

Mikrobiyal Ekolojiye Yolculuk: Metagenomik'e Genel Bakış ve Analiz Araçları

Gölnur Uzun^Y, Fatma Duran^E

Giriş

İlk defa **1998'de Handelsman** ve **çalışma arkadaşları** tarafından tanımlanan, doğrudan çevresel örneklerden elde edilen genetik materyallerin incelenmesi ile ilgilenen **metagenomik** kavramı; nükleotit dizisinin analizi vasıtasıyla çevresel örneklerde yaşayan karmaşık mikrop popülasyonunun araştırılmasında kullanılmaktadır¹. Günümüzde mikrobiyal ekolojinin gelişmeye devam eden önemli bir alanı olarak karşımıza çıkmaktadır²⁻⁷. Ana amacı spesifik bir ortamda bulunabilen mikrobiyal popülasyonu belirlemek olan metagenomik, mikrobiyal popülasyona ait çevresel bir örnekteki genetik materyalin incelenmesine atıfta bulunan teknik ya da teknikler bütünüdür. Laboratuvarıda kültür yapılmasına ihtiyaç duymadan mikrobiyal toplulukların taksonomik ve fonksiyonel profillerini ortaya çıkarabilen, yüksek verimli dizilemenin yeni bir metodolojisidir^{2,8-10}.

Bir ortamdaki mikroorganizmaları laboratuvara aktarma sırasında geleneksel tekniklerin yarattığı zorluklar sebebiyle genellikle **yeni nesil dizileme (NGS)** yöntemleri tercih edilmektedir⁸. **Bakteriler, arkeler, virüsler** ve **tek hücreli ökaryotlardan** oluşan çeşitli mikrobiyal topluluklar yüzyıllardır önemini korumaktadır ve 20. yy'ın başlarında ortaya çıkan metagenomik, günümüze kadar bilim dünyasının **tıp, biyoteknoloji, genetik, tarım** ve **gıda mikrobiyolojisi** gibi farklı fakat birbiri

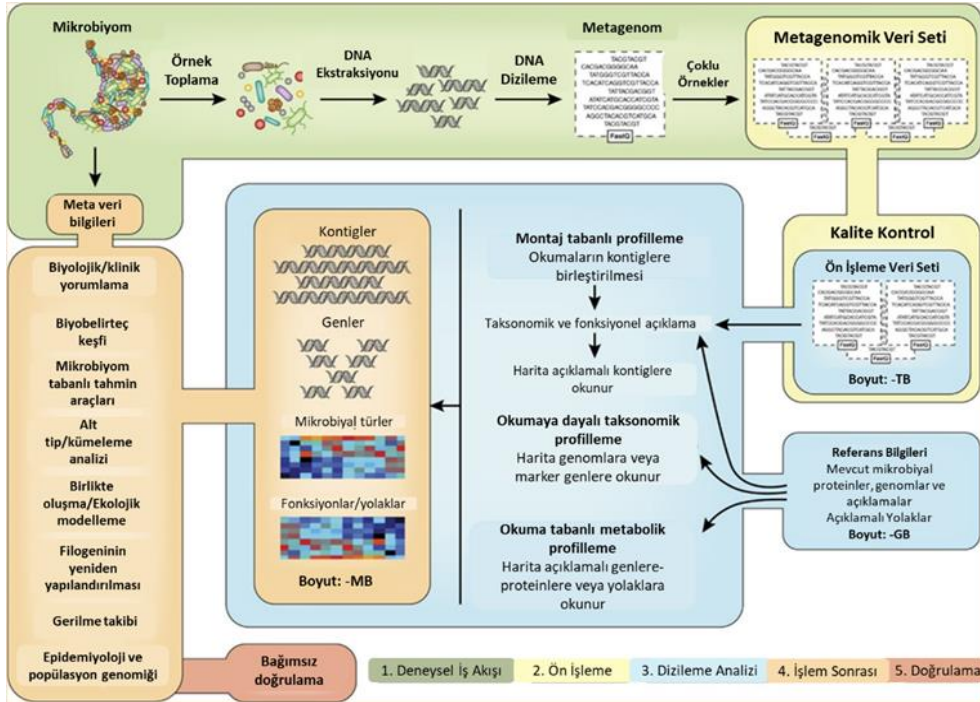
ile ilişkili alanlarında kendini göstermeye ve ilgisini artırmaya devam etmiştir^{8,11}. Moleküler biyolojideki gelişmeler ve mikrobiyal toplulukların anlaşılması konusunda yapılan çalışmalarda omik yöntemlerin geliştirilmesini de beraberinde getirmiştir^{3,12}. Metagenomik ve diğer omik disiplinler, gelecekte; **gıda güvenliği, kamu sağlığı** ve **klirik mikrobiyoloji** araştırmaları için çözüm potansiyeli bulundurmaktadır¹³.

Metagenomik, 5S ve 16S rRNA'nın DNA'sının üzerinde durulduğu **filogenetik bakış açısı** ve farklı enzim ve metabolik ürünlerin çalışıldığı **fonksiyonel bakış açısı** olmak üzere 2 temel bakış açısından anlaşılabilir. Bu yaklaşımlardan ilki bir ekosistem üyelerini tanımlamayı ikincisi ise işlevsel genlerin tanımlanmasını amaçlamaktadır^{8,14}. Nadir görülen, tanı konulması zor, komplike enfeksiyonların ve bulaşıcı hastalıkların anlaşılması konusunda gitgide daha çok tercih edilen bir teknik olarak karşımıza çıkan **metagenomik yeni nesil dizileme (mNGS)**, yalnızca bir dizilemeden klirik bir örnekteki tüm patojen türlerini tespit edebilen yüksek verimlilikte bir tekniktir. Tedavi seçenekleri için yol gösterici ve antibiyotik yöntemlerini iyileştirmek için tercih edilebilecek bilgiler sunmaktadır^{15,16}. Günümüzde metagenomik karaktere ilişkin pek çok çalışma yürütölmekte ve çalışmaların büyük çoğunluğu **tarım** ve **gıda** alanında süregelmektedir; bunun yanında

farmakoloji ve insan sağlığı ile alakalı gerçekleştirilen metagenomik çalışmalar zaman içinde artış göstermiştir⁸.

Metagenomik İş Akışı

Metagenomik iş akışı; **deneysel iş akışı, ön-ışleme, dizileme analizi, işlem sonrası ve doğrulama** aşamaları olmak üzere temelde 5 adımdan oluşmaktadır (**Şekil1**)¹¹.



