



## **WAWASAN BIOINFORMATIKA *NEXT GENERATION* SEQUENCING DALAM SAMPEL METAGENOMIK MIKROBIOMA USUS**

**Alfria Margaret<sup>a</sup>**

<sup>a</sup>Program Studi Teknologi Laboratorium Medis, Sekolah Tinggi Ilmu Kesehatan Kesetiakawanan Sosial Indonesia, Jakarta, Indonesia  
e-mail : [alfrianamargareta@stikeskesosi.ac.id](mailto:alfrianamargareta@stikeskesosi.ac.id)  
No Telepon WA : 081394471726

### **ABSTRACT**

*The gut microbiota is composed of bacteria, fungi and viruses that play an important role in human health and pathogenesis. The intestinal bacterial community has been studied more widely than fungi and viruses because the gut is dominated by bacteria. Although, the diversity of intestinal fungal is lower than that of the bacterial community, it is known that fungi also play an important role in human intestinal disease. Several studies have demonstrated the involvement of fungi in gastrointestinal and lung tumors. It is known, the DNA of tumor-associated fungi can serve as a diagnostic or prognostic biomarker. Culture methods to study fungal communities are challenging to implement. Next generation sequencing (NGS) is a widely appropriate sequencing technique for studying the gut microbiome. NGS data can be analyzed based on 18S rRNA, internally transcribed spacer (ITS) and shotguns metagenomics which require bioinformatics insight in operating the data. Therefore, bioinformatics tools were summarized for the metagenomic analysis of fungi. This research used the literature review method by reviewing 65 NCBI and Frontiers journals from 2000 to 2022. The results of this review journal provided information about bioinformatics tools for metagenomic fungal analysis. Fungal metagenomic analysis based on amplicon and shotgun sequencing can help researchers in studying fungal profiles. In addition, it can also predict particular genes or proteins that play a role in specific pathways so that it can be useful for studying a disease.*

**Keywords:** microbiota, fungi; 18S rRNA, ITS, shotgun, next generation sequencing

### **ABSTRAK**

Mikrobiota usus terdiri dari bakteri, fungi dan virus yang memiliki peran penting dalam kesehatan manusia dan patogenesis penyakit. Komunitas bakteri usus telah banyak dipelajari dibandingkan fungi dan virus karena usus didominasi oleh bakteri. Meskipun, keragaman fungi atau mikrobiom usus lebih rendah dibandingkan dengan komunitas bakteri, diketahui fungi juga berperan penting dalam penyakit usus manusia. Beberapa penelitian menunjukkan keterlibatan fungi pada tumor gastrointestinal dan paru-paru. Diketahui, DNA fungi terkait tumor tersebut dapat berfungsi sebagai biomarker diagnostik atau prognostik. Metode kultur untuk mempelajari komunitas fungi secara komprehensif sulit untuk dilakukan. *Next generation sequencing* (NGS) merupakan teknik sekuensing yang dapat diterapkan secara luas untuk mempelajari mikrobioma usus. Data NGS dapat dianalisis berdasarkan marker-gene 18S rRNA, internal transcribed spacer (ITS) dan shotgun metagenomik yang membutuhkan wawasan bioinformatik didalam mengoperasikan datanya. Oleh karena itu, pada ulasan ini, dirangkum *pipeline* bioinformatika yang umum digunakan untuk analisis data mikrobioma usus. Penelitian ini menggunakan metode *literature review* dengan cara mengulas jurnal-jurnal NCBI dan Frontiers sebanyak 65 jurnal dari tahun 2000 sampai dengan 2022. Hasil penelitian ini memberikan

*Received December 30, 2022; Revised January 15, 2023; Accepted January 27, 2023*

informasi tentang wawasan bioinformatika dalam analisis metagenomik fungi. Analisis metagenomik fungi berdasarkan amplikon dan *shotgun* memudahkan peneliti dalam mempelajari taksonomi dan fungsional fungi. Selain itu, juga dapat memprediksi gen atau protein tertentu sehingga dapat berguna untuk mempelajari suatu penyakit.

**Kata Kunci:** mikrobiota, fungi, 18S rRNA, ITS, *shotgun*, *next generation sequencing*

## 1. PENDAHULUAN

Mikrobiota merupakan komunitas sel mikroba yang kompleks, terdiri dari bakteri, archaea, virus, protista dan fungi (Pérez, 2021). Kompleksitas mikroba berdampak besar bagi kesehatan inang (Gorkiewicz & Moschen, 2018; Heintz-Buschart & Wilmes, 2018) karena dapat mempengaruhi fisiologis manusia, metabolisme (Guinane & Cotter, 2013), penyerapan nutrisi dan mineral, fermentasi serat menjadi asam lemak rantai pendek, sintesis vitamin, pemecahan komponen beracun, dan pengaturan sistem kekebalan tubuh (Gao *et al.*, 2021). Dibandingkan dengan anggota mikrobiota lainnya, bakteri lebih banyak dipelajari, khususnya mengenai penyakit dan kesehatan usus manusia. Fungi juga berperan penting dalam penyakit usus manusia dan bertindak sebagai reservoir patogen oportunistik pathobiont (Soverini *et al.*, 2019). Beberapa penelitian menunjukkan keterlibatan fungi pada tumor gastrointestinal dan paru-paru. Diketahui, DNA fungi terkait tumor tersebut dapat berfungsi sebagai biomarker diagnostik atau prognostik (Anders B Dohlman *et al.*, 2022).

Penelitian mengenai profiling komunitas fungi masih bergantung pada cara tradisional dengan metode kultur. Namun demikian, perkembangan platform *Next Generation Sequencing* (NGS), terutama sekuensing metagenomik dan *tools* bioinformatika telah membantu di dalam studi komunitas fungi. Metagenomik adalah analisis sampel genomik dari lingkungan kompleks untuk mempelajari profil mikroorganisme secara keseluruhan dengan skrining gen fungsional atau analisis sekuensing (Thomas *et al.*, 2012a). NGS merupakan standar gold untuk penelitian yang mengeksplorasi keanekaragaman mikrobiota (Chin *et al.*, 2020). Strategi NGS yang lebih sederhana dan akuisisi *sequencer* yang terjangkau, memungkinkan peneliti untuk mempelajari genom mikrobiota. Analisis data NGS dapat dilakukan dengan dua metode, yaitu sekuensing yang ditargetkan (amplikon atau *marker gene*

\*Margareta, A./ Jurnal Medical Laboratory Vol 1. No. 2 (2023) 41-57

*metagenomics*) dan *shotgun metagenomics*. Pemilihan metode ini bergantung pada jenis studi lingkungan yang akan dilakukan (X. Zhang *et al.*, 2019).

