T.C
ERCİYES ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİNFORMATİK SİSTEMLER BİYOLOJİSİ ANABİLİM DALI

İNSAN MİKROBİYOMUNDA BULUNAN BAKTERİYOFAJ KOMÜNİTESİNİN KONAKÇI HASTALIKLARIYLA OLAN İLİŞKİSİNİN ARAŞTIRILMASI

Hazırlayan
Mehmet HORA

Danışman
Prof. Dr. Aycan GÜNDOĞDU

Doktora Tezi

Ocak, 2025
KAYSERİ

T.C
ERCİYES ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİNFORMATİK SİSTEMLER BİYOLOJİSİ ANABİLİM DALI

İNSAN MİKROBİYOMUNDA BULUNAN BAKTERİYOFAJ KOMÜNİTESİNİN KONAKÇI HASTALIKLARIYLA OLAN İLİŞKİSİNİN ARAŞTIRILMASI

Doktora Tezi

Hazırlayan
Mehmet HORA

Danışman
Prof. Dr. Aycan GÜNDOĞDU

Bu Tez TÜBİTAK Tarafından 323S130 No'lu Proje ile Desteklenmiştir.

Ocak, 2025
KAYSERİ

BİLİMSEL ETİĞE UYGUNLUK

Bu tezin kendi çalışmam olduğunu, tüm bilgilerin akademik ve etik kurallara uygun bir şekilde elde edildiğini beyan ederim. Aynı zamanda akademik ve etik kuralların gerektirdiği gibi tüm materyal ve sonuçları tam olarak aktardığımı, başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel kurallara uygun olarak atıfta bulunduğumu ve kaynaklar listesinde gösterdiğimi belirtirim.

Adı-Soyadı: Mehmet HORA

İmza:

YÖNERGEYE UYGUNLUK ONAYI

‘İnsan Mikrobiyomunda Bulunan Bakteriyofaj Komünitesinin Konakçı Hastalıklarıyla Olan İlişkisinin Araştırılması ’adlı **Doktora tezi,** Erciyes Üniversitesi Lisansüstü Tez Önerisi ve Tez Yazma Yönergesi’ne uygun olarak hazırlanmıştır.

Tezi Hazırlayan	**Danışman**
Mehmet HORA	**Prof. Dr. Aycan GÜNDOĞDU**

Anabilim Dalı Başkanı

Doç. Dr. Özkan Ufuk NALBANTOĞLU

KABUL VE ONAY SAYFASI

Prof. Dr. Aycan GÜNDOĞDU danışmanlığında **Mehmet Hora** tarafından hazırlanan **"İnsan Mikrobiyomunda Bulunan Bakteriyofaj Komünitesinin Konakçı Hastalıklarıyla Olan İlişkisinin Araştırılması"** adlı bu çalışma jürimiz tarafından Erciyes Üniversitesi Sağlık Bilimleri Enstitüsü Biyoinformatik Sistemler Biyolojisi Anabilim Dalında **doktora tezi** olarak kabul edilmiştir.

14 / 01 / 2025

JÜRİ **İmza**

Danışman : Prof. Dr. Aycan GÜNDOĞDU

Üye : Doç.Dr.Özkan.Ufuk NALBANTOĞLU

Üye : Prof. Dr. Ayşegül ULU KILIÇ

Üye : Prof. Dr. Barış OTLU

Üye : Prof. Dr. Çağrı ERGİN

ONAY:
Bu tezin kabulü Enstitü Yönetim Kurulunun ……...… tarih ve ……..…… sayılı kararı ile onaylanmıştır.

……. /……../ ……
Prof. Dr. Bilal AKYÜZ
Enstitü Müdürü

TEŞEKKÜR

Öncelikle, değerli danışmanım Sayın Prof. Dr. Aycan GÜNDOĞDU'ya, akademik bilgi ve yönlendirmeleri ile bu çalışmanın başarılı bir şekilde tamamlanmasında gösterdiği katkılardan dolayı en içten teşekkürlerimi sunarım. Bilimsel yaklaşımı, rehberliği ve motivasyonu sayesinde yalnızca tezimi şekillendirmekle kalmayıp, bilimsel düşünce yapımı geliştirmemde ve gelecekteki çalışmalarım için sağlam bir temel oluşturmamda bana büyük katkı sağlamıştır. Kendisi, bilim aşkını ve araştırmaya olan tutkusunu benimle paylaşarak, bu süreçte hem akademik hem de kişisel anlamda önemli bir ilham kaynağı olmuştur.

