

متاآنلیکس و کارکردهای آن در شناخت محیط زیست پیرامون ما

محمدحسین همت جو¹، داود نامدار خجسته^{2*}، آرزو طهمورث پور³، اصغر میرزایی اصل⁴

1. دانش آموخته دکتری بیولوژی و بیوتکنولوژی خاک، گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه بوعلی سینا، همدان

2. استادیار گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه شاهد، تهران

3. دانشیار گروه بیوتکنولوژی پزشکی، دانشگاه آزاد اسلامی واحد اصفهان (خوراسگان)، ایران

4. دانشیار گروه بیوتکنولوژی کشاورزی، دانشکده کشاورزی، دانشگاه بوعلی سینا، همدان

*نویسنده مسئول : d.namdar@shahed.ac.ir

پذیرش: 1400/7/24

دریافت: 1399/9/4

چکیده

شمار سلول‌های میکروبی در سیاره زمین از ستاره‌هایی که در کهکشان‌ها می‌شناسیم، بسیار بیشتر است. گوناگونی میکروبی و شبکه بوم‌شناختی آنها با اینکه کارکردهای ویژه‌ای در زیست‌بوم‌های کره زمین دارند، ناشناخته مانده است. فناوری‌های آمیکس مانند متاژنومیکس ابزارهایی را برای شناخت بخش بزرگی از این نادیده‌ها با دقت زیادی فراهم کرده‌اند که ژرفای آگاهی‌هایی که نسبت به آنچه از روش‌های مبتنی بر کشت میکروارگانیسم‌ها به دست می‌آید، بسیار بیشتر است. یک نمونه در مورد گوناگونی و مپایرلیس‌ها است که از آغازیان و یوکاریوت‌های میکروسکوپی هستند. روش‌های متاآمیکس آشکار کرده‌اند که گوناگونی درون این گروه از میکروارگانیسم‌ها با گوناگونی همه سلسله قارچ‌ها برابری می‌کند. آنها در هر گوشه‌ای از طبیعت، از اقیانوس‌ها گرفته تا خاک‌ها دیده می‌شوند و تنها یکی از هفت گروه آغازیان هستند. در این مقاله افزون بر آشنایی با فناوری‌های آمیکس به پژوهش‌های کلان در این زمینه نیز پرداخته شده است تا دستاوردهای آنها برای شناخت میکروارگانیسم‌ها و کاربرد آنها ارائه شود. در متاژنومیکس توالی‌یابی مستقیم، شناسایی ژن‌ها و ژنوم‌های موجود در زیست‌بوم‌های پیچیده میکروبی انجام می‌شود. ویرومیکس پژوهش بر متاژنوم‌های ویروسی است. در متاترانسکرپتومیکس mRNA آنالیز می‌شود که به دلیل ناپایداری بودن آن در نمونه‌های زیست‌محیطی، گردآوری و رمزگشایی از آن چالش اصلی این آمیکس است. شناسایی و اندازه‌گیری پروتئین‌های گوناگون که می‌تواند به‌طور مستقیم فعالیت میکروبی را اندازه‌گیری کند، در متاپروتئومیکس انجام می‌شود. متابولومیکس زیست‌محیطی، مطالعه متابولیت‌های مولکولی با وزن کم تولیدشده از برهم‌کنش‌های میان میکروارگانیسم‌ها همانند یوکاریوت‌های کوچک، گیاهان، جانوران، شکارچیان و تنش‌های غیرزنده و دیگر محرک‌ها است. هم‌اکنون، روش‌های متاآمیکس به‌سرعت در حال توسعه و تکامل فناوری‌هایی هستند که دستیابی ما را هم به گوناگونی زیستی و هم به بیان ژن در تمامی محیط‌زیست‌ها از بدن انسان تا خاک و اقیانوس، شدنی می‌کنند. به این ترتیب قوانینی را که میلیون‌ها سال است حاکم هستند، درک کرده و از رازهای پنهان آنها رمز خواهیم گشود.

واژه‌های کلیدی: متاآمیکس، متاژنومیکس، ویرومیکس، متاترانسکرپتومیکس، متاپروتئومیکس، متابولومیکس و گوناگونی میکروبی.

مقدمه

در واژه متاآنلیکس پیشوند متا¹ به معنی فرا یا ماورا است و در اینجا به فراتر از یک جاندار یا پدیده اشاره می‌کند. بنابراین نشان‌دهنده فراتر یا آن سوی جاندار یا پدیده، یعنی محیط‌زیست یا زیستگاه موجودات زنده است. پس زیستگاه و همه جاندارانی را که در آن زندگی می‌کنند، در بر می‌گیرد. واژه آنلیکس پدیدآمده از پسوند ome در زیست‌شناسی است که تمامیت یا کلیت هر آنچه را که پژوهش می‌شود، نشان می‌دهد. برای نمونه، به مجموعه‌ای از ارگانیسم‌های زنده بیوم، به مجموعه ژن‌ها در یک ارگانیسم، ژنوم و به مجموعه میکروارگانیسم‌ها در یک زیستگاه میکروبیوم می‌گویند [1]. پس آنلیکس پژوهش و بررسی همزمان همه این مجموعه‌هاست که در آن افزون بر ساختار موجود، ویژگی‌های مولکول‌های گوناگون درگیر در آنها را نیز در بر می‌گیرد. مطالعه گسترده ژنوم‌ها یا هریک از ارگانیسم‌ها و یا متاژنوم‌های جوامع برای درک برهم‌کنش‌های فیلوژنتیکی و ارزیابی‌های وراثتی همچون انتقال افقی ژن و ... ارزش بسیار دارند و در این سطح بررسی، آنلیکس و متاآنلیکس درحقیقت به بررسی DNA می‌پردازند که شامل ژنومیکس و متاژنومیکس (یا ژنومیکس جامعه، بررسی ژنوم‌های آمیخته به‌دست‌آمده از یک محیط‌زیست مانند خاک) می‌شوند. مفهوم آنلیکس از DNA فراتر رفته و شامل رونوشت‌های RNA (ترانسکریپتومیکس یا متاترانسکریپتومیکس)، پروتئین‌ها (پروتئومیکس یا متاپروتئومیکس) و متابولیت‌ها (متابولومیکس یا متابولومیکس زیست‌محیطی) هم می‌شود که به آن نظام آبشاری آنلیکس گفته می‌شود. در این نظام آنلیکس، گوناگونی‌های آنلیکسی دیگری نیز در سال‌های پیشین پدیدار شده است که بخشی از این نظام آبشاری آنلیکس هستند همانند گلیکومیکس، لیپیدومیکس و اینترکتومیکس [2]. هدف این مقاله آشنایی با روش‌ها و

فنون متاآنلیکس برای شناخت میکروبیوم‌ها در محیط‌زیست‌های گوناگون کره زمین است که در سال‌های پیشین دستاوردهای شگرفی داشته‌اند. این روش‌ها در برگیرنده استخراج، بررسی و تجزیه و تحلیل اسیدهای نوکلئیک مانند DNA و RNA، پروتئین‌ها و متابولیت‌ها از محیط‌زیست هستند که به ترتیب متاژنومیکس، متاترانسکریپتومیکس، متاپروتئومیکس و متابولومیکس زیست‌محیطی نامیده می‌شوند و در ادامه آورده شده‌اند. سپس، پژوهش‌های کلان که در این زمینه‌ها طراحی و اجرا شده‌اند، نیز معرفی و برخی از دستاوردهای آنها آورده شده‌اند.

متاژنومیکس

هندلسمن و همکاران [3] نخستین بار واژه متاژنومیکس را برای ژن‌های گردآوری‌شده از میکروارگانیسم‌های یک جامعه (همانند یک نمونه خاک) به کار بردند. سپس، تعریف متاژنومیکس گسترش پیدا کرد و دربرگیرنده همه روش‌هایی گردید که برپایه آنالیز DNA استخراج‌شده از نمونه‌های زیست‌محیطی باشد. این تعریف گسترش پیدا کرده و شامل توالی‌یابی ژن 16S rRNA و روش‌های انگشت‌نگاری فیلوژنتیکی مربوط به آن می‌شود [4,5]. درحقیقت، متاژنومیکس به معنای توالی‌یابی مستقیم و شناسایی ژن‌ها و ژنوم‌های موجود در زیست‌بوم‌های پیچیده میکروبی (متاژنوم) است که با گذار از جداسازی کشت خالص در روش‌های میکروبیولوژی زیست‌محیطی، انقلابی برپا کرده است. متاژنومیکس نوید پیشرفت درک ما از گوناگونی، کارکرد و تکامل ارگانیسم‌های کشت‌ناپذیر را که نسبت به کشت‌پذیرها بسیار بیشترند، می‌دهد و رابطه تنگاتنگی با دردسترس قرارگرفتن توالی‌یابی نسل جدید (NGS)² در تمامی مراحل اجرایی آن دارد [6]. کاربرد روش‌های متاژنومیک دست‌یابی به

2 Next generation sequencing

1 Meta

اطلاعات ژنتیکی یا ژنومیک از ویروس‌ها، باکتری‌ها، آرکی‌ها، قارچ‌ها و آغازیانی را میسر می‌سازد که اجتماعات پیچیده‌ای را می‌سازند. متاژنومیکس پرسش‌های بنیادینی را همچون اینکه در یک زیستگاه کدامیک از میکروب‌ها وجود داشته و ژن‌های آنها توان انجام چه کارهایی را دارند، به بحث می‌گذارد و به این ترتیب درک ما از محیط‌زیست و زیست‌شناسی دگرگون می‌کند [2، 7]. برای نمونه متاژنومیکس نشان داده است که آرکی‌ها در اکسیداسیون آمونیاک در بسیاری از زیست‌بوم‌ها به‌ویژه زیست‌بوم‌هایی که نیتروژن کمی دارند، مانند زیست‌بوم‌های طبیعی، چراگاه‌ها و همچنین خاک‌های زیرین زمین‌های کشاورزی، همکاری به‌سزایی دارند [1]. امروزه در بیشتر پژوهش‌های متاژنومیکس به‌طور مستقیم DNA را به‌منظور توالی‌یابی استخراج می‌کنند. اگر ژن‌های ویژه‌ای هدف ما باشند (همانند 16S rRNA) می‌توان آنها را پیش از توالی‌یابی تکثیر کرد. همچنین DNA استخراج‌شده را می‌توان بدون تکثیر کردن هر ژن ویژه‌ای توالی‌یابی کرد. این روش به‌عنوان توالی‌یابی شات‌گان شناخته می‌شود. در این فرایند، DNA جامعه استخراج‌شده و به بخش‌های کوچکی شکسته می‌شود و به‌طور مستقیم به‌وسیله پلات‌فرم‌های توالی‌یابی توانمند مانند ایلومینا¹ تعیین توالی می‌شود. پس از توالی‌یابی و کنترل کیفیت آن، می‌توان خوانش‌ها را:

- به‌طور مستقیم با پایگاه‌های داده‌ای تاکسونومیک مقایسه و یا از نظر کارکردی تفسیر کرد، یا
- با هم جمع کرد تا DNA بلندتری که بتواند اطلاعات بهتری را از بخش‌های بیشتری از ژنوم (ها) نشان دهد، فراهم شود [2]. البته مونتاژ طول‌های به‌نسبت کوتاه (کمتر از 500 جفت باز) در پلات‌فرم‌های توالی‌یابی نسل دوم (همانند ایلومینا) نه‌تنها چالش‌برانگیز است بلکه تفسیر مستقیم از خوانش‌ها را دشوارتر می‌کند. پیشرفت

فناوری‌های نوین تر توالی‌یابی (نسل سوم) به‌وسیله شرکت سیفیک بیوساینس² که نوید فراهم کردن خواندن‌های بیشتر را داده‌اند (بیش از 3000 جفت باز)، به‌طوری که کل ژن‌ها و احتمالاً اوپرون‌ها را شامل می‌شود. بنابراین اجازه طبقه‌بندی‌های بهتر تاکسونومیک و یا پیش‌بینی‌های کارکردی بهتر را می‌دهند مانند توالی‌یابی SMRT که این چالش را برطرف می‌کند [6]. سپس از پایگاه‌های داده‌ای که در NCBI و MG-RAST هستند، استفاده می‌شود. اگر شناسایی کارکردی بالاتری مدنظر است، می‌توان ژن‌ها را به‌وسیله پایگاه‌های داده‌ای همانند KEGG دسته‌بندی کرد. این پایگاه‌های داده، شناسایی مسیرهای آنزیمی و کارکردی را آسان می‌کند [2].

یکی از مهم‌ترین مواردی که باید در متاژنومیکس مدنظر قرار داد، مقدار داده‌های تولیدشده به‌وسیله نسل جدید توالی‌یاب‌هاست که باید مدیریت شوند. پلات‌فرم راش 454 ظرفیت بیش از یک میلیون باز خوانده‌شده را دارد که 85 درصد از خوانش‌هایش بلندتر از 500 جفت باز در هر اجرا هستند. رقابت میان پلات‌فرم‌ها شامل سیستم‌های آنالیزگر ژنومی ایلومینا، سولکسا³ و سولید⁴ می‌شود که خروجی‌های بزرگی دارند، اما طول‌های خوانده‌شده کوتاه‌تری دارند که مونتاژ ژنوم را سخت‌تر می‌کند و پتانسیل بیشتری برای پاسخ‌های جانبدارانه یا اریب⁵ دارند [4]. لو و همکاران [4] توانایی پلات‌فرم‌های راش 454 و ایلومینا را برای شناسایی نمونه DNA استخراج‌شده از یک جامعه پلانکتونی آب شیرین با یکدیگر مقایسه کردند. آنها دریافتند که مونتاژهای ژنی به‌دست‌آمده وقتی که در 90 درصد کل توالی‌هایشان همپوشانی دارند، قادر به برآورد تشابهات فراوانی‌های ژنی و ژنومی خواهند بود. راش 454 یک توالی با طول

2Pacific Biosciences (PacBio)

3 Solexa

4 SOLiD

5 bias

1Illumina

پارامترهای گوناگونی (گوناگونی آلفا، بتا و فیلوژنتیکی) دنبال می‌شود.

باید در نظر داشت که کارکردهای ژن‌ها وقتی از داده‌های متاژنومیکس به دست می‌آیند باید به لحاظ بیوانفورماتیکی تفسیر شوند تا اینکه به شیوه آماری بررسی شوند. این ویژگی تفسیر داده‌های متاژنومیکس به وسیله بیوانفورماتیک، منجر به پیش‌بینی توانایی‌های ژنومی جامعه میکروبی به‌عنوان پلات فرمی برای به‌دست‌آوردن درک بهتری از ساختار، گوناگونی، پتانسیل ژنتیکی و کارکرد جامعه میکروبی می‌شود که برای دانستن جزئیات بیشتر به همراه بهره‌گیری از سایر ارزیابی‌های آمیکسی همانند ترانسکریپتومیکس و پروتئومیکس، خواهند توانست ژن‌های مفروض بیان‌شده و پروتئین‌های پیش‌بینی‌شده را مشخص کنند [1].

ویرومیکس

ژنوم‌های ویروسی موجود در هر نمونه را متاژنوم ویروسی یا ویروم می‌نامند. برآورد می‌شود که ویروم سیاره ما از 10^{31} ذره ویروسی همراه با میزبان (که شامل میزبان انسانی هم می‌شود) یا بدون میزبان تشکیل شده باشد. ویروم دارای فراوان‌ترین و سریع‌ترین عناصر ژنتیکی جهشی را در زمین دارد [2]. ویروس‌ها با میزبان انسانی از راه جوامع باکتریایی آن (ویروس‌های پروکاریوتی) و با سکونت‌داشتن در سلول‌های انسانی یا برهم‌کنش با سلول‌های انسانی در عفونت‌های حساس، پایدار یا پنهان دارای برهم‌کنش هستند. شناخت ژنتیکی ویروم به‌عنوان بخش اصلی میکروبیوم در پشت سر آنالیزهای میکروبیوم باکتریایی حرکت می‌کند که بخشی از آن به دلیل کمبود یک توالی حفاظت شده است که بتوان از آن برای تعیین تاکسونومی استفاده کرد و بخشی از آن هم به دلیل بسیاری از ژن‌های رمزکننده در ویروم است که پیش از این تفسیر نشده‌اند [12]. پژوهش‌های پیشگام

502 مگاجفت باز (با طول خوانده شده 450 جفت باز) تولید کرد درحالی‌که ایلومینا یک توالی با طول 2460 مگاجفت باز (با طول 100 جفت باز یا bp) را تولید کرد. ولی زمانی‌که مونتاژ انجام شد، این اعداد به‌ترتیب 46 Mbp و 57 Mbp برای کل توالی‌های یگانه به دست می‌آیند. با وجود طول‌های خوانده‌شده به‌مراتب کوتاه‌تر در پلات فرم ایلومینا، این پلات فرم کانتینگ‌های بلندتر و دقیق‌تری تولید می‌کند و هزینه آن تنها 25 درصد پلات فرم راش 454 است.