Şekil 1. Metagenomik İş Akışı¹¹. **Adım(1):** Çalışma tasarımı ve deney protokolü. **Adım(2):** Hesaplamalı ön-ışleme. Kalite düzeltme, sıralama kopyalarının kaldırılması. Yabancı veya hedef olmayan DNA dizileri filtrelenmektedir ve takson veya fonksiyon çeşitliliği karşılaştırılırsa okunan sayıları normalleştirmek için örnekler alt örneklere ayrılmaktadır. **Adım(3):** Dizi analizi. Bu aşamada deneysel hedeflere bağlı olarak “okumaya dayalı” ve “montaja dayalı” yaklaşımların kombinasyonu uygulanmaktadır. **Adım(4):** İşlem sonrası. Verileri yorumlamak için farklı istatistiksel yöntemler kullanılmaktadır. **Adım(5):** Doğrulama. Yüksek boyutlu biyolojik verilerden elde edilen sonuçlar analizlerde büyük önem taşımaktadır.

Örnek Toplama, DNA Ekstraksiyonu ve Örnek İşleme Aşamaları

Örnek toplanması ve korunması süreçleri metagenomik verilerin doğruluğunu ve kalitesini direkt olarak etkilemektedir ve bu aşamada elde edilebilecek sonuçlar farklı durumlardaki biyolojik değişkenlerin etki büyüklüğünden daha büyük olabilmektedir. Örnek işleme prosedürlerindeki varyasyonlar farklı çalışmalardan elde edilen veri setlerinin metagenomik analizinde de örnek toplama ve

örneklerin korunmasında olduğu kadar önemlidir^{11,17-19}. Tüm örnek türleri için doğrulanmış toplama ve koruma tekniklerinin optimal olduğu varsayılmamaktadır; spesifik her örnek için farklılık gösterebilmektedir¹⁷. Bu sebeple, örnek türleri için optimize deney koşullarına ulaşabilme konusunda dikkatli bir ön çalışma gerekmektedir; bu aşamada örnek toplama ve saklama prosedürü ve koşulların kaydedilmesi gereken hususlardır. Temel hedefler, dizileme için yeterli mikrobiyal

biyokütleyle ulaşmak ve örneklerin kontaminasyon düzeyini en aza indirmektir. Zenginleştirme teknikleri mikroorga- nizmaların daha az olduğu ortamlar için tercih edilebilmektedir^{11,19,20}.

DNA ekstraksiyonu metagenomik iş akışında oldukça önemlidir ve bu aşamada tercih edilen tekniğin çeşitli mikrobiyal taksonlar için etkili olması gerekmektedir; aksi takdirde dizileme sonuçlarına sadece kolay parçalanmış ve mikroorganizmalardan türetilen DNA hakim olmaktadır. DNA parçalarının kısılmasına sebep olması sebebi ile mekanik lizise dayalı yöntemler verimlilikte farklılık göstermekte ve DNA kaybına yol açabilmektedir¹⁹⁻²¹.

Örnek işleme aşamalarında kit ya da laboratuvar reaktiflerinden kaynaklı değişken miktarlarda mikrobiyal kontaminantlarla karşılaşma sonucu kontaminasyon görülebilmektedir. Düşük biyokütle örneklerinden, örneğin deri sürüntüsü, alınan metagenomik veri setleri, bu probleme karşı hassasiyet taşımaktadır, bunun sebebi düşük düzeyde kontaminasyonla rekabet edebilecek daha az gerçek “**sinyal**” olmasıdır. Düşük biyokütleli numunelerle çalışan araştırmacılara **ultraclean reaktifleri** kullanmaları ve reaktiflerin örnek şablonu eklemeyen sekanslandığı “**boş**” dizileme kontrollerini de dahil etmeleri önerilmektedir^{11-13,19-21}.

Metagenomik Dizileme

Metagenomik dizileme, karmaşık mikrobiyal popülasyonların kültüre bağımlı olmadan çalışıldığı bir yaklaşımdır; genellikle *in silico* ile birleştirilmektedir²³. Mikrobiyal toplulukların

yüksek verimli sıralaması, konakçı mikrobiyom ile sağlık durumu arasındaki ilişkileri karakterize etmek, patojenleri tespit etmek ve bir organizmanın mikrobiyomunun yapıları çevre ile etkileşimini belirlemek için bir araç sağlamaktadır²⁴.

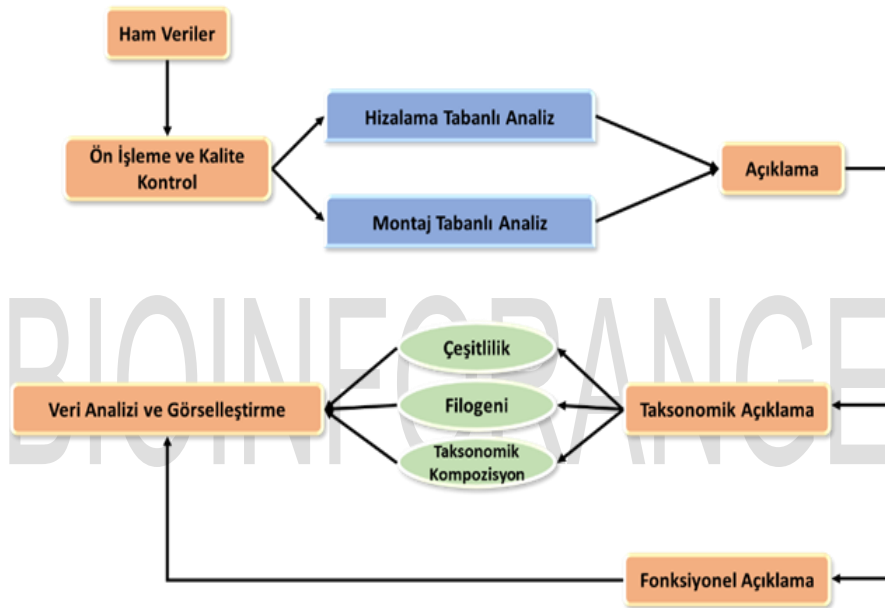
Metagenomikte örnekleme ve dizilemenin tanımına yönelik standartlar, daha geleneksel genomikteki standartlardan gelişmiştir. Bahsedilen standartlarda hala eksiklikler olsa da örnekleme ve dizilemeye ilişkin meta veriler genellikle metagenomik çalışma için toplanmaktadır²⁵.