Araştırma süreçlerine farklı bakış açılarıyla yaklaşarak ufkumu genişleten ve tezime önemli katkılarda bulunan Sayın hocam Doç. Dr. Özkan Ufuk NALBANTOĞLU'na içten teşekkürlerimi sunarım. Biyoinformatik ve Sistem Biyolojisi alanında kazandığım bilgi ve becerilerin temelini kendisinden aldığım eğitim ve rehberlik oluşturmuştur. Disiplini, detaylara verdiği önem ve öğretim tarzıyla bu alandaki yetkinliğimi geliştirmemde büyük rol oynamıştır. Kendisine, emeği ve desteği için minnettarım.

Tez çalışmamda yardımlarını esirgemeyen arkadaşlarım Ayşenur SOYTÜRK PATAT, Umay Gülfem ELGÜN ÇİFTÇİOĞLU ve Samed SAKA'ya değerli katkıları için minnettarım.

Hayatımda her zaman desteğini ve sevgisini hissettiğim kıymetli eşim Esra AKYÜREK HORA'ya, sabrı ve anlayışı için sonsuz teşekkürlerimi sunarım.

Son olarak, bana her zaman güvenen ve desteğini esirgemeyen aileme, bana kattıkları tüm değerler için teşekkür ederim.

İNSAN MİKROBİYOMUNDA BULUNAN BAKTERİYOFAJ KOMÜNİTESİNİN KONAKÇI HASTALIKLARIYLA OLAN İLİŞKİSİNİN ARAŞTIRILMASI

Mehmet HORA

Erciyes Üniversitesi, Sağlık Bilimleri Enstitüsü
Biyoinformatik Sistemler Biyolojisi Anabilim Dalı
Doktora Tezi, Ocak 2025
Danışman Prof. Dr. Aycan GÜNDOĞDU

ÖZET

Bağırsak mikrobiyomu son derece çeşitli bir ekosistemdir. Bakteriler en çok temsil edilen üye olmanın yanında mantarlar, parazitler, arkeler ve virüsler mikrobiyomu oluşturan bileşenlerdir. Ancak, mutlak miktarlarının düşük olması ve biyoinformatik analizler için özelleşmiş araçların eksikliği gibi nedenlerle bakteri dışındaki üyeler daha seyrek olarak çalışmalara konu olmaktadır. Bu çalışmada, bağırsak mikrobiyomunun önemli bir bileşeni olan bakteriyofaj ekosisteminin iki kronik hastalıktaki (kolorektal kanser (KK) ve inflamatuar bağırsak hastalığı (İBH)) kompozisyonunun değerlendirilmesi ve hastalık spesifik biyobelirteçlerin keşfedilmesi amaçlanmıştır. Bu amaç doğrultusunda, KK ve İBH hastalarına ve sağlıklılara ait bağırsak metagenom veri setlerinin, veri madenciliği yaklaşımlarıyla incelenmesi ve bakteriyofaj kompozisyonlarının analiz edilmesi hedeflenmiştir. Çalışma kapsamında 6 ayrı veri setinden KK grubu için 383 hasta ve 446 sağlıklı kontrol, İBH için 468 hasta ve 267 sağlıklı kontrolün bağırsak mikrobiyomuna ait shotgun metagenom dizileme verileri bilimsel kullanıma açık veri tabanlarından temin edilmiştir. Tüm veri setleri için kalite-kontrol analizleri yapılmış ve sonrasında IDBA_UD programı ile de novo assembly (birleştirme) gerçekleştirilmiştir. Birleştirilmiş metagenom verileri içerisindeki bakteriyofajların tespitinde PhaBOX iş akışı takip edilmiştir. Ilımlı fajlar filtrelendikten sonra her bir kohorttaki bakteriyofaj profillerinin hastalıklarla ilişkisi, makine öğrenme ve transfer öğrenme yöntemleri kullanılarak değerlendirilmiştir. Bu amaçla üç farklı yapay zekâ algoritması kullanılmıştır. Rastgele Orman (Random Forest), Gradyan Güçlendirilmiş Karar Ağaçlar (XGBoost) ve Yapay Sinir Ağı (Multi Layer Perceptron) algoritmaları ile regresyon analizi gerçekleştirilmiştir. İstatistiksel analizde Python Scipy modülleri kullanılmış olup $p<0.05$ değeri istatistiksel olarak anlamlı kabul edilmiştir. Nicel değişkenler için bağımsız t-testi kullanılmıştır. Toplamda 1564 kişiye ait 7365 Gbp veri çalışma kapsamına dahil edilmiştir. Kalite kontrol aşamasını geçemeyen 320 Gbp veri çalışma dışı bırakıldıktan sonra toplamda yaklaşık 7.0 Terabyte'lık metagenom verisi için biyoinformatik analizler yürütülmüştür. Takip edilen biyoinformatik iş akışı sonucunda *Bacteroides fragilis*, *Bacteroides faecis*, *Achromobacter xylosoxidans* bakteriyofajlarının KK hastalarında, *Ruminococcaceae bacterium*, *Coprobacillus sp. AF13-15* fajlarının ise İBH hastalarında sağlıklılara göre istatistiki olarak anlamlı oranda ($p <0.05$) fazla olduğu tespit edilmiştir. İBH veri setlerinde KK veri setlerine göre daha fazla sayıda faj türü hastalıkla ilişkilendirilmiştir (KK için dört veri setinde toplam 21 faj çeşidi, İBH için iki veri setinde toplam 42 faj çeşidi). *Eubacterium rectale* bakterisinin İBH hasta grubunda