Pengurutan yang ditargetkan dengan marker gen adalah wilayah DNA spesifik yang digunakan untuk mengidentifikasi keberadaan mikroba dalam sampel metagenomik. Penanda atau marker gen yang digunakan untuk studi taksonomi fungi resolusi tinggi adalah 18S rRNA dan ITS banyak digunakan untuk menganalisis keragaman fungi dalam sampel lingkungan. Sedangkan, sekuensing *shotgun metagenomic* merupakan teknik pengambilan sampel semua gen di semua organisme yang ada dalam sampel kompleks tertentu. Oleh karena itu, dapat mengevaluasi keanekaragaman dan mendeteksi kelimpahan mikroba di berbagai lingkungan (Soverini *et al.*, 2019; Thomas *et al.*, 2012b). Data NGS berdasarkan marker ataupun *shotgun metagenomik* ini sangat membantu untuk mengetahui zat bioaktif, gen fungsional dan metabolit mikroba (L. Zhang *et al.*, 2021). Pada ulasan ini, akan fokus membahas alat komputasi dan *pipeline* bioinformatika yang biasa digunakan untuk analisis metagenomik fungi dari data NGS berdasarkan *marker-gene* dan *shotgun sequencing*. Melalui studi bioinformatika ini, maka akan membantu kita dalam mempelajari profil, keragaman dan peran fungi terkait dengan patogenesis dan kesehatan usus.

## 2. METODOLOGI PENELITIAN

Penelitian ini merupakan penelitian *literature review*, dengan mengumpulkan jurnal-jurnal yang berkaitan dengan wawasan bioinformatika dalam menganalisis sampel metagenomik fungi. Metode yang digunakan adalah meringkas jurnal-jurnal tersebut untuk ditulis kembali berdasarkan pendapat peneliti. Data dan informasi yang diperoleh dari jurnal NCBI dan Frontiers sebanyak 100 jurnal dan disaring menjadi 65 jurnal dari jurnal tahun 2000 sampai dengan 2022. Adapun kata kunci yang digunakan di dalam pencarian jurnal adalah metagenomik fungi, NGS, dan *pipeline* bioinformatika.

## 3. HASIL DAN PEMBAHASAN

Setelah melakukan penyaringan jurnal sebanyak 65 jurnal penelitian terpilih, mengenai wawasan bioinformatika, maka perlu melakukan sekuensing pada sampel DNA terlebih dahulu. Salah satu teknologi yang memproses sampel DNA metagenom fungi adalah NGS.

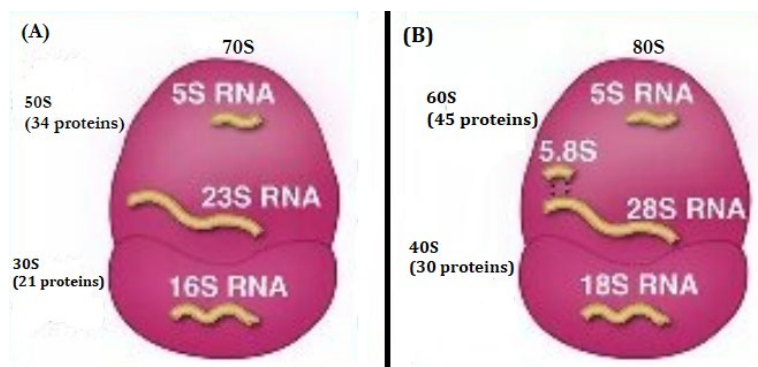
### 3.1 Sekuensing Fragmen Gen 18S rRNA dan *Internal Transcribed Spacer* (ITS)

NGS merupakan teknologi sekuensing yang memungkinkan para peneliti melakukan sekuensing jutaan hingga milyaran nukleotida DNA dalam satu kali pengerjaan. Metode ini bertujuan untuk diagnosis penyakit genetik yang lebih cepat dan akurat, penetapan taksonomi spesies mikroba, studi evolusi, analisis ekspresi gen, dan analisis korelasi. Adapun strategi NGS dalam analisis metagenomik terbagi dua, yaitu *marker-gene metagenomics* dan *shotgun metagenomics* (Navgire *et al.*, 2022). Teknologi NGS berbasis *marker* dapat mendeteksi metataksonomi yang cepat dan akurat. Penanda atau *marker* yang paling banyak digunakan untuk identifikasi fungi adalah fragmen gen 18S rRNA dan ITS yang merupakan daerah konservatif dari RNA ribosom. Kedua fragmen ini menggunakan PCR untuk amplifikasi DNA dengan primer yang spesifik, dilakukan proses sekuensing, analisis sekuens dan membandingkan hasil dengan *database* yang diketahui. Oleh karena itu dapat diketahui profil organisme hingga ke tingkat spesies berdasarkan keluaran data sekuens yang besar (Nilsson *et al.*, 2019; Schoch *et al.*, 2012).

Gen 18S rRNA adalah komponen dari RNA dan struktural subunit ribosom eukariotik kecil (40S), dan 40S dan 60S. 18S rRNA mirip dengan 16S rRNA prokariotik (Gambar 1) yang merupakan salah satu komponen dasar dari semua sel eukariotik. Gen ini terdiri dari daerah konservatif dan hipervariabel, yaitu memiliki sembilan wilayah hipervariabel (V1-V9). Subunit ribosom kecil ini merupakan penanda penting untuk PCR target acak dalam skrining keanekaragaman dan sering digunakan dalam studi filogenetik (Banos *et al.*, 2018).

Selain 18S rRNA, ITS juga merupakan penanda lain yang sering digunakan dalam studi filogenetik, urutan DNA ribosom inti pada 500-700 pasang basa (bp). Daerah ITS dipisahkan menjadi ITS 1 (antara 18S dan 5.8S) dan ITS2 (antara 5.8S dan 28S).

ITS1 dan ITS2 telah ditemukan sebagai penanda yang paling cocok untuk analisis filogenetik fungi karena urutannya yang bervariasi, primer yang dilestarikan, dan sifat multikopi. Penanda ITS biasanya paling berguna untuk sistematika molekuler pada tingkat spesies, dibandingkan wilayah rDNA lainnya seperti sub unit kecil (SSU) dan sub unit besar (LSU) rDNA karena tingkat variasinya yang tinggi (Hoggard *et al.*, 2018).