Shotgun Metagenomik

Shotgun metagenomikte, mikrobiyal topluluklar, direkt DNA dizilimi temel alınarak incelenmektedir (**Şekil 2**). Üretilen sıralama okumalarından tahmin edilen gen bolluklarını karşılaştırarak, topluluklar arasındaki fonksiyonel farklılıklar tanımlanmaktadır. Bununla birlikte, gen bolluğu verileri, istatistiksel gücü büyük ölçüde azaltabilen ve yanlış pozitiflere yol açabilen yüksek düzeyde sistematik değişkenlikten etkilenmektedir. Bu nedenle sistematik değişkenliğin belirlendiği ve düzenlendiği süreç olan **normalizasyon** veri analizinin önemli bir adımıdır. Günümüze kadar yapılan çalışmalarda yüksek boyutlu sayım verileri için farklı normalizasyon yöntemleri önerilmiştir; ancak shotgun metagenomik dizileme verilerinin analizi üzerindeki performansları değerlendirilmemiştir²⁶.

Shotgun metagenomik dizileme mikroorganizmaların insan sağlığı ve çevre üzerindeki önemli işlevsel görevleriyle bağlantılı bir bakış açısı sağlamaktadır. Metagenomik verilerin yorumlanmasında temel analiz adımlarından biri; bireysel okumaları taksonomik sınıflandırma olarak adlandırılan spesifik taksonlara atamaktır^{27,28}. Büyük ölçekli **metagenomik montaj (metagenomic assembly)** çalışmalarının çoğunluğu insan bağırsağı gibi belli yaşam alanlarında bilinen

mikrobiyom çeşitliliğini büyük oranda genişleyen metagenom örneklerinden çok sayıda kültürlenmemiş yeni türü tekrar yapılandırmaktadır²⁷. Bahsedilen yeni kültürlenmemiş organizma konakçılarının sağlık durumundaki rolünü araştırmak için hedefe yönelik teknikler geliştirmek, mikrobiyom araştırmalarında önemli ve çözülmemiş bir sorun olmaya devam etmektedir²⁷⁻³¹.



Şekil 2. Shotgun Metagenomik Dizileme Analizi İş Akışı³².

Shotgun dizileme, hedeflenen dizileme teknikleri ile karşılaştırıldığında nispeten tarafsızdır. 16S ribozomal RNA dizileme ve 16S rRNA genine sahip olmayan ökaryotlar da dahil DNA tabanlı bir genoma sahip olan tüm türleri yakalayabilme avantajını sağlamaktadır²⁸ ve her taksona ait okuma sayısını belirleyerek orijinal örnekteki her taksonun yoğunluğunu tahmin etmede de kullanılabilir⁴. Shotgun metagenomik verilerin analizinde meydana gelebilecek zorluklara karşılık geliştirilen modüler ve açık

kaynaklı bir yazılımlardan biri olan **metaWRAP** (<https://github.com/bxlab/metaWRAP>), ham dizileme okumalarından başlayarak metagenomik **kutular (bins)** ve analizlerinde metagenomik verileri işlemek için güncel bir teknoloji kullanmaktadır. Metagenomik verilerden yüksek kaliteli kutuları ayıklamak ve iyileştirmek için çeşitli yazılımların güçlü yönlerinden yararlanan hibrit algoritmalar içermektedir. Kurulum ve kullanımı kolay olan metaWRAP, analiz sonuçları için kontrol gerektirebilmektedir^{29,31}.

Metagenomik Verilerin İstatistiksel Analizi

Dizi analizi temelde “okumaya dayalı” ve “montaja dayalı” yaklaşımlar olarak 2

yöntemle incelenmektedir. Okumaya ve montaja dayalı yaklaşımların birbirinden farklı özellikleri bulunmaktadır (Tablo 1)¹¹.

Tablo 1. Montaj Tabanlı Analiz ve Okumaya Dayalı Analiz Karşılaştırma Tablosu¹¹.

	Montaj Tabanlı Analiz	Okumaya Dayalı Analiz ('Eşleme')
Kapsamlılık	Sadece bir araya getirilip gruplanacak kadar yeterli kapsama sahip organizmalar için birden fazla tam genom oluşturabilmektedir.	Topluluk işlevinin veya yapısının toplu bir resmini sağlayabilmektedir, ancak sadece veritabanlarına başvurmak için etkili bir şekilde haritalanan okumaların bir kısmına dayanmaktadır.
Topluluk Karmaşıklığı	Karmaşık topluluklarda, genomların yalnızca bir kısmı montaj yoluyla çözülebilmektedir.	Yeterli sıralama derinliği ve tatmin edici referans veri tabanı kapsamı verildiğinde karmaşıklığa sahip topluluklarla başa çıkabilmektedir.
Yenilik	Dizilenmiş, akrabaları olmayan, tamamen yeni organizmaların genomlarını çözebilmektedir.	Yakın akrabalarının genomları bilinmeyen organizmaları çözmemektedir.
Hesaplama Yüğü	Hesaplama açısından maliyetli montaj, haritalama ve gruplama gerektirmektedir.	Verimli bir hesaplama gerçekleştirilebilmektedir ve büyük meta analizlere olanak tanımaktadır.
Genom Çözülmüş Metabolizma	Yeni çeşitlilik için bile tamamen birleştirilmiş genomlar aracılığıyla metabolizma ilişkilendirebilmektedir.	Tipik olarak yalnızca topluluğun toplam metabolizmasını çözebilmektedir ve filogeni ile bağlantılar sadece bilinen referans genomları bağlamında mümkündür.
Uzman Manuel Denetim	Doğru gruplama, yapı iskelesi ve yanlış montaj tespiti için manuel düzenleme gerektirmektedir.	Genellikle manuel iyileştirme gerektirmemektedir, ancak kullanılacak referans genomların seçimi insan gözetimi gerektirmektedir.
Mikrobiyal Genomik ile Entegrasyon	Mekanizmalar, saf kültürlenmiş izolatlardan genomların analizi için tasarlanmış mikrobiyal genomik boru hatlarına beslenebilmektedir.	Elde edilen profiller, saf kültürlenmiş izolatlardan türetilen genomların bağlamına doğrudan konulamamaktadır.

Metagenomik veri analizinin en kritik adımı, metagenomik yaklaşımlar ve dizileme araçları tarafından üretilen küçük dizi içeren DNA parçalarını bir araya getiren hesaplama programları kullanılarak topluluklardaki mikroorganizmaların bireysel genlerinin ve genomlarının yeniden yapılandırılmasıdır^{7,25,32,33}. Metagenomik veri kümelerinin tekrardan üretilebilirliği, kullanılabilirliği ve daha iyi paylaşımını sağlamak için standartlaştırılmış yöntemler önerilmektedir^{7,31}, ancak şu anda metagenomik veri setlerinin nasıl analiz edildiğini raporlamak için herhangi bir standart bulunmamaktadır^{25,32}.