sağlıklı kontrollere kıyasla istatistiki olarak anlamlı oranda fazla bulunduğu, buna karşın *Eubacterium rectale* bakteriyofajının sağlıklı kontrol grubunda istatistiki olarak anlamlı oranda fazla olduğu tespit edilmiştir. *Roseburia intestinalis* bakterisinin KK hastalarında, bu bakteriye ait fajın ise sağlıklı kontrol grubunda istatistiki olarak anlamlı oranda daha fazla bulunduğu tespit edilmiştir. Faj bağıl bollukları üzerinden yapay zekâ algoritmalarıyla ve öznitelik seçimleriyle yapılan analizler sonucunda, fajların hastalık ve sağlık durumunu ayırt etme gücü KK için en yüksek %74, İBH için en yüksek %66 olarak hesaplanmıştır. Sınıflandırma performansını artırmak için kullanılan transfer öğrenme modelleriyle fajların hastalık ve sağlık durumunu ayırt etme gücü KK için en yüksek %78, İBH için en yüksek %74 sınıflandırma başarısına ulaşmıştır.

Bu tez çalışmasında, farklı veri setlerine ait mikrobiyom faj kompozisyonları incelenerek KK ve İBH için nispeten kararlı biyobelirteçler keşfedilmiştir. Bu biyobelirteçler üzerinden gerçekleştirilecek ileri çalışmalar ile KK ve İBH'nin tanı/tedavisinde kullanılabilecek yöntemlerin geliştirilme potansiyeli bulunmaktadır.

Anahtar Kelimeler: Bağırsak Viromu; Bakteriyofaj; Biyoinformatik; Metagenom; Virom; Yapay Zekâ.

INVESTIGATION OF THE RELATIONSHIP BACTERIOPHAGE COMMUNITY IN THE HUMAN MICROBIOME WITH HOST DISEASES

Mehmet HORA

Erciyes University, Health Sciences Institute
Department of Bioinformatics and System Biology
Ph.D. Thesis, January 2025
Supervisor: Prof. Dr. Aycan GÜNDOĞDU

ABSTRACT

The gut microbiome is an extremely diverse ecosystem. While bacteria are the most represented members, fungi, parasites, archaea, and viruses also contribute to the microbiome. However, due to their relatively low absolute abundance and the lack of specialized bioinformatics tools, non-bacterial components are less frequently studied. This study aims to evaluate the composition of the bacteriophage ecosystem an essential component of the gut microbiome in two chronic diseases, colorectal cancer (CRC) and inflammatory bowel disease (IBD), and to identify disease-specific biomarkers. To achieve this, publicly available gut metagenomic datasets from CRC and IBD patients, as well as healthy controls, were analyzed using data mining approaches to investigate the bacteriophage compositions.

