مبنایی برای تدوین استراتژی‌های درمانی متناسب با پروفایل‌های متابولیکی افراد باشد (۲۲).

پیشرفت‌های الگوریتم‌های

یادگیری ماشین در تحلیل داده‌های میکروبیوم (۲۰۲۴)

چن و همکاران، یک الگوریتم جدید یادگیری ماشین را به‌طور اختصاصی برای تحلیل مجموعه‌داده‌های میکروبیوم طراحی کردند که قادر است پیش‌بینی‌های دقیق‌تری از وضعیت بیماری‌ها بر اساس ترکیب و عملکرد اجتماعات میکروبی ارائه دهد. این مطالعه با استفاده از پیشرفت‌های اخیر در زمینه یادگیری عمیق و هوش مصنوعی، مدل‌هایی را ایجاد کرد که نسبت به روش‌های آماری سنتی، پیش‌بینی‌های دقیق‌تری فراهم می‌کنند. کاربرد این الگوریتم در تحلیل داده‌های میکروبیوم منجر به کشفیات جدیدی درباره ارتباط میان ترکیب و عملکرد اجتماعات میکروبی بیماری‌ها شد. به‌عنوان نمونه، نقش اجتماعات میکروبی روده در بیماری‌های خودایمنی نظیر آرتریت روماتوئید (Rheumatoid arthritis) و اختلالات متابولیکی مانند دیابت نوع دو روشن‌تر گردید. این مطالعه، به‌عنوان یکی از پژوهش‌های پیشرو در استفاده از بیوانفورماتیک و هوش مصنوعی در مطالعات میکروبیومی، گامی بزرگ در جهت تحقق پزشکی شخصی بر پایه میکروبیوم و کشف اهداف درمانی جدید محسوب می‌شود (۲۳).

روش تحقیق

روش‌های بیوانفورماتیکی برای تحلیل مجموعه داده‌های پیچیده‌ای که توسط پژوهش‌های نوین میکروبیوم تولید می‌شوند، ضروری و اجتناب‌ناپذیرند. این روش‌ها مراحل مختلفی از جمع‌آوری و پیش‌پردازش داده‌ها تا تحلیل‌های پیشرفته را در بر می‌گیرند و به پژوهشگران اجازه می‌دهند تا بینش‌های تاکسونومیک (طبقه‌بندی میکروبی) و عملکردی (نقش میکروارگانیسم‌ها) در اجتماعات میکروبی را به‌طور دقیق آشکار سازند. در این بخش، مروری جامع بر روش‌های بیوانفورماتیکی مورد استفاده در پژوهش‌های میکروبیومی ارائه می‌شود و بر فناوری‌های توالی‌یابی، کنترل کیفیت داده‌ها و تکنیک‌های تحلیل داده تأکید خواهد شد.

روش‌های استانداردشده برای نمونه‌گیری، توالی‌یابی و تحلیل داده‌های میکروبیوم، موجب پیشرفت چشمگیر در حوزه تحقیقات میکروبیومی شد. همچنین، این پروژه پایه‌گذار تحقیق درباره تغییرات ترکیب میکروبی در شرایط مختلف بیماری گردید (۲).

کاربرد

یادگیری ماشین در تحلیل داده‌های میکروبیوم (۲۰۱۶)

نایتس و همکاران، تکنیک‌های یادگیری ماشین را روی داده‌های میکروبیوم بکار بردند تا فنوتیپ‌های میزبان، از جمله وضعیت‌های بیماری را بر اساس ترکیب اجتماعات میکروبی پیش‌بینی کنند. این مطالعه توان بالقوه روش‌های یادگیری ماشین را در تشخیص الگوهای از داده‌های میکروبیوم نشان داد که به کمک آن‌ها می‌توان وضعیت سلامتی و بیماری میزبان را پیش‌بینی کرد. یافته‌های این پژوهش بیانگر این موضوع بود که الگوریتم‌های یادگیری ماشین قادرند شرایط بیماری‌هایی مانند دیابت نوع دو و بیماری کرون را با دقت بالایی پیش‌بینی کنند. این مطالعه نقطه عطفی در پژوهش‌های میکروبیوم به شمار می‌رود و زمینه را برای تحقیقات بعدی در راستای ایجاد مدل‌های پیش‌بینی‌کننده و درمان‌های شخصی‌سازی‌شده بر اساس اطلاعات میکروبیوم فراهم ساخته است (۲۱).

یکپارچه‌سازی رویکردهای چند-اومیکس (Multi-Omics

Integration) در پژوهش‌های میکروبیوم (۲۰۲۳)

لی و همکاران، با استفاده از رویکردهای چند-اومیکس شامل متازنومیکس، متابولومیکس و پروتئومیکس به بررسی نقش میکروبیوم روده در بیماری‌های متابولیکی پرداختند. در این مطالعه، ابزارهای بیوانفورماتیکی به‌منظور تلفیق لایه‌های مختلف داده‌های اومیکس به‌کار گرفته شد و در نتیجه، دیدگاهی جامع از تعاملات ژن‌ها، متابولیت‌ها و پروتئین‌های میکروبی با محیط میزبانی فراهم گردید. این رویکرد یکپارچه منجر به شناسایی نشانگرهای زیستی و اهداف درمانی نوینی در بیماری‌هایی مانند دیابت نوع دو و بیماری التهابی روده شد. یافته‌های این پژوهش گام مهمی در پیشرفت علم میکروبیوم محسوب می‌شوند و قدرت روش‌های چند-اومیکس را در درک بهتر تعاملات پیچیده میان میزبان و میکروبیوم به اثبات رساندند. همچنین، نتایج به‌دست‌آمده از این پژوهش، رویکردهای چند-اومیکس را

پیش‌بینی ژن‌ها و تفسیر عملکردی است که مبنایی برای تحلیل عمیق‌تر ساختار و عملکرد اجتماعات میکروبی به‌شمار می‌روند (۳۴-۳۸).

در نهایت، یکپارچه‌سازی داده‌ها به‌عنوان یکی از جنبه‌های اساسی پیش‌پردازش در پژوهش‌های میکروبیوم مطرح است. با گسترش استفاده از داده‌های چند-اومیکسی، یکپارچه کردن انواع مختلف داده‌ها شامل داده‌های ژنومیکس، ترانسکریپتومیکس، پروتئومیکس و متابولومیکس چالشی جدی محسوب می‌شود. برای هر یک از انواع داده‌ها، مراحل پیش‌پردازش اختصاصی طراحی می‌شود تا بتوان این داده‌ها را به نحو مؤثری با یکدیگر تلفیق کرد و نمای جامعی از اکوسیستم میکروبی ارائه نمود. در این مرحله معمولاً از الگوریتم‌های پیشرفته برای هم‌ترازسازی و نگاشت داده‌ها در پلتفرم‌های گوناگون استفاده می‌شود تا تمامی فعالیت‌های مرتبط با اجتماعات میکروبی به‌طور یکپارچه بررسی و تحلیل شوند (۳۹-۴۳).