Metagenomik analizlerde büyük miktarda veri hedeflenmekte ve yoğun bir hesaplama gerekmektedir. Yazılımların çoğunlukla sadece Linux sistemleri için bulunması ve büyük

miktarda bilgi işlem kaynağına ihtiyaç duyulması ise dezavantaj yaratmaktadır^{31,33}. Birbirinden bağımsız çalışmalardan elde edilen metagenomik veri analizleri sadece analiz iş akışları uyumlu bir biçimde açıklandığında karşılaştırılabilmektedir. Metagenomikteki temel adımlar; **malzeme örnekleme, malzeme dizileme, veri analizi ve veri arşivleme-yayınlama** şeklinde sıralanmaktadır²⁵. Birincil metagenomik dizi analizlerinde kullanılan tekniklerden bağımsız olarak, çıktılar örneklerin veri matrislerine karşı mikrobiyal özelliklerini içermektedir. İşlem sonrası analiz, bahsedilen matrisleri yorumlamak ve elde edilen bulguların örnek meta verilerle ne şekilde ilişkili olduğunu anlamlandırmak için istatistiksel araçlar barındırmaktadır. Analizlerde kullanılan

istatistiksel bakış açılarının çoğu metagenomik için spesifik olarak geliştirilmemiştir^{11,25}.

Metagenomik Analiz Araçları

Yalnızca **en düşük ortak atayı (LCA-Lowest Common Ancestor)** doğru bir şekilde bildiren **Kraken**, kullanıcının seçtiği herhangi bir türü içerebilecek, özel olarak oluşturulmuş bir veri tabanı kullanarak metagenomik örnekteki her okuma için bir taksonomi etiketi atamaya çalışmaktadır. Bu veri tabanı kullanıcı tarafından belirlenen bir genom kütüphanesine dayalı olarak oluşturulmuştur ve *k-mer* ile alakalı en spesifik taksonomik ağacın düğümünün eş zamanlı olarak aranmasına izin vermektedir. Sekans sınıflandırılması, *k-mer*'leri referans sekansla ilişkilendirmek için veri tabanının sorgulanması yolu ile gerçekleştirilmektedir. Elde edilen LCA verileri, sekansları benzerliğe göre sınıflandırma sırasında kullanılmaktadır. Veri tabanında tanımlanabilir *k-mer*'leri bulunmayan sekanslar, Kraken tarafından sınıflandırılmamış şekilde bırakılmaktadır. Her okumanın kesin olarak bir türe atanması mümkün olmasa da her türün bolluğunu saptamak adına **MetaPhlAn**, **ConStrains**, **GAAS**, **GASiC**, **TAEC** ve **GRAMMy** gibi çeşitli yazılım araçları geliştirilmiştir; fakat bu araçlar Kraken'in *k-mer* yaklaşımı kadar güvenilirlik taşımayan farklı stratejiler kullanmaktadır^{4,34,35}. Kraken'in *k-mer* tabanlı yaklaşımı, metagenomik dizileme verilerinin hızlı bir taksonomik sınıflandırmasını sağlamaktadır ancak büyük bellek gereksinimleri bazı uygulamalar için sınırlayıcı özellik

taşıyabilmektedir. Buna karşılık **Kraken2** bellek kullanımını %85 oranında azaltıp Kraken1'i daha kullanılabilir hale getirmektedir. Aynı zamanda yüksek doğruluk özelliğini korumakta, hızı 5 kat artırmakta ve daha fazla miktarda referans genomik verinin kullanılmasına olanak tanımaktadır. Buna ek olarak Kraken2, viral metagenomik analizde artan hassasiyet sağlayan bir arama modu sunmaktadır^{36,37}.

MetaWRAP veya **DASTool**, rafine gruplama sonuçları ve daha az kontaminasyonla daha eksiksiz genomlar elde etmek için birkaç gruplama yazılım paketini entegre etmektedir. Bu işlem hatları ayrıca değerlendirme ve görselleştirme için yararlı komut dosyaları sağlamaktadır³³.

BusyBee Web, metagenomik verilerin popülasyon düzeyinde çözümlenmiş analizlerini kolaylaştırmaktadır böylelikle çeşitli ortamlardan ve dizileme teknolojilerinden türetilen karmaşık mikrobiyal toplulukların incelenmesini sağlamaktadır³⁸.

HOME-BIO çeşitli referans veri tabanlarından yararlanmaktadır; bu sayede kullanıcıların biyolojik örneklerin içeriğini detaylı araştırmasına olanak tanımaktadır. Konakla ilişkili olan mikrobiyal popülasyonların bileşimini karakterize etmek için NGS verilerini işleyen HOME-BIO, sonuçların eş zamanlı ve daha kolay yorumlanabilmesi için tablolar ve çizelgeler üreten, hem metagenomik analizlerde kullanılan araçlar hem de özel **Python** kodları kullanılarak uygulanmaktadır. **Kalite Kontrol**, **Shotgun Metagenomik** ve

De-novo Montaj olmak üzere 3 ana bloktan oluşan modüler bir yapısı bulunmaktadır⁶.

Yaygın olarak kullanılan fonksiyonel profil oluşturma yazılımı **HUMAnN2**, numune içi ve numuneler arası katkı çeşitliliğini (türlerin belirli bir işleve katkıları) keşfetmek için de kullanılmaktadır. **MEGAN**, taksonomik ve işlevsel analizler gerçekleştiren platformlar arası bir **grafik kullanıcı arayüzü (GUI)** yazılımıdır. Bu özelleştirilmiş veri tabanları, uygun çalışma alanında taksonomik ve işlevsel açıklama için kullanılabilir ve verimli, kesin, hızlı analize olanak tanımaktadır. **AntiSMASH** veri tabanı, **ikincil metabolit biyosentezinde** yer alan gen kümelerini tanımlamak, açıklama eklemek ve görselleştirmek için kullanılmaktadır. Mevcut gruplandırma araçları arasında **CONCOCT**, **MaxBin 2** ve **MetaBAT2** yer almaktadır. **MetaPhlAn2**, taksonomik sınıflandırma gerçekleştirmek için metagenom okumalarını önceden tanımlanmış bir işaretleyici gen veritabanına hizalayan, yaygın olarak kullanılan bir taksonomik profil oluşturma aracıdır³³.