کیفیت داده‌ها در همه آنالیزهای مربوط به آمیکس، به‌ویژه در بررسی‌های جامعه میکروبی خاک، مسئله اصلی است [9]. پتانسیل وجود پاسخ‌های جانبدارانه در تمام مراحل پژوهش‌های متاژنومیکس خاک کم نیستند. این پاسخ‌های جانبدارانه شامل جانبداری در نمونه به‌دلیل نمونه‌گیری و پردازش آن (که به‌ویژه در مورد نمونه‌برداری از خاک و استخراج اسیدهای نوکلئیک رخ می‌دهد)، جانبداری در PCR، جانبداری در توالی‌یابی، جانبداری در مونتاژ ژنوم و جانبداری در ساخت کتابخانه می‌باشند [10، 2]. ویتروفسکی و بالدین [11] از مجموعه نرم‌افزارهای SEED برای کمک به کنترل‌کردن خطاهای ویژه روش در آنالیز داده‌های پیروسیکوئنسینگ (پلات فرم راش 454) استفاده کردند و برای انجام آنالیز داده‌ها کارهای زیر را به‌ترتیب انجام دادند: حذف توالی‌های با کیفیت پایین، حذف اخلاص¹ از توالی‌ها، حذف توالی‌های کایمری (توالی‌های دورگه‌ای هستند که از دو الگوی DNA طی PCR ممکن است، پدید آیند)، پیرایش توالی‌هایی که دارای DNA الگو هستند و خوشه‌بندی توالی‌ها در گروه‌هایی بر مبنای تشابه توالی‌ها. این فرایند با هم‌ردیف‌سازی توالی‌ها، شناسایی توالی‌ها، آنالیز ترکیب جامعه و نمونه‌گیری دوباره تصادفی پیش از محاسبه

نشان داده‌اند که گوناگونی ویروسی بسیار کم برآورد شده است، زیرا تنها در 200 لیتر آب دریا بیش از 7000 ژنوم ویروسی مختلف یافت شده است و درصد بالایی (66 درصد) از توالی‌های ویروسی ناشناخته هستند [12]. این درجه بالای گوناگونی ژنتیکی ویروسی به‌وسیله پژوهش‌های متاژنومیک روی نمونه‌های دریایی که پس از آن انجام گرفتند، نیز اثبات شده است. نخستین پژوهش‌ها پس از کشف ویروس‌های با ژنوم DNA در یک نمونه مدفوع از یک فرد سالم به دست آمدند و نتایج نشان داد که توالی‌های تولیدشده، ناشناخته‌اند. برآورد شده است که در میان توالی‌های ویروسی قابل شناسایی، بیشترین تطابق با باکتریوفاژها و جامعه آنها که دارای غنا (در حدود 1200 ژنوتیپ) و گوناگونی بالایی هستند، وجود دارد. یافته‌های قابل مقایسه‌ای نیز در دو پژوهش بر DNA ویروم روده انسان به دست آمد که گزارش کردند درصدهای توالی‌های ناشناخته به ترتیب 81 و 98 درصد بوده و فاژها در جامعه ویروسی غالب بوده‌اند [2].

آنالیز متاژنومیکس ویروسی نشان داد که بیشتر از 60 درصد توالی‌ها در پژوهش‌های روی ویروس‌ها منحصر به فرد هستند. بنابراین بسیاری از گونه‌های ویروسی همچنان ناشناخته هستند. ظرفیت حاضر شناخت اعضای ویروم به‌طور انحصاری تقریباً به شباهت توالی‌های اسیدهای نوکلئیکی و پروتئینی قابل شناسایی در اطلاعات انتشار یافته برمی‌گردد. یک مانع اصلی در تعریف یک ویروس، استفاده از هم‌ردیف‌های توالی اسید نوکلئیک داده شده است که تغییرپذیری بالایی از ارگانیسم‌ها در آن سطح دارد. پس، بسیار محتمل است که ویروم دربرگیرنده ویروس‌های جدیدی باشد که هنوز شناسایی نشده‌اند [12].

متاترانسکریپتومیکس

هرچند، روش‌های جدید ژنومیکس همانند متاژنومیکس داده‌های بسیار زیادی را تولید می‌کنند و گرچه این داده‌ها

پتانسیل DNA یک سیستم زیستی را نشان می‌دهند ولی به‌طور لزوم فنوتیپ بیان‌شده به‌وسیله آن را نشان نمی‌دهند. این داده‌ها برای بازکردن قفل بخش ژنومیکس باید به RNA یا پروتئین بیان شوند که به آنها به‌ترتیب ترانسکریپتومیکس یا متاترانسکریپتومیکس و پروتئومیکس یا متاپروتئومیکس می‌گویند. بنابراین، RNA به‌ویژه mRNA که رونوشت DNA است، هدف آنالیزهای ترانسکریپتومیکس می‌باشد. بسیاری از آنالیزهای متاترانسکریپتومیکس در طبیعت جنبه اکتشافی دارند. از طرف دیگر، برخی پژوهش‌های ترانسکریپتومیکس بر بیان ژن‌های هدف متمرکز شده‌اند و افزون بر آن برای کامل کردن این تصویر بر سایر آمیکس‌ها تکیه دارند. بیشتر پژوهش‌های اولیه ترانسکریپتومیکس روی نمونه‌های کلینیکی مدفوع انجام شده‌اند [13] که از آنها می‌توان به‌عنوان پیش‌بینی‌کننده‌ای برای نمونه‌های زیست‌محیطی استفاده کرد و قالبی برای آنالیز خاک و آب و ریزوسفر فراهم می‌کند. برخلاف DNA، mRNA در نمونه‌های زیست‌محیطی بیشتر از چند دقیقه دوام نمی‌آورد. افزون بر آن، نیمه عمر mRNA در محیط‌ها و میکروارگانیسم‌های گوناگون و به‌دلیل کارکرد ژنی که از آن رونویسی می‌شود، متفاوت است (ژن‌های خانه‌دار mRNA پایداری تولید می‌کنند). به همین دلیل، نمونه‌ها باید ظرف چند دقیقه (بهتر است ظرف چند ثانیه) جمع‌آوری شوند. دستورکارهای آزمایشگاهی¹ زیادی شامل استفاده از کیت‌های تجاری با استانداردهای ساده و روش‌های دستی متداول که مایه کمیت و کیفیت بالای RNA گردآوری شده می‌شوند، وجود دارند، هرچند که مایه سخت‌ترشدن پژوهش‌های گسترده جغرافیایی می‌شوند [2].

پژوهش انجام‌شده توسط یوریچ و همکاران [14] به‌عنوان نخستین پژوهش انجام‌شده متاترانسکریپتومیکس

شات گان بر خاک بسیار برجسته و مهم است. این پژوهشگران کل خزانه RNA را استخراج کرده و سپس rRNA و mRNA بازیابی شده را بدون تکتیر کردن این مولکول‌ها آنالیز کرده و در پایان ساختار و همچنین کارکرد جامعه خاک را بررسی کردند. در مطالعه آنها یوکاریوت‌ها 10/3 درصد زیر واحد کوچک RNA ریبوزومی (SSU) و 13/3 درصد زیر واحد بزرگ RNA ریبوزومی (LSU) را از RNA بازیابی شده با استفاده از برچسب ریبوزومی نشان دادند، در حالی که 87/2 درصد SSU و 83/8 درصد LSU از باکتری‌ها و 1/5 درصد SSU و 1/4 درصد LSU از آرکی‌ها بود. این نسبت‌ها با نتایج گزارش شده توسط ترنر و همکاران [15] که پس از آن انجام گرفت، همخوانی دارند. این پژوهش‌ها نشان دادند که کرن/آرکیوتا¹ از آرکی‌ها تقریباً 0/01 درصد علامت‌های توالی rRNA را تشکیل می‌دهند. این نتایج با یافته‌های تروش و همکاران [16] و نیکول و اسکلیپر [17] که با استفاده از روش‌های اولیه همسانه‌سازی و توالی‌یابی نقش معنادار کرناکیوتا را در اکسیداسیون آمونیاک در بسیاری از خاک‌ها شناسایی کردند، اثبات می‌شود. در میان باکتری‌ها، پروتئوباکتريا و اکتینوباکتريا فراوان‌ترین بوده و پس از آنها به ترتیب فیرمیکات‌ها، اسیدوباکتريا و پلانکتومایست‌ها فراوان می‌باشند که این یافته نیز با نتایج ترنر و همکاران [15] هماهنگ است. در میان یوکاریوت‌ها با استفاده از شناسایی برچسب‌های ریبوزومی به ترتیب سلسله قارچ‌ها (به‌طور عمده آسکومایکوتا و گلوومرومایکوتا) با 50 درصد فراوانی، سلسله گیاهان (گیاهان آونددار) با 20 درصد فراوانی و سلسله آغازیان (متازوا یا پروتیست‌ها) با فراوانی 10 درصد شناسایی شده‌اند. 1370 عدد برچسب ریبوزومی به‌عنوان آغازیان بسیار جالب و مهمی شناسایی شدند که بزرگ‌ترین دیتاست مولکولی یک جامعه از آغازیانی را که

تا آن زمان تولید شده بود، نمایندگی می‌کردند و نخستین پنجره را به سوی ساختار جامعه خود در خاک گشودند. کپک‌های لزج (مایستوزوا، 25 درصد) بیشترین فراوانی را داشته و پس از آنها به ترتیب سرکوزوا (آمیب‌ها و تاژکداران، 17 درصد)، پلاسمودیوفورا (بیشتر انگل‌های گیاهی، 16 درصد) و آلویولوتا (همانند مژک‌داران و دینوفلاژلاتا، 15 درصد) فراوان‌ترین هستند. تمامی این آغازیان میل سیری‌ناپذیری در شکار و تغذیه از باکتری‌های خاک داشته و بنابراین در برگرداندن مواد آلی به خاک و کانی‌سازی نیتروژن، فسفر و گوگرد در خاک نقش زیادی دارند.

متاپروتئومیکس

هرچند، روش‌های مبتنی بر DNA و RNA اطلاعات کافی از نقش‌هایی که میکروارگانیسم‌ها در محیط‌زیست بازی می‌کنند، فراهم می‌کنند، اما این پروتئین‌ها و نه ژن‌ها هستند که به‌طور مستقیم مسئول بیشتر فرایندهای میکروبی محسوب می‌شوند. بنابراین، اندازه‌گیری پروتئین‌ها (مانند آنزیم‌ها) می‌تواند به‌طور مستقیم فعالیت میکروبی را اندازه‌گیری کند. به پروتئین‌های تولیدشده به‌وسیله گروهی از میکروارگانیسم‌ها در شرایط ویژه پروتئوم گویند. پروتئوم در مقایسه با ژنوم، متغیرتر است (همانند ترانسکریپتوم) که تفاوت در پروتئین‌های تولیدشده بستگی به مرحله متابولسیم سلولی و محرک‌های زیست‌محیطی دارد. اصطلاح متاپروتئومیکس یا پروتئومیکس جامعه برای جداسازی بسیاری از پروتئین‌های جامعه میکروبی به‌منظور درک بهتر نقش میکروارگانیسم‌ها در جامعه طراحی شده است. پژوهش‌های متاپروتئومیکس به این ترتیب انجام می‌شوند: قراردادن میکروارگانیسم‌ها در معرض شرایط ویژه، جداسازی پروتئین‌ها از جمعیت‌های میکروارگانیسمی، جداسازی هر پروتئین، شناسایی پروتئین‌ها [1].

متابولومیکس زیست‌محیطی

متابولومیکس مطالعه متابولیت‌های مولکولی با وزن کم است. متابولومیکس زیست‌محیطی شامل متابولیت‌های تولیدشده ناشی از برهم‌کنش‌های میان میکروارگانیسم‌ها همانند یوکاریوت‌های کوچک، گیاهان، جانوران، شکارچیان و تنش‌های غیرزنده و محرک‌ها می‌شود. متابولیت‌های متداول (کمتر از 1500 دالتون) شامل اسیدهای آلی، اسیدهای آمینه و ساکاریدهای گوناگون می‌شوند. همانند ژنومیکس، ترانسکریپتومیکس و پروتئومیکس در بیشتر موارد هدف متابولومیکس روشن‌ساختن کارکرد یک میکروارگانیسم یا جامعه میکروبی است. البته آشکار است که پروتئومیکس و متابولومیکس اطلاعات مربوط به تولید نهایی ژنوم را آشکار می‌کنند، به‌گونه‌ای مشابه، متابولومیکس برهم‌کنش‌های میان بخش‌های سازنده میکروبی و محیط‌زیست آنها یا میان ارگانیسم‌های میکروبی و ارگانیسم‌های پیشرفته‌تر نظیر گیاهان و جانوران را تعیین می‌کند. متابولومیکس به‌عنوان یک ابزار اکتشافی برای کشف وضعیت کارکردی جمعیت‌های میکروبی و هریک از سلول‌ها در محیط‌زیستشان استفاده می‌شود که ساختار جامعه و بوم‌شناسی آنها را آشکار می‌کند. این امر به متابولومیکس توانایی آن را می‌دهد که بر یک متابولیت ویژه تمرکز کند. برای نمونه، زمانی که یک تیمار بالا و پایین‌رفتن میزان یک متابولیت را دیکته می‌کند. این درحالی است که متابولومیکس جهانی کلیت سیستم زیستی و همه متابولیت‌های آن را می‌بیند [1].

متابولیت‌ها به دو گروه تقسیم می‌شوند: متابولوم‌های درونی و متابولوم‌های بیرونی که به‌ترتیب شامل متابولیت‌های درون‌سلولی و برون‌سلولی هستند. همانند ترانسکریپتومیکس، مطالعه متابولیت‌های درون‌سلولی دشوارتر است، زیرا این مولکول‌ها در یک جریان زودگذر یا ناپایدار (یعنی با عمر کم) قرار دارند. پیچیدگی

دو مرحله نخست در مطالعات پروتئومیکس به‌نسبت ساده هستند. روش‌های کارآمد بسیاری برای جداسازی و خالص‌سازی پروتئین‌های ناهمگن و مخلوط پیچیده آنها که به‌وسیله میکروارگانیسم‌ها ساخته می‌شوند، وجود دارد. اما جداسازی هریک از پروتئین‌ها از این آمیخته پیچیده پروتئینی چالش‌برانگیزترین گام در پروتئومیکس است. از دو راه برای جداسازی پروتئین‌ها بیشتر از دیگر روش‌ها استفاده می‌شود که یکی الکتروفورز دوبعدی بر ژل پلی‌اکریل آمید (2D-PAGE)¹ و دیگری کروماتوگرافی مایع-اسپکترومتری جرمی (LC-MS)² است. در 2D-PAGE پروتئین‌ها نخست بر مبنای نقاط ایزوالکتریک خود جدا می‌شوند که به معنای pH است که در آن، پروتئین بدون بار است. بعد دوم 2D-PAGE پروتئین‌ها را بر پایه جرمشان با استفاده از ژل پلی‌اکریل آمید جدا می‌کند. در نتیجه این کار لکه‌های پروتئینی بسیاری به دست می‌آیند که هرکدام نشانگر یک پروتئین هستند و سپس توسط روش اسپکترومتری جرمی شناسایی می‌گردند [2]. پژوهش‌ها نشان داده‌اند که توان عملیاتی بالا و جامع‌بودن روش‌های پروتئومیکس همچنین می‌توانند مایه روشن‌شدن مسیرهای تجزیه زیستی شوند. کیم و همکاران [18] مسیرهای تجزیه زیستی ترکیب‌های آلی آروماتیک به‌وسیله سویه *Pseudomonas* sp. K82 را با استفاده از روش 2D-PAGE و سپس اسپکترومتری جرمی برای شناسایی پروتئین‌های درگیر در آن آزمایش کردند. نتایج آنها نشان داد که سه مسیر متابولیکی که هرکدام شامل یک ترکیب متفاوت آروماتیکی می‌باشد، وجود دارد. کاربرد پروتئومیکس در زیست‌بوم‌های میکروبی همانند متاژنومیکس نوپدید است و رابطه میان‌گونه‌های میکروبی و جوامع پیچیده آنها به‌منظور درک کارکرد آنها را نشان می‌دهد [19].

1two-dimensional polyacrylamide gel electrophoresis (2D-PAGE)

2 liquid chromatography-mass spectrometry (LC-MS)

متابولوم‌ها و اهداف پژوهش همانند متابولیت‌های برون‌سلولی و درون‌سلولی، نوع استخراج و فرایند پردازش آن را تعیین می‌کنند. زمانی که متابولیت‌ها استخراج شدند، شناسایی آنها به وسیله کروماتوگرافی گازی و مایع - اسپکترومتری جرمی، اسپکتروسکوپی رامان و رزونانس مغناطیسی هسته (NMR) انجام می‌شود. در بسیاری از موارد بسته به پیچیدگی سیستم زیستی، ترکیبی از دو یا بیشتر از این روش‌ها انجام می‌گیرد [2].