Gambar 1. Struktural subunit ribosom prokariotik 70S (A) dan ribosom eukariotik 80S (B)

Dibandingkan dengan pengurutan ITS, salah satu keuntungan pengurutan 18S rRNA adalah memungkinkan penyelarasan sekuens DNA lintas taksa. Pengurutan ITS tidak dapat melakukannya karena kurangnya referensi sekuens nukleotida atau *database* sekuens ITS. Namun, 18S rRNA juga memiliki kelemahan untuk analisis pada tingkat keragaman spesies. Hal ini karena sekuensing 18S rRNA hanya dapat memberikan informasi mengenai tingkat taksonomi. Sedangkan pengurutan ITS dapat memberikan informasi tingkat lebih rendah pada tingkat spesies dan subspecies karena terdapat lebih banyak variasi di wilayah ITS1 dan ITS2 daripada wilayah 18S rRNA (Ghosh *et al.*, 2018).

### 3.2 Sekuensing *Shotgun* Metagenomik

Sekuensing *shotgun* metagenomik dapat mengatasi keterbatasan dari sekuensing berdasarkan penanda gen. Prinsip sekuensing berbasis penanda, hanya menargetkan wilayah dari suatu gen. Sedangkan *shotgun* metagenomik melakukan sekuensing secara acak untuk mendeteksi semua fragmen DNA atau semua metagenom sampel yang terukur tanpa menggunakan primer DNA spesifik di dalam sekali proses. Beberapa dari bacaan *shotgun* akan dihasilkan dari wilayah yang

\*Margareta, A./ Jurnal Medical Laboratory Vol 1. No. 2 (2023) 41-57

informatif secara taksonomi seperti 16S, 18S, dan ITS 1 dan 2. Melalui pembacaan pengkodean sekuens DNA, akan mendapatkan wawasan tentang fungsi biologis yang dikodekan dalam genom dan juga memberikan wawasan tentang keanekaragaman hayati dan fungsi dari komunitas mikroba (Sharpton, 2014).

Hasil dari analisis data sekuensing *shotgun* memberikan profil komunitas yang lebih terperinci dibandingkan sekuensing berbasis gen penanda karena terdapat informasi mengenai komposisi gen pada karakterisasi taksonomi komunitasnya dan kapasitas fungsional mikroba dari sampel lingkungan tertentu, misalnya mikrobioma usus. Komposisi gen yang diperoleh, digunakan untuk merumuskan *pathway* fungsional yang diduga, misalnya respon terhadap bahan kimia, *coenzyme binding* dan *quorum sensing*. Namun, pemilihan sekuensing dengan “shotgun” lebih mahal dan membutuhkan waktu yang lebih lama dari pada analisis metagenomik berdasarkan penanda gen. Oleh karena itu, pemilihan teknik sekuensing sangat perlu untuk diperhatikan. Apabila hanya ingin menganalisa sampel lingkungan, untuk mengetahui keragaman pada tingkat spesies, maka pemilihan sekuensing berbasis penanda gen sudah cukup. Namun, untuk mendeteksi organisme dari semua domain kehidupan, sekuensing *shotgun* metagenomik merupakan teknik yang paling efektif dan komprehensif untuk mendapatkan data struktural dan fungsional (Ghosh *et al.*, 2018; Sharpton, 2014). Analisis metagenomik dengan dengan teknik sekuensing *shotgun* diterapkan untuk mempelajari perubahan fungsional mikrobioma usus pada berbagai penyakit, seperti penyakit radang usus (Franzosa *et al.*, 2019), sindrom iritasi usus (Mei *et al.*, 2021), penyakit hati terkait alkohol (Gao *et al.*, 2020), penyakit hati berlemak nonalkohol (Loomba *et al.*, 2017), dan tuberculosis (Hu *et al.*, 2019).

### 3.3 Alat Bioinformatika untuk Analisis *Marker-Gene Metagenomics*

*Software* bioinformatika secara terus-menerus dikembangkan untuk dapat menganalisis metagenomik berdasarkan penanda gen. Analisis secara umum, terdiri dari tiga langkah, antara lain pra-pemrosesan data dan kontrol kualitas, penetapan taksonomi, karakterisasi komunitas dan visualisasi dengan analisis statistik. Pra-pemrosesan pembacaan data mentah sekuens yang dikonversi ke data keluaran berkualitas tinggi sangat penting dan harus dilakukan dengan hati-hati karena terdapat sekuens bias yang harus dikurangi tetapi penyaringan sekuens dapat

\*Margareta, A./ Jurnal Medical Laboratory Vol 1. No. 2 (2023) 41-57

menyebabkan OTU tertentu menjadi berkurang (Soverini *et al.*, 2019b). *Pipeline* yang digunakan untuk ampikon ITS dan 18S rRNA mirip dengan *pipeline* 16S rRNA. Perbedaan terdapat pada pemrosesan data dan strategi pengelompokan sekuen, tercermin dalam kecepatan eksekusi yang berbeda dan interpretasi ampikon yang dihasilkan. Selain itu menggunakan basis data yang berbeda, seperti basis data untuk ITS adalah UNITE (Nilsson, Larsson, *et al.*, 2019), sedangkan 18S rRNA adalah pangkalan data SILVA (Kataoka & Kondo, 2019).

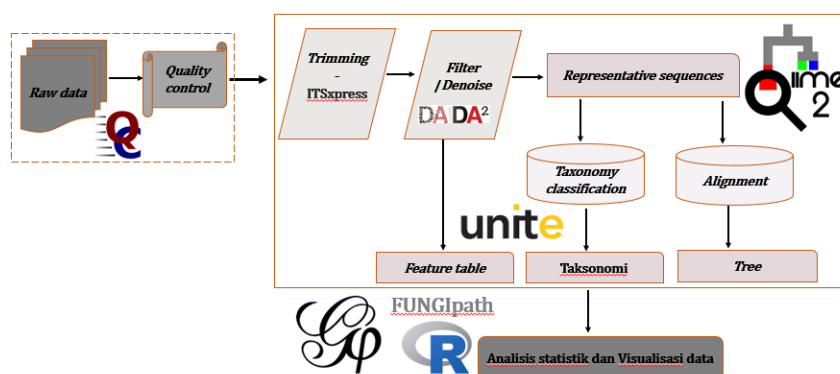
Anotasi taksonomi adalah langkah kunci dalam *pipeline* analisis data sekuensing berbasis penanda gen 18S rRNA dan ITS. Penetapan taksonomi didalam analisis data ITS dan 18S rRNA memiliki kemiripan *pipeline* dengan data sekuensing 16S rRNA. Terdapat paket *software* yang dapat digunakan untuk bakteri (16S rRNA) dan fungi (18S rRNA dan ITS), antara lain QIIME (Bolyen *et al.*, 2019), LotuS 2 (Özkurt *et al.*, 2022), MICCA (Albanese *et al.*, 2015), dan PEMA (Zafeiropoulos *et al.*, 2020). Selain itu, terdapat beberapa paket *software* yang dirancang hanya untuk data ITS, seperti DANIEL, ITSscan (Ferro *et al.*, 2014), ITSx (Bengtsson-Palme *et al.*, 2013), dan ITSxpress (Rivers *et al.*, 2018). Basis data yang umum digunakan untuk analisis fungi adalah UNITE (Nilsson, Larsson, *et al.*, 2019b), ITSoneDB (Santamaria *et al.*, 2018) dan EukRef (del Campo *et al.*, 2018).