Genom, transkriptom ve epigenom için olanlar da dahil olmak üzere diğer yüksek verimli dizileme analizlerinde, keşif amaçlı hesaplamalı ve görsel analizi entegre eden yeni nesil genom tarayıcılarının etkili analiz araçları olduğu kanıtlanmıştır. Bununla birlikte, mikrobiyom verileri için keşif analiz araçları daha azdır bu durum kısmen mikrobiyom özelliklerinin, ölçüm ve analiz birimlerinin taksonomik bir hiyerarşide organize

edilmesinin zorluğundan kaynaklanmaktadır. Metagenomik özelliklerin etkili etkileşimli keşfi, analizi ve veri görselleştirmesi için tercih edilen **Metaviz** aracı uzunlamasına veri kümelerinin analizinde de tercih edilmektedir. Metaviz etkileşimli görselleştirme aracıdır ve **R/Bioconductor** paketleri kullanılarak hesaplamalı ve istatistiksel analizlerin entegrasyonunu sağlamaktadır²⁴.

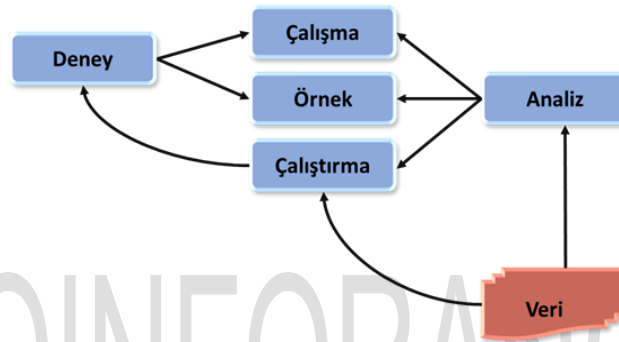
Metagenomikte yaygın olarak kullanılan bir başka platform ise entegre analiz araçlarını birlikte kullanan **Mothur**'dur. Buna ek olarak Mothur C++ ile yazılmış olması sebebi ile daha fazla sayıda dizi analiz edilebilmektedir. **MEGAN (Metagenome Analyzer)** programı, metagenomların taksonomik ve fonksiyonel analizi için çok önemli bir hesaplama aracıdır. Avantajı, taksonomik analizin metagenom düzeneği gerektirmemesidir; çünkü program okuma veya bitiş dizileriyle çalışmaktadır. Çalışması, bir hizalama aracı olan **BLAST** aracılığıyla okumaların bilinen dizilerin veritabanlarıyla karşılaştırıldığı bir ön işleme aşamasına dayanmaktadır. Çıktı, sonuçları özetlemek ve sıralamak için NCBI tabanlı veri kümesinin taksonomik içeriğini etkileşimli olarak tahmin eden ve kullanan MEGAN'da görselleştirilmektedir. Program, bir referans takson ile okumanın her isabetinden her okumaya LCA atayan basit bir algoritma kullanmaktadır. Sonuç olarak, uç dallarda türe özgü dizilerin gruplandığı bir ağaçtır. Üretilen ağaç düğümleri, farklı taksonomik seviyeleri özetlemek için daraltılabilmekte veya genişletilebilmektedir. Son zamanlarda bu araç, **InterPro2GO**, gen merkezli okuma

derlemesi ve taksonomi ve fonksiyon özelliklerinin ana koordinat analizini birleştirerek fonksiyonel analizin entegrasyonuna izin vermiştir. Bu yeni sürüm **MEGAN Community Edition (MEGAN-CE)** adı verilmektedir ve açık kaynaklıdır³⁴.

METAnnotatorX2, kısa ve uzun okumaların karıştırılmamış ve hibrit düzenekleri için eksiksiz bir iş akışı sağlayarak, metagenomik verilerden mikrobiyal genomların türe özgü yeniden yapılandırılmasını ve gen

açıklamasını sağlamaktadır. Aynı zamanda bir dizi gelişmiş özellik içeren ve **arke, bakteri, mantar, protista ve virüsler** için özelleştirilmiş veri tabanları barındırmaktadır³⁹.

NGS teknolojilerinin gelişmesi ile birlikte büyük ölçekte dizileme verileri üretilmesi de sağlanmış, meydana gelen DNA dizileri ve birbiri ile ilişkili bilgiler aşağıdaki kavramları temsil eden bir süreci beraberinde getirmiştir (**Şekil 3**)^{25,39,40}.



Şekil 3. Okuma verileri ve ilişkili meta veriler için ortak bir veri modeli²⁵. Okuma verileri ve ilişkili meta veriler için ortak bir veri modeli. **1) Çalışma:** projedeki tüm verilerin bir araya geldiği bir dizileme kapsamı hakkında bilgi. **2) Örnek:** Dizilenen örneklerin menşei ve özellikleri hakkında bilgi. **3) Deney:** Kitaplık ve araç ayrıntıları da dahil, dizileme deneyleri hakkında bilgiler. **4) Çalıştırma:** Veri dosyalarında temsil edilen dizileme okumalarını içeren bir dizileme deneyinin çıktısı. **5) Analiz:** Dizi düzenekleri, fonksiyonel ve taksonomik ek açıklamalar da dahil birincil dizileme sonuçlarından hesaplanan bir dizi çıktı.

Metagenomikte Zorluklar, Sınırlamalar ve Gelişim Potansiyelleri

NGS teknolojilerinin azalan maliyeti, karmaşık popülasyonların analizini kolaylaştırmış; büyük ölçekli verilerin üretildiği toprak üzerindeki metagenomik çalışmalar mümkün hale gelmiştir. Bahsedilen avantajlarla beraber metagenomik veri üretmenin kolaylığı, iş akışı ve analizi için yeni zorluklar ön plana çıkmıştır². NGS teknolojisinde gerçekleşen büyük ilerlemeye karşın sürekli güncellenmekte olan veri tabanlarına, özel biyoenformatik araçlara

ve işlevsel korelasyonlara duyulan ihtiyaç ile alakalı teknik sorunlar hala görülmektedir. Dizileme teknolojilerindeki gelişmeler iş hacmini artırması ve dizileme maliyetini büyük oranda düşürmesi sebebi ile kullanım, analiz ve depolanma konusu göz önüne alındığında dikkatli değerlendirme yapılmasını mecbur kılan büyük miktarda veri birikmektedir^{36,37,40}.

Uzmanlık, programlama, tekrarlı analiz, dizi tekrarlarının ortadan kaldırılması ve veri akışları biyoenformatik yaklaşımların kullanılmasındaki kalıcı zorluklar olarak

tanımlanmaktadır. Dizilemede meydana gelen gelişmeler yüksek tanımlı transkripsiyonlar elde edilmesini sağlasa da bozulmaya uğramamış ortamların zengin mikrobiyal kaynakların metagenomunun, biyoaktivitelere sahip yeni enzimlerin biyolojik araştırılması için kullanılması gerekmektedir buna ek, metagenomik bakış açısı, iklim değişikliği ve bunun toprak üzerinde oluşturduğu etkileri anlamada mikrobiyomun faydalı özelliklerine dikkat çekerek olumsuz sonuçları en aza indirme fırsatlarını iyileştiren toprak mikrobiyomu konusunda büyük ölçekli bir bilgi sağlayacaktır¹⁸.