Within the scope of this study, metagenomic shotgun sequencing data from six different datasets were obtained from scientific repositories, comprising 383 CRC patients and 446 healthy controls, as well as 468 IBD patients and 267 healthy controls. Quality control analyses were conducted for all datasets, followed by *de novo* assembly using the IDBA_UD program. The detection of bacteriophages within the assembled metagenomic data was performed using the PhaBOX workflow. After filtering out temperate phages, the relationship between bacteriophage profiles and diseases in each cohort was evaluated using machine learning and transfer learning techniques.

Through the bioinformatics workflow, it was determined that *Bacteroides fragilis*, *Bacteroides faecis*, and *Achromobacter xylosoxidans* bacteriophages were significantly more abundant ($p < 0.05$) in colorectal cancer (CRC) patients compared to healthy controls. In contrast, *Ruminococcaceae bacterium* and *Coprobacillus sp. AF13-15* phages were found to be significantly more abundant in inflammatory bowel disease (IBD) patients. Notably, a greater number of phage species were associated with disease in the IBD datasets compared to the CRC datasets (21 phage species across four CRC datasets vs. 42 phage species across two IBD datasets). Additionally, *Eubacterium rectale* was significantly more abundant in IBD patients than in healthy controls, whereas its corresponding bacteriophage was significantly more prevalent in the healthy control group. Similarly, *Roseburia intestinalis* was significantly more abundant in CRC patients, while its phage was more prevalent in healthy controls. Machine learning-based analyses using phage relative abundances and feature selection methods revealed that the predictive power of phages in distinguishing disease and health conditions was up to 74% for CRC and 66% for IBD. Transfer learning models further improved classification performance, achieving a maximum accuracy of 78% for CRC and 74% for IBD.

In this thesis, the microbiome phage compositions of different datasets were analyzed, leading to the identification of relatively stable biomarkers for CRC and IBD. Further studies focusing on these biomarkers may contribute to the development of novel diagnostic and therapeutic strategies for CRC and IBD.
Keywords: Artificial Intelligence; Bacteriophage; Bioinformatics; Gut Virome; Metagenomics.

İÇİNDEKİLER

İÇ KAPAK .....
BİLİMSEL ETİĞE UYGUNLUK i
YÖNERGEYE UYGUNLUK ONAYI ii
KABUL VE ONAY SAYFASI iii
TEŞEKKÜR iv
ÖZET v
ABSTRACT vii
İÇİNDEKİLER ix
KISALTMALAR ve SİMGELER xi
TABLO LİSTESİ xiii
ŞEKİL LİSTESİ xiv
1. GİRİŞ VE AMAÇ 1
2. GENEL BİLGİLER 4
2.1. Bağırsak Mikrobiyomu 4
2.1.1. İnsan Viromu 6
2.1.1.1. Bakteriyofajlar 7
2.1.1.1.1. Bakteriyofajların Keşfi 7
2.1.1.1.2. Bakteriyofajların Yapıları ve Yaşam Döngüleri 8
2.1.1.1.3. Bakteriyofajların Uygulama Alanları 9
2.1.1.2. Ökaryotik Hücreleri Enfenkte Eden Virüsler (Ökaryotik Virüsler) ... 10
2.1.1.3. Endojen Retrovirüsler 12
2.2. Hastalık Mikrobiyom İlişkisi 12
2.2.1. İnflamatuar Bağırsak Hastalığı ve Mikrobiyom ilişkisi 13
2.2.2. Kolorektal Kanser ve Mikrobiyom ilişkisi 15
2.3. Metagenomik Yaklaşımlar 18
2.3.1. Viral Metagenomik 21
2.3.1.1. Viral Metagenomikte Taksonomik Tanımlama ve Zorluklar 23
2.4. Mikrobiyom ve Makine Öğrenimi 24
3. GEREÇ VE YÖNTEM 30
3.1. Gereç 30
3.2. Yöntem 30