مقادیر زیادی از داده‌های متابولیکی بدون در نظر گرفتن پلات فرم به کار گرفته شده، تولید شده‌اند. متابولیت‌ها در بسیاری از نمونه‌های بررسی شده در مقیاس جهانی ناشناخته‌اند، از این رو به پایگاه‌های داده‌ای جستجوگر به منظور تفهیم، استنباط، کارکرد و هدف متابولیت نیاز است. اغلب متابولیت‌ها به عنوان فراورده‌ها یا حدواسط‌های جمعیت‌های زیست محیطی در شرایط تنش سلامتی کلی یک سیستم زیستی شناخته می‌شوند. به دلیل به نسبت نوین بودن متابولومیکس به ویژه در علوم زیست محیطی، پایگاه‌های داده‌ای بسیار کمی برای شناسایی متابولیت‌های زیست محیطی وجود دارند. پایگاه‌های داده‌ای معمول متابولومیکس شامل پایگاه داده‌ای متابولوم انسانی¹ و دانشنامه ژن‌ها و ژنومیکس کیوتو (KEGG)² هستند [1].

کاربردهای گوناگون متا‌میکس با تمرکز بر متاژنومیکس

برای داشتن درک درستی از میکروبیوم موجود در کل سیاره زمین و برهم‌کنش آن با محیط زیست، به پژوهش‌های بزرگ مقیاس در زیست‌شناسی ارگانیسم‌ها و مدل‌های بوم‌شناختی که روابط فیلوژنتیکی و کارکردی میان ارگانیسم‌ها را با یکدیگر ترکیب کنند، نیاز داریم.

برخی از این پژوهش‌های بزرگ عبارتند از پژوهش MetaHIT، پژوهش روده آمریکایی، پژوهش غذای انسانی، پژوهش میکروبیوم خانگی، پژوهش سرشماری جهانی میکروب‌های دریایی، پژوهش اقیانوس‌های تارا³، پژوهش میکروبیوم انسانی، پژوهش میکروبیوم سیاره زمین و پژوهش مرکوری⁴ برای آنالیز میکروبیوم ایستگاه فضایی. در ادامه به برخی از دستاوردهای برخی از این پژوهش‌ها پرداخته می‌شود.

رم و همکاران [20] از متا‌پروتئومیکس برای بررسی و شناسایی جامعه بیوفیلم یک معدن اسیدی استفاده کردند. در بیشتر روش‌های پروتئومیکس به کاررفته در این پژوهش، داده‌های توالی ژنومیکس کار پروتئومیکس را آسان می‌کنند. این پژوهشگران به صورت اختصاصی پایگاه داده‌ای از 12148 توالی پروتئینی پیش‌بینی شده را از جامعه بیوفیلم مشابهی ساختند که پیش‌تر با استفاده از متاژنومیکس شناسایی شده بود [21]. با استفاده از این پایگاه داده و روش کروماتوگرافی مایع - اسپکترومتری جرمی برای شناسایی پروتئین‌ها، 2033 پروتئین شناسایی شدند که بیشتر آنها به وسیله گونه‌های جنس *Leptospirillum* تولید شده بودند و در سازگاری با شرایط اسیدی شدید (pH=0/8) و سرشار از فلزات سنگین نقش داشتند. کارکرد بسیاری از این پروتئین‌هایی که بسیار هم فراوان بودند، تعیین نشد. یکی از آنها که به وسیله روش‌های متاژنومیکس به دست آمده بود، به عنوان پروتئینی که احتمالاً در اکسیداسیون آهن نقش دارد، شناسایی شده و مشخص شد سیتوکروم جدیدی است که در اکسیداسیون آهن و تشکیل اسید زهکش معدن نقش دارد. پژوهش بعدی نشان داد که پروتئوم زمان گسترش بیوفیلم تغییر می‌کند [22]. برای نمونه، باکتری غالب، یعنی لپتوسپیریلیوم گروه II آنزیم بیشتری برای متابولیسم

ترکیب‌های 1 و 2 کربنی، سنتز پروتئین طی توسعه اولیه بیوفیلیم و پروتئین‌های مرتبط با تنش و مرتبط با اکسیداسیون آهن تولید می‌کند که احتمالاً تشکیل‌دهنده اسید زهکش معدن هستند و دلیل آن گسترش بیوفیلیم، کمتر شدن و محدودتر شدن سرچشمه‌های غذایی آنهاست. پژوهش در میکروبیوم انسان آشکار کرده است که حتی افراد سالم هم به‌طور چشمگیری در میکروب‌هایی که در زیستگاه‌های بدنی خود مانند روده، پوست و واژن دارند، متفاوتند. بیشتر این گوناگونی بدون توضیح باقی مانده است، هرچند می‌دانیم که رژیم غذایی، محیط، ژنتیک میزبان و نخستین قرارگیری در معرض میکروب دلایلی بر وجود این گوناگونی هستند. از این‌رو، برای شناخت بوم‌شناسی جوامع میکروبی همراه با انسان، پژوهش میکروبیوم انسانی (HMP)¹ با بررسی 300 مرد و زن بزرگسال سالم بزرگ‌ترین آنالیزها را سازماندهی کرده و انجام داده است که تاکنون در مجموعه زیستگاه‌های بدنی بسیار متمایز و به لحاظ کلینیکی متضاد انجام گرفته‌اند. دستاوردهای پیچیده آن به روشن شدن سازوکار و درمان بیماری‌های بسیار پیچیده انسانی مانند چاقی مفرط و دیابت نوع دو، کمک شایانی می‌کند [23].

ترنر و همکاران [15] از متاترانسکریپتومیکس مقایسه‌ای یا تطبیقی برای آزمایش جوامع خاک ریزوسفری و ناریزوسفری گندم، نخودفرنگی، یولاف و یک یولاف جهش‌یافته در ژن *sad1* (که توانایی تولید ماده پادقارچی اونا سین² را نداشت) استفاده کردند. آنها دریافتند که تفاوت‌های ژرف (در سطح قلمرو تاکسونومیک) در جوامعی که در ریزوسفر گونه‌های مختلف گیاهی در همان خاک توسعه پیدا کرده‌اند، اثباتی بر درست بودن تمامی کارهای پیشین است که نشان داده‌اند گونه‌های گیاهی محرک اصلی در شکل‌دهی به

جوامع میکروبی خاک هستند. بیشترین بخش توالی‌هایی که آنها استخراج کردند، از باکتری‌ها بود که از 91 درصد برای خاک ناریزوسفری تا 73/7 درصد کل توالی‌های ریزوسفر نخود فرنگی را در بر می‌گیرد. پروتئوباکتیریا، اکتینوباکتیریا، فیرمیکات‌ها، اسیدوباکتیریا، پلانکتومایست‌ها و باکتروایدت‌ها به بهترین وجه ممکن در توالی‌های بازبایی شده نمایش داده شدند. جوامع باکتریایی در ریزوسفر گیاهان از جوامع باکتریایی موجود در خاک ناریزوسفری متفاوت بوده است. همچنین میان گونه‌های گیاهی هم در سطح جنس تفاوت وجود داشت؛ هرچند که در سطح شاخه، جوامع پروکاریوتی ریزوسفر گندم از جوامع پروکاریوتی خاک ناریزوسفری تفاوتی نداشتند. آرکی‌ها تقریباً در همه نمونه‌ها 5 درصد توالی‌های بازبایی شده را تشکیل می‌دادند، درحالی‌که یوکاریوت‌ها 2/8 درصد جامعه ناریزوسفری خاک و 3/3 درصد جامعه ریزوسفر گندم را تشکیل می‌دادند، ولی در ریزوسفر نخودفرنگی 20/7 درصد و در ریزوسفر یولاف 16/6 درصد جوامع را تشکیل می‌دادند. ریزوسفر نخودفرنگی به‌شدت به‌وسیله قارچ‌ها اشغال شده بود. نماتدها و پروتوزوئترهای باکتری‌خوار در هم‌ریزوسفرهای بررسی شده فراوانی بالایی داشتند.

نخستین یافته‌ها درباره میکروبیوم روده که از پژوهش MetaHIT با بهره‌گیری از روش‌های متاژنومیکس به دست آمدند، نشان دادند که تعداد ژن‌های میکروبی بیش از 100 برابر تعداد ژن‌های انسانی است و این پژوهش تا بیش از 3 میلیون ژن باکتریایی را تنها در روده انسان پیش‌بینی نمود [24].

کشاورزی و میکروبیولوژی غذایی می‌توانند از پیشرفت‌های متاژنومیکس با بهبود در ایمنی و امنیت غذایی، بهبود در تشخیص تهدیدهای تولید و تدارک غذا و افزایش بهره‌وری از جانوران و گیاهان بومی سود ببرند. استفاده از روش‌های متاژنومیکس در صنایع غذایی برای

درک بهتر و جوامع میکروبی همراه با غذا می‌تواند منجر به بهبود در بهره‌وری، کیفیت و امنیت غذایی شود. روش‌های متاژنومیکس امکان مطالعه‌های جامع‌تری را از تأثیر پروبیوتیک فراهم می‌سازند. همراهی متاژنومیکس با سایر رویکردهای آمیکس توانایی شناسایی متابولیت‌ها و مسیرهای متابولیکی را دارد، از جمله سازوکارهای بازخورد میزبان- میکروبیوتا که به‌وسیله پروبیوتیک موجب تنظیم سلامتی می‌شود. پوشش عمیق و بالای NGS، متاژنومیکس را به ابزاری توانمند در تشخیص و نظارت بر پاتوژن‌های غذایی، تشخیص طغیان‌ها و مسیرهای انتقال بیماری‌های ناشی از غذا، آزمایش غذاها و محیط‌های مرتبط با غذا و شناسایی میکروبیوتایی که امکان دارد در برابر بیماری‌های ناشی از غذا کار محافظتی را انجام دهند، تبدیل کرده است [6].

روش‌های متاژنومیکس امکان آنالیز مقایسه عمیق مکان‌های چندگانه در یک فرد و در کل جمعیت‌ها را فراهم می‌کنند همانند پژوهش‌های جغرافیای زیستی در دریای سارگاسو که این جغرافیای زیستی با حضور شبکه‌ای از روابط و ساختارهای گوناگون همراه می‌شود. میکروبیولوژی سنتی نشان داده است که اینگونه روابط می‌توانند منجر به برهم‌کنش‌های فیزیکی مستقیم به همراه توالی تشکیل بیوفیلم شده که سرانجام منجر به یک ایترکتوم¹ می‌شوند. این شبکه در زمان آنالیز داده‌های توالی‌یابی نسل جدید به شبکه‌هایی که به‌طور هم‌زمان با هم رخ می‌دهند، گسترش پیدا می‌کند که به‌طور معمول ولی نه همیشه، فیلوتیپ‌هایی دارند؛ یعنی با همدیگر در یک مکان حضور دارند. هرچند ما از درک تمامی این روابط دور هستیم، یک وابستگی متابولیکی به دلیل وجود شبکه تجزیه‌ای آبشارمانند از مواد مغذی که هم میکروبیوم و هم میزبان را تحت تأثیر قرار می‌دهند، وجود دارد. به نظر می‌رسد که انتقال افقی ژن در یک محیط‌زیست

کوچک، یک انتخاب رقابتی برای تکمیل پائل قابلیت‌های کارکردی است که به‌وسیله آنالیز ژنوم‌های در دسترس نشان داده می‌شود. در مورد باکتری اختصاصی روده انسانی، یعنی *Bacteroides plebius* تبادل ژنتیکی با یک باکتری دریایی رخ می‌دهد. این انتقال افقی ژن، گوارش جلبک دریایی را در برخی از ژاپنی‌هایی که حامل *B. plebius* هستند، آسان می‌کند. در جوامع چندگونه‌ای، انتخاب دیگر استفاده از رابطه همزیستی هم برای افزایش جذب مواد مغذی و هم محافظت از خود در برابر میزبان است [24].

میرته و همکاران [25] با به‌کارگیری روش‌های آمیکس به بررسی وجود آنزیم‌های جدید درگیر در سازگاری مولکولی و پروتیین‌های درگیر در شرایط محیط‌های دشوار پرداختند. آنها با بررسی کارهای دیگر پژوهشگران گزارش کردند که سلولازها و استرازهای گوناگونی که کاربردهای بسیاری در بیوتکنولوژی دارند، در زیست‌گاه‌های با درجه حرارت بالا و پایین، بسیار شور و اسیدی وجود دارند که با شناسایی آنها و انتقال ژن‌های آنها به میزبان‌های مزوفیل می‌توان از آنها بهره گرفت.

پانلی و همکاران [26] گزارش کردند که مدیریت کشاورزی و فصل سال (بهار و پاییز) می‌تواند به‌طور عمیق بر جوامع قارچی موجود در خاک تأثیرگذار باشد. سیستم کشاورزی مبتنی بر به‌کارگیری مواد آلی و بدون شخم حتی در نخستین سال اجرای آن جوامع قارچی متمایزی را در خاک پرورش داد. *Ascomycota* (کلاس *Sordariomycetes*)، *Basidiomycota* (کلاس *Agaricomycetes*)، *Zygomycota* (رده *Mortierellales*)، *Chytridiomycota*، *Glomeromycota* و *Rozellomycota* به‌ترتیب فراوانی، تشکیل‌دهنده جامعه قارچی این خاک هستند. چنین اطلاعاتی می‌تواند در ارزیابی مدیریت موفق خاک سودمند باشد و با کاهش

تاکسای بیماریزا و افزایش جمعیت تاکسای مفید به تولیدات کشاورزی و کاهش آلودگی زیست‌محیطی (برای مثال با مصرف کمتر آفت‌کش‌ها) کمک شایانی کند. استفاده از روش‌های متاژنومیکس و دیگر آمیکس‌ها فراوانی و گوناگونی بسیار بالای انواع ژن‌های پایداری به آنتی‌بیوتیک‌ها (ARGs) را در محیط‌زیست‌های متفاوتی نشان داده است که خطر بالقوه‌ای محسوب می‌شوند. یانگ و همکاران [27] با بررسی ریزگردهای معلق در هوا که از دامداری‌ها و تصفیه‌خانه‌های فاضلاب برخاسته بودند، دریافتند که ARGs در این ریزگردها فراوانی و گوناگونی بالایی دارند. ژن‌های پایداری به آمینوگلیکوزید، پایداری به ¹MLS و پایداری به تتراسایکلین در ریزگردهای برخاسته از دامداری‌ها غالب بودند. این در حالی است که ژن‌های پایداری به باسیتراسین و داروهای چندگانه² در ریزگردهای برخاسته از تصفیه‌خانه‌های فاضلاب غالب بودند. نتایج این پژوهشگران خطر وجود و انتشار ARGs را به‌وسیله این ریزگردها نشان می‌دهد که سلامت عمومی جامعه را به‌شدت تهدید می‌کند. زنگ و همکاران [28] با آنالیز انواع محیط‌های آلوده‌شده به محتویات روده انسان، خوک و مرغ به‌عنوان نمونه‌های آلوده و تحت تأثیر فعالیت‌های انسان و مقایسه آنها با نمونه‌های آب اقیانوس و چندین نمونه خاک دست‌نخورده به‌عنوان نمونه‌های بکر و طبیعی نشان دادند که کل ژن‌های پایداری به آنتی‌بیوتیک در محیط‌های آلوده بسیار بیشتر از محیط‌های بکر است و باکتری‌های *Escherichia*, *Bacteroides* و *Clostridium* باکتری‌های غالب حامل این ژن‌ها هستند. این در حالی است که در محیط‌های بکر و دست‌نخورده باکتری‌های *Alteromonas*, *Vibrio* و *Proteobacteria* حامل ARGs هستند. این پژوهش از جمله پژوهش‌های پیشگام در زمینه شناسایی

ARGs در آشیان‌های بوم‌شناختی گوناگون و میزبان‌های متفاوت است. در پژوهشی دیگر، ونگ و همکاران [29] توانستند شمار بالایی از ARGs را در ژنوم‌های فاژهای (فاژوم³) موجود در محیط‌های تحت تأثیر انسان و مزارع پرورش حیوانات شناسایی کنند. در فاضلاب این مناطق، فاژهای خانواده *Siphoviridae* و *Myoviridae* و *Podoviridae* فاژهای غالبی بودند که ARGs را با خود حمل می‌کردند. فراوانی بالای ژن‌های پایداری به تتراسایکلین و MLS در این فاژها و باکتری‌های میزبان آنها شناسایی شد. این درحالی است که فراوانی سایر ARGs تغییری نداشته و ثابت بوده است. وجود هسته پایداری به آنتی‌بیوتیک‌ها در چنین شرایطی خطرناک است و باید این نکته را در نظر داشت که این ژن‌ها - که بیشتر به‌وسیله فاژها حمل می‌شوند - به‌راحتی انتشار پیدا می‌کنند.

رادان و رویز [30] توانستند 120 ژنوم میکروبی را از متاژنوم جوامع میکروبی همراه با مواد فیبری پلاستیکی که به‌مدت 14 ماه در شرایط گرمسیری قرار داشتند، جدا کنند که می‌توانند در تجزیه آلاینده‌های پلاستیکی در محیط‌زیست راهگشا باشند.