Yeni araçlar yaratıldıkça ve yeni keşifler yapıldıkça dizileme verilerinin daha önce bilinmeyen ve kullanılmadıkları şekillerde kullanılabilmesi metagenomiklerin avantajlarından biri olarak değerlendirilmektedir. Buna ek olarak uzun okumalar veren tek molekül dizileme gibi yeni dizileme teknolojileri gelişerek ortaya çıkmakta ve yerel araçlar yeni veri türleri ile karşılaştırılabilmektedir. Konu hakkında yapılan detaylı araştırmalar ve klinik laboratuvarlar dizileme verilerine gitgide daha bağımlı oldukça bu konuda çalışma yapan bilim insanları ve biyoenformatikçiler de daha önemli hale gelmektedir⁴⁰.

Klinik ve gıda mikrobiyolojisi laboratuvarlarında metagenomik yaklaşımının kullanımı, klinik metagenomiklerin tanısal olarak uygulanabilirliğindeki eksiklikler sebebiyle sınırlanmaktadır. Düzenlenmiş ve güncel veri tabanlarının eksikliğinde, doğru teşhis yapmak

mümkün olmayacak buna bağlı olarak yanlış teşhis olması halinde araştırma sonucuna zarar vererek tutarsız bulgular elde edilecektir. Mikrobiyal teşhis için temelde biyoenformatik ya da hesaplamalı biyoloji uzmanlığı önemlidir fakat bunun yanında çeşitli analiz platformları ve farklı yazılımların önemi de yadsınamayacak kadar büyüktür. Metagenomik teknolojisinin pratik kullanımı, verilerin hatasız ve klinik olarak hayata geçirilebilir bir sürede analiz etmenin farklı hesaplama ve teknik sınırlılıkları sebebiyle engellenmektedir. Hesaplama ile ilgili performansların zorluğunun yanında numune hazırlama ve dizileme prosedürlerinin de önemi büyüktür ve standartlaştırılmaları büyük bir ihtiyaçtır. Uygulanan dizileme yönteminin tercihi, her yöntemin avantaj ve dezavantajları dikkate alınmaktadır ve önemli ölçüde araştırma ve tanı hedeflerine bağımlıdır¹³.

Son yıllarda evrimsel tarihin, yapısal, fonksiyonel ve ekolojik biyoçeşitliliğin anlaşılmasında çevresel örneklerin kullanıldığı genomik çalışmalar kayda değer bir zemin oluşturmaktadır. NGS platformları ile verilerin genişliği ve sayısındaki artışlar sebebiyle, ekolojik çalışmalar büyük hacimli sekans verilerinin kullanımına yönlendirilmektedir. Bu yaklaşımlar PCR'da hata ve sapma bulgularının azaltılmasına yardımcı olmuş ve çalışma sürelerinde önemli oranda azalma gözlemlenmiştir. Son zamanlarda avantaj ve dezavantajlara sahip çok sayıda platform keşfedilmiştir ve bu platformların hazırlanmasında farklı şablon ve sıralama

sinyali algılamada farklı yöntemler tercih etmektedir³.

Metagenomik veri kümelerinin nasıl analiz edildiği, veri analizi çıktılarının nasıl arşivlenmesi ve yayınlanması gerektiği ile ilgili standartların eksikliği günümüzde hala vurgulanmakta olan bir problem olarak karşımıza çıkmaktadır^{7,25,32,39}.

Metagenom düzeneklerinin zorlukları; montaj sürecinin zorlukları, montaj sonuçları oluşturmanın zorlukları ve sonuç kullanımının zorlukları olmak üzere 3 ana başlıkta incelenmektedir⁷.

Nanopore ve **PacBio** gibi uzun okumalar oluşturan yeni DNA dizileme teknolojilerinin yakın zamanda piyasaya sürülmesi, kısa okuma verilerinin yanı sıra entegre hibrit analiz gerçekleştirmek için özel araçların eksikliğinden hâlâ muzdarip olan değerli bir veri kaynağını temsil etmektedir. Metagenomların fonksiyonel profillemesi ve hibrit montaj yaklaşımlarına yönelik artan bilimsel ilgiye rağmen, rehberli veri analizi için kapsamlı ve kullanıma hazır bir biyoinformatik platform hala bulunmamaktadır. Şu an analizlerde kullanılan yazılım araçlarının çoğu bilgi işlem sürelerini en aza indirmek için milyonlarca okuma girdisine gerek duymakta ve yaklaşık değerlere güvenmektedir. Bahsedilen sorunlar okuma analizi yoluyla taksonomik veya işlevsel profillerin rekonstrüksiyonu ya da metagenomik düzenek tarafından tekrardan oluşturulan genomların analizi yoluyla kullanıldığı zaman doğruluk,

duyarlılık ve özgüllük gibi kavramlarca yeterli görülmeyen sonuçlara sebep olmaktadır^{32,39}.

Metagenomik, kompleks vakalarda ve pandemilerde geleneksel tanının etkileyici bir tamamlayıcısı olarak karşımıza çıksa da metagenomiklerin yaygınlaşp benimsenmesinden önce açıklanması gereke pek çok etken vardır bu sebeple biyolojik örneklerin yüksek verimli dizilemesinin umut vadeden bir geleceği var olsa da kullanımının standart ve onaylanmış bir **Kültürden Bağımsız Tanı Testleri (Culture-Independent Diagnostic Tests)** tekniği durumuna gelmesi için zaman ve araştırma ihtiyacı söz konusudur¹¹.

Son geliştirilen analiz araçları, çok değişkenli istatistikler ve makine öğrenmesi algoritmalarını içermektedir. Denetimsiz yöntemler, basit kümeleme ve örneklerin korelasyonunu, ısı haritalarını, sıralama veya verilerde modellerin grafiksel şekilde ortaya çıkarılmasına olanak tanıyan ağlar gibi görselleştirme tekniklerini içermektedir. Denetimli yöntemler, gruplar arası farklılıkların doğrudan hipotez testi için **çok değişkenli varyans analizi (ANOVA)** gibi istatistiksel yöntemleri ya da rastgele ormanlar veya destek vektör makineleri (NS) gibi örnek gruplarını etiketlemek için modelleri eğiten makine öğrenmesi sınıflandırıcıları içermektedir. Denetimsiz ve denetimli yöntemler popülasyonu bir bütün olarak incelemektedir. Tamamlayıcı bir strateji, örnek türleri ya da hasta gruplar arasında hangi spesifik taksonların veya fonksiyonel genlerin

istatistiksel olarak farklı olduğunu sormaktır. Çoğu durumda uygulama sonrası sonuçları görselleştirmek için özel grafik araçları ve görselleştirme yaklaşımları gerekmektedir^{11,40,41}.