Pada pemilihan dan penetapan analisis OTU, ada dua strategi berbeda, yaitu analisis berbasis unit taksonomi operasional (OTU) dan analisis berbasis varian urutan ampikon (ASV). OTU ditentukan oleh kesamaan urutan sekuens nukleotida. Bacaan dianggap sebagai OTU yang sama ketika kesamaan sekuens mencapai ambang kesamaan yang telah ditentukan, paling sering 97%. Umumnya, analisis berbasis OTU pertama mengelompokkan urutan ke dalam OTU yang berbeda dan kemudian memproses penetapan taksonomi. Banyak metode berbasis OTU yang telah dikembangkan, seperti UCLUST (Edgar, 2010) di dalam QIIME sebagai *pipeline* utamanya, UPARSE (Edgar, 2013), CD-HIT pada paket *software* LotuS2 (W. Li & Godzik, 2006), hc-OTU (Park *et al.*, 2018), dan PIPITS pada *software* DANIEL (Gweon *et al.*, 2015).

\*Margareta, A./ Jurnal Medical Laboratory Vol 1. No. 2 (2023) 41-57

Analisis ASV dapat memberikan keuntungan yang signifikan untuk identifikasi mikroba yang lebih tepat. Selain itu, dapat memberikan gambaran yang lebih rinci tentang keragaman dalam suatu sampel. Hal ini karena analisis berbasis ASV menggunakan pendekatan *denoising* untuk menyimpulkan sekuens biologis dalam sampel sebelum pengenalan amplifikasi dan mengurangi kesalahan pembacaan sekuens, yang memungkinkan untuk menyelesaikan sekuens yang berbeda hanya dengan satu nukleotida. Oleh karena itu, analisis berbasis ASV mampu memberikan hasil taksonomi beresolusi lebih tinggi. Beberapa metode berbasis ASV telah dikembangkan, termasuk DADA2 (Callahan et al., 2016), UNOISE 2 (Edgar, n.d.), dan Deblur (Amir *et al.*, 2017).

Alur kerja analisis *marker – gene metagenomics* diperlihatkan pada Gambar 2 yang menggunakan QIIME sebagai *pipeline* utamanya. QIIME merupakan *software* yang terus melakukan pengembangan rutin untuk memudahkan dalam analisis sampel metagenomik. Pengecekan sekuens merupakan langkah pertama yang dilakukan, sebelum diproses pada *software* QIIME. Pengecekan data mentah sekuens dalam format FASTQ menggunakan FastQC. Kemudian *output* dari FastQC dikompilasi menggunakan MultiQC agar dapat lebih mudah untuk dianalisis (Ewels *et al.*, 2017.). Apabila terdapat bagian sekuens dengan nilai rerata *phred score* di bawah 20, maka akan dilakukan pemotongan (*trimming*) di bagian tersebut.



Gambar 2. Alur kerja bioinformatika menggunakan QIIME 2 sebagai *pipeline* utama. *Pre-processing* ditunjukkan dengan garis putus-putus, sedangkan *processing* ditunjukkan dengan garis tegas.

Sebelum memproses data ke QIIME2, maka file FASTQ terlebih dahulu diimpor ke format QZA agar dapat dianalisis menggunakan QIIME2 (Bolyen *et al.*, 2019).



\*Margareta, A./ Jurnal Medical Laboratory Vol 1. No. 2 (2023) 41-57

ITSxpress yang diasosiasikan dengan QIIME2 dirancang untuk memotong fragmen sekuen yang memiliki kualitas buruk sehingga dapat membantu klasifikasi taksonomi yang lebih akurat (Rivers *et al.*, 2018). Setelah dilakukan *trimming*, dilakukan proses penyaringan dan *denoising* melalui plugin DADA2 dalam QIIME2, dengan parameter yang disesuaikan kualitas sekuen. Hasil dari tahapan ini adalah sekuen representatif dan tabel fitur (Callahan *et al.*, 2016). Sekuen representative dianalisis menggunakan Scikit-learn (Pedregosa *et al.*, 2011) dalam QIIME2 untuk menentukan anotasi taksonomi. Acuan taksonomi yang digunakan adalah basis data UNITE. Hasil dari proses ini adalah tabel taksonomi. Selain anotasi taksonomi, penyelarasan sekuen juga dilakukan pada sekuen representatif menggunakan MAFFT dalam QIIME2, dengan *output* berupa berkas *tree* berformat Newick (Rozewicki *et al.*, 2019).

Data sekuen representatif, tabel fitur, tabel taksonomi dan *tree* digunakan pada *plugin taxa bar plot* dalam QIIME2 untuk analisis kelimpahan relatif, dan dimodifikasi dengan program R untuk membantu analisis statistik (Brennan, 2017.). Sedangkan Genepiper (Tong & Chan, 2020) digunakan untuk analisis dan visualisasi kurva rarefaksi, diversitas alfa, diversitas beta, diagram venn ASV, pohon filogenetik dan *taxonomical Heatmap*. Selain itu, data sekuen representatif dan tabel fitur juga digunakan sebagai masukan pada *software* FUNGIpath untuk proses analisis prediksi fungsional. Analisis perediksi dilakukan untuk perbandingan metabolisme dalam spesies berdasarkan kelompok protein ortologis dalam genom fungi (Grossetête *et al.*, 2010). Grup ortologi dianotasi dengan *pipeline* MaRiO berdasarkan data Swiss-Prot dan MetaCyc (Pereira *et al.*, 2014). Sedangkan, data tabel fitur dan tabel taksonomi dianalisis berdasarkan Teori Matriks Acak (RMT) untuk menggambarkan *network* dan memberikan *output* berupa visualisasi analisis *co-occurrence network* (Deng *et al.*, 2012). Oleh karena itu, dapat diketahui dominansi *keystone taxa* pada jaringan taksa tersebut, yang divisualisasikan dengan *software* Gephi (Bastian *et al.*, 2009).