مهم‌ترین و پرکاربردترین نرم‌افزارها و پایگاه‌های داده بیوانفورماتیکی مورد استفاده در پژوهش‌های میکروبیوم

امروزه مجموعه متنوعی از ابزارهای بیوانفورماتیکی و پایگاه‌های داده برای پردازش، تحلیل و تفسیر داده‌های میکروبیوم در دسترس است. این ابزارها در انجام وظایف مختلف از جمله تنظیم توالی‌ها، طبقه‌بندی تاکسونومیک، تفسیر عملکردی و بصری‌سازی داده‌ها به محققان کمک می‌کنند. همچنین، پایگاه‌های داده تخصصی این امکان را به پژوهشگران می‌دهند که داده‌های میکروبیومی را ذخیره کرده و به آن‌ها دسترسی داشته باشند که این امر به اشتراک‌گذاری و یکپارچه‌سازی داده‌ها میان مطالعات گوناگون کمک می‌کند. در این بخش برخی از مهم‌ترین و پرکاربردترین نرم‌افزارها و پایگاه‌های داده بیوانفورماتیکی مورد استفاده در پژوهش‌های میکروبیوم بررسی می‌شوند:

نرم‌افزار QIIME 2

نرم‌افزار QIIME 2 (Quantitative Insights Into Microbial Ecology) یکی از رایج‌ترین و محبوب‌ترین مجموعه ابزارهای بیوانفورماتیکی در تحلیل داده‌های میکروبیوم است. این نرم‌افزار شامل مجموعه جامعی از ابزارها برای پردازش داده‌های توالی‌یابی خام است که شامل کنترل کیفیت توالی‌ها، طبقه‌بندی تاکسونومیک، تحلیل تنوع

نخستین مرحله در تحلیل بیوانفورماتیکی پژوهش‌های میکروبیوم، جمع‌آوری و پیش‌پردازش داده‌ها است که زیربنای تحلیل‌های بعدی را تشکیل می‌دهد. فرایند جمع‌آوری داده‌ها شامل نمونه‌گیری از منابع مختلف میکروبیوم نظیر مدفوع، بزاق یا نمونه‌های بافتی است که به دنبال آن، استخراج DNA، RNA یا پروتئین از این نمونه‌ها انجام می‌گیرد. این مواد زیستی، اطلاعات لازم برای بررسی دقیق اجتماعات میکروبی را فراهم می‌کنند. پیش‌پردازش صحیح این داده‌ها موجب تضمین دقت و کیفیت تکنیک‌های توالی‌یابی و پروفایلینگ می‌شود و به حذف نویز و آمادگی داده‌ها برای تحلیل‌های پیشرفته کمک می‌کند. انجام مؤثر پیش‌پردازش داده‌ها جهت اطمینان از صحت نتایج به‌دست‌آمده از مطالعه امری حیاتی است (۲۸-۲۴).

یکی از متداول‌ترین روش‌های جمع‌آوری داده‌ها در پژوهش‌های میکروبیوم، استفاده از فناوری‌های توالی‌یابی است که شامل توالی‌یابی ژن 16S rRNA، توالی‌یابی متازنومیکس شاتگان و متاترانسکریپتومیکس می‌شود. هر یک از این روش‌ها نوع متفاوتی از اطلاعات را ارائه می‌دهند؛ توالی‌یابی ژن 16S rRNA عمدتاً بر طبقه‌بندی تاکسونومیک تمرکز دارد، در حالی که توالی‌یابی شاتگان امکان تحلیل گسترده‌تر از جمله بررسی عملکرد ژن‌ها را فراهم می‌کند. متاترانسکریپتومیکس نیز یک گام فراتر رفته و بیان فعال ژن‌ها را در اجتماعات میکروبی مورد مطالعه قرار می‌دهد. اقدامات پیش‌پردازشی نظیر کنترل کیفیت، حذف خوانش‌های کم‌کیفیت و حذف آلودگی‌های ناشی از DNA میزبان برای تضمین صحت و دقت تحلیل‌های بعدی بیوانفورماتیکی ضروری هستند (۲۹-۳۳).

پیش‌پردازش همچنین شامل تکنیک‌های نرمال‌سازی است که به‌منظور کاهش سوگیری‌های ناشی از فرایند توالی‌یابی مانند تفاوت در عمق توالی‌یابی مورد استفاده قرار می‌گیرند. تکنیک‌های رایج نرمال‌سازی در مطالعات میکروبیوم شامل رریفکشن (Rarefaction) و مقیاس‌بندی فراوانی نسبی است. این تکنیک‌ها باعث می‌شوند که داده‌های به‌دست‌آمده از نمونه‌های مختلف قابل مقایسه باشند و به پژوهشگران اجازه می‌دهند تا الگوهای زیستی معناداری را شناسایی کنند. سایر مراحل مهم پیش‌پردازش شامل خوشه‌بندی واحدهای تاکسونومیک عملیاتی،

فراهم می‌آورد. این روش در تمایز گونه‌های نزدیک به هم بسیار مؤثر است و دقت بالاتری را در پروفایل‌بندی اجتماعات میکروبی ارائه می‌دهد. نرم‌افزار DADA2 می‌تواند به‌عنوان ابزاری مستقل استفاده شود یا در پلتفرم‌هایی مانند QIIME 2 و دیگر نرم‌افزارهای بیوانفورماتیکی ادغام گردد. توانایی منحصر به فرد DADA2 در اصلاح خطاها باعث شده تا این نرم‌افزار به‌عنوان یکی از دقیق‌ترین ابزارهای موجود برای پردازش داده‌های توالی آمپلیکون شناخته شود (۵۲-۵۰).