Biyolojik veriler genel olarak karmaşık, yorumlanması güç ve heterojen olduklarından

Derin Öğrenme (DL) algoritmalarını temel alan tekniklerin kullanılması iyi bir test ortamını da beraberinde getirmektedir. DL yaklaşımları daha iyi performans elde etme ya da anlamlı biyobelirteçleri tespit etme konusunda metagenomikte bir avantaj gösterme noktasında başarısız olsa da umut vadeden girişimler son zamanlarda ortaya çıkmıştır^{42,43}.

BIOINFORANGE

Referanslar:

- Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chemistry and Biology*, 5(10). [https://doi.org/10.1016/S1074-5521\(98\)90108-9](https://doi.org/10.1016/S1074-5521(98)90108-9)
- Leite, M. F. A., van den Broek, S. W. E. B., & Kuramae, E. E. (2022). Current Challenges and Pitfalls in Soil Metagenomics. *Microorganisms*, 10(10), 1–8. <https://doi.org/10.3390/microorganisms10101900>
- Fadiji, A. E., & Babalola, O. O. (2020). Metagenomics methods for the study of plant-associated microbial communities: A review. *Journal of Microbiological Methods*, 170(February), 105860. <https://doi.org/10.1016/j.mimet.2020.105860>
- Lu, J., Breitwieser, F. P., Thielen, P., & Salzberg, S. L. (2017). Bracken: Estimating species abundance in metagenomics data. *PeerJ Computer Science*, 2017(1), 1–17. <https://doi.org/10.7717/peerj-cs.104>
- Harris, Z. N., Dhungel, E., Mosior, M., & Ahn, T. H. (2019). Massive metagenomic data analysis using abundance-based machine learning. *Biology Direct*, 14(1), 1–13. <https://doi.org/10.1186/s13062-019-0242-0>
- Ferravante, C., Memoli, D., Palumbo, D., Ciaramella, P., Di Loria, A., D'Agostino, Y., Nassa, G., Rizzo, F., Tarallo, R., Weisz, A., & Giurato, G. (2021). HOME-BIO (sHOTgun MEtagenomic analysis of BIOlogical entities): a specific and comprehensive pipeline for metagenomic shotgun sequencing data analysis. *BMC Bioinformatics*, 22(Suppl 7), 1–10. <https://doi.org/10.1186/s12859-021-04104-y>
- Lapidus, A. L., & Korobeynikov, A. I. (2021). Metagenomic Data Assembly – The Way of Decoding Unknown Microorganisms. *Frontiers in Microbiology*, 12(March). <https://doi.org/10.3389/fmicb.2021.613791>
- Garrido-Cardenas, J. A., & Manzano-Agugliaro, F. (2017). The metagenomics worldwide research. *Current Genetics*, 63(5), 819–829. <https://doi.org/10.1007/s00294-017-0693-8>
- Borroni, D., Romano, V., Kaye, S. B., Somerville, T., Napoli, L., Fasolo, A., Gallon, P., Ponzin, D., Esposito, A., & Ferrari, S. (2019). Metagenomics in ophthalmology: Current findings and future perspectives. *BMJ Open Ophthalmology*, 4(1). <https://doi.org/10.1136/bmjophth-2018-000248>
- Ramachandran, P. S., & Wilson, M. R. (2020). Metagenomics for neurological infections — expanding our imagination. *Nature Reviews Neurology*, 16(10), 547–556. <https://doi.org/10.1038/s41582-020-0374-y>
- 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. <https://doi.org/10.1038/nbt.3935>
- Alawiye, T. T., & Babalola, O. O. (2019). Bacterial diversity and community structure in typical plant rhizosphere. *Diversity*, 11(10). <https://doi.org/10.3390/d11100179>
- Forbes, J. D., Knox, N. C., Ronholm, J., Pagotto, F., & Reimer, A. (2017). Metagenomics: The next culture-independent game changer. *Frontiers in Microbiology*, 8(JUL), 1–21. <https://doi.org/10.3389/fmicb.2017.01069>
- Wang, W. L., Xu, S. Y., Ren, Z. G., Tao, L., Jiang, J. W., & Zheng, S. Sen. (2015). Application of metagenomics in the human gut microbiome. *World Journal of Gastroenterology*, 21(3), 803–814. <https://doi.org/10.3748/wjg.v21.i3.803>
- Lin, C., Bonsu, A. A. F. K., Li, J., Ning, C., Chen, L., Zhu, S., Zhong, Q., Shen, D., & Huang, G. (2022). Application of metagenomic next-generation sequencing for suspected infected pancreatic necrosis.