### 3.4 Alat Bioinformatika untuk Analisis *Shotgun Metagenomics*

Sekuensing berdasarkan ampikon atau *marker-gene metagenomics* digunakan untuk mengkarakterisasi diversitas mikroorganisme yang seringkali hanya

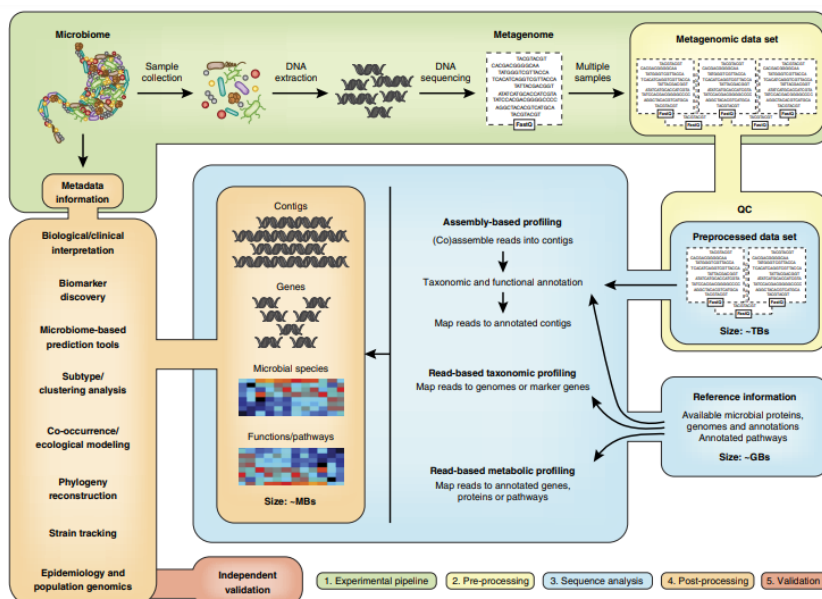
menargetkan satu gen. Hasil dari pengurutan ini, menyediakan informasi berupa komposisi keragaman mikroba tanpa fungsi biologis dari taksa terkait. Sedangkan metode sekuensing *shotgun* mampu melakukan pembacaan secara acak seluruh metagenom tanpa primer spesifik setelah DNA difragmentasi, yang mengurangi bias dari pilihan primer. Oleh karena itu, sekuensing metagenomik *shotgun* menyediakan analisis yang lebih terperinci karena dapat memberikan informasi tentang komposisi gen dan kapasitas fungsional mikrobioma usus. Namun, metode ini membutuhkan biaya yang lebih mahal dibandingkan dengan analisis metagenomik berdasarkan gen yang ditargetkan (Quince *et al.*, 2017).

Metode *shotgun* metagenomik telah digunakan untuk analisis diversitas metabolisme (L. Li *et al.*, 2022; Sarah R Carrino-Kyke *et al.*, 2013), investigasi *outbreak* patogen, deteksi biomarker (Anders B Dohlman *et al.*, 2022), dan analisis mikrobiom usus (Salamon *et al.*, 2021). Alur kerja analisis metagenomik *shotgun* secara umum ditampilkan pada gambar 3. Alur kerja dapat diringkas sebagai berikut: pengumpulan dan penyimpanan sampel fungi usus, ekstraksi asam nukleat, persiapan perpustakaan metagenomik, kontrol kualitas, dan analisis data. Kontrol kualitas adalah langkah pertama dalam proses *pipeline* bioinformatik. FASTQC merupakan *software* yang digunakan untuk evaluasi sekuens DNA. Informasi yang didapat dari FASTQC meliputi kualitas per basa nukleotida, kualitas per sekuen, konten basa nitrogen, konten GC, konten basa N, distribusi panjang sekuen, tingkat duplikasi sekuen, *overrepresented sequence*, dan konten adaptor (Andrews, 2010). Hasil dari kontrol kualitas ini melibatkan berbagai alat seperti MultiQC (Ewels *et al.*, 2016), Trimmomatic, Ktrim, dan Cutadapt untuk analisis lebih lanjut dan visualisasi yang lebih baik (Gao *et al.*, 2021).

Bacaan berkualitas tinggi dari metagenomik *shotgun* dapat dikategorikan menjadi dua pendekatan, yaitu *read-based analysis* dan *assembly based analysis*. Pendekatan berbasis *read-based analysis* mengidentifikasi sekuensing taksonomi bacaan dan profil fungsional melalui pemetaan bacaan ke genom referensi mikroba yang diketahui atau mencari database keluarga protein yang dikarakterisasi oleh *software*, seperti Bowtie2 (Langmead & Salzberg, 2012), DIAMOND, dan BMAP (Quince *et al.*, 2017). Pemetaan sekuen oleh DIAMOND misalnya, diproses dengan

\*Margareta, A./ Jurnal Medical Laboratory Vol 1. No. 2 (2023) 41-57

prinsip *alignment* untuk analisis taksonomi dan fungsional dengan MEGAN6-LR. Adapun *database* untuk anotasi taksonomi dan fungsional, dapat menggunakan Kyoto Encyclopedia of gene and genomes (KEGG) (Kanehisa & Goto, 2000), anotasi keluarga protein (PFAM) (Mi *et al.*, 2019), ontologi gen (GO), *cluster of orthologous groups* (COG), genealogi evolusi gen: Non-supervised Orthologous Groups (eggNOG) dan UniProt Reference Clusters (UniRef) (Gao *et al.*, 2020, 2021).



Gambar 3. Gambaran umum alur kerja analisis taksonomi dan fungsional mikrobiom usus

Pendekatan berbasis *assembly* digunakan untuk mendapatkan *contig* yang kemudian digunakan untuk anotasi. Prinsipnya adalah dengan merekonstruksi banyak genom untuk mengumpulkan bacaan pendek ke dalam *contigs*, yang memungkinkan penyelarasan beberapa urutan bacaan relatif terhadap urutan konsensus, dan kemudian mengelompokkan *contigs* (Peng *et al.*, 2011). Proses ini dapat menggunakan *software* Meta-IDBA, IDBA-UD, MetaVelvet (Namiki *et al.*, 2012) dan MegaHit (Liu *et al.*, 2014), dan lain-lain. *Final contigs* yang diperoleh dari *software* ini, digunakan untuk evaluasi hasil *contigs*, *alignment*, dan prediksi gen. *Final contigs* kemudian dievaluasi dengan MetaQUAST dan BMap untuk mendapatkan jumlah *contigs* dari panjang basa tertentu sehingga didapatkan *output* anotasi. Proses anotasi dapat dilakukan dengan MEGAN6-LR, yang kemudian dapat

\*Margareta, A./ Jurnal Medical Laboratory Vol 1. No. 2 (2023) 41-57

dimodifikasi dan divisualisasi berdasarkan analisis statistik dari tujuan penelitian dengan menggunakan STAMP (Parks et al., 2014) dan R (Sylvia Tippmann, 2015).