از جمله پژوهش‌های متاژنومیکس که در کشور انجام شده است، می‌توان به پایان‌نامه کارشناسی ارشد موسوی ملکی [31] اشاره کرد. وی با بررسی گوناگونی میکروبی اطراف کانسار کرومیت فرومد در استان سمنان جنس‌های باکتریایی غالب را *Rheinheimera* و *Cedecae* گزارش کرد. البته فراوانی باکتریایی کم بود که علت آن وجود غلظت‌های بالای فلزهای سنگین، رطوبت کم خاک و تعداد کم کلون‌های تعیین توالی شده بود. از سوی دیگر، گوناگونی بالایی وجود داشته که دلیل آن به شرایط قلیایی این کانسار در مقایسه با شرایط اسیدی بیشتر کانسارها نسبت داده شد. همچنین فروزان [32] در پایان‌نامه دکتری

1macrolide-lincosamide-streptogramin

2 multidrug and bacitracin

3 phageome

خود، متازنوم خاک‌های کانسارهای آهن استان سمنان را بررسی، تجزیه و تحلیل کرده و گزارش مفصلی از ژن‌های پایداری در برابر غلظت‌های بالای فلزهای سنگین به‌ویژه کادمیوم ارائه کرد. راهبردهای میکروارگانسم‌ها که در این متازنوم‌ها شناسایی شد، بسیار جالب بوده و کاربردهای مهمی در بیوتکنولوژی زیست‌محیطی از جمله معدن‌کاری زیستی¹ و زیست‌پالایی² خواهند داشت. در پژوهشی دیگر که به‌وسیله همت‌جو و همکاران [1] بر روی خاک‌های کانسار سرب و روی باما در اصفهان انجام گرفت، گوناگونی بالا و فراوانی بالایی از باکتری‌ها به دست آمد. فهرست باکتری‌های فراوان به‌دست‌آمده اینها هستند: *Solirubrobacter* از اکتینوباکتیریا، *Geobacter* از دلتا پروتئوباکتیریا، *Edaphobacter* از اسیدوباکتیریا، *Pseudomonas* از گاما پروتئوباکتیریا، *Gemmatimonas* از جماتیمونادتها، *Nitrosomonas* از بتا پروتئوباکتیریا، *Xanthobacter* و *Sphingomonas* از آلفا پروتئوباکتیریا، *Pedobacter* از باکترواییدت‌ها، *Ktedonobacter* از کلروفلکسی، *Verrucomicrobium* از ورووکومیکروبییا، *Sporosarcina* از باکترواییدت‌ها، *Beggiatoa* از گاما پروتئوباکتیریا و *Thiobacillus* از بتا پروتئوباکتیریا. این پژوهشگران در پژوهش تکمیلی خود افزون بر شناسایی تاکسونومی باکتری‌ها، آرکی‌ها و قارچ‌های موجود در خاک‌های بسیار آلوده کانسار باما، مسیرهای متابولیکی، ژن‌های پایداری به غلظت‌های بالای فلزهای سنگین سرب، روی، مس، کادمیوم، نیکل، آهن و آرسنیک را شناسایی کردند که شامل 52 ژن هستند. این ژن‌ها سازوکارهای سامانه‌های پایداری به آلودگی این فلزها را مشخص کردند که مهم‌ترین آنها به این ترتیب هستند: سامانه پایداری ATP‌آزهای نوع P، سامانه پایداری کمک‌رسان‌های پخشیدگی کاتیونی و سامانه پایداری

HME-RNDs [34]. شناسایی این جوامع میکروبی که بیشترشان کشت‌ناپذیر هستند، با بهره‌گیری از متازنومیکس می‌تواند به درک چگونگی انجام فرایندهای زیستی در چنین شرایط دشواری به دلیل غلظت بالای فلزهای سنگین منتهی شود که اهمیت بسیاری برای حفاظت از محیط‌زیست و مدیریت خاک دارند [35].

در پژوهشی دیگر آریایی‌نژاد و همکاران [36] نشان دادند که داده‌های حاصل از متازنومیکس می‌تواند در شناسایی آنزیم‌هایی که در شکمبه شتر با خاستگاه میکروبی می‌تواند وجود داشته باشد، راهگشا باشد. با توجه به اینکه بیشتر میکروبیوتا کشت‌ناپذیر هستند، این پژوهشگران توانستند یک آنزیم زایلاناز جدید پایداری به حرارت بالا را شناسایی کنند که در محدوده وسیعی از pH نیز می‌تواند فعال باشد. این پژوهشگران در پژوهشی دیگر نشان دادند که آنالیز *In Silico* داده‌های متازنومیکس همچنین وجود آنزیم‌های زایلاناز قلیایی پایداری به حرارت را در شکمبه شتر تأیید می‌کند که کاربردهای بسیاری می‌توانند در صنعت و کشاورزی داشته باشند [37]. همچنین معتمدی و همکاران [38] یک آنزیم لاکاز به‌دست‌آمده از پساب صنایع نساجی را شناسایی کردند که پایداری به شوری و گرمای بالا است. این آنزیم همچنین در محدوده گسترده‌ای از pH همچنان فعال است و می‌تواند رنگ‌های به‌کارگرفته شده در صنعت نساجی را که با ورود به پساب حاصل منجر به آلودگی محیط‌زیست می‌شوند، حذف کند. در پژوهشی دیگر رضایی و همکاران [39] با بررسی نمونه‌های آب و رسوب خلیج فارس که همیشه در معرض خطر آلودگی به هیدروکربن‌های نفتی به دلیل ترافیک بالای عبور نفتکش‌ها قرار دارد، گزارش کردند که آلودگی اندک ولی مداوم موجب رشد بسیار بالای برخی از باکتری‌های تجزیه‌کننده نفت خام مانند *Rhodobacterales*، *Oceanospirillales*، *Alteromonadales* و *Flavobacteriales* شده است. نوع هیدروکربن نفتی (آلیفاتیک یا آروماتیک)، مدت زمان

آلودگی و عمق رسوبات از عوامل اصلی تأثیرگذار بر چنین تغییراتی هستند.

ساخت نقشه زیستی میکروب‌ها برای سیاره زمین

پژوهشی بلندپروازانه در زمینه متاژنوم خاک با نام پژوهش میکروبیوم سیاره زمین (EMP)¹ آغاز شده است که هدف آن شناخت متاژنوم‌های خاک، از بیوم‌های اصلی زمین است و با نمونه‌گیری از آنها (200000 نمونه) و انجام مقایسه‌ها میان این بیوم‌ها به پیش می‌رود. از اهداف دیگر این پژوهش تولید داده‌های توالی متاژنومیکس شات‌گان آمپلیکون‌های 16S rRNA از نمونه‌های زیست‌محیطی است [4]. سایر اهداف EMP، به‌دست‌آوردن اطلس جهانی از کارکردهای پروتئین‌ها، ایجاد فهرستی برای پراکندگی تاکسونومیکی ژنوم‌های مشابه و به‌این‌ترتیب استفاده از داده‌های تولیدشده برای توسعه مدل‌های پیش‌بینی‌کننده اکوسیستمی هستند. درواقع پژوهش میکروبیوم سیاره زمین، تلاشی سیستماتیک در جهت شناخت گوناگونی تاکسونومیک و کارکردی جهانی میکروب‌ها با هدف بهره‌وری برای انسان و زمین است. این پژوهش یک تلاش چندبعدی کلان به‌منظور آنالیز جوامع میکروبی در سرتاسر کره زمین است. در این پژوهش چهار دستاورد کلیدی وجود خواهد داشت [40]:

- 1- اطلس ژن (GA): یک ذخیره‌گاه مرکزی و پایگاه داده برای تمامی اطلاعات به‌دست‌آمده در طول این پژوهش است. این ذخیره‌گاه فرمتی قابل جستجو برای نگهداری تمامی اطلاعات در رابطه با تفسیر آنها، متادیتای زیست‌محیطی و توالی‌ها فراهم می‌کند. به‌این‌ترتیب، محلی برای ذخیره همه اطلاعات، چه شناخته‌شده و چه شناخته‌نشده که به آن ماده سیاه هم می‌گویند، پدید خواهد آمد؛

- 2- ژنوم‌های مونتاژشده میکروبیوم زمین (EM-AG): دارای تمامی ژنوم‌های مونتاژشده از داده‌های EMP خواهند بود که با استفاده از مجموعه‌ای از نرم‌افزارها به‌صورت خودکار و در ذخیره‌گاه‌های همگانی آماده شده و یک پلات‌فرم تحلیلی مشتق‌شده از KBASE تفسیر خواهند شد. این کار، آنالیز ژنومیک تطبیقی را برای تمامی ژنوم‌های شناخته‌شده تاکنون و همچنین ژنوم‌های مشتق‌شده از EMP و متاژنوم‌ها امکانپذیر خواهد ساخت؛
- 3- پروتال نمایش میکروبیوم زمین (EM-VIP): این دستاورد کارشناسان را برای ساخت نرم‌افزار نمایش برهم‌کنش‌ها جهت سنتز دیدگاه یگانه ما به فرمتی قابل دسترس برای همه، به‌شدت درگیر خواهد کرد. به‌این‌ترتیب، چشم‌اندازی از زمین در فضای میکروبی پدید خواهد آمد که پارامترهای مکانی زیست‌محیطی و فضای ژنومیک کارکردی را به تصویر می‌کشد و اجازه بازبینی داده‌های EMP و ارائه تئوری‌های جدید بوم‌شناختی را می‌دهد؛

- 4- بازسازی متابولیکی میکروبیوم زمین (EMMR): بر مبنای تشریح و پیش‌بینی نرم‌افزاری متابولوم متاژنومیک همچون modelSEED و جریان نسبی متابولیک، تغییرات در متابولیت‌ها و نیمرخ‌های متابولیکی در کل فضای بیورژئوگرافیک تشریح خواهد شد. این کار برای ارائه توضیحاتی درخصوص تولید متابولیت‌ها در بیوم‌های ویژه مورد استفاده قرار خواهد گرفت که معیار اندازه‌گیری دیگری برای بازتعریف ویژگی‌های هر بیوم فراهم می‌کند.

برخی از دستاوردهای EMP در اینجا آورده شده‌اند: داده‌های EMP با آنالیز رسوبات آب شیرین رودخانه در امتداد مسیر آن به جهت بررسی تأثیرات انسان نشانه‌های تأثیر بسیار زیاد فعالیت‌های انسانی بر جوامع میکروبی را نشان دادند. گوناگونی مکان‌های مطالعه و پرسش‌های نهفته در پژوهش در این 30000 نمونه اولیه، شگفت‌آور است و تنها قله کوه یخ را آشکار می‌کنند. آنالیز مقدماتی

10000 نمونه، در حدود 6 میلیون یگان تاکسونومیکی باکتریایی¹ (در سطح تاکسای جنس یا گونه) را شناسایی کرد که تنها بخش کوچکی از آنچه که می شد برای فیلورنی های شناخته شده با استفاده از پایگاه های داده ای 16S rRNA مانند GreenGenes به صورت نقشه نمایش داد، بودند. بسامد و پراکنش این گونه ها ما را قادر به پاسخ گویی به پرسش های جالبی می سازد، همچون پراکنش تاکسا در کل زیست بوم های گوناگون. همچنین دیتاست های EMP نشان می دهند که هم پوشانی چشمگیری در تاکسا میان مکان های مختلف وجود دارد و ارگانسیم هایی که در یک محل فراوانند، در محل دیگری بسیار کمیاب هستند. چنین پراکنشی، پیش تر در مکان های دریایی نشان داده شده است [40].

در سیستم خاک، جوامع میکروبی خاک های علفزار در ناحیه غرب میانه ایالات متحده آمریکا که بیشتر آن با زمین های کشاورزی جایگزین شده است، تغییراتی را در ترکیب بیشتر این جوامع و به ویژه بر فراوانی نسبی *Verrucomicrobia* (شاخه باکتریایی که بسیار کم مورد پژوهش بوده ولی در کل نمونه های خاک این زیست بوم با فراوانی بیشتر از 50 درصد توالی های 16S rRNA غالب است) و تأثیر پویایی کربن نشان داده است. غالب بودن وروکومیکروبی در خاک های این زیست بوم آشکار است ولی بوم شناسی آنها همچنان به این دلیل که کشت اعضای این گروه و مطالعه آنها در آزمایشگاه دشوار است، به درستی فهمیده نشده است. تمامی تاکسای فراوان وروکومیکروبیایی در کلاس اسپارتوباکتیریا طبقه بندی شدند ولی این تاکسا به جدایه های وروکومیکروبی که پیشتر کشت شده، ربطی نداشته اند که تعیین ویژگی های بوم شناختی این تاکسا را دشوار می سازد. همچنین مشخص شد که تغییر پذیری فضایی در فراوانی وروکومیکروبی از شرایط آب و هوایی قابل پیش بینی است

و اینکه این گروه در خاک هایی که در معرض درجه حرارت و بارش متوسط بودند، فراوان ترین گروه بودند. این آنالیزها می توانند در کمک به بهبود تلاش برای ترمیم زیست بوم علفزار سودمند واقع شوند. با بررسی گوناگونی میکروبی خاک (که زمانی در این بیوم وجود داشته است) مشخص شد که الگوهای زیست جغرافیایی فراوانی نسبی وروکومیکروبی در بسیاری از خاک های این چمنزار به شدت تغییر کرده است. داده های متازنومیکس شات گان از این فرضیه که وروکومیکروبی یک تاکسای به نسبت کند رشد است، حمایت می کند زیرا که از شرایط محدودیت عناصر غذایی در دسترس بهره می برد. به طور ویژه، فراوانی های وروکومیکروبیایی با ژن های گوناگون مربوط به متابولیسم کربوهیدرات ها همبستگی مثبت دارد ولی با ژن های مربوط به متابولیسم نیتروژن و تقسیم سلولی همبستگی منفی دارد. بنابراین امکان دارد که وروکومیکروبی بخش بزرگی از جوامع زیر سطحی را در مناطقی که تغییرات در کیفیت و کمیت مواد آلی گیاهی وارد شده رشد بیشتر تاکسای کوپوتروف را تحمیل می کند، نمایندگی کند. این فرضیه با نتایجی همخوانی دارد که بیانگر کاهش مداوم در فراوانی های نسبی وروکومیکروبی هستند هنگامی که به خاک های کل شمال آمریکا عناصر غذایی (با کوددهی) افزوده می شود. همچنین، این فرضیه با اطلاعات ژنومیک اخیر به دست آمده از *Spartobacteria aquaticum* که یک وروکومیکروبیایی آبزی است و در همان کلاس وروکومیکروبیایی غالب مشاهده شده در خاک در این پژوهش قرار داشته و در تجزیه ترکیبات کربنی مقاوم به تجزیه میکروبی²، ویژه و خاص است، مطابقت دارد. همچنین این داده ها نشان دادند که الگوهای فضایی با تغییرات شدید در پویایی کربن همراه اند. به این ترتیب بهره گیری از روش های متازنومیکس برای بازسازی

گوناگونی بیوژئوشیمیایی زیرسطحی و شیب تغییرات در زیست‌بوم‌های در خطر کاربردی‌بودن آن را نیز به نمایش می‌گذارد. چنین اطلاعاتی به‌ویژه اگر بازگرداندن کارکردهای کلیدی زیست‌بوم مانند کربن اندوزی¹ در خاک که به‌شدت به‌وسیله جوامع زیرسطحی کنترل می‌شوند، هدف اصلی باشند، اهمیت پیدا می‌کنند. همچنین انحراف در جوامع میکروبی خاک از حالت پیش‌بینی‌شده پیش از کشاورزی را می‌توان برای تعیین گستردگی میزان تجزیه به‌وسیله خاک‌ها در کل مرتع چمنزار بومی به کار برد. به‌طور کلی، این کار نشان می‌دهد که ما می‌توانیم از پیشرفت‌های پیشین در شناخت همه‌جانبه جوامع میکروبی برای بازسازی الگوهای جغرافیای زیستی در گوناگونی و توانایی‌های کارکردی میکروب‌ها در کل یک زیست‌بوم تقریباً از بین‌رفته بهره ببریم. این رویکرد می‌تواند برای تعیین چگونگی تغییرات تاریخی در شرایط محیطی که گوناگونی و کارکرد جوامع زیرسطحی را در سیستم‌های دیگر تغییر داده‌اند و یا برای تعیین چگونگی تغییر اقلیم روی داده شده به‌وسیله انسان و ویژگی‌های زیست‌بوم در آینده، گسترش بیشتری پیدا کند [41].