نرم‌افزار HUMAnN

نرم‌افزار HUMAnN (The HMP Unified Metabolic Analysis Network) ابزاری بیوانفورماتیکی است که برای استنباط ظرفیت عملکردی اجتماعات میکروبی از داده‌های متانومیکس و متاترانسکریپتومیکس طراحی شده است. این نرم‌افزار به پژوهشگران اجازه می‌دهد تا مسیرهای متابولیکی موجود در اجتماعات میکروبی را از طریق نگاشت داده‌های توالی‌یابی به پایگاه‌های مرجع مانند UniRef شناسایی کنند. نرم‌افزار HUMAnN به‌ویژه برای پیش‌بینی ظرفیت عملکردی اکوسیستم‌های میکروبی و درک فرآیندهای بیوشیمیایی در این اجتماعات سودمند است. HUMAnN معمولاً همراه با نرم‌افزار MetaPhlan برای تعیین پروفایل تاکسونومیک به کار می‌رود که این ترکیب، تحلیل جامعی از طبقه‌بندی و عملکرد اجتماعات میکروبی را امکان‌پذیر می‌سازد (۵۳-۵۵).

پایگاه داده GMrepo

پایگاه داده GMrepo یک منبع جامع و تخصصی است که داده‌های میکروبیوم استخراج شده از پژوهش‌های منتشر شده در مقالات علمی را گردآوری و ارائه می‌کند. این پایگاه به پژوهشگران امکان می‌دهد تا به جست‌وجو و تحلیل تاکسون‌های میکروبی، عملکرد ژن‌ها و ارتباطات میکروبی با شرایط مختلف سلامت و بیماری بپردازند. GMrepo با گردآوری و تجمیع اطلاعات از مطالعات مختلف، بررسی الگوهای میکروبیوم در مجموعه داده‌های متنوع را تسهیل می‌کند. این پایگاه داده برای شناسایی میکروارگانیسم‌های کلیدی دخیل در سلامت و بیماری بسیار مفید بوده و از پژوهش‌ها در حوزه‌های میکروبیوم انسانی، حیوانی و محیطی پشتیبانی می‌کند (۵۶-۵۸).

میکروبی و نیز بصری‌سازی داده‌ها می‌شود. QIIME 2 به‌طور گسترده‌ای در تحلیل داده‌های توالی‌یابی 16S rRNA مورد استفاده قرار می‌گیرد و همچنین از روندهای تحلیل داده‌های توالی‌یابی دیگر نیز پشتیبانی می‌کند. ساختار ماژولار و مستندات جامع این نرم‌افزار باعث شده تا استفاده از آن ساده باشد و به پژوهشگران امکان دهد تحلیل‌ها را متناسب با نیازهای خود شخصی‌سازی کنند. قابلیت ایجاد نمودارهای تعاملی توسط QIIME 2 به محققان کمک می‌کند تا به‌طور مؤثرتری به کاوش و تفسیر داده‌های میکروبیوم بپردازند (۴۴-۴۶).

نرم‌افزار Mothur

نرم‌افزار Mothur یک بسته نرم‌افزاری بیوانفورماتیکی است که به‌طور گسترده برای پردازش داده‌های توالی‌یابی 16S rRNA و سایر داده‌های مرتبط با بوم‌شناسی میکروبی به‌کار می‌رود. این نرم‌افزار نیز ابزارهایی را برای هم‌ردیفی توالی‌ها، طبقه‌بندی تاکسونومیک، خوشه‌بندی توالی‌ها و تحلیل تنوع میکروبی ارائه می‌کند. یکی از نقاط قوت کلیدی نرم‌افزار Mothur توانایی آن در پردازش کارآمد مجموعه داده‌های حجیم است که باعث می‌شود این نرم‌افزار برای تحلیل داده‌های حاصل از پلتفرم‌های توالی‌یابی پربازده مناسب باشد. همچنین این نرم‌افزار از روندهای تحلیل شخصی‌سازی شده پشتیبانی می‌کند و قابلیت یکپارچه‌سازی با سایر ابزارها را برای انجام پژوهش‌های تکرارپذیر در حوزه میکروبیوم فراهم می‌آورد. Mothur به‌طور معمول در مطالعاتی همچون «پروژه میکروبیوم انسانی» استفاده می‌شود و بستری قابل اعتماد برای درک ترکیب و عملکرد جوامع میکروبی به شمار می‌رود (۴۷-۴۹).

نرم‌افزار DADA2

نرم‌افزار DADA2 (Divisive Amplicon Denoising Algorithm 2) ابزاری پیشرفته در حوزه بیوانفورماتیک است که برای پردازش داده‌های توالی آمپلیکون (Amplicon)، به‌ویژه توالی‌یابی 16S rRNA طراحی شده است. برخلاف روش‌های سنتی خوشه‌بندی توالی‌ها، DADA2 از یک رویکرد مبتنی بر مدل برای اصلاح خطاهای موجود در آمپلیکون‌ها استفاده می‌کند و به این ترتیب امکان شناسایی دقیق گونه‌های توالی واقعی را تا سطح تک‌نوکلئوتیدی

پایگاه داده SILVA

پایگاه داده SILVA مجموعه‌ای جامع از توالی‌های RNA ریبوزومی است که عمدتاً برای طبقه‌بندی تاکسونومیک میکروارگانیسم‌ها بر اساس ژن‌های 16S، 18S و 23S rRNA به کار می‌رود. این پایگاه نیز ابزارهایی برای هم‌ردیفی توالی‌ها، تخصیص تاکسونومیک و ساخت درخت فیلوژنتیکی فراهم می‌آورد. SILVA به‌طور گسترده در پژوهش‌های میکروبیوم برای شناسایی و طبقه‌بندی تاکسون‌های میکروبی مورد استفاده قرار می‌گیرد و به‌ویژه در مطالعات مبتنی بر توالی‌یابی آمپلیکون ژن‌های rRNA ابزاری بسیار کارآمد محسوب می‌شود (۵۹-۶۱).

پایگاه داده SRA

پایگاه داده SRA (Sequence Read Archive) یکی از بزرگ‌ترین مخازن داده‌های خام توالی‌یابی، از جمله داده‌های متانومیک و داده‌های مرتبط با میکروبیوم است. SRA داده‌های خام توالی‌یابی تولیدشده توسط پلتفرم‌های مختلف را ذخیره کرده و دسترسی به این مجموعه داده‌ها را برای تحلیل‌های بعدی فراهم می‌آورد. این پایگاه به‌طور گسترده توسط پژوهشگران میکروبیوم مورد استفاده قرار می‌گیرد تا به داده‌های توالی‌یابی از مطالعات گوناگون در محیط‌های مختلف، از جمله میکروبیوم انسانی، حیوانی و محیطی دسترسی پیدا کنند و همچنین داده‌های خود را به اشتراک بگذارند (۶۲-۶۴).