#### 4. KESIMPULAN DAN SARAN

Pada jurnal *review* ini memberikan informasi tentang wawasan bioinformatik untuk mempelajari komunitas mikroba, khususnya analisis metagenomik fungi. Analisis metagenomik fungi berdasarkan amplicon atau *marker gene* dan *shotgun* memudahkan peneliti dalam mempelajari taksonomi dan fungsional mikroba. Selain itu, juga dapat memprediksi gen atau protein tertentu yang berperan di dalam *pathway* tertentu sehingga dapat berguna untuk mempelajari penyakit tertentu. Saran untuk penelitian selanjutnya adalah menganalisa perbedaan *software* bioinformatik pada *pipeline* analisis metagenomik fungi.

#### 5. UCAPAN TERIMA KASIH

Ucapan Terimakasih disampaikan kepada Yayasan Kesetiakawanan Sosial Indonesia yang telah memberikan dana penelitian sehingga penelitian ini dapat dilaksanakan dengan baik.

#### 6. DAFTAR PUSTAKA

- Albanese, D., Fontana, P., de Filippo, C., Cavalieri, D., & Donati, C. (2015). MICCA: A complete and accurate software for taxonomic profiling of metagenomic data. *Scientific Reports*, 5. <https://doi.org/10.1038/srep09743>.
- Amir, A., McDonald, D., Navas-Molina, J. A., Kopylova, E., Morton, J. T., Zech Xu, Z., Kightley, E. P., Thompson, L. R., Hyde, E. R., Gonzalez, A., & Knight, R. (2017). Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *MSystems*, 2(2). <https://doi.org/10.1128/msystems.00191-16>.
- Anders B Dohlman, Jared Klug, Marissa Mesko, Iris H Gao, Steven M Lipkin, Xiling Shen, & Iliyan D Iliev. (2022). A pan-cancer mycobiome analysis reveals fungal involvement in gastrointestinal and lung tumors. *National Laboratory of Medicine*.
- Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data [online]*.
- Banos, S., Lentendu, G., Kopf, A., Wubet, T., Glöckner, F. O., & Reich, M. (2018). A comprehensive fungi-specific 18S rRNA gene sequence primer toolkit suited for diverse research issues and sequencing platforms. *BMC Microbiology*, 18(1). <https://doi.org/10.1186/s12866-018-1331-4>.
- Bastian, M., Heymann, S., & Jacomy, M. (2009). *Gephi: An Open Source Software for Exploring and Manipulating Networks Visualization and Exploration of Large Graphs*. [www.aiai.org](http://www.aiai.org).
- Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., de Wit, P., Sánchez-García, M., Ebersberger, I., de Sousa, F., Amend, A., Jumpponen, A., Unterseher, M., Kristiansson, E., Abarenkov, K., Bertrand, Y. J. K., Sanli, K., Eriksson,

\*Margareta, A./ Jurnal Medical Laboratory Vol 1. No. 2 (2023) 41-57

- K. M., Vik, U., Nilsson, R. H. (2013). Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in Ecology and Evolution*, 4(10), 914–919. <https://doi.org/10.1111/2041-210X.12073>.
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., ... Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. In *Nature Biotechnology* (Vol. 37, Issue 8, pp. 852–857). Nature Publishing Group. <https://doi.org/10.1038/s41587-019-0209-9>.
- Brennan, P. (n.d.). *Data visualization with the programming language R*. [https://doi.org/10.1042/bio\\_2021\\_174/921287/bio\\_2021\\_174.pdf](https://doi.org/10.1042/bio_2021_174/921287/bio_2021_174.pdf).
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581–583. <https://doi.org/10.1038/nmeth.3869>.
- Chin, V. K., Yong, V. C., Chong, P. P., Amin Nordin, S., Basir, R., & Abdullah, M. (2020). Mycobioome in the Gut: A Multiperspective Review. In *Mediators of Inflammation* (Vol.2020). Hindawi Limited.
- Campo, J., Kolisko, M., Boscaro, V., Santoferrara, L. F., Nenarokov, S., Massana, R., Guillou, L., Simpson, A., Berney, C., de Vargas, C., Brown, M. W., Keeling, P. J., & Wegener Parfrey, L. (2018). EukRef: Phylogenetic curation of ribosomal RNA to enhance understanding of eukaryotic diversity and distribution. *PLoS Biology*, 16(9). <https://doi.org/10.1371/journal.pbio.2005849>.
- Deng, Y., Jiang, Y.-H., Yang, Y., He, Z., Luo, F., & Zhou, J. (2012). *Molecular ecological network analyses*. <http://www.biomedcentral.com/1471-2105/13/113>
- Edgar, R. C. (2018). *UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing*. <https://doi.org/10.1101/081257>.
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19), 2460–2461.
- Edgar, R. C. (2013). UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, 10(10), 996–998. <https://doi.org/10.1038/nmeth.2604>.
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (n.d.). *Data and text mining MultiQC: Summarize analysis results for multiple tools and samples in a single report*. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047–3048.
- Ferro, M., Antonio, E. A., Souza, W., & Bacci, M. (2014). ITScan: A web-based analysis tool for Internal Transcribed Spacer (ITS) sequences. *BMC Research Notes*, 7(1). <https://doi.org/10.1186/1756-0500-7-857>.
- Franzosa, E. A., Sirota-Madi, A., Avila-Pacheco, J., Fornelos, N., Haiser, H. J., Reinker, S., Vatanen, T., Hall, A. B., Mallick, H., McIver, L. J., Sauk, J. S., Wilson, R. G., Stevens, B. W., Scott, J. M., Pierce, K., Deik, A. A., Bullock, K., Imhann, F., Porter, J. A., ... Xavier, R. J. (2019). Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nature Microbiology*, 4(2), 293–305. <https://doi.org/10.1038/s41564-018-0306-4>.