EMP نشان داده است که در خاک‌های قطبی شمالگان واقع در تونداری سیبری فراوانی باکتری‌ها در خاک‌های مدفون‌شده همچون خاک‌های سطحی مدفون‌شده زیاد است. از سویی دیگر، فراوانی‌های قارچی با افزایش عمق کاهش پیدا می‌کند و به‌گونه معناداری در خاک‌های مدفون‌شده کمتر از خاک‌های مدفون نشده است و در نتیجه نسبت قارچ‌ها به باکتری‌ها در خاک‌های مدفون‌شده به‌گونه چشمگیری بسیار کم است. نیمرخ‌سازی جامعه قارچی کاهش در قارچ‌های اکتومایکوریزایی را نشان داد. شرایط فیزیکی نامناسب (دماهای زیر صفر و کمی اکسیژن) و کاهش فراوانی

قارچ‌ها احتمالاً یک آشیان بوم‌شناختی برای باکتری‌های بی‌هوازی تجزیه‌کننده مواد آلی خاک (SOM)² مانند اعضای خانواده اکتینوباکتیریا که فراوانی‌های نسبی معنادار بالاتری را در خاک‌های مدفون‌شده نسبت به مدفون‌نشده دارند، فراهم کرده‌اند. این پژوهش، اطلاعات ما درباره ساختار جامعه میکروبی در خاک‌های نواحی شمالی دارای پرمافراست³ (دارای یخبندان همیشگی) را گسترش داده و نشانه‌هایی از تجزیه به تأخیر افتاده SOM را در خاک‌های مدفون‌شده برای تاکسای میکروبی ویژه و به‌خصوص کاهش در فراوانی و فعالیت قارچ‌های اکتومایکوریزایی می‌باشند و مقادیری را که تجزیه‌کننده‌های باکتریایی می‌توانند به‌عنوان جانشین‌های کارکردی آنها فعالیت کنند، نشان می‌دهد [42].

EMP پژوهشی را انجام داده است که در آن از آب بخشی از کانال انگلیسی مانس نمونه‌برداری شده و با اجرای روش‌های متاژنومیکس با پوشش بسیار عمیق توالی‌یابی انجام شده و سپس با دیتاست‌های سرشماری جهانی میکروب‌های دریایی (ICoMM) که از سراسر جهان از نمونه‌های دریایی و رسوبات گرفته تا محیط‌های مربوط به اسفنج‌های دریایی به دست آمده بودند، مقایسه کرده و به این ترتیب مشخص کرده است که در مکان نمونه‌برداری شده با افزایش عمق توالی‌یابی، همپوشانی فیلوژنتیکی بیشتری با داده‌های جهانی ICoMM به دست می‌آید. این ناحیه دارای 31/7 تا 66/2 درصد یگان‌های بهره‌برداری تاکسونومیک (OTU) شناسایی‌شده در یک بیوم مشخص از ICoMM بود. برون‌یابی این همپوشانی به $10^{11} \times 1/93$ توالی از ناحیه نمونه‌برداری شده اشاره می‌کند که تمامی گوناگونی فیلوژنتیکی باکتریایی ICoMM، یعنی گوناگونی تمام اقیانوس‌های کره زمین را در خود دارد. این نتایج به‌طور قوی تأکید می‌کنند که زیست‌کره دریایی

از یک بانک بذر میکروبی پایدار که پیش‌تر از وجود آن ناآگاه بودیم، نگهداری می‌کند و گفته باس بکینگ¹ را تأیید می‌کند: در بوم‌شناسی میکروبی همه چیز همه جا وجود دارد و این محیط است که از میان آنها گزینش می‌کند [43].

رویکردی جامع نسبت به زیست‌شناسی زیست‌بوم‌های دریایی

زیست‌بوم‌های دریایی در تنظیم چرخه‌های بیوژئوشیمیایی سیاره زمین و آب‌وهوای آن حیاتی هستند. هنوز هم سازماندهی، تکامل و پویایی اقیانوس‌ها به‌درستی درک نشده است. پژوهش اقیانوس‌های تارا در سپتامبر سال 2009 برای یک پژوهش سه ساله که زیست‌بوم‌های اقیانوسی را از روی کشتی تارا مطالعه می‌کند، آغاز شد و یک برنامه نمونه‌برداری منحصربه‌فرد که روش‌های مشاهده‌ای و ژنومیک را در برمی‌گیرد برای تشریح ویروس‌ها، باکتری‌ها، آرکی‌ها، آغازیان تک‌سلولی و متازوآ در محیط زیست فیزیکوشیمیایی طبیعی هریک اجرا شده است. هدف این پژوهش میان‌رشته‌ای، تولید سیستماتیک و دیتاست‌های با دسترسی آزاد برای کاوش کردن ساختمان و ترکیب مرفولوژیکی و مولکولی، گوناگونی، تکامل، بوم‌شناسی و تأثیرات جهانی پلانکتون‌ها بر سیستم سیاره زمین است که قابل استفاده و کاربردی باشند [2].

پژوهش اقیانوس‌های تارا دربرگیرنده پژوهش کلان و همه‌جانبه پنج رده بزرگ حیات، یعنی ویروس‌ها، گیروس‌ها² (این واژه چکیده ویروس غول‌پیکر است که به دلیل DNA نوکلئوپلاسمی سلولی بزرگی که دارند به این نام خوانده می‌شوند و ژن‌های ویژه‌ای دارند که در دیگر اشکال حیات یافت نمی‌شوند)، باکتری‌ها، آغازیان و زوپلانکتون‌ها برای دیتاست‌های آماری در حال توسعه

است که آگاهی‌بخش الگوهای مفهومی درباره پیچیدگی برهم‌کنش میان ارگانیسم‌ها است که جریان‌های انرژی، بیوژئوشیمیایی و گردش مولکول‌ها در زیست‌بوم‌های اقیانوسی را به پیش می‌راند. چنین دیتاستی درک ما از اصول حاکم بر زیست‌بوم‌های دریایی و تکامل حیات در اقیانوس را بهبود خواهد بخشید و به‌این‌ترتیب مایه بالابردن ظرفیت ما از ارزیابی خدمات زیست‌بوم‌ها و تواناکردن ما را در پیش‌بینی‌های بهتر پراکندگی ماهی‌های موجود و تأثیرات دگرگونی‌های جهانی آب‌وهوایی می‌گردد. همچنین ارگانیسم‌های پلانکتونی منبعی بزرگ و دست‌نخورده از ترکیبات زیست‌فعال برای صنایع دارویی، غذایی و آرایشی هستند، همان‌گونه که مسیرهای متابولیکی آنها ممکن است نیازهای آینده‌ما به انرژی را برطرف کنند [44].

آشنایی با برخی از ابزارهای محاسباتی موردنیاز در متاژنومیکس

با توجه به بزرگی بسیار داده‌های به‌دست‌آمده از روش‌های متاژنومیکس که آن را وارد بحث داده‌های کلان³ می‌کند، رویکردهای محاسباتی در پردازش و تحلیل این داده‌ها نقش پررنگی بازی می‌کنند. پرداختن به جزئیات نقش محاسبات و الگوریتم‌های مربوط به آن از اهداف مقاله کنونی نبوده است. بنابراین در اینجا تنها به چکیده‌ای درباره ابزارهای محاسباتی برای نیم‌رخ‌سازی تاکسونومیک میکروبیوم متاژنوم‌های شات‌گان بسنده می‌شود که در حوزه متاژنومیکس قرار دارد.

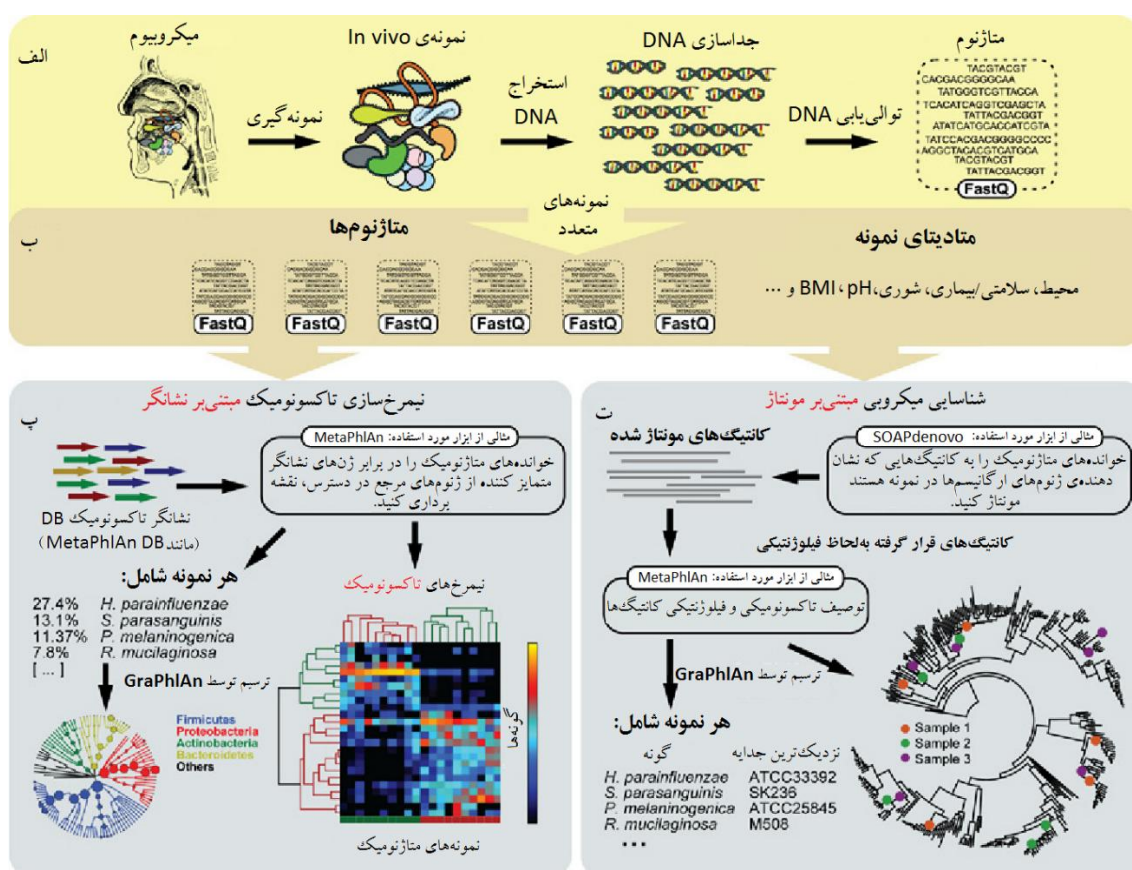
متاژنومیکس شات‌گان، دیتاست‌های بسیار بزرگی از توالی‌های کوتاه تولید می‌کند که آنالیز کردن آنها بسیار چالش‌برانگیز است. به این منظور، روش‌های محاسباتی مختلفی برای کاوش در داده‌های شات‌گان (WMS)⁴ از دیدگاه‌های مختلف و حتی مکمل یکدیگر (همانند ترکیب

3 Big Data
4 whole-metagenome shotgun

3Baas Becking
2 Girus: Gigantic Virus

تاکسونومیک، پتانسیل متابولیکی و گوناگونی فیلوژنتیکی) پیشنهاد شده‌اند. نیمرخ‌سازی تاکسونومیک یک کار محاسباتی برای استنباط کلادهای تاکسونومیک است که جمعیت یک جامعه میکروبی و نسبت آنها (فراوانی نسبی) را تشکیل می‌دهند. به این منظور، از دیتابیس‌های مرجع برای اختصاص دادن برجسب‌های تاکسونومیک به توالی‌های WMS استفاده می‌شود. البته اندازه بزرگ متاژنوم و دیتابیس‌های مرجع چالش مهمی را پیش روی کار نیمرخ‌سازی تاکسونومیک قرار می‌دهند، برای مثال اندازه نمونه‌های WMS در پژوهش‌های مربوط بین 10^9 و 10^{10} باز بوده که در حال افزایش نیز است و در حال حاضر شمار ژنوم‌های توالی‌یابی‌شده در دیتابیس‌های همگانی بیش از 400000 (بیشتر از 2000 گیگابایت) عدد هستند که به سرعت هم در حال افزایش هستند. نخستین روش‌های نیمرخ‌سازی تاکسونومیک براساس نقشه‌برداری با BLAST بودند به گونه‌ای که هر خوانش توالی در طول هر ژنوم میکروبی برای نسبت‌دادن تاکسونومی (معمولاً به وسیله بهترین خط برخورد) جستجو می‌شود، ولی این روش برای WMS و دیتاست‌های مرجع ژنوم‌های متداول حتی با وجود ابزارهای محاسباتی بسیار پیشرفته موجود عملی نمی‌باشد. چالش‌های دیگر آن شامل جانبداری‌ها در دیتابیس‌های

مرجع می‌شوند که به طور گسترده به وسیله ژنوم ارگانیزم‌های کشت‌پذیر و به خوبی شناخته شده پُر شده‌اند. همچنین برخی از بخش‌های درخت زندگی نماینده‌های اندکی دارند. این کمبود ژنوم‌های مرجع که وابستگی نزدیکی هم داشته باشند، مانع شناسایی توالی‌های محافظت‌شده فیلوژنتیکی و ارزیابی تأثیر رخداد‌های انتقال افقی ژن می‌گردند. افزون بر آن، آنالیز داده‌های WMS رودرروی مسائل معمول همچون سایر پژوهش‌های توالی‌یابی توانمند نسل جدید قرار می‌گیرند که شامل طول خوانش‌های به نسبت کوتاه (به طور عمده بین 100 تا 250 نوکلئوتید) در مقایسه با توالی‌یابی سنتی سنگر (در حدود 1000 نوکلئوتید) می‌شوند و دارای نرخ خطای نه‌چندان کمی هستند که در کل دیتاست توالی‌یابی وجود دارد. روش‌های متداول برای نیمرخ‌سازی تاکسونومیک WMS را می‌توان بر مبنای چگونگی استفاده مستقیم آنها از ژنوم‌های مرجع گروه‌بندی کرد. این روش‌ها عبارتند از روش مونتاژ، روش نقشه‌برداری، ترکیبی از این دو روش و روش مبتنی بر نشانگر، شمای کلی دو تا از مشهورترین این روش‌ها، یعنی روش مبتنی بر نشانگر و روش مبتنی بر مونتاژ در شکل 1 به ترتیب در بخش‌های پ و ت ارائه شده است [45,46].



شکل 1 روند کلی کار به دست آوردن نیمرخ تاکسونومیک از گردآوری نمونه تا نیمرخ‌سازی تاکسونومیک مبتنی بر نشانگر یا مبتنی بر مونتاژ. الف) استخراج شده و از نمونه‌های میکروبیوم جدا شده و به وسیله توالی‌یابی شات‌گان تعیین توالی شده که میلیون‌ها خوانش کوتاه در هر نمونه WMS تولید می‌کند؛ ب) یک دیتاست معمول متاژنومیک دارای نمونه‌های توالی‌یابی شده به فرمت FastQ (شامل توالی نوکلئوتیدی و نمره‌های کیفیت مربوطه) و متادیتای مربوط به آن است. با استفاده از یک (پ) روش مبتنی بر نشانگر همچون استفاده از MetaPhlAn، خوانش‌های کوتاه را به وسیله ژن‌هایی که از نظر تاکسونومیک مهم هستند (ژن‌های آگاهی بخش تاکسونومی) و ژن‌های نماینده یا شاخص به صورت نقشه در می‌آورند. این نتایج در نیمرخ‌های فراوانی در سطح گونه برای هر نمونه که بعد می‌تواند ادغام شده و برای خوشه‌بندی یا آنالیز گوناگونی استفاده شود، کاربرد دارند؛ ت) خوانش‌های کوتاه را می‌توان به طور متناوب نخست به کانتیگ‌های بلندتر مونتاژ کرد (برای مثال با استفاده از SOAPdenovo) که سپس برای کشف گوناگونی زیستی برای هر نمونه در یک درخت فیلوژنتیکی قرار می‌گیرند (برای مثال با استفاده از PhyloPhlAn). سایر نیمرخ‌های جایگزین در متن به بحث گذارده شده‌اند.

در برخی از موارد امکان بازسازی ژنوم‌ها از اعضای غالب‌تر جامعه وجود دارد ولی مونتاژ همه اعضای یک جامعه شدنی نیست. پس از آغاز مرحله مونتاژ، در مرحله دوم با مقایسه توالی با ژنوم‌های مرجع، اطلاعات تاکسونومیک یا فیلوژنتیکی به هر کانتیگ نسبت داده می‌شود. چندین انتخاب برای هر دو مرحله در دسترس بوده و چند نرم‌افزار نیز مانند MetAMOS، MOCAT و

روش نیمرخ‌سازی تاکسونومیک مبتنی بر مونتاژ میکروبیوم

برای دستیابی به یک تصویر کلی و کامل ژنومی از نمونه‌های زیست‌محیطی، نیاز به بازیابی توالی با طول کامل ژنوم برای هر میکروب داریم. به این منظور، با استفاده از روش‌های مونتاژ de novo متاژنومیک، نخست خوانش‌های WMS را می‌توان به کانتیگ‌ها مونتاژ کرد و

Ray Meta هر دو مرحله را با هم در یک مرحله ترکیبی یکی می‌کند [46].