پایگاه داده ENA

پایگاه داده ENA (European Nucleotide Archive) یک پایگاه جامع است که امکان دسترسی به داده‌های توالی‌یابی نوکلئوتیدی شامل مجموعه داده‌های ژنومیک، ترانسکریپتومیک و متانومیک را فراهم می‌کند. این پایگاه بخشی از کنسرسیوم بین‌المللی پایگاه داده‌های توالی نوکلئوتیدی (INSDC) است و به‌عنوان مرکزی برای دسترسی به داده‌های خام و داده‌های توالی‌یابی مونتاژشده عمل می‌کند. پژوهشگران حوزه میکروبیوم از ENA برای دریافت مجموعه داده‌های متانومیک جهت انجام مطالعات مقایسه‌ای یا پروفایلینگ عملکردی اجتماعات میکروبی استفاده می‌کنند (۶۵-۶۷).

پایگاه داده DDBJ

بانک داده DNA ژاپن (DNA Data Bank of Japan) یکی از مهم‌ترین مخازن بین‌المللی داده‌های توالی نوکلئوتیدی است که در چارچوب همکاری بین‌المللی پایگاه‌های داده توالی نوکلئوتیدی (INSDC)، در کنار پایگاه‌های ENA و GenBank فعالیت می‌کند. این پایگاه داده، اطلاعات توالی نوکلئوتیدی شامل داده‌های ژنومیکس، متانومیکس و ترانسکریپتومیکس را ذخیره می‌کند و از پژوهش‌های حوزه میکروبیوم با ارائه بستری برای ثبت و دسترسی به داده‌های خام توالی‌یابی پشتیبانی می‌نماید (۶۸-۷۰).

نتایج تحقیق**اهمیت****بیوانفورماتیک در پژوهش‌های میکروبیوم و میکروبیوتا**

همانگونه که پیش‌تر در این پژوهش بیان شد، بیوانفورماتیک نقشی حیاتی در تحلیل مجموعه داده‌های گسترده و پیچیده‌ای دارد که در پژوهش‌های مرتبط با میکروبیوم و میکروبیوتا تولید می‌شوند. این داده‌ها که حاصل فناوری‌های توالی‌یابی پربازده همچون متانومیکس، متاترانسکریپتومیکس و توالی‌یابی ژن rRNA 16S هستند، برای تفسیر صحیح و دقیق به ابزارهای محاسباتی پیشرفته‌ای نیاز دارند. این بخش به بیان ضرورت بکارگیری بیوانفورماتیک در پژوهش‌های میکروبیومی، مزایای قابل توجه آن نسبت به روش‌های سنتی و نیز محدودیت‌ها و چالش‌هایی که همچنان در این حوزه وجود دارد، می‌پردازد.

ضرورت بکارگیری روش‌های بیوانفورماتیک

ضرورت استفاده از روش‌های بیوانفورماتیک در پژوهش‌های میکروبیوم به حجم عظیم داده‌های تولیدشده توسط فناوری‌های مدرن توالی‌یابی باز می‌گردد. بدون استفاده از ابزارهای بیوانفورماتیک، پردازش میلیون‌ها توالی به دست آمده از تکنیک‌هایی نظیر توالی‌یابی ژن rRNA 16S، متانومیکس شانگان و متاترانسکریپتومیکس عملاً ناممکن خواهد بود. روش‌های بیوانفورماتیک به پژوهشگران امکان می‌دهند تا میکروب‌ها را طبقه‌بندی کنند، تنوع میکروبی را مورد تحلیل قرار دهند و ظرفیت‌های عملکردی اجتماعات میکروبی را پیش‌بینی

می‌دهند تا داده‌های حجیم حاصل از نمونه‌ها و مطالعات مختلف را تحلیل کنند. این مقیاس‌پذیری به‌ویژه برای مطالعات مقایسه‌ای بسیار مهم است که در آن‌ها پژوهشگران قصد بررسی تغییرات جوامع میکروبی را در جمعیت‌ها، مکان‌ها یا شرایط مختلف دارند. روش‌های سنتی قادر به مدیریت و پردازش حجم عظیم داده‌های تولیدشده در چنین مطالعاتی نیستند. علاوه بر این، بیوانفورماتیک انجام مطالعات طولی را تسهیل می‌کند که در آن‌ها تغییرات جوامع میکروبی در طول زمان ردیابی می‌شوند و بینش ارزشمندی از تأثیر دینامیک میکروبی بر سلامت، بیماری یا فرآیندهای محیطی فراهم می‌آید (۸۲-۸۰). ابزارهای بیوانفورماتیکی همچنین موجب افزایش قابلیت بازتولیدپذیری در مطالعات میکروبیوم می‌شوند. این ابزارها به پژوهشگران امکان می‌دهند تا مسیرهای محاسباتی را به اشتراک گذاشته و آن‌ها را در پروژه‌های مختلف به‌کار گیرند و از این طریق تحلیل‌های مشابه را بر روی مجموعه داده‌های متفاوت به شکلی یکنواخت و استاندارد اجرا کنند. این بازتولیدپذیری برای ایجاد دانش مستحکم در زمینه میکروبیوم و اعتباربخشی به یافته‌ها در مطالعات مستقل بسیار حیاتی است. انتظار می‌رود با تکامل ابزارهای بیوانفورماتیکی، توانایی آن‌ها در ارائه بینش‌های دقیق‌تر، جامع‌تر و بازتولیدپذیرتر در پژوهش‌های میکروبیوم افزایش یابد (۸۳-۸۵).

معایب و محدودیت‌های روش‌های بیوانفورماتیکی

با وجود مزایای چشمگیر روش‌های بیوانفورماتیکی در پژوهش‌های مرتبط با میکروبیوم، چندین محدودیت و چالش نیز در این حوزه وجود دارد که باید مورد توجه قرار گیرند. یکی از مهم‌ترین محدودیت‌ها، وابستگی زیاد به پایگاه‌های داده مرجع برای طبقه‌بندی تاکسونومیک و حاشیه‌نویسی عملکردی است. این پایگاه‌ها معمولاً ناقص هستند و به سمت موجودات کاملاً شناخته‌شده تمایل دارند و در نتیجه، دانش ما در خصوص میکروارگانیسم‌های کمتر شناخته‌شده یا تازه کشف‌شده همچنان ناقص است. این مسئله باعث بروز خطا در طبقه‌بندی یا نادیده گرفتن گونه‌های جدید می‌شود؛ به‌خصوص در محیط‌های کمتر بررسی‌شده که دقت تحلیل‌های میکروبیوم را محدود می‌کند (۸۶-۸۸). چالش دیگر بیوانفورماتیک، نیاز به توان محاسباتی بسیار زیاد برای تحلیل مجموعه داده‌های حجیم