- Pancreatology, 22(7), 864–870.
<https://doi.org/10.1016/j.pan.2022.07.006>
16. Fomsgaard, A. S., Rasmussen, M., Spiess, K., Fomsgaard, A., Belsham, G. J., & Fonager, J. (2022). Improvements in metagenomic virus detection by simple pretreatment methods. *Journal of Clinical Virology Plus*, 2(4), 100120.
<https://doi.org/10.1016/j.jcvp.2022.100120>
17. Tian, L., Heymsfield, G., & Li, L. (2005). Observations of melting layer from 94 GHz and 10 GHz airborne Doppler radars. 11th Conference on Mesoscale Processes and the 32nd Conference on Radar Meteorology, 2009–2011.
<https://doi.org/10.1101/q.r.151803.112.Freely>
18. Azeem, M., Soundari, P. G., Ali, A., Tahir, M. I., Imran, M., Bashir, S., Irfan, M., Li, G., Zhu, Y. G., & Zhang, Z. (2022). Soil metaphenomics: a step forward in metagenomics. *Archives of Agronomy and Soil Science*, 68(12), 1645–1663.
<https://doi.org/10.1080/03650340.2021.1921743>
19. Felczykowska, A., Krajewska, A., Zielińska, S., & Łos, J. M. (2015). Sampling, metadata and DNA extraction - Important steps in metagenomic studies. *Acta Biochimica Polonica*, 62(1), 151–160.
<https://doi.org/10.18388/abp.2014.916>
20. Marotz, C., Amir, A., Humphrey, G., Gaffney, J., Gogul, G., & Knight, R. (2017). DNA extraction for streamlined metagenomics of diverse environmental samples. *BioTechniques*, 62(6), 290–293.
<https://doi.org/10.2144/000114559>
21. Bramucci, A. R., Focardi, A., Rinke, C., Hugenholtz, P., Tyson, G. W., Seymour, J. R., & Raina, J.-B. (2021). Microvolume DNA extraction methods for microscale amplicon and metagenomic studies. *ISME Communications*, 1(1), 1–5.
<https://doi.org/10.1038/s43705-021-00079-z>
22. Motley, S. T., Picuri, J. M., Crowder, C. D., Minich, J. J., Hofstadler, S. A., & Eshoo, M. W. (2014). Improved multiple displacement amplification (iMDA) and ultraclean reagents. *BMC Genomics*, 15(1), 1–10.
<https://doi.org/10.1186/1471-2164-15-443>
23. Singh, N., Singh, V., Rai, S. N., Mishra, V., Vamanu, E., & Singh, M. P. (2022). Deciphering the gut microbiome in neurodegenerative diseases and metagenomic approaches for characterization of gut microbes. *Biomedicine and Pharmacotherapy*, 156, 113958.
<https://doi.org/10.1016/j.biopha.2022.113958>
24. Wagner, J., Chelaru, F., Kancherla, J., Paulson, J. N., Zhang, A., Felix, V., Mahurkar, A., Elmqvist, N., & Bravo, H. C. (2018). Metaviz: interactive statistical and visual analysis of metagenomic data. *Nucleic Acids Research*, 46(6), 2777–2787.
<https://doi.org/10.1093/nar/gky136>
25. ten Hoopen, P., Finn, R. D., Bongo, L. A., Corre, E., Fosso, B., Meyer, F., Mitchell, A., Pelletier, E., Pesole, G., Santamaria, M., Willassen, N. P., & Cochrane, G. (2017). The metagenomic data life-cycle: Standards and best practices. *GigaScience*, 6(8), 1–11.
<https://doi.org/10.1093/gigascience/gix047>
26. Pereira, M. B., Wallroth, M., Jonsson, V., & Kristiansson, E. (2018). Comparison of normalization methods for the analysis of metagenomic gene abundance data. *BMC Genomics*, 19(1), 1–17.
<https://doi.org/10.1186/s12864-018-4637-6>
27. Liang, Q., Bible, P. W., Liu, Y., Zou, B., & Wei, L. (2020). DeepMicrobes: taxonomic classification for metagenomics with deep learning. *NAR Genomics and Bioinformatics*, 2(1), 1–13.
<https://doi.org/10.1093/nargab/lqaa009>
28. Cheng, W. Y., Liu, W.-X., Ding, Y., Wang, G., Shi, Y., Chu, E. S. H., Wong, S., Sung, J. J. Y., & Yu, J. (2022). High Sensitivity of Shotgun Metagenomic Sequencing in Colon Tissue Biopsy by Host DNA Depletion. *Genomics, Proteomics & Bioinformatics*.

- <https://doi.org/10.1016/j.gpb.2022.09.003>
29. Xu, Y., & Zhao, F. (2018). Single-cell metagenomics: challenges and applications. *Protein and Cell*, 9(5), 501–510. <https://doi.org/10.1007/s13238-018-0544-5>
 30. Chiang, A. D., & Dekker, J. P. (2020). From the pipeline to the bedside: Advances and challenges in clinical metagenomics. *Journal of Infectious Diseases*, 221(3), S331–S340. <https://doi.org/10.1093/infdis/jiz151>
 31. Uritskiy, G. V., DiRuggiero, J., & Taylor, J. (2018). MetaWRAP— a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome*, 6(1), 1–13. <https://doi.org/10.1186/s40168-018-0541-1>
 32. Gao, B., Chi, L., Zhu, Y., Shi, X., Tu, P., Li, B., Yin, J., Gao, N., Shen, W., & Schnabl, B. (2021). An introduction to next generation sequencing bioinformatic analysis in gut microbiome studies. *Biomolecules*, 11(4), 1–22. <https://doi.org/10.3390/biom11040530>
 33. Liu, Y. X., Qin, Y., Chen, T., Lu, M., Qian, X., Guo, X., & Bai, Y. (2021). A practical guide to amplicon and metagenomic analysis of microbiome data. *Protein and Cell*, 12(5), 315–330. <https://doi.org/10.1007/s13238-020-00724-8>
 34. Almeida, O. G. G., & De Martinis, E. C. P. (2019). Bioinformatics tools to assess metagenomic data for applied microbiology. *Applied Microbiology and Biotechnology*, 103(1), 69–82. <https://doi.org/10.1007/s00253-018-9464-9>
 35. Lindgreen, S., Adair, K. L., & Gardner, P. P. (2016). An evaluation of the accuracy and speed of metagenome analysis tools. *Scientific Reports*, 6, 1–14. <https://doi.org/10.1038/srep19233>
 36. Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, 20(1), 1–13. <https://doi.org/10.1186/s13059-019-1891-0>
 37. D'Argenio, V. (2018). Human microbiome acquisition and bioinformatic challenges in metagenomic studies. *International Journal of Molecular Sciences*, 19(2), 1–12. <https://doi.org/10.3390/ijms19020383>
 38. Laczny, C. C., Kiefer, C., Galata, V., Fehlmann, T., Backes, C., & Keller, A. (2017). BusyBee Web: Metagenomic data analysis by bootstrapped supervised binning and annotation. *Nucleic Acids Research*, 45(W1), W171–W179. <https://doi.org/10.1093/nar/gkx348>
 39. Hildebrand, F. (2021). Ultra-resolution Metagenomics: When Enough Is Not Enough. *MSystems*, 6(4), 1–5. <https://doi.org/10.1128/msystems.00881-21>
 40. Mulcahy-O'Grady, H., & Workentine, M. L. (2016). The challenge and potential of metagenomics in the clinic. *Frontiers in Immunology*, 7(FEB), 1–8. <https://doi.org/10.3389/fimmu.2016.00029>
 41. Sudarikov, K., Tyakht, A., & Alexeev, D. (2017). Methods for The Metagenomic Data Visualization and Analysis. *Current Issues in Molecular Biology*, 24, 37–58. <https://doi.org/10.21775/cimb.024.037>
 42. Fioravanti, D., Giarratano, Y., Maggio, V., Agostinelli, C., Chierici, M., Jurman, G., & Furlanello, C. (2018). Phylogenetic convolutional neural networks in metagenomics. *BMC Bioinformatics*, 19(Suppl 2), 1–13. <https://doi.org/10.1186/s12859-018-2033-5>
 43. Kukkar, D., Sharma, P. K., & Kim, K. H. (2022). Recent advances in metagenomic analysis of different ecological niches for enhanced biodegradation of recalcitrant lignocellulosic biomass. *Environmental Research*, 215(P2), 114369. <https://doi.org/10.1016/j.envres.2022.114369>