همچنین ابزارهای مونتاژ تک‌ژنوم همچون SOPAdenovo به‌طور مستقیم برای داده‌های متاژنومیک به کار گرفته شده‌اند و موفقیت‌هایی نیز به‌دست آورده‌اند. البته، این ابزارها برای پژوهش‌های متاژنومیک که آمیخته‌ای از ژنوم‌ها را در خود دارند، بهینه‌سازی نشده‌اند. چندین بسط و توسعه در متاژنومیکس برای برآمدن از عهده توالی‌های متاژنومیک رخ داده است که شامل رفع مشکلات حافظه رایانه به دلیل اندازه بزرگ نمونه‌های WMS و برطرف کردن خطر تولید کانتیگ‌ها و اسکافلدهای کایمری ترا ارگانیسمی است [46].

MetaVelvet و Meta-IDBA دو نرم‌افزار مشهور مونتاژ مبتنی بر دوبراین¹ هستند که برای فراوان‌ترین ارگانیزم‌های موجود در یک نمونه WMS و مونتاژهای با کیفیت بالای به‌دست‌آمده به‌وسیله اکستنشن‌های *ad hoc* در روش‌های موجود، استفاده می‌شوند. لازم به توضیح است که مونتاژ داده‌های حاصل از توالی‌یابی شات‌گان را می‌توان به دو روش کلی انجام داد: روش همپوشانی مورد توافق² (OLC) و روش نمودار دوبراین. در این دو روش از چندین مونتاژگر متفاوت ژنوم استفاده می‌شود. به‌طور اساسی، مونتاژ OLC مقایسه جفتی خوانش‌های توالی و تطبیق دادن آنها در یک نمودار همپوشانی است. سپس، این توالی‌های همپوشان به هم متصل می‌شوند تا یک توالی مورد توافق به دست آید. در روش مونتاژ دوبراین هر خوانش توالی در نموداری که همه *k-mer*های ممکن وجود دارند، نشان داده می‌شود. دو *k-mer* زمانی به هم متصل می‌شوند که موقعیت‌های همپوشان و پشت‌سرهم داشته باشند. پس، همه خوانش‌ها در یک دیتاست با اتصال در یک نمودار نشان داده می‌شوند و کانتیگ‌های

مونتاژشده به‌وسیله پیمایش این اتصالات برای به‌دست‌آمدن یک توالی از *k-mer*ها تولید می‌شوند [46,47].

قرارگیری تاکسونومیک کانتیگ‌های بازسازی‌شده بارها به‌وسیله نقشه‌برداری توالی در ژنوم‌های مرجع انجام می‌شود. هرچند این راهبرد نجات‌بخش و راهنما می‌تواند در برخی از موارد کارآمد باشد، ابزارهای خودکار و دقیق‌تری به‌طور اختصاصی برای این کار توسعه پیدا کرده‌اند همانند MetaPhyler و PhyloPhlAn. این ابزارها علاوه بر کارکرد تاکسونومیک خود، یک ارزیابی فیلوژنومیک هم از کانتیگ‌ها با قراردادن آنها در درخت زندگی میکروبی انجام می‌دهند (شکل 1 ت). به‌طور ویژه، PhyloPhlAn از بیش از 400 پروتئین به‌شدت محافظت‌شده در اعضای توالی‌یابی‌شده فیلوژنی میکروبی برای پی‌بردن به جایگاه فیلوژنتیکی کانتیگ‌های مونتاژی ژنوم‌ها یا متاژنوم‌های جدید استفاده می‌کند. مشخص شده است که حتی از کانتیگ‌های کوچک هم که تنها 1 درصد کل ژنوم را در خود دارند، به‌دقت می‌توان برای ریشه‌یابی فیلوژنی میکروبی استفاده کرد و سپس تاکسونومی را به‌طور خودکار استنباط کرد و یا به‌طور دستی انجام داد [48].

به‌طور ویژه، روش‌های مبتنی بر مونتاژ برای میکروبیوم‌هایی که بخش بزرگی از آنها را میکروب‌های کشف‌نشده (توالی‌یابی‌نشده) تشکیل می‌دهند، مناسب‌اند. مزیت استفاده از روش‌های مبتنی بر مونتاژ برای چنین متاژنوم‌هایی که تنها بخشی از آنها به‌وسیله ژنوم‌های مرجع پوشش داده می‌شوند، این است که این روش‌ها بیشتر بر استفاده غیرمستقیم از ژنوم‌های مرجع در قیاس با سایر روش‌های نیم‌رخ‌سازی تاکسونومیک که ممکن است بخش‌های نوین جامعه را از دست بدهند، متکی‌اند. از طرف دیگر، برای محیط‌زیست‌هایی همچون بدن انسان تلاش‌های پرهزینه‌ای که برای توالی‌یابی ژنوم‌ها انجام

1de Bruijn

1 overlap-layout consensus

شده‌اند، سبب به‌دست‌آمدن نماینده‌هایی برای همه گوناگونی میکروبی موجود در آن شده‌اند. به‌طور معمول روش‌های نیمرخ‌سازی تاکسونومیکی بدون مونتاژ، در مورد توالی‌های اعضا با فراوانی کم جامعه که مونتاژ آنها به‌طور عمده چالش‌برانگیز است، بهتر عمل می‌کنند. هم‌اکنون، مونتاژ متاژنومیک یک حوزه فعال پژوهشی برای ارگانسیم‌های بسیار نزدیک به یکدیگر، نواحی به‌شدت حفاظت‌شده DNA و انتقال افقی ژن است که منجر به پدیدآمدن چالش‌های مهمی در دستیابی به مونتاژهای دقیق می‌شوند [46].

روش‌های ترکیبی برای تعیین جایگاه¹ متاژنومیک

روش‌های ترکیبی ویژگی‌های ذاتی توالی‌ها را بدون اتکا به نوکلئوتیدهای مستقیم یا هم‌ردیفی توالی پروتئین مورد مقایسه قرار می‌دهند. این ویژگی‌های ذاتی به دلیل اثرهای مفید ارگانسمی خود مشهورند که شامل تغییرات در مقدار GC، جانبداری کاربرد کدون و پراکنش k-merهای با طول متغیر می‌شوند به‌ویژه مورد آخر که مهم‌ترین ویژگی این مقایسه در نظر گرفته می‌شود. نخستین گام در یک روش ترکیبی، ساخت یک مدل آماری از ویژگی‌های ذاتی گونه یا جنس به‌وسیله پیش‌پردازش ژنوم‌های مرجع است (می‌توان آن را مرحله یادگیری یا آموزش هم نامید). گام دوم، استفاده از این مدل برای مقایسه و طبقه‌بندی خوانش‌های متاژنومیک است. روش‌های چندی برای دستیابی به این اهداف وجود دارند؛ برای نمونه PhyloPythia یا PhyloPythiaS از پشتیبانی یک ماشین بُرداری طبقه‌بندی‌کننده براساس آمار k-merها کمک می‌گیرند. دیگر روش‌ها از دیگر ابزارهای جدید و ترکیبی از نوین‌ترین ایده‌ها و پیشرفت‌های یادگیری ماشین استفاده می‌کنند که شامل Phymm و NBC هستند و بر پایه‌ی مدل‌های Bayesian و TACOا که راهبردی

مبتنی بر نزدیک‌ترین k را در پیش می‌گیرند، کار می‌کنند [45].

به دلیل اینکه روش‌های ترکیبی از هم‌ردیف‌سازی توالی محاسباتی پرهزینه دوری می‌کنند معمولاً زمان اجرای سریعی دارند. در این روش‌ها به‌گونه‌ای مشابه با روش‌های مبتنی بر مونتاژ، قابلیت‌های کلی نشان‌دادن ویژگی‌های خوبی در طبقه‌بندی خوانش‌ها بدون استفاده از ژنوم‌های مرجع بسیار نزدیک وجود دارد. این قابلیت به دلیل این حقیقت است که اطلاعات ذاتی توالی از نظر تکاملی محافظت‌شده‌تر از همولوژی توالی نوکلئوتیدی آن است. البته، این توانایی زمانی که توالی‌های نزدیک مربوطه در دیتابیس‌های مرجع وجود دارند، به دلیل هزینه‌های پایین در قدرت تمایز به دست می‌آید. به این دلیل، دقت نیمرخ‌سازی تاکسونومیکی ترکیبی معمولاً تا سطح جنس محدود می‌شود. افزون بر آن، قدرت تفکیک پایین با خوانش‌های کوتاه توالی‌یابی کمتر می‌شود. ترکیب کردن این روش‌ها با روش‌های مبتنی بر نقشه‌برداری می‌تواند کاستی‌های هر دو روش را کمتر کند [49].

بهره‌گیری مبتنی بر نقشه‌برداری از خوانش‌های متاژنومیک

روش‌های مبتنی بر نقشه‌برداری یا هم‌ردیف‌سازی، خوانش‌های متاژنومیک را بر پایه شباهت توالی با ژنوم‌های مرجع دسته‌بندی می‌کنند. در حال حاضر، پیشرفته‌ترین ابزارها براساس پیشرفت‌های اخیر در ابزارهای مبتنی بر نقشه‌برداری DNA خوانش ژنوم‌ها می‌باشند که در مقایسه با ابزارهای تولیدشده نسل اول شبیه BLAST، به‌مراتب بسیار سریع‌تر بوده و می‌توانند میلیون‌ها خوانش نقشه‌برداری شده را در طول ژنوم‌های انسانی در عرض تنها چند دقیقه نقشه‌برداری و ترسیم کنند. این ابزارها از شاخص‌های فشرده (مانند آنهایی که براساس تبدیل Burrows-Wheeler هستند) برای

شناسایی کارآمد محدود شده به مجموعه‌های بعدی ژنوم‌های مرجع که هم‌ردیف‌سازی کامل برای آنها انجام گرفته است، استفاده می‌کنند. هرچند، برخی از روش‌های نیم‌رخ‌سازی تاکسونومیکی هنوز از BLASTN به‌عنوان یک موتور نقشه‌برداری استفاده می‌کنند و به‌روزرسانی آنها برای اینکه الگوریتم‌های سرعت بالایی همچون SOAP2، Bowtie2 یا BWA را در برگیرند، هم ناچیز بوده است. در برخی از ابزارها، از نقشه‌برداری خام به‌طور مستقیم با نماینده‌ای برای ترکیب جامعه میکروبی به‌گونه‌ای ساده با تعیین یک برچسب تاکسونومیکی برای هر خوانش متاژنومیک بر مبنای بهترین برخورد استفاده می‌شود. هرچند در بیشتر موارد، خروجی خام برای رفع ابهامات در نقشه‌برداری به پردازش نهایی (پساپردازش) نیاز دارد که به‌وسیله نواحی حفاظت‌شده ژنومی، ژنوم‌های مرجع متعدد موجود در دیتابیس از همان کلاد تاکسونومیکی یا نقشه‌برداری خوانش‌ها برای ژنوم‌های دهنده نواحی انتقال‌یافته به شیوه‌ی افقی به وجود می‌آیند. این خوانش‌ها نمایندگان مبهمی هستند که به‌طور معمول می‌توان با استفاده از روش دورترین نیای مشترک¹ که خوانش‌ها را به کوچک‌ترین کلاد تاکسونومیکی ممکن دسته‌بندی می‌کند و شامل تمامی برخوردهای معنادار بوده و به‌وسیله ابزارهای مبتنی بر وب مانند MG-RAST و MEGAN انجام می‌شود، آنها را برچسب زد. ابزارهای توانمند و پیشرفته‌تر مبتنی بر فیلوژنتیک در دسترس هستند و همچنین روش‌های ترکیبی هم پیشنهاد شده‌اند که راهبردهای ترکیبی و راهبردهای مبتنی بر نقشه را گرد هم می‌آورند همچون PhymmBL که مدل‌های مارکوف را با نقشه‌برداری توالی ترکیب می‌کند، RITA که نظام آبشاری نقشه‌برداری مستقیم و ترجمه‌شده و «طبقه‌بندی‌کننده ترکیبی Naïve Bayes» را پیاده می‌کند و SPHINX که فضای جستجوی بسامدهای تترانوکلوئیدها

را که با نقشه‌برداری ترجمه‌شده دنبال می‌شوند، محدود می‌کند. اینگونه روش‌های ترکیبی، کارهای محاسباتی و قابلیت عمومی‌سازی برای سطوح بالای تاکسونومیکی، ابزارهای ترکیبی و قابلیت تفکیک‌پذیری برای سطوح پایینی تاکسونومیکی (مانند گونه) و ابزارهای مبتنی بر نقشه‌برداری را با یکدیگر ترکیب می‌کنند [45,50,51].

روش نیم‌رخ‌سازی تاکسونومیکی مبتنی بر نشانگر

بخش بزرگی از اطلاعات ژنومیکی موجود در ژنوم‌های مرجع برای اهداف نیم‌رخ‌سازی تاکسونومیکی مفید نبوده (مانند توالی‌های حفاظت‌شده در طول تاکسون‌های متعدد) و گهگاه همراه‌کننده (مانند ژن‌هایی که به شیوه افقی منتقل شده‌اند) هم هستند. روش‌های مبتنی بر نشانگر، ژنوم‌های مرجع را برای حذف توالی‌های فراوان و غیرتفکیک‌کننده پردازش نموده و بر نشانگرهای دارای اطلاعات مفید از نظر تاکسونومیکی تمرکز می‌کنند (شکل 1 پ). در نتیجه این کار، اندازه دیتابیس ژنوم‌های مرجع کوچک شده و بنابراین نیاز محاسباتی نمونه‌های WMS تنها برای بخشی از هر ژنوم که مجموعه نشانگر خوانده می‌شود، نیز کمتر شده که بسیار از نظر زمانی سریع‌تر انجام خواهد گرفت. تا به حال، از دو جور نشانگر در نیم‌رخ‌سازی تاکسونومیکی استفاده شده است: نشانگرهای جهانی یا فراگیر² و نشانگرهای ویژه کلاد³ [45].

نشانگرهای فراگیر توالی‌هایی هستند که در تمامی میکروب‌ها وجود داشته باشند و همچنین دارای نواحی متغیری باشند که بتوان از آنها به‌عنوان نشانه‌های تاکسونومیکی یا فیلوژنتیکی بهره‌برداری کرد. ژن 16S ریبوزومی برجسته‌ترین نمونه از نشانگرهای فراگیر است که دهه‌هاست از آن برای بررسی‌های فیلوژنتیکی و تاکسونومیکی استفاده می‌شود و اکنون گوناگونی

2 universal markers

3clade-specific markers

1 the lowest (least) common ancestor

زیرمجموعه‌ای از نه ناحیه به شدت متغیر آن، هدف چندین روش توالی‌یابی توانمند قرار دارند. ژن 16S rRNA در نمونه‌های WMS هم به خوبی وجود دارد (در حدود 0/1 درصد توالی‌های باکتریایی). بنابراین می‌توان از آن برای نیمرخ‌سازی تاکسونومیک با ابزارهایی همچون PhyloOTU استفاده کرد. برای مشخص‌تر کردن علامت تاکسونومیک 16S rRNA، از ژن‌های اضافی به شدت حفاظت شده می‌توان استفاده کرد (همانند *hsp65* و *rpoB*). همچنین ابزارهایی مانند AMPHORA و MetaPhyler مجموعه نشانگرهای فراگیر را به تعداد بسیار بیشتری گسترش داده‌اند که ژن‌های باکتریایی و ژن‌های آرکی‌یابی را هم در بر می‌گیرند. بنابراین دقت استنباط نیمرخ‌های تاکسونومیک جامعه را بالا می‌برد. PhyloPhlAn با استفاده از 400 نشانگر شبه‌فراگیر¹ برای استنباط جایگاه فیلوژنتیکی ارگانیسم‌ها در میکروبیوم پس از یک مرحله موتاژ جزئی ما را یک گام به پیش می‌برد. پس، نشانگرهای فراگیر از توالی‌های ژنومی اندکی که به‌طور فراگیر محافظت شده‌اند، استفاده می‌کنند و انتظار می‌رود در میکروب‌هایی که تاکنون توالی‌یابی شده‌اند، حضور داشته باشند ولی با آنها نمی‌توان از ژن‌هایی که فراگیر نیستند و بیشتر ژنوم‌های میکروبی را تشکیل می‌دهند، سود جست [52].

از نواحی نافراگیر ژنوم‌های میکروبی می‌توان با تمرکز بر ژن‌های نشانگر ویژه کلاد استفاده کرد که به‌طور انحصار در هر کلاد تاکسونومیک (همانند گونه) حضور دارند. این ژن‌ها به‌عنوان هسته ژن‌های درون یک کلاد تعریف می‌شوند که هیچ توالی مشابهی با سایر ژن‌های خارج از کلاد نداشته باشند. بنابراین، این ژن‌ها همچون اثرهای انگشت بی‌همتایی از هر کلاد تاکسونومیک هستند و می‌توانند ارگانیسم‌های بسیار نزدیک را با دقت بالایی و تنها با بررسی بود یا نبود آنها در متاژنوم از

یکدیگر تفکیک کنند. در MetaPhlAn تقریباً از 400000 ژن نشانگر ویژه کلاد که نمایانگر کل درخت زندگی برای ویژگی‌های تاکسونومیک ارگانیسم‌ها از نمونه‌های WMS هستند، استفاده می‌شود که مایه دقت بالا، برآورد کمی و تفکیک تا حد زیرگونه می‌شود. این ابزارها کارآمدترین استفاده را از ژنوم‌های مرجع در دسترس میسر ساخته و بنابراین چنین روش‌هایی، بیشترین پتانسیل را برای توسعه سریع و دقیق نیمرخ‌سازی متاژنومیک دارند [33].