نماینده به‌عنوان مثال، ابزارهایی مانند QIIME 2 و DADA2 به‌طور گسترده برای طبقه‌بندی تاکسونومیک میکروبیوم استفاده می‌شوند و نرم‌افزارهایی مانند MetaPhlan و HUMAnN برای تحلیل عملکردی داده‌های متانومیک کاربرد دارند. این ابزارها علاوه بر کمک به شناسایی گونه‌ها، این امکان را فراهم می‌کنند تا نقش‌های اکولوژیکی و عملکردی میکروبیوم‌ها در اکوسیستم میکروبی به دقت بررسی شود و از این رو، بیوانفورماتیک نقشی اساسی و غیرقابل چشم‌پوشی در این حوزه ایفا می‌کند (۷۱-۷۳).

مزایای

روش‌های بیوانفورماتیکی نسبت به روش‌های سنتی

بیوانفورماتیک در مقایسه با روش‌های سنتی میکروبیولوژی مزایای متعددی دارد. روش‌های سنتی اغلب بر تکنیک‌های کشت متکی هستند که تنها بخشی از تنوع میکروبی موجود در یک نمونه را پوشش می‌دهند. بسیاری از میکروارگانیسم‌های موجود در میکروبیوم با روش‌های آزمایشگاهی استاندارد قابل کشت نیستند و در نتیجه بررسی کامل پیچیدگی اجتماعات میکروبی با این روش‌ها دشوار است. در مقابل، ابزارهای بیوانفورماتیکی به پژوهشگران این امکان را می‌دهند تا توالی‌های DNA را مستقیماً از نمونه‌های محیطی یا بالینی تحلیل کنند و به این ترتیب نیاز به کشت را از میان بردارند. این رویکرد به نمایش دقیق‌تر و جامع‌تری از میکروبیوم و تنوع آن منجر می‌شود و عمق تحلیل را به میزان قابل توجهی افزایش می‌دهد (۷۴-۷۶). علاوه بر آشکارسازی تنوع میکروبی، بیوانفورماتیک امکان تحلیل گسترده و هم‌زمان مجموعه داده‌های متعدد را فراهم می‌سازد که انجام آن با روش‌های سنتی عملاً غیرممکن است. این ابزارها قادرند داده‌های حاصل از فناوری‌های مختلف اومیکس، نظیر ژنومیکس، ترانسکریپتومیکس، پروتئومیکس و متابولومیکس را تلفیق کرده و دیدگاهی جامع از اجتماعات میکروبی ارائه دهند. چنین یکپارچه‌سازی نه‌تنها به شناسایی گونه‌های میکروبی موجود در یک محیط کمک می‌کند، بلکه فعالیت‌ها و نحوه تعامل آن‌ها با میزبان یا محیط را نیز مشخص می‌سازد. توانایی تولید بینش‌های عملکردی در کنار پروفایل‌های تاکسونومیک، یکی از نقاط قوت اصلی بیوانفورماتیک در پژوهش‌های میکروبیوم است (۷۹-۷۷). همچنین روش‌های بیوانفورماتیکی مقیاس‌پذیر هستند و به پژوهشگران امکان

بحث و نتیجه‌گیری

مطالعه ساختار، عملکرد و دینامیک میکروبیوم و میکروبیوتا با اتکا به فناوری‌های توالی‌یابی و ابزارهای بیوانفورماتیکی، به یکی از محورهای اصلی پژوهش‌های زیست‌پزشکی و زیست‌محیطی تبدیل شده است. در این مقاله، نشان داده شد که ابزارهای بیوانفورماتیکی مانند QIIME2، و DADA2 و پایگاه‌های داده‌ای چون SILVA و ENA، نه تنها در شناسایی تاکسونومی و تحلیل تنوع میکروبی نقش کلیدی ایفا می‌کنند، بلکه در شبیه‌سازی مسیرهای متابولیکی و عملکردی اجتماعات میکروبی نیز بسیار مؤثر هستند. به‌ویژه، رویکردهای چند-اومیکس با ادغام داده‌های ژنومیکس، پروتئومیکس و متابولومیکس، درک جامعی از برهم‌کنش‌های میزبان-میکروبی را فراهم می‌سازند.

یکی از مهم‌ترین پیامدهای پیشرفت در بیوانفورماتیک، امکان بهره‌گیری از داده‌های میکروبیوم در توسعه درمان‌های مبتنی بر

پزشکی فردمحور است. به‌طور خاص، استفاده از مدل‌های یادگیری ماشین در تحلیل داده‌های پیچیده میکروبی، توانایی پیش‌بینی وضعیت‌های بیماری، شناسایی نشانگرهای زیستی و طراحی مداخلات درمانی شخصی‌سازی‌شده را فراهم کرده است. این دستاوردها در زمینه‌هایی مانند بیماری‌های گوارشی، متابولیکی و حتی روان‌پزشکی چشمگیر بوده و آینده روشنی را برای کاربردهای بالینی ترسیم می‌کند.

با وجود این پیشرفت‌ها، چالش‌هایی همچون نیاز به منابع محاسباتی پرقدرت، دشواری در تحلیل گونه‌های با فراوانی کم، و نبود استانداردهای تحلیلی یکپارچه همچنان پابرجاست. برای غلبه بر این موانع، توسعه ابزارهای کاربرپسند، آموزش پژوهشگران در حوزه‌های بین‌رشته‌ای، و گسترش پایگاه‌های جامع و به‌روز، ضروری به نظر می‌رسد. این مقاله با مرور جامع ادبیات علمی موجود، نقشه راهی برای پژوهش‌های آینده در راستای بکارگیری مؤثر بیوانفورماتیک در مطالعات میکروبیومی فراهم می‌سازد.