- 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. In *Biomolecules* (Vol. 11, Issue 4). MDPI AG. <https://doi.org/10.3390/biom11040530>.
- Gao, B., Emami, A., Zhou, R., Lang, S., Duan, Y., Wang, Y., Jiang, L., Loomba, R., Brenner, D. A., Stärkel, P., & Schnabl, B. (2020). Functional Microbial Responses to Alcohol Abstinence in Patients With Alcohol Use Disorder. *Frontiers in Physiology*, *11*. <https://doi.org/10.3389/fphys.2020.00370>.
- Ghosh, A., Mehta, A., & Khan, A. M. (2018). Metagenomic analysis and its applications. In *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics* (Vols. 1–3, pp. 184–193). Elsevier. <https://doi.org/10.1016/B978-0-12-809633-8.20178-7>.
- Gorkiewicz, G., & Moschen, A. (2018). Gut microbiome: a new player in gastrointestinal disease. *Virchows Archiv*, *472*(1), 159–172. <https://doi.org/10.1007/s00428-017-2277-x>.
- Grossetête, S., Labedan, B., & Lespinet, O. (2010). FUNGIpath: A tool to assess fungal metabolic pathways predicted by orthology. *BMC Genomics*, *11*(1). <https://doi.org/10.1186/1471-2164-11-81>.
- Guinane, C. M., & Cotter, P. D. (2013). Role of the gut microbiota in health and chronic gastrointestinal disease: Understanding a hidden metabolic organ. *Therapeutic Advances in Gastroenterology*, *6*(4), 295–308. <https://doi.org/10.1177/1756283X13482996>.
- Gweon, H. S., Oliver, A., Taylor, J., Booth, T., Gibbs, M., Read, D. S., Griffiths, R. I., & Schonrogge, K. (2015). PIPITS: An automated pipeline for analyses of fungal internal transcribed spacer sequences from the Illumina sequencing platform. *Methods in Ecology and Evolution*, *6*(8), 973–980. <https://doi.org/10.1111/2041-210X.12399>.
- Heintz-Buschart, A., & Wilmes, P. (2018). Human Gut Microbiome: Function Matters. In *Trends in Microbiology* (Vol. 26, Issue 7, pp. 563–574). Elsevier Ltd. <https://doi.org/10.1016/j.tim.2017.11.002>.
- Hoggard, M., Vesty, A., Wong, G., Montgomery, J. M., Fourie, C., Douglas, R. G., Biswas, K., & Taylor, M. W. (2018). Characterizing the human mycobiota: A comparison of small subunit rRNA, ITS1, ITS2, and large subunit rRNA genomic targets. *Frontiers in Microbiology*, *9*(SEP).
- Hu, Y., Feng, Y., Wu, J., Liu, F., Zhang, Z., Hao, Y., Liang, S., Li, B., Li, J., Lv, N., Xu, Y., Zhu, B., & Sun, Z. (2019). The gut microbiome signatures discriminate healthy from pulmonary tuberculosis patients. *Frontiers in Cellular and Infection Microbiology*, *9*(APR). <https://doi.org/10.3389/fcimb.2019.00090>.
- Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. In *Nucleic Acids Research* (Vol. 28, Issue 1). <http://www.genome.ad.jp/kegg/>.
- Kataoka, T., & Kondo, R. (2019). Data on taxonomic annotation and diversity of 18S rRNA gene amplicon libraries derived from high throughput sequencing. *Data in Brief*, *25*. <https://doi.org/10.1016/j.dib.2019.104213>.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. <https://doi.org/10.1038/nmeth.1923>.
- Li, L., Yang, K., Li, C., Zhang, H., Yu, H., Chen, K., Yang, X., & Liu, L. (2022). Metagenomic shotgun sequencing and metabolomic profiling identify specific human gut

\*Margareta, A./ Jurnal Medical Laboratory Vol 1. No. 2 (2023) 41-57

- microbiota associated with diabetic retinopathy in patients with type 2 diabetes. *Frontiers in Immunology*, 13.
- Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13), 1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>.
- Liu, C.-M., Luo, R., & Lam, T.-W. (2014). *MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph*. <http://arxiv.org/abs/1401.7457>.
- Loomba, R., Seguritan, V., Li, W., Long, T., Klitgord, N., Bhatt, A., Dulai, P. S., Caussy, C., Bettencourt, R., Highlander, S. K., Jones, M. B., Sirlin, C. B., Schnabl, B., Brinkac, L., Schork, N., Chen, C. H., Brenner, D. A., Biggs, W., Yooseph, S., ... Nelson, K. E. (2017). Gut Microbiome-Based Metagenomic Signature for Non-invasive Detection of Advanced Fibrosis in Human Nonalcoholic Fatty Liver Disease. *Cell Metabolism*, 25(5), 1054-1062.e5.
- Mei, L., Zhou, J., Su, Y., Mao, K., Wu, J., Zhu, C., He, L., & Cui, Y. (2021). Gut microbiota composition and functional prediction in diarrhea-predominant irritable bowel syndrome. *BMC Gastroenterology*, 21(1). <https://doi.org/10.1186/s12876-021-01693-w>.
- Mi, H., Muruganujan, A., Ebert, D., Huang, X., & Thomas, P. D. (2019). PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research*, 47(D1), D419–D426. <https://doi.org/10.1093/nar/gky1038>.
- Namiki, T., Hachiya, T., Tanaka, H., & Sakakibara, Y. (2012). MetaVelvet: An extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Research*, 40(20). <https://doi.org/10.1093/nar/gks678>.
- Navgire, G. S., Goel, N., Sawhney, G., Sharma, M., Kaushik, P., Mohanta, Y. K., Mohanta, T. K., & Al-Harrasi, A. (2022). Analysis and Interpretation of metagenomics data: an approach. In *Biological Procedures Online* (Vol. 24, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s12575-022-00179-7>.
- Nilsson, R. H., Anslan, S., Bahram, M., Wurzbacher, C., Baldrian, P., & Tedersoo, L. (2019). Mycobiome diversity: high-throughput sequencing and identification of fungi. In *Nature Reviews Microbiology* (Vol. 17, Issue 2, pp. 95–109). Nature Publishing Group. <https://doi.org/10.1038/s41579-018-0116-y>.
- Nilsson, R. H., Larsson, K. H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., Kennedy, P., Picard, K., Glöckner, F. O., Tedersoo, L., Saar, I., Kõljalg, U., & Abarenkov, K. (2019a). The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Research*, 47(D1), D259–D264. <https://doi.org/10.1093/nar/gky1022>.
- Nilsson, R. H., Larsson, K. H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., Kennedy, P., Picard, K., Glöckner, F. O., Tedersoo, L., Saar, I., Kõljalg, U., & Abarenkov, K. (2019b). The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Research*, 47(D1), D259–D264. <https://doi.org/10.1093/nar/gky1022>.
- Özkurt, E., Fritscher, J., Soranzo, N., Ng, D. Y. K., Davey, R. P., Bahram, M., & Hildebrand, F. (2022). LotuS2: an ultrafast and highly accurate tool for amplicon sequencing analysis. *Microbiome*, 10(1). <https://doi.org/10.1186/s40168-022-01365-1>.