3.2.1. Veri Setleri 30
3.2.1.1. Kolorektal Kanser Veri Setleri 31
3.2.1.2. İnflamatuar Bağırsak Hastalığı Veri Setleri 31
3.2.2. Biyoinformatik Analizler 31
3.2.2.1. Ham Verilerin Kalite Kontrolü 31
3.2.2.2. Verilerin Birleştirilmesi (Assembly) 31
3.2.2.3. Bakteriyel Taksonomik Atama 32
3.2.2.4. Bakteriyofaj İdentifikasyonu 32
3.2.2.5. Bakteriyofajların Taksonomik Sınıflandırılması ve Yaşam Döngüsü 33
3.2.2.6. Faj-Konak İlişkilendirmesi 33
3.2.3. Makine Öğrenmesi, Derin Öğrenme ve Transfer Öğrenme Yöntemleriyle Sınıflandırma 34
3.2.3.1 Makine öğrenmesi 34
3.2.3.2. Transfer Öğrenme 35
3.2.3.3. Öznitelik Seçimi (Feature Selection) 35
3.2.4. İstatistiksel Analiz 35
4.BULGULAR 36
4.1. Biyoinformatik Analizler 37
4.1.1. Kolorektal Kanser Veri Setleri İçin Bakteriyel Çeşitlilik Karşılaştırması 37
4.1.1.1. Veri Seti 1 (PRJEB7774) 37
4.1.1.2. Veri Seti 2 (PRJEB27928) 38
4.1.1.3. Veri Seti 3 (PRJEB12449) 39
4.1.1.4. Veri Seti 4 (PRJEB10878) 39
4.1.2. İnflamatuar Bağırsak Hastalığı Veri Setleri İçin Bakteriyel Çeşitlilik Karşılaştırması 40
4.1.2.1. Veri Seti 1 (PRJEB1220) 41
4.1.2.2. Veri Seti 2 (PRJNA400072) 41
4.1.3. Kolorektal Kanser Veri Setleri İçin Litik Bakteriyofaj Karşılaştırması ... 42
4.1.3.1. Veri Seti 1 (PRJEB7774) 42
4.1.3.2. Veri Seti 2 (PRJEB27928) 43
4.1.3.3. Veri Seti 3 (PRJEB12449) 44
4.1.3.4. Veri Seti 4 (PRJEB10878) 45

4.1.4. İnflamatuar Bağırsak Hastalığı Veri Setleri İçin Litik Bakteriyofaj Karşılaştırması 46
4.1.4.1. Veri Seti 1 (PRJEB1220) 46
4.1.4.2. Veri Seti 2 (PRJNA400072) 47
4.1.5. Tüm veri setlerinin her bir hastalık grubunda ortak analizi 48
4.2. Makine Öğrenmesi Yöntemleri ile Veri Setlerinin Sınıflandırma Sonuçları .. 49
4.2.1. Kolorektal Kanser Veri Setlerine Ait Sınıflandırma Sonuçları 49
4.2.2. İnflamatuar Bağırsak Hastalığı (İBH) Veri Setlerinin Sınıflandırma Sonuçları 50
4.3. Transfer Öğrenme Modeli ile Veri Setlerinin Sınıflandırma Sonuçları, ... 51
4.3.1. Kolorektal Kanser Veri Setlerinin Sınıflandırma Sonuçları 51
4.3.2. İnflamatuar Bağırsak Hastalığı (İBH) Veri Setlerinin Sınıflandırma Sonuçları 52
5. TARTIŞMA 54
6. SONUÇ VE ÖNERİLER 61
7. KAYNAKÇA 64
EKLER
ÖZGEÇMİŞ

KISALTMALAR ve SİMGELER

CH :	Crohn Hastalığı
DL :	Derin Öğrenme
DNA :	Deoksiribonükleik Asit
Enet :	Elastik Ağlar
GB :	Gradyan Artırma
ICTV :	Uluslararası Virüs Taksonomi Komitesi
İBH :	İnflamatuar Bağırsak Hastalığı
KK :	Kolorektal Kanser
KVH :	Kardiyovasküler Hastalıklar
LR :	Lojistik Regresyon
ML :	Makine Öğrenmesi
MLP :	Yapay Sinir Ağları
MS :	Multipl Skleroz
NCBI :	Ulusal Biyoteknoloji Bilgi Merkezi
ORF :	Açık okuma çerçeveleri
PZR :	Polimeraz Zincir Reaksiyonu
RA :	Romatoid Artrit
RF :	Rastgele Orman Algoritması
RNA :	Ribonükleik Asit
SCFA :	Kısa Zincirli Yağ Asitleri
SVM :	Destek Vektör Makinesi Algoritması
ÜK :	Ülseratif Kolit
VLP :	Virüs Benzeri Partikül