است. توالی‌یابی پُربازده حجم عظیمی از داده‌ها تولید می‌کند و تحلیل این داده‌ها اغلب نیازمند منابع محاسباتی قابل‌توجه و همچنین تخصص بالا است. بسیاری از ابزارهای بیوانفورماتیکی به منابع سخت‌افزاری گسترده‌ای نیاز دارند که ممکن است برای گروه‌های پژوهشی کوچک‌تر در دسترس نباشد و این مسئله، مانعی جدی در مسیر انجام تحلیل‌های پیشرفته میکروبیوم به شمار می‌رود (۹۱-۸۹). علاوه بر این، شناسایی گونه‌های میکروبی با فراوانی کم با ابزارهای بیوانفورماتیکی فعلی دشوار است. بسیاری از این ابزارها برای تشخیص تاکسون‌های غالب در جوامع میکروبی بهینه‌سازی شده‌اند و در شناسایی گونه‌های نادر که ممکن است نقش اکولوژیکی بسیار مهمی داشته باشند، دچار مشکل می‌شوند. گونه‌های کم‌فراوان معمولاً در تحلیل‌ها توسط گونه‌های پرشمارتر پوشیده می‌شوند و این موضوع منجر به نمایه ناقصی از جامعه میکروبی می‌گردد. علاوه بر این، اغلب مسیرهای تحلیل بیوانفورماتیکی شامل مراحل متعدد هستند و تغییر در انتخاب ابزارها، پارامترها و پایگاه‌های داده می‌تواند منجر به نتایج متناقض در مطالعات مختلف شود و به این ترتیب بازتولیدپذیری در پژوهش‌های میکروبیوم را به چالشی مهم بدل می‌سازد (۹۴-۹۲). در نهایت، اگرچه ابزارهای بیوانفورماتیکی حجم زیادی از اطلاعات را در اختیار پژوهشگران قرار می‌دهند، تفسیر نتایج به‌دست آمده اغلب نیازمند دانش تخصصی عمیق در حوزه‌های زیست‌شناسی محاسباتی و میکروبیولوژی است. پژوهشگران نه تنها باید با عملکرد و جزئیات ابزارهای محاسباتی آشنایی داشته باشند، بلکه لازم است اهمیت زیستی نتایج را نیز به‌خوبی درک کرده و آن‌ها را در بستر درست علمی قرار دهند. این دانش بین‌رشته‌ای در بسیاری از گروه‌های پژوهشی محدود است و بر ضرورت ارائه آموزش‌های تخصصی و همچنین توسعه ابزارهای بیوانفورماتیکی کاربرپسند برای محققانی که تخصص کمتری دارند تأکید می‌کند (۹۵-۹۷).

منابع

1. Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R., & Gordon, J.I. (2007). The

Human Microbiome Project. *Nature*, 449(7164), 804-810.

2. Human Microbiome Project Consortium. (2012). Structure, function, and diversity of the healthy human microbiome. *Nature*, 486(7402), 207-214.
3. Lynch, S.V., & Pedersen, O. (2016). The human intestinal microbiome in health and disease. *New England Journal of Medicine*, 375(24), 2369-2379.
4. Gill, S.R., Pop, M., Deboy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S., ... & Fraser-Liggett, C.M. (2006). Metagenomic analysis of the human distal gut microbiome. *Science*, 312(5778), 1355-1359.
5. Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., ... & Bork, P. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285), 59-65.
6. Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., & Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature*, 489(7415), 220-230.
7. Kau, A.L., Ahern, P.P., Griffin, N.W., Goodman, A.L., & Gordon, J.I. (2011). Human nutrition, the gut microbiome and the immune system. *Nature*, 474(7351), 327-336.
8. Qin, N., Yang, F., Li, A., Prifti, E., Chen, Y., Shao, L., ... & Ehrlich, S.D. (2014). Alterations of the human gut microbiome in liver cirrhosis. *Nature*, 513(7516), 59-64.
9. Ridaura, V.K., Faith, J.J., Rey, F.E., Cheng, J., Duncan, A.E., Kau, A.L., ... & Gordon, J.I. (2013). Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science*, 341(6150), 1241214.
10. Cryan, J.F., & Dinan, T.G. (2012). Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nature Reviews Neuroscience*, 13(10), 701-712.
11. Shreiner, A.B., Kao, J.Y., & Young, V.B. (2015). The gut microbiome in health and in disease. *Current Opinion in Gastroenterology*, 31(1), 69-75.
12. Quince, C., Walker, A.W., Simpson, J.T., Loman, N.J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, 35(9), 833-844.
13. O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufu, S., Haddad, D., McVeigh, R., ... & Pruitt, K.D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 44(D1), D733-D745.
14. Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.F., ... & Woyke, T. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, 499(7459), 431-437.
15. Li, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., ... & Wang, J. (2014). An integrated catalog of reference genes in the human gut microbiome. *Nature Biotechnology*, 32(8), 834-841.
16. Fierer, N., & Jackson, R.B. (2006). The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences*, 103(3), 626-631.
17. Mendes, R., Kruijt, M., de Bruijn, I., Dekkers, E., van der Voort, M., Schneider, J.H., ... & Raaijmakers, J.M. (2011). Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science*, 332(6033), 1097-1100.
18. Berendsen, R.L., Pieterse, C.M., & Bakker, P.A. (2012). The rhizosphere microbiome and plant health. *Trends in Plant Science*, 17(8), 478-486.
19. Falkowski, P.G., Fenchel, T., & Delong, E.F. (2008). The microbial engines that drive Earth's biogeochemical cycles. *Science*, 320(5879), 1034-1039.
20. Blaser, M.J., & Falkow, S. (2009). What are the consequences of the disappearing human microbiota? *Nature Reviews Microbiology*, 7(12), 887-894.
21. Knights, D., Silverberg, M.S., Weersma, R.K., et al. (2016). Prediction of host phenotype based on microbial community structures in the human microbiome. *PLOS Computational Biology*, 7(3), e1002124.
22. Li, J., Jia, H., Cai, X., et al. (2023). Integration of metagenomics, metabolomics, and proteomics reveals the role of the gut microbiome in metabolic diseases. *Nature Communications*, 14, 3056.
23. Chen, X., Wang, Y., Huang, Z., et al. (2024). Advancing machine learning algorithms for microbiome research: Predicting disease states and therapeutic outcomes. *Bioinformatics*, 40(2), 1234-1246.
24. Bolger, A.M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120.
25. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, 17(1), 10-12.
26. Langmead, B., & Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357-359.
27. Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27(6), 863-864.
28. Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data.
29. Caporaso, J.G., Lauber, C.L., Walters, W.A., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME Journal*, 6(8), 1621-1624.
30. Quince, C., Walker, A.W., Simpson, J.T., et al. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, 35(9), 833-844.
31. Franzosa, E.A., McIver, L.J., Rahnava, G., et al. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. *Nature Methods*, 15(11), 962-968.
32. Zhernakova, A., Kurilshikov, A., Bonder, M.J., et al. (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science*, 352(6285), 565-569.
33. Choi, J., Kim, S., Lee, J.Y., et al. (2020). Integrated metagenomics and metabolomics reveal correlations between gut microbiome composition and host metabolism in health and disease. *Nature Communications*, 11(1), 1-13.
34. McMurdie, P.J., & Holmes, S. (2013). Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLOS One*, 8(4), e61217.
35. Love, M.I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion

- for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550.
36. Edgar, R.C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, 10(10), 996-998.
 37. Haas, B.J., Gevers, D., Earl, A.M., et al. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research*, 21(3), 494-504.
 38. Schloss, P.D., Westcott, S.L., Ryabin, T., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 75(23), 7537-7541.
 39. Franzosa, E.A., Sirota-Madi, A., Avila-Pacheco, J., et al. (2019). Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nature Microbiology*, 4(2), 293-305.
 40. Knight, R., Vrbanac, A., Taylor, B.C., et al. (2018). Best practices for analysing microbiomes. *Nature Reviews Microbiology*, 16(7), 410-422.
 41. Li, D., Liu, C.M., Luo, R., et al. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, 31(10), 1674-1676.
 42. Noecker, C., Eng, A., & Borenstein, E. (2016). A method for inferring the contribution of mutualistic microbes to host traits. *PLoS Computational Biology*, 12(9), e1004889.
 43. Luo, Y., & Ye, Y. (2021). Multi-omics integration reveals novel host-microbiome interactions in microbial communities. *Bioinformatics Advances*, 37(3), 292-299.
 44. Bolyen, E., Rideout, J.R., Dillon, M.R., et al. (2019). Reproducible, interactive, scalable, and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, 37(8), 852-857.
 45. Caporaso, J.G., Kuczynski, J., Stombaugh, J., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7(5), 335-336.
 46. Estaki, M., Jiang, L., Bokulich, N.A., et al. (2020). QIIME 2 enables comprehensive end-to-end analysis of diverse microbiome data and comparative studies with publicly available data. *Scientific Data*, 7(1), 412.
 47. Schloss, P.D., Westcott, S.L., Ryabin, T., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 75(23), 7537-7541.
 48. Kozich, J.J., Westcott, S.L., Baxter, N.T., et al. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology*, 79(17), 5112-5120.
 49. Schloss, P.D. (2010). The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Computational Biology*, 6(7), e1000844.
 50. Callahan, B.J., McMurdie, P.J., Rosen, M.J., et al. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581-583.
 51. Callahan, B.J., Sankaran, K., Fukuyama, J.A., et al. (2016). Bioconductor workflow for microbiome data analysis: from raw reads to community analyses. *F1000Research*, 5, 1492.
 52. Nearing, J.T., Douglas, G.M., Comeau, A.M., & Langille, M.G.I. (2018). Denoising the denoisers: an independent evaluation of microbiome sequence error-correction approaches. *6*, e5364.
 53. Franzosa, E.A., McIver, L.J., Rahnnavard, G., et al. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. *Nature Methods*, 15(11), 962-968.
 54. Truong, D.T., Franzosa, E.A., Tickle, T.L., et al. (2015). MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nature Methods*, 12(10), 902-903.
 55. Abubucker, S., Segata, N., Goll, J., et al. (2012). Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Computational Biology*, 8(6), e1002358.
 56. Wu, L., Zeng, T., Delgadillo, L.F., et al. (2020). GMrepo: a database of curated and consistently annotated human gut metagenomes. *Nucleic Acids Research*, 48(D1), D545-D553.
 57. Li, F., Ning, L., Zeng, T., & Zhou, L. (2021). Benchmarking human gut microbiome gene function annotation methods using GMrepo. *BMC Genomics*, 22(1), 543.
 58. McDonald, D., Hyde, E., Debelius, J.W., et al. (2018). American gut: an open platform for citizen science microbiome research. *mSystems*, 3(3), e00031-18.
 59. Quast, C., Pruesse, E., Yilmaz, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1), D590-D596.
 60. Yilmaz, P., Parfrey, L.W., Yarza, P., et al. (2014). The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Research*, 42(D1), D643-D648.
 61. Glöckner, F.O., Yilmaz, P., Quast, C., et al. (2017). 25 years of serving the community with ribosomal RNA gene reference databases and tools. *Journal of Biotechnology*, 261, 169-176.
 62. Leinonen, R., Sugawara, H., Shumway, M., & The International Nucleotide Sequence Database Collaboration. (2011). The Sequence Read Archive. *Nucleic Acids Research*, 39(S1), D19-D21.
 63. Kodama, Y., Shumway, M., & Leinonen, R. (2012). The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Research*, 40(D1), D54-D56.
 64. Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., & Sayers, E.W. (2016). GenBank. *Nucleic Acids Research*, 44(D1), D67-D72.
 65. Silvester, N., Alako, B., Amid, C., et al. (2021). The European Nucleotide Archive in 2021. *Nucleic Acids Research*, 50(D1), D106-D110.
 66. Leinonen, R., Sugawara, H., Shumway, M., & The International Nucleotide Sequence Database Collaboration. (2011). The Sequence Read Archive. *Nucleic Acids Research*, 39(S1), D19-D21.
 67. Harrison, P.W., Ahamed, A., Aslam, R., et al. (2021). The European Nucleotide Archive in 2020. *Nucleic Acids Research*, 49(D1), D82-D85.