روش‌های مبتنی بر نشانگر، توالی‌های مرجع را به‌منظور کاهش اندازه و افزایش قدرت جداسازی آن که از ابزارهای بسیار دقیق و سریع نتیجه می‌شوند، مورد پیش‌پردازش قرار می‌دهند. روش‌های مبتنی بر ترکیب و هم‌ردیف‌سازی تلاش می‌کنند تا چشم‌اندازی جامع و کامل با تعیین تاکسونومی برای بیشتر خوانش‌هایی که میسر است، ارائه کنند ولی راهبردهای آنها در مورد کارآمدی شرایط گوناگون و دقت تفکیک تاکسونومیک تفاوت می‌کنند [45].

نیمرخ‌سازی تاکسونومیک بر ژنوم‌های مرجعی که به‌خوبی شناسایی شده باشند، تکیه می‌کند. پس صحت نیمرخ به‌طور عمیقی از چگونگی در دسترس بودن دیتابیس‌های ژنوم‌های مرجع تأثیر می‌پذیرد که گوناگونی زیستی را در نمونه‌های متاژنومیک تحت بررسی پوشش می‌دهند. میکروبیوم‌های همراه انسان و به‌ویژه میکروبیوم روده در دیتابیس‌های ژنوم‌های مرجع به‌خوبی نمایش داده شده‌اند که به دلیل نقش مهم آنها در سلامت و بیماری‌های انسانی است. در روش‌های نوین توالی‌یابی ژنوم و مهم‌تر از آن در توالی‌یابی تک‌سلول، توجهی دوباره به جانبداری‌ها و توسعه مجموعه‌های مرجع برای بیشتر محیط‌زیست‌های گوناگون در حال انجام است. بنابراین، نیمرخ تاکسونومیک متاژنومیک می‌تواند از مجموعه غنی‌تری از ژنوم‌هایی که به‌سرعت در حال تولید هستند، بهره‌بردار، اگرچه چالش‌های محاسباتی در استفاده

¹quasi-universal markers

از مجموعه بزرگ‌تر توالی‌های مرجع وجود داشته و احتمالاً یکی از عوامل محدودکننده استفاده از آن خواهند بود [45].

نوروزی بیرامی و همکاران [53] با معرفی ابزار بیوانفورماتیکی CAMAMED که از روش مبتنی بر نقشه‌برداری استفاده می‌کند، گام مهمی در آنالیز داده‌های متاژنومیکس برداشتند. با توجه به اینکه اگر فهرست‌های ژنی در دسترس باشند، روش‌های مبتنی بر نقشه‌برداری به روش‌های مبتنی بر مونتاژ برتری داده می‌شوند که به‌ویژه در مورد آنالیز کارکردی داده‌های متاژنومیکس صادق است، CAMAMED معرفی شد. در این ابزار، توالی‌های متاژنومیکس با فهرست‌های ژنی غیرتکراری مطابقت داده شده و بسامد ژنی در نمونه‌ها به دست می‌آید. با توجه به ماهیت ترکیبی داده‌های متاژنومیکس روش مقیاس‌بندی تجمعی هم در سطح تاکسا و هم در سطح ژن انجام می‌گیرد. همچنین با منطبق کردن ژن‌ها بر پایگاه داده‌ای KEGG تفسیرهای ژنی در سطوح مختلف کارکردی مانند آنزیمی به دست می‌آیند. افزون بر این، CAMAMED کاربر را قادر می‌سازد تا با بررسی موارد نشانگرهای زیستی بالقوه در نمونه، شناسایی را انجام دهد.

متاژنومیکس شات‌گان پتانسیل شناسایی اعضای غیرباکتریایی را که در گوناگونی زیستی جوامع میکروبی حضور دارند، نیز دارد. برای نمونه، آرکی‌ها را می‌توان با همه ابزارهایی که توضیح داده شدند، شناسایی کرد درحالی‌که اکستشن‌های نیمرخ‌سازی تنوع ویروسی و میکروویوکاریوتی (مانند قارچ‌ها) هم‌اکنون تنها در چند روش نیمرخ‌سازی در دسترس قرار دارند. هرچند، نیمرخ‌سازی ارگانیزم‌های غیرباکتریایی به دلیل کمبود توالی‌های مرجع در دسترس و کمبود دستورکارهای بهینه استخراج DNA، چالش‌های جدیدی را به چالش‌های موجود اضافه می‌کنند. توسعه نیمرخ‌سازی تاکسونومیک

تا آنجا که تمامی قلمروهای حیات را در بر گیرد، ما را به مسائل دقت تفکیک تاکسونومیک و کارآمدی محاسباتی رهنمون می‌سازد که مهم‌ترین چالش‌هایی هستند که خواستار روش‌های نوین محاسباتی برای رویارویی با آنالیز WMS هستند، بنابراین این زمینه پژوهشی را برای پژوهش‌های بیشتر همچنان فعال نگه می‌دارند. مد نظر قرارداد این چالش‌ها برای شناخت بهتر و درک چنین جوامع میکروبی پیچیده‌ای بسیار حیاتی است [45].

یکپارچه‌سازی داده‌های آمیکس

توانایی تولید دیتاست‌های چندگانه آمیکس برای یک سیستم در یک مکان و زمان، پتانسیل فراهم کردن تصویری با جزئیات کامل از سیستم‌های زیستی و بوم‌شناختی را دارد. تلاش‌هایی برای یکی کردن فناوری‌های چندگانه آمیکس با فناوری‌های دیگر به‌ویژه در جوامع پیچیده میکروبی هنوز بسیار اندک بوده و اغلب بر لایه‌های داده‌های آمیکس در مسیرهای جامع یا همبستگی یک دیتاست آمیکس با دیگری استوار است. یکی از راه‌های این یکپارچه‌سازی استفاده از پایگاه‌های داده‌ای مرجع همانند KEGG، BioCyc و MetaCyc است [54]. این پایگاه‌ها با گذر زمان تأیید شده‌اند و دربردارنده اطلاعات مسیرهای متابولیکی و آنزیم‌ها، ترکیب‌های شیمیایی و ژن‌هایی هستند که امکان کارکرد برای آنها را فراهم می‌کنند. البته بسیاری از مسیرهای اصلی متابولیکی به‌طور عمده در سرتاسر درخت زندگی رخ می‌دهند و این پایگاه‌های داده‌ای مرجع همچنین شامل اطلاعاتی درباره پراکنش تاکسونومیک (شناخته‌شده) ژن‌های ویژه و مسیرهای ویژه هستند [55].

پایگاه‌های داده‌ای مرجع می‌توانند به‌عنوان پلاتفرمی برای پشتیبانی از داده‌های اکتشافی آمیکس به کار روند و نرم‌افزارهای در حال پدیدارشدن به این پایگاه‌های داده این امکان را می‌دهند که به‌عنوان پلاتفرمی برای یکی کردن داده‌های آمیکس به کار گرفته شوند. از پایگاه

داده‌ای KEGG می‌توان به‌طور مستقیم به‌وسیله اینترنت استفاده کرد. این پایگاه داده به کاربران اجازه کاوش ژن‌ها، ترکیب‌های شیمیایی و مسیرهای متابولیکی را می‌دهد. همچنین اطلاعات پایه‌ای درباره ژن‌ها و ترکیب‌های شیمیایی گوناگون شناخته‌نشده را برای تعیین مسیرهای کارکردی تحت تأثیر آنها فراهم می‌کند. کاوشگر مسیرهای برهم‌کنشی¹ نیز یک ابزار مبتنی بر وب است که به تصویرکشیدن و آنالیزکردن اطلاعات مسیرها را انجام می‌دهد. همچنین ابزار پاتویو² نیز یکپارچه‌سازی داده‌های متعدد آمیکس را انجام می‌دهد [56].

روش‌های مبتنی بر شبکه و همبستگی

لایه‌بندی داده‌های آمیکس در مسیرهای متابولیکی اصلی و موجود در همه موجودات زنده، امکان شناخت این دیتاست‌ها در زمینه‌های معین، مسیرها و واکنش‌های به‌خوبی شناخته‌شده را می‌دهد. در مورد پژوهش‌های آمیکسی چندگانه امکان بررسی کارکردی انواع داده‌های مربوطه (همانند ژن‌ها و ترکیب‌های درگیر در همان واکنش) برای در نظر گرفتن در زمینه واکنش‌های زیستی و بیوشیمیایی شناخته‌شده وجود دارد. زمانی نگرش‌های لایه‌ای می‌توانند بسیار مفید باشند که این روش‌ها به‌طور لزوم برای الگوهای جهانی و روابط درون و میان دیتاست‌های آمیکس طراحی نشده باشند [1].

در برابر آن، روش‌های همبستگی و مبتنی بر شبکه به‌منظور شناسایی الگوهای هم‌رخدادی و یا هم فراوانی برای دیتاست‌های آمیکس قرار دارند. این روش‌ها به‌نسبت ساده بوده و اغلب بدون خطاهای ذاتی در اندازه‌گیری‌های آمیکس و در زیست‌شناسی که توصیف می‌شوند، هستند. در این روش‌ها اغلب از همبستگی‌های پیرسون یا اسپیرمن استفاده می‌کنند و می‌توانند به آسانی این پرسش باشند که آیا ژن A می‌تواند به اندازه فراوانی

ژن B و C یا متابولیت D در همه نمونه‌ها حضور داشته باشد یا برخی از باکتری‌ها همیشه (یا هیچ‌وقت) با یکدیگر در نمونه‌ها حضور دارند یا خیر؟ گرچه این پرسش‌ها به نظر ساده می‌آیند، پتانسیلی برای شناسایی همبستگی‌های به‌دست‌آمده از آنالیز داده‌ها به جای بررسی روابط زیستی واقعی فراهم می‌کنند. همچنین روش‌های پیشرفته‌تری پیشنهاد شده‌اند که شامل رگرسیون جزئی حداقل مربعات، همبستگی‌های اسپارز برای داده‌های ترکیبی و مدل‌های خطی ارتقایافته هستند [57].

با وجود امکان به تله‌افتادن در آنالیزهای رگرسیونی، از آنها با موفقیت در بسیاری از پژوهش‌ها استفاده شده است و این توانایی را دارند که بینش‌های جدیدی را در زیست‌شناسی سیستم مورد علاقه ما آشکار کنند [55]. برای مثال با استفاده از همبستگی‌های آمیکسی تعیین کردن زمینه کارکردی خاستگاه ژن‌های ناشناخته، شناسایی ژن‌ها، متابولیت‌ها یا باکتری‌های همراه یا ویژگی‌های بوم‌شناختی یا زیست‌محیطی مربوط به هر محیط‌زیست و فراهم کردن اطلاعاتی درباره ارگانسیم‌های کشت‌ناپذیر ممکن می‌شود [1، 4، 2]. مثال‌هایی هم از همبستگی‌های کراس‌آمیکس (مانند آمیکس‌های چندگانه) وجود دارد و تلاش‌های اخیر برای یکی‌کردن دیتاست‌های ترانسکریپتوم و متابولیت‌ها از آزمایشگاه شیمی ایستایی [58] و ترکیب جامعه در دستگاه گوارش انسانی [59] در منابع علمی آمده‌اند که در این مقاله به برخی از آنها اشاره شد. یکپارچه‌سازی دیتاست‌های آمیکسی طلایه‌دار دانش بوده و توانایی بسیار بالایی برای شناخت میکروارگانسیم‌های زیست‌محیطی داشته و اگرچه یک چالش تکنولوژیکی را نمایان می‌کند ولی این چالش به‌طور کامل به‌ویژه در پژوهش بر جوامع میکروبی برطرف خواهد شد [2].

نتیجه‌گیری

ابزارهای مولکولی فرصت بی‌مانندی برای تشخیص و شناخت بیوتای محیط‌زیست سیاره ما به ما پیشکش

بتوانند متاژنومیکس، پروتئومیکس، لیپیدومیکس، متابولومیکس و دیگر آمیکس‌ها را در یک چارچوب مفهومی جامع و یکپارچه به‌گونه‌ای شایسته یکی کنند، لازم هستند. نخستین گام در این مسیر تفسیرهای مناسبی از دیتاست‌ها با به‌کارگیری استانداردهای حداقل اطلاعات در زمان ذخیره‌کردن این دیتاست‌ها و تفسیرها و استانداردسازی نام‌های مکان‌های بدنی و همچنین سایر بخش‌های تشریحی است. هم‌اکنون این فرصت برای همه ما فراهم شده است.

منابع

[1] Hollister, E.B., Brooks, J.P., and Gentry, T.J. (2015). Bioinformation and 'Omic Approaches for Characterization of Environmental Microorganisms. In: Pepper, I.L., Gerba, C.P., and Gentry, T.J., editors. Environmental Microbiology, 3th ed. Academic Press is an imprint of Elsevier, 525 B Street, Suite 1800, San Diego, CA 92101-4495, USA, pp. 482-505.

[2] همت جو م. ح. (1396) متاژنومیکس؛ پرده‌برداری از رازهای سیاره میکروبی ما، انتشارات چاپار- اساطیر پارسی، تهران، 493 ص.

[3] Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and Goodman, R.M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology*, 5: 248-249.

[4] Thies, J.E. (2015). Molecular Approaches to Studying the Soil Biota. In: Paul, E.A., editor. *Soil Microbiology, Ecology, and Biochemistry*, 4th ed. USA: Academic Press, Elsevier, pp. 151-185.

[5] طهمورث‌پور آ، صالحی ر، کرمانشاهی ک، و اسلامی گ. (1388) «استخراج RNA از سلول‌های بیوفیلم استریپتوکوکوس موتانس»، مجله میکروبیولوژی پزشکی ایران، 4: 9-14.

[6] Rivera, M.C., and Izard, J. (2015). Promises and Prospects of Microbiome Studies. In: Izard, J., and Rivera, M.C., editors. *Metagenomics for microbiology*. 3rd ed. Academic Press is an imprint

کرده‌اند. این ابزارها به ما اجازه طرح پرسش‌هایی را در مقیاس‌های جغرافیایی گسترده‌تری نسبت به آنچه که پیش از این امکانپذیر بود، داده‌اند (مانند پژوهش میکروبیوم سیاره زمین). برای نمونه، اکنون ما می‌توانیم به آزمون مسائلی همچون چگونگی تغییر جمعیت‌های میکروبی در انواع مختلف خاک‌ها و نواحی آب و هوایی، به همراه تغییر در ریشه‌های گیاه و همچنین میان‌گونه‌های مختلف گیاهی در پاسخ به مدیریت و آلودگی خاک بپردازیم. هم‌اکنون، روش‌های متاآنالیز به سرعت در حال توسعه و تکامل فناوری‌هایی هستند که دستیابی ما را هم به گوناگونی زیستی و هم به بیان ژن در تمامی محیط‌زیست‌ها از بدن انسان تا خاک و اقیانوس، شدنی می‌کنند و همچنین می‌توانند سرچشمه‌های طبیعی و بی‌پایانی از ارگانیسم‌ها با کاربردهای پزشکی و درمانی را به ما نشان دهند. البته می‌رود و نانپیری [60] خاطرنشان می‌کنند که پژوهش‌های آمیکس باید به فراسوی تشریح کنندگی و شناسایی گذر کنند و بیشتر فرضیه‌ساز باشند و همچنین بیشتر بر طراحی آزمایش‌ها تمرکز کنند به جای اینکه تنها بر نوآوری‌های تکنولوژیکی تأثیر بگذارند و به ما این نکته را یادآور می‌شوند که یک نگرش چندفازی برای درک جامعی از بیوژئوشیمی محیط‌زیست همچنان مورد نیاز است. فناوری‌هایی همچون آنالیز ایزوتوپی RFLP، تکثیر توالی ژن‌های *nirS*، *amoA*، *nifH* و *pmoA* و کاوشگرهای ایزوتوپی پایدار DNA، RNA و پروتئین¹ به ما اجازه هدف قرار دادن ارگانیسم‌ها یا گروهی از ارگانیسم‌های مسئول در کارکردهای ویژه در محیط‌زیست به‌خصوص آنهایی را که در واکنش‌های چرخه‌های کربن، نیتروژن و گوگرد درگیرند، با حساسیت بالا می‌دهند [61، 62، 1]. به این ترتیب قوانینی را که میلیون‌ها سال است، حاکم هستند، درک خواهیم کرد و از رازهای پنهان آنها رمز خواهیم گشود. در پایان، روش‌های نوینی که

1 DNA-, RNA-, and protein-SIP

International Society for Microbial Ecology Journal, 7: 2248-2258.

[16] Treusch, A.H., Leininger, S., Kletzin, A., Schuster, S.C., Klenk, H.P., and Schleper, C. (2005). Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic Crenarchaeota in nitrogen cycling. *Environmental Microbiology*, 7: 1985-1995.

[17] Nicol, G.W., and Schleper, C. (2006).