TABLO LİSTESİ

Tablo 4.1. Çalışmaya dahil edilen KK veri setleri 36

Tablo 4.2. Çalışmaya dahil edilen İBH veri setleri 36

Tablo 4.3. KK veri setleri için sınıflandırma performansı sonuçları 50

Tablo 4.4. İBH veri setleri için sınıflandırma performansı sonuçları 51

Tablo 4.5. KK veri setlerinin transfer öğrenme modeli ile gerçekleştirilmiş sınıflandırma performansı sonuçları 52

Tablo 4.6. İBH veri setlerinin transfer öğrenme modeli ile gerçekleştirilmiş sınıflandırma performansı sonuçları 53

ŞEKİL LİSTESİ

Şekil 2.1. Bakteriyofaj yapısının çizimi 7

Şekil 2.2. Litik ve lizojenik döngüde faj 8

Şekil 4.1. KK 1 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyel taksonlar.. 37

Şekil 4.2. KK 2 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyel taksonlar.. 38

Şekil 4.3. KK 3 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyel taksonlar.. 39

Şekil 4.4. KK 4 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyel taksonlar.. 40

Şekil 4.5. İBH 1 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyel taksonlar.. 41

Şekil 4.6. İBH 2 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyel taksonlar.. 422

Şekil 4.7. KK 1 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyofajlar..... 433

Şekil 4.8. KK 2 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyofajlar..... 44

Şekil 4.9. KK 3 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyofajlar..... 455

Şekil 4.10. KK 4 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyofajlar..... 466

Şekil 4.11. İBH 1 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyofajlar..... 47

Şekil 4.12. İBH 2 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyofajlar..... 48

1. GİRİŞ VE AMAÇ

Bağırsak mikrobiyomu bağırsaklarda yaşayan mikroorganizmaların, bunların genomlarının, genom ürünlerinin, kendi aralarındaki ve konak organizmayla olan etkileşimlerinin oluşturduğu dinamik bir ekosistem olarak tanımlanmaktadır (Holmes ve ark. 2012). Gelişen analitik yöntemlerle birlikte, bağırsak mikrobiyomunun kompozisyonunun detaylı bir şekilde haritalanması sağlanmış ve bu ekosistemin metabolik hastalıklar (Qin ve ark., 2012), kanser (Yachida ve ark., 2019), gastrointestinal sistem hastalıkları (Hodgkinson ve ark., 2023), nörodejeneratif hastalıklar (Zhu ve ark., 2021) ve otoimmün hastalıklar (Miyauchi ve ark., 2023) gibi farklı sistem ve organları tutan hastalıklar ile karmaşık ilişkiler içerisinde olduğu ortaya konulmuştur.

Mikrobiyom genellikle bakteri ağırlıklı bir ekosistem olarak düşünülse de yapılan çalışmalar bu ekosistemin yalnızca bakterilerden ibaret olmadığını; mantarlar, parazitler, arkeler ve virüslerin de dahil olduğu farklı mikrobiyal bileşenleri içerdiğini ortaya koymuştur (Fusco ve ark., 2023). Bununla birlikte bakteriler dışındaki mikrobiyal üyeler, genomik analizler için gerekli olan uygun araçların eksikliği ve bu üyelerin mikrobiyom içerisindeki nispi düşük miktarları nedeniyle uzun süre bilimsel araştırmalarda arka planda kalmıştır (Conceição-Neto ve ark., 2015). Bu bağlamda, mikrobiyomun genellikle ihmal edilen üyelerinden biri olan virüsler, doğada en yaygın bulunan mikroorganizmaları temsil eder ve her ekosistemde bulunabilirler.