- Park, S., Choi, H. S., Lee, B., Chun, J., Won, J. H., & Yoon, S. (2018). Hc-OTU: A Fast and Accurate Method for Clustering Operational Taxonomic Units Based on Homopolymer Compaction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *15*(2), 441–451. <https://doi.org/10.1109/TCBB.2016.2535326>.
- Parks, D. H., Tyson, G. W., Hugenholtz, P., & Beiko, R. G. (2014). STAMP: Statistical analysis of taxonomic and functional profiles. *Bioinformatics*, *30*(21), 3123–3124. <https://doi.org/10.1093/bioinformatics/btu494>.
- Pedregosa FABIANPEDREGOSA, F., Michel, V., Grisel OLIVIERGRISEL, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., Cournapeau, D., Pedregosa, F., Varoquaux, G., Gramfort, A., Thirion, B., Grisel, O., Dubourg, V., Passos, A., Brucher, M., Perrot and Édouard and, M., Duchesnay, and Édouard, & Duchesnay EDOUARDDUCHESNAY, Fré. (2011). Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot. In *Journal of Machine Learning Research* (Vol. 12). <http://scikit-learn.sourceforge.net>.
- Peng, Y., Leung, H. C. M., Yiu, S. M., & Chin, F. Y. L. (2011). Meta-IDBA: A de Novo assembler for metagenomic data. *Bioinformatics*, *27*(13). <https://doi.org/10.1093/bioinformatics/btr216>.
- Pereira, C., Denise, A., & Lespinet, O. (2014). A meta-approach for improving the prediction and the functional annotation of ortholog groups. *BMC Genomics*, *15*(6). <https://doi.org/10.1186/1471-2164-15-S6-S16>.
- Pérez, J. C. (2021). Fungi of the human gut microbiota: Roles and significance. *International Journal of Medical Microbiology*, *311*(3). <https://doi.org/10.1016/j.ijmm.2021.151490>.
- Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. In *Nature Biotechnology* (Vol. 35, Issue 9, pp. 833–844). Nature Publishing Group. <https://doi.org/10.1038/nbt.3935>.
- Rivers, A. R., Weber, K. C., Gardner, T. G., Liu, S., & Armstrong, S. D. (2018). ITSxpress: Software to rapidly trim internally transcribed spacer sequences with quality scores for marker gene analysis [version 1; peer review: 2 approved]. *F1000Research*, *7*.
- Rozewicki, J., Li, S., Amada, K. M., Standley, D. M., & Katoh, K. (2019). MAFFT-DASH: Integrated protein sequence and structural alignment. *Nucleic Acids Research*, *47*(W1), W5–W10. <https://doi.org/10.1093/nar/gkz342>.
- Salamon, D., Sroka-Oleksiak, A., Gurgul, A., Arent, Z., Szopa, M., Bulanda, M., Małcki, M. T., & Gosiewski, T. (2021). Analysis of the gut mycobiome in adult patients with type 1 and type 2 diabetes using next-generation sequencing (Ngs) with increased sensitivity—pilot study. *Nutrients*, *13*(4). <https://doi.org/10.3390/nu13041066>.
- Santamaria, M., Fosso, B., Licciulli, F., Balech, B., Larini, I., Grillo, G., de Caro, G., Liuni, S., & Pesole, G. (2018). ITSoneDB: A comprehensive collection of eukaryotic ribosomal RNA Internal Transcribed Spacer 1 (ITS1) sequences. *Nucleic Acids Research*, *46*(D1), D127–D132. <https://doi.org/10.1093/nar/gkx855>.
- Sarah R Carrino-Kyke, Kurt A Smemo, & David J Burke. (2013). Shotgun metagenomic analysis of metabolic diversity and microbial community structure in experimental vernal pools subjected to nitrate pulse. *BMC Microbiology*. <https://doi.org/10.1186/1471-2180-13-78>.
- Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., Chen, W., Bolchacova, E., Voigt, K., Crous, P. W., Miller, A. N., Wingfield, M. J., Aime, M. C., An, K.



\*Margareta, A./ Jurnal Medical Laboratory Vol 1. No. 2 (2023) 41-57

- D., Bai, F. Y., Barreto, R. W., Begerow, D., Bergeron, M. J., Blackwell, M., ... Schindel, D. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(16), 6241–6246. <https://doi.org/10.1073/pnas.1117018109>.
- Sharpton, T. J. (2014). An introduction to the analysis of shotgun metagenomic data. In *Frontiers in Plant Science* (Vol. 5, Issue JUN). Frontiers Research Foundation. <https://doi.org/10.3389/fpls.2014.00209>.
- Soverini, M., Turrone, S., Biagi, E., Brigidi, P., Candela, M., & Rampelli, S. (2019a). HumanMycobiomeScan: A new bioinformatics tool for the characterization of the fungal fraction in metagenomic samples. *BMC Genomics*, *20*(1). <https://doi.org/10.1186/s12864-019-5883-y>.
- Soverini, M., Turrone, S., Biagi, E., Brigidi, P., Candela, M., & Rampelli, S. (2019b). HumanMycobiomeScan: A new bioinformatics tool for the characterization of the fungal fraction in metagenomic samples. *BMC Genomics*, *20*(1). <https://doi.org/10.1186/s12864-019-5883-y>.
- Sylvia Tippmann. (2015). *Programming tools: Adventures with R*. [www.r-bloggers.com](http://www.r-bloggers.com).
- Thomas, T., Gilbert, J., & Meyer, F. (2012a). *Metagenomics-a guide from sampling to data analysis*. <http://www.microbialinformatics.com/content/2/1/3>.
- Thomas, T., Gilbert, J., & Meyer, F. (2012b). *Metagenomics-a guide from sampling to data analysis*. <http://www.microbialinformatics.com/content/2/1/3>.
- Tong, W. M., & Chan, Y. (2020). GenePiper, a Graphical User Interface Tool for Microbiome Sequence Data Mining. *Microbiology Resource Announcements*, *9*(1). <https://doi.org/10.1128/mra.01195-19>.
- Zafeiropoulos, H., Viet, H. Q., Vasileiadou, K., Potirakis, A., Arvanitidis, C., Topalis, P., Pavlodi, C., & Pafilis, E. (2020). PEMA: A flexible Pipeline for Environmental DNA Metabarcoding Analysis of the 16S/18S ribosomal RNA, ITS, and COI marker genes. *GigaScience*, *9*(3). <https://doi.org/10.1093/gigascience/giaa022>.
- Zhang, L., Chen, F. X., Zeng, Z., Xu, M., Sun, F., Yang, L., Bi, X., Lin, Y., Gao, Y. J., Hao, H. X., Yi, W., Li, M., & Xie, Y. (2021). Advances in Metagenomics and Its Application in Environmental Microorganisms. In *Frontiers in Microbiology* (Vol. 12). Frontiers Media S.A. <https://doi.org/10.3389/fmicb.2021.766364>.
- Zhang, X., Li, L., Butcher, J., Stintzi, A., & Figeys, D. (2019). Advancing functional and translational microbiome research using meta-omics approaches. In *Microbiome* (Vol. 7, Issue 1). BioMed Central Ltd.