Ammonia-oxidising Chrenarchaeota: important players in the nitrogen cycle? *Trends in Microbiology*, 14: 207-212.

[18] Kim, S.I., Kim, J.Y., Yun, S.H., Kim, J.H., Lee, S.H., and Lee, C. (2004). Proteome analysis of *Pseudomonas* sp. K82 biodegradation pathways. *Proteomics*, 4: 3610-3621.

[19] Hettich, R.L., Pan, C., Chorney, K., and Giannone, R.J. (2013). Metaproteomics: harnessing the power of high performance mass spectrometry to identify the suite of proteins that control metabolic activities in microbial communities. *Analytical Chemistry*, 85: 4203-4214.

[20] Ram, R.J., VerBerkmoes, N.C., Thelen, M.P., Tyson, G.W., Baker, B.J., and Blake, R.C. (2005).

Community proteomics of a natural microbial biofilm. *Science*, 308: 1915-1920.

[21] Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., and Richardson, P.M. (2004). Community structure and metabolism

through reconstruction of microbial genomes from the environment. *Nature*, 428: 37-43.

[22] Mueller, R.S., Dill, B.D., Pan, C., Belnap, C.P., Thomas, B.C., and VerBerkmoes, N.C. (2011.) Proteome changes in the initial bacterial

colonist during ecological succession in an acid mine drainage biofilm community. *Environmental Microbiology*, 13: 2279-2292.

[23] HMP Consortium. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, 486: 207-214.

[24] Castillo, E.D., and Izard, J. (2015).

Metagenomics for Bacteriology. In: Izard, J., and Rivera, M.C., editors. *Metagenomics for microbiology*. 3rd ed. Academic Press is an imprint of Elsevier, 32 Jamestown Road, London NW1 7BY, UK, pp. 112-134.

[25] Mirete, S., Morgante, V., and Gonzalez-Pastor, J.E. 2016. Functional metagenomics of extreme environments. *Current Opinion in Biotechnology*, 38: 143-149.

of Elsevier, 32 Jamestown Road, London NW1 7BY, UK, pp. 144-159.

[7] همت جو م. ح.، صفری سنجانی ع. ا.، میرزایی اصل ا.،

طهمورث پور آ. (1397) «بهره‌گیری از انگشت‌نگاری

PCR-DGGE و تنفس خاک برای شناخت دگرگونی‌های

باکتریایی در خاک‌های آلوده به فلزهای سنگین»، مجله

زیست‌شناسی میکروارگانیسم‌ها، 29: 41-56.

[8] Luo, C., Tsementzi, D., Kyrpides, N., Read, T., and Konstantinidis, K.T. (2012). Direct

comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS One*, 7, 1-12.

[9] Hazen, T.C., Rocha, A.M., and Techtman, S.M. (2013). Advances in monitoring

environmental microbes. *Current Opinion in Biotechnology*, 24: 526-533.

[10] Tahmourespour, A., Salehi, R., and Kasra-Kermanshahi, R. (2011). *Lactobacillus acidophilus*-

derived biosurfactant effect on GTFB and GTFC expression level in *Straptococcus* mutants' biofilm cells. *Brazilian Journal of Microbiology*, 42: 330-339.

[11] Vetrovsky, T., and Baldrian, P. (2013). The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One*, 8: 1-10.

[12] Ajami, N.J., and Petrosino, J.F. (2015).

Toward the human virome. In: Izard, J., and Rivera, M.C., editors. *Metagenomics for microbiology*. 3rd ed. Academic Press is an imprint of Elsevier, 32 Jamestown Road, London NW1 7BY, UK, pp. 112-134.

[13] Gosalbes, M.J., Durban, A., Pignatelli, M., Abellan, J.J., Jimenez-Hernandez, N., and Perez-Cobas, A.E. (2011). Metatranscriptomic approach

to analyze the functional human gut microbiota. *PLoS One*, 6: 1-9.

[14] Urich, T., Lanze'n, A., Qi, J., Huson, D.H., Schleper, C., and Schuster, S.C. (2008).

Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One*, 3: 1-13.

[15] Turner, T.R., Ramakrishnan, K., Walshaw, J., Heavens, D., Alston, M., Swarbreck, D., Osbourn, A., Grant, A., and Poole, P.S. (2013). Comparative

metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants.

- [34] Hemmat-Jou, M. H., Safari-Sinegani, A. A., Che, R., Mirzaie-Asl, A., Tahmourespour, A and Tahmasbian, I. (2020). Toxic trace element resistance genes and systems identified using the shotgun metagenomics approach in an Iranian mine soil. *Environmental Science and Pollution Research*, 28(4):4845-4856.
- [35] همت‌جو م. ح. (1397) بهره‌گیری از متاژنومیکس برای شناخت دگرگونی ریزجانداران در خاک‌های آلوده به فلزهای سنگین. پایان‌نامه دکترای بیولوژی و بیوتکنولوژی خاک. دانشگاه بوعلی سینا.
- [36] Ariaeenejad, S., Maleki, M., Hosseini, E., Kavousi, K., Moosavi-Movahedi, A.A., and Hosseini Salekdeh, G. (2019a). Mining of camel rumen metagenome to identify novel alkali-thermostable xylanase capable of enhancing the recalcitrant lignocellulosic biomass conversion. *Bioresource Technology*, 281: 343-350.
- [37] Ariaeenejad, S., Hosseini, E., Maleki, M., Kavousi, K., Moosavi-Movahedi, A.A., and Hosseini Salekdeh, G. (2019b). Identification and characterization of a novel thermostable xylanase from camel rumen metagenome. *International Journal of Biological Macromolecules*, 126: 1295-1302.
- [38] Motamedi, E., Kavousi, K., Sadeghian-Motahar, S.F., Ghaffari, M.R., Sheykh-Abdollahzadeh-Mamaghani, A., Hosseini-Salekdeh, G., and Ariaeenejad, S. (2021). Efficient removal of various textile dyes from wastewater by novel thermo-halotolerant laccase. *Bioresource Technology*, 337: 125468.
- [39] Rezaei-Somee, M., Dastgheib, S.M.M., Shavandi, M., Ghanbari-Maman, L., Kavousi, K., and Amoozegar, M.A. (2021). Distinct microbial community along the chronic oil pollution continuum of the Persian Gulf converge with oil spill accidents. *Scientific reports*, 11: 11316.
- [40] Gilbert, A., Jansson, J.K., and Knight, R. (2014). The Earth Microbiome project: successes and aspirations. *BMC Biology*, 12: 1-4.
- [41] Fierer, N., Ladau, J., Clemente, J.C., Leff, J.W., Owens, S.M., Pollard, K.S., Knight, R., Gilbert, J.A., and McCulley, R.L. (2013). Reconstructing the Microbial Diversity and
- [26] Panelli, S., Capelli, E., Comandatore, F., Landinez-Torres, A., Granata, M.U., Tosi, S., and Picco, A.M. (2017). A metagenomic-based, cross-seasonal picture of fungal consortia associated with Italian soils subjected to different agricultural managements. *Fungal Ecology*, 30: 1-9.
- [27] Yang, Y., Zhou, R., Chen, B., Zhang, T., Hu, L., and Zou, S. (2018). Characterization of airborne antibiotic resistance genes from typical bioaerosol emission sources in the urban environment using metagenomic approach. *Chemosphere*, 213: 463-471.
- [28] Zeng, J., Pan, Y., Yang, J., Hou, M., Zeng, Z., and Xiong, W. (2019). Metagenomic insights into the distribution of antibiotic resistome between the gut-associated environments and the pristine environments. *Environment International*, 126: 346-354.
- [29] Wang, M., Sun, Y., Zeng, Z., and Wang, Z. (2021). Metagenomics of wastewater phageome identifies an extensively cored antibiotic resistome in a swine feedlot water treatment environment. *Ecotoxicology and Environmental Safety*, 222: 112552.
- [30] Radwan, O., and Ruiz, O.N. (2020). Shotgun metagenomic data of microbiomes on plastic fabrics exposed to harsh tropical environments. *Data in Brief*, 32: 106226.
- [31] موسوی ملکی م. س. (1391) بررسی تنوع میکروبی محیط اطراف معدن کرومیت فرومد در استان سمنان به روش متاژنومیکس، پایان‌نامه کارشناسی ارشد بیوتکنولوژی کشاورزی، دانشگاه پیام‌نور مرکز تهران شرق.
- [32] فروزان، ا. (1396) جستجوی ژن‌های مقاومت به فلز سنگین در متاژنوم معادن آهن استان سمنان، پایان‌نامه دکترای بیوتکنولوژی کشاورزی، پژوهشگاه ملی مهندسی ژنتیک و زیست‌فناوری.
- [33] Hemmat-Jou, M. H., Safari-Sinegani, A. A., Mirzaie-Asl, A., and Tahmourespour, A. (2018). Analysis of Microbial Communities in Heavy Metals-Contaminated Soils Using Metagenomic Approach. *Ecotoxicology*, 27(9):1281-1291.

and genomes from metagenomes using PhyloPhlAn 3.0. *Nature Communications*, 11, 2500.

[49] Markowitz, V.M., Chen, I.M., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., Jacob, B., Huang, J., Williams, P., Huntemann, M., Anderson, I., Mavromatis, K., Ivanova, N.N., Kyrpides, N.C. (2012). IMG: The Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Researches*, 40(Database issue): D115-22.

[50] Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T. Rodriguez, A. Stevens, R. Wilke, A. Wilkening, and J. Edwards, R.A. (2008). The metagenomics RAST

server –a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, 9: 386.

[51] Huson, D.H., Auch, A.F., Qi, J., Schuster, S.C. (2007). MEGAN analysis of metagenomic data.

Genome Research, 17(3): 377–86.

[52] Sharpton, T.J., Riesenfeld, S.J., Kembel, S.W., Ladau, J., O'Dwyer, J.P., Green, J.L., Eisen, J.A., Pollard, K.S. (2011). PhylOTU: a high-throughput procedure quantifies microbial community diversity and resolves novel taxa from metagenomic data. *PLoS Computational Biology*, 7(1): e1001061.

[53] Norouzi-Beirami, M.H., Marashi, S.A., Banaei-Moghaddam, A.M., Kavousi, K. (2021).

CAMAMED: a pipeline for composition-aware mapping-based analysis of metagenomic data. *NAR Genomics and Bioinformatics*, 3(1): lqaa107.

[54] Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research*, 40: 109-114.

[55] میرزایی اصل ا، گودرزی ص،، خدائی ل. (1394)

تجزیه داده‌های زیستی (ابزارهای داده‌پردازی زیستی)،

انتشارات دانشگاه بوعلی سینا، همدان، 304 ص.

[56] Luo, W., and Brouwer, C. (2013). Pathview: An R/bioconductor package for pathway-based data integration and visualization. *Bioinformatics*, 29: 1830-1831.

[57] Friedman, J., and Alm, E.J. (2012). Inferring correlation networks from genomic survey data. *PLOS Computational Biology*, 8: 1-11.

[58] Pir, P., Kirdar, B., Hayes, A., Onsan, Z.I., Ulgen, K., and Oliver, S. (2006). Integrative

Function of Pre-Agricultural Tallgrass Prairie Soils in the United States. *Science*, 342: 621-624.

[42] Gittel, A., Barta, J., Kohoutova, I., Mikutta R., Owens, S., Gilbert, J., Schneck, J., Wild, B., Hannisdal, B., Maerz, J., Lashchinskiy, N., Capek, P., Santruckova, H., Gentsch, N., Shibistova, O., Guggenberger, G., Richter, A., Torsvik, V., Schleper, C., and Urich, T. (2014). Distinct

microbial communities associated with buried soils in the Siberian tundra. *International Society for Microbial Ecology Journal*, 8: 841-853.

[43] Gibbons, S.M., Caporaso, J.G., Pirrung, M., Fiedle, D., Knight, R., and Gilbert, J.A. (2012).

Evidence for a persistent microbial seed bank throughout the global ocean. *Proceedings of the National Academy of Sciences of the United States of America. Ecology*, 110: 4651-4655.

[44] Karsenti, E., Acinas, S.G., Bork, P., Bowler, C., Vargas, C.D., Raes, J., Sullivan, M., Arendt, D., Benzon, F., Claverie, J.M., Follows, M., Gorsky, G., Hingamp, P., Iudicone, D., Jaillon, O., Kandels-Lewis, S., Krzic, U., Not, F., Ogata, H., Pesant, S., Reynaud, E.G., Sardet, C., Sieracki, M.E., Speich, S., Velayoudon, D., Weissenbach, J., and Wincker, P. (2011). A Holistic Approach to Marine Eco-

Systems Biology. *PLoS Biology*, 9: 1-5.

[45] Scholz, M., Tett, A., and Segata, N. (2015).

Computational Tools for Taxonomic Microbiome Profiling of Shotgun Metagenomes. In: Izard, J., and Rivera, M.C., editors. *Metagenomics for microbiology*. 3rd ed. Academic Press is an imprint of Elsevier, 32 Jamestown Road, London NW1 7BY, UK, pp. 67-80.

[46] Segata, N., Boernigen, D., Tickle, T.L., Morgan, X.C., Garrett, W.S., and Huttenhower, C. (2013). Computational meta'omics for microbial community studies. *Molecular System Biology*, 9: 666.

[47] Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W., and Nielsen, P.H. (2013). Genome sequences of rare, uncultured

bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnology*, 31(6): 533–8.

[48] Asnicar, F., Thomas, A.M., Beghini, F., Mengoni, C., Manara, S., Manghi, P., Zhu, Q., Bolzan, M., Cumbo, F., May, U., Sanders, J., Zolfo, M., Kopylova, E., Pasolli, E., Knight, R., Mirarab, S., Huttenhower, C., and Segata, N. (2020).

Precise phylogenetic analysis of microbial isolates

editors. Omics in Soil Science. 1st ed. Caister Academic Press, Portland, OR, pp. 179-187.

[61] VerBerkmoes, N.C., Denef, V.J., Hettich, R.L., and Banfield, J.F. (2009). Functional analysis of natural microbial consortia using community proteomics. *Nature Reviews Microbiology*, 7: 196-205.

investigation of metabolic and transcriptomic data. *BMC Bioinformatics*, 7: 1-12.

[59] McHardy, I., Goudarzi, M., Tong, M., Ruegger, P., Schwager, A.E., and Weger, J. (2013).

Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships. *Microbiome*, 2618: 1-17.

[60] Myrold, D.D., and Nannipieri, P. Classical techniques versus omics approaches. (2014). In: Nannipieri, P., Pietramellara, G., and Renella, G.,

Omics-Meta and its applications our of recognition to environment

Mohammad Hossein Hemmat-Jou¹, Davood Namdar Khojasteh^{2*}, Arezoo Tahmourespour³, Asghar Mirzaie-Asl⁴

1-Graduated student in soil biology and biotechnology, Department of Soil Science, Faculty of Agriculture, Bu-Ali Sina University of Hamedan, Iran.

2-Assistant Professor of Soil Science, Department of Soil science, Faculty of Agriculture, Shahed University, Tehran, Iran.

3-Associate Professor of Microbiology, Department of Basic Medical Sciences, Isfahan (Khorasgan) branch, Islamic Azad University, Isfahan, Iran.

4-Associate Professor of Biotechnology, Department of Agricultural Biotechnology, Faculty of Agriculture, Bu-Ali Sina University of Hamedan, Iran.

*Corresponding author: d.namdar@shahed.ac.ir

Received: 2020/11/24

Accepted: 2021/10/16

Abstract

The number of microbial cells on the planet is much larger than the stars we know in the galaxies. However, the microbial diversity and their ecological network remain unknown, they have key roles on the Earth's ecosystems. Omics technologies such as metagenomics provide tools for recognizing a large part of these cryptic forms of life accurately, which are much higher than the uncultivated majority. One example is the diversity of the *Vampyrellids* from *protista* and micro-eukaryotes. Using meta-omics technologies, it found that the diversity within this one group equals that of the entire kingdom of fungi, and they are found in all corners of nature, from the oceans to terrestrial soils. It is noticeable that they are only one of the seven *protista* groups. In this article, in addition to introducing Omics technologies, some of big relevant projects and their results have also been discussed covering all of the Earth's environment. Metagenomics is the direct sequencing and characterization of genes and genomes present in complex microbial ecosystems (e.g. metagenomes). Viromics is the research of viral metagenome. In metatranscriptomics the mRNA is being analyzed which is due to its notoriously labile nature in environmental samples, its conservation and analyzing are the main challenges in this omics. Identification and measurement of various proteins that can directly measure microbial activity is performed in metaproteomics. Environmental metabolomics includes the study of low molecular weight metabolites generated from interactions between microorganisms, such as small eukaryotes, plants, animals, predators, abiotic stresses, and other stimulants. The meta-omics approaches are rapidly developing and evolving technologies that are expanding our access to both biotic diversity and gene expression in all environments from the human body to the soils and the oceans. Eventually, we will understand the millions of years' ecosystem's rules in microbial world.

Keywords: Meta-omics, Metagenomics, Viromics, Metatranscriptomics, Metaproteomics, Metabolomics and Microbial Diversity.