İnsan bağırsak mikrobiyomunda yer alan virüs ekosisteminde, ökaryotik hücre virüsleriyle (ökaryotik virüsler) birlikte, prokaryotik hücreleri hedefleyen bakteriyofajlar (Faj) en yüksek bollukta bulunmaktadır (Shkoporov ve ark., 2019). Fajlar, yalnızca bakterileri ve arkeleri enfekte edebilen virüsler olup

bakterilerin/arkelerin varlığına bağlı olarak çeşitlilik ve bollukları değişebilen, insan hücreleri ile doğrudan bir etkileşime girmeyen ve son yıllarda tedavi ajanı olarak kullanım potansiyeli üzerine kapsamlı çalışmalar yapılan mikroorganizmalardır. (Holtappels ve ark., 2023; Lopetuso ve ark., 2016). Bu özellikleriyle bakteriyofajlar, mikrobiyal ekosistem dinamiklerinde ve insan sağlığına odaklanan çalışmalarda önemli bir araştırma konusu haline gelmiştir.

Metagenomik yaklaşımlar, herhangi bir örnekteki genetik materyalin tamamını kültürleme gereksinimi olmaksızın analiz etmeye olanak sağlayan yeni nesil teknolojilerdir (Lazarevic ve ark., 2009). Son 15 yıl içerisinde yeni nesil dizileme teknolojilerinin hızla gelişmesi ve maliyetlerinin düşmesi, metagenomik araştırmaların sayısını artırmış ve yeni bilimsel soruların ortaya çıkmasına zemin hazırlamıştır.

Bir habitattaki viral bileşenlerin tamamı "virom" olarak adlandırılmaktadır. İnsan viromu genel olarak, ökaryotik hücreleri enfekte eden virüsler, prokaryotik hücreleri enfekte eden bakteriyofajlar ve endojen retrovirüsler gibi insan genomunda yer alan virüsten türemiş genetik elementlerden oluşmaktadır (Haynes ve ark., 2011). İnsan mikrobiyomunda bulunan virüslerin kapsamlı olarak çalışıldığını gösteren bir çok yayın mevcuttur (Cadwell ve ark., 2015; Schoggins ve ark., 2014; Lecuit ve ark., 2013; Wylie ve ark., 2012). Bu çalışmalarda, bakteriyofajlar hem insan viromunun en kalabalık elamanları olmaları hem de mikrobiyomun baskın üyesi olan bakterilerin dengeleyici parazitleri olarak işlev görmeleri nedenleriyle virom çalışmalarının temel araştırma konusunu oluşturmaktadır. Bakteriyofajların, bağırsak mikrobiyotasının stabilitesine katkı sağlayan ve insan sağlığı üzerinde olumlu etkileri bulunan kommensal üyeler olduğunun bilinmesine karşın (Holtappels ve ark., 2023; Lopetuso ve ark., 2016), kompozisyon değişikliklerinin KK ve İBH'nin içinde bulunduğu birçok hastalığın patogenezinde rol oynayabileceği bildirilmiştir (Claesson ve ark., 2017; Liang ve Bushman 2021;Norman ve ark., 2015;). Bu çalışmalarda bağırsak faj ekosisteminin hastalık durumunda belirli imzalar taşıyabileceği hipotezi ortaya atılsa da, bu imzaları bulmaya yönelik çalışmalar oldukça sınırlıdır.

Bu tez çalışması kapsamında, KK ve İBH hastalarının bağırsak mikrobiyomlarına ait metagenom verileri üzerinden bakteriyofaj ekosistemi incelenmiştir. Çalışmanın temel amacı, KK ve İBH hastalarının ve sağlıklı kontrollerin bağırsak mikrobiyomundaki bakteriyofaj ekosistem dinamiklerini karşılaştırmalı olarak analiz etmek ve hastalık ilişkili potansiyel biyobelirteçleri keşfetmektir.

2. GENEL BİLGİLER

2.1. Bağırsak Mikrobiyomu

Vücudumuzda yaşayan mikroorganizmaların sayısının yaklaşık 38 milyar olduğu ve bunun insan hücrelerinin sayısından yaklaşık 1.3 kat fazla olduğu tahmin edilmektedir (Sender ve ark., 2016). İnsan genomundan yaklaşık 100 kat daha fazla benzersiz gene sahip olduğu öne sürülen bu mikroorganizmaların büyük bir kısmı bağırsakta yer alır. Mikrobiyota, bir habitatta yerleşmiş ve fonksiyonel bir ekolojik denge (homeostazis) içinde bulunan mikrobiyal topluluğu ifade ederken, mikrobiyom ise bu topluluk ile birlikte mikroorganizmaların genomlarını, genom ürünlerini ve hem birbirleriyle hem de konak organizma ile olan etkileşimlerini kapsayan bir kavramdır (Marchesi ve ark., 2015).

İnsan mikrobiyomunun en büyük bileşenlerinden biri olan bağırsak mikrobiyomu, insan ile çevresel faktörler (örneğin, besinler, patojenik organizmalar, toksinler vb.) arasında bir arayüz görevi üstlenmekte ve konağıyla birçok metabolik faaliyette ortaklık yapmaktadır. Yapılan araştırmalar, besin sindirimi (Malongane ve ark., 2024), vitamin biyosentezi (Bedani ve ark., 2024), davranışsal yanıtlar (de Noronha ve ark., 2024), ve patojenlere karşı savunma (Lin ve ark., 2024) gibi çeşitli temel biyolojik süreçlerin mikrobiyom tarafından desteklendiğini ortaya koymuştur (Light ve ark., 2024).

Son dönem çalışmaları, insan-mikrobiyota etkileşimindeki homeostatik dengenin bozulması (disbiyozis) ile birçok kompleks hastalığın ilişkili olduğunu göstermiştir (Sung ve ark., 2024). Mikrobiyota disbiyozisinin, belirli hastalıkların patogenezinde ve prognozunda önemli bir rol oynadığına dair kanıtların artması, hastalıkların tanı ve tedavisine yönelik yeni fırsatlar sunan bir araştırma alanının ortaya çıkmasına

neden olmuştur (Dorobisz ve ark., 2024). Bu durum, mikrobiyota ve mikrobiyomun araştırmalarını, yüksek translasyonel potansiyele sahip öncelikli bir bilimsel çalışma sahası haline getirmiştir.

Kompleks hastalıklar bağlamında, mikrobiyota disbiyozisinin hastalığın kötüleşen prognozuna ve hatta patogenezine katkıda bulunabileceğinin gözlemlenmesi, mikrobiyota homeostazisinin yeniden sağlanmasının veya desteklenmesinin terapötik bir müdahale stratejisi olabileceği fikrini ortaya çıkarmıştır (Saedi ve ark., 2024). Mikrobiyom biliminin translasyonel yönelimi doğrultusunda, hastalıkların tedavisi veya önlenmesi amacıyla fekal transplantasyon, hedefe yönelik bakteriyofaj terapileri, prebiyotikler, probiyotikler, simbiyotikler, postbiyotikler ve diyet müdahaleleri, mikrobiyomun manipülasyonu için kullanılan/gelecekte potansiyel olarak kullanılabilecek başlıca yöntemler arasında yer almaktadır (Leon-Derecho, 2025).

Son yıllarda insan mikrobiyomunun temel bileşenleri arasında bakteriyom, mikobiyom ve virom öne çıkmaktadır (Shay ve ark., 2020). Bakteriyom, mikrobiyomun en fazla araştırılan bileşeni olup, belli bir habitattaki bakteri topluluklarının tamamını ifade eder. Benzer şekilde mikobiyom, mantar topluluklarının tamamı, virom ise virüslerin tamamı olarak tanımlanır ve bu mikrrorganizmaların mikrobiyomda denge içinde bulunması beklenir (Chen ve ark., 2022). Ancak, mikobiyomda bir dengesizlik (disbiyozis) ortaya çıktığında, herhangi bir mikroorganizmanın aşırı çoğalması, bağışıklık sistemi bozukluklarına, enfeksiyonlara ve diğer sağlık sorunlarına yol açabilir. Mikobiyom, yalnızca bağışıklık yanıtları ve mikrobiyal çeşitliliğin korunmasında değil, aynı zamanda metabolizma, sindirim ve patojen savunmasında da önemli roller üstlenir. Son yıllarda yapılan araştırmalar, mikobiyomun çeşitli hastalıklarla ilişkisini ortaya koyarak, bu topluluklarının sağlığın korunmasında kritik bir öneme sahip olduğunu göstermektedir (Belibasakis ve ark., 2024). Dolayısıyla, bu üç bileşenin her biri, mikrobiyomun insan sağlığı üzerindeki çok boyutlu etkilerini anlamamıza katkı sağlar.

2.1.1. İnsan Viromu