68. Ogasawara, O., Kodama, Y., Mashima, J., et al. (2021). DDBJ Database Updates and Computational Infrastructure. *Nucleic Acids Research*, 50(D1), D109-D114.
69. Kodama, Y., Mashima, J., Kosuge, T., et al. (2018). DNA Data Bank of Japan: 30th anniversary. *Nucleic Acids Research*, 46(D1), D30-D35.
70. Kaminuma, E., Mashima, J., Kodama, Y., & Ogasawara, O. (2020). DDBJ new system and service refactoring for enhancing usability. *Nucleic Acids Research*, 48(D1), D45-D49.
71. Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., ... & Caporaso, J.G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, 37(8), 852-857.
72. Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., ... & Huttenhower, C. (2015). MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nature Methods*, 12(10), 902-903.
73. Franzosa, E.A., McIver, L.J., Rahnavard, G., Thompson, L.R., Schirmer, M., Weingart, G., ... & Huttenhower, C. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. *Nature Methods*, 15(11), 962-968.
74. Hugerth, L.W., Larsson, J., Alneberg, J., Lindh, M.V., Legrand, C., Pinhassi, J., & Andersson, A.F. (2015). Metagenome-assembled genomes uncover a global bacterial functional guild. *Nature*, 528(7583), 504-509.
75. Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.F., ... & Woyke, T. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, 499(7459), 431-437.
76. Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J., ... & Knight, R. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*, 551(7681), 457-463.
77. Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., ... & Bork, P. (2015). Structure and function of the global ocean microbiome. *Science*, 348(6237), 1261359.
78. Gilbert, J.A., Jansson, J.K., & Knight, R. (2014). The Earth Microbiome project: successes and aspirations. *BMC Biology*, 12(1), 69.
79. Salazar, G., Cornejo-Castillo, F.M., Benitez-Barrios, V., Fraile-Nuez, E., Álvarez-Salgado, X.A., Duarte, C.M., ... & Acinas, S.G. (2016). Global diversity and biogeography of deep-sea pelagic prokaryotes. *The ISME Journal*, 10(3), 596-608.
80. Chen, J., Wang, Y., Zhang, L., & Xu, J. (2024). Statistical approaches for longitudinal microbiome data analysis: Challenges and opportunities. *Genes*, 15(1), 51.
81. Martínez Arbas, S., Narayanasamy, S., Herold, M., & Vogel, J. (2021). Longitudinal microbiome studies: Challenges, strategies, and perspectives. *Frontiers in Genetics*, 12, 666244.
82. Asgari, E., McHardy, A.C., & Mofrad, M.R. (2015). A novel scalable computational framework for high-throughput microbiome analysis. *arXiv preprint*, arXiv:1512.00397.
83. Keefe, C.R., Dillon, M.R., Herman, C., Jewell, M., Wood, C.V., Bolyen, E., & Caporaso, J.G. (2023). Facilitating bioinformatics reproducibility. *arXiv preprint*, arXiv:2305.11198.
84. Hothorn, T., & Leisch, F. (2011). Case studies in reproducibility. *Briefings in Bioinformatics*, 12(3), 288-300.
85. Sandve, G.K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten simple rules for reproducible computational research. *PLoS Computational Biology*, 9(10), e1003285.
86. Wright, A.P., Williams, H., & Chen, J. (2022). Ten common issues with reference sequence databases and how to address them. *Frontiers in Microbiology*, 13, 10978663.
87. Martínez-Porchas, M., & Vargas-Albores, F. (2017). Microbial metagenomics in aquaculture: a potential tool for a deeper insight into the activity. *Reviews in Aquaculture*, 9(1), 42-56.
88. Balvočiūtė, M., & Huson, D.H. (2017). SILVA, RDP, Greengenes, NCBI and OTT—how do these taxonomies compare? *BMC Genomics*, 18(1), 114.
89. Petrillo, U.F., Sorella, M., Cattaneo, G., Giancarlo, R., & Rombo, S. (2018). Analyzing big datasets of genomic sequences: Fast and scalable collection of k-mer statistics. *BMC Bioinformatics*, 19(1), 199. *arXiv+1BioMed Central+1*
90. Aluru, S. (2015). Modelling computational resources for next generation sequencing bioinformatics analysis of 16S rRNA samples. *arXiv preprint*, arXiv:1503.02974.
91. Sandve, G.K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten simple rules for reproducible computational research. *PLoS Computational Biology*, 9(10), e1003285.
92. Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., & Madsen, K.L. (2016). Characterization of the gut microbiome using 16S or shotgun metagenomics. *Frontiers in Microbiology*, 7, 459.
93. Nearing, J.T., Douglas, G.M., Comeau, A.M., & Langille, M.G. (2018). Denoising the denoisers: an independent evaluation of microbiome sequence error-correction approaches. *PeerJ*, 6, e5364.
94. Pollock, J., Glendinning, L., Wisedchanwet, T., & Watson, M. (2018). The madness of microbiome: attempting to find consensus “best practice” for 16S microbiome studies. *Applied and Environmental Microbiology*, 84(7), e02627-17.
95. Attwood, T.K., Blackford, S., Brazas, M.D., Davies, A., & Schneider, M.V. (2017). Bioinformatics training: a review of challenges, actions and support requirements. *Briefings in Bioinformatics*, 18(4), 558-565.
96. Mulder, N., Schwartz, R., Brazas, M.D., Brooksbank, C., Gaeta, B., Morgan, S.L., & Schneider, M.V. (2018). The development and application of bioinformatics core competencies to improve bioinformatics training and education. *PLoS Computational Biology*, 14(2), e1005772.
97. Wilson Sayres, M.A., Hauser, C., Sierk, M., Robic, S., Rosenwald, A.G., Smith, T.M., Triplett, E.W., Williams, J.J., Dinsdale, E., Morgan, W.R., Burnette, J.M., Donovan, S.S., Goller, C.C., Haverkamp, C.R., Hughes, L.E., Kelso, A.A., Drew, J.C., & Elgin, S.C.R. (2018). Bioinformatics core competencies for undergraduate life sciences education. *PLoS One*, 13(6), e0196878.

Bioinformatics Approaches to Microbiome and Microbiota Analysis: A Comprehensive Review

Noohi B. and Zahedi M.

Faculty of Chemistry and Petroleum Sciences, Shahid Beheshti University, Tehran, I.R. of Iran

Abstract

In recent decades, the study of the microbiome and microbiota has emerged as a pivotal field within systems biology, attracting increasing attention across medical, agricultural, and environmental sciences. The rapid advancement of next-generation sequencing (NGS) technologies has enabled in-depth and accurate analysis of microbial communities. However, the interpretation and extraction of meaningful insights from the massive and complex datasets generated by these technologies would be unfeasible without the aid of bioinformatics tools. This review provides a comprehensive overview of bioinformatics applications in microbiome analysis, introducing and comparing key software tools and databases such as QIIME2, DADA2, ENA, and SILVA. Additionally, multi-omics approaches—including metagenomics, metatranscriptomics, metaproteomics, and metabolomics—are explored as innovative methods for assessing microbial function. The article further highlights how machine learning techniques and predictive modeling can transform biological data into powerful tools for early disease detection, personalized therapeutics, and biomarker discovery. Finally, the study addresses critical challenges in the field, such as the high computational demands, the complexity of data interpretation, and the limitations of existing reference databases. It underscores the pressing need for developing advanced computational infrastructure and interdisciplinary training programs. Overall, this review offers a scientific roadmap for researchers aiming to decode the biological intricacies of the microbiome through the lens of bioinformatics.

Keywords: Microbiome, Microbiota, Bioinformatics, Databases and Software Tools, Multi-Omics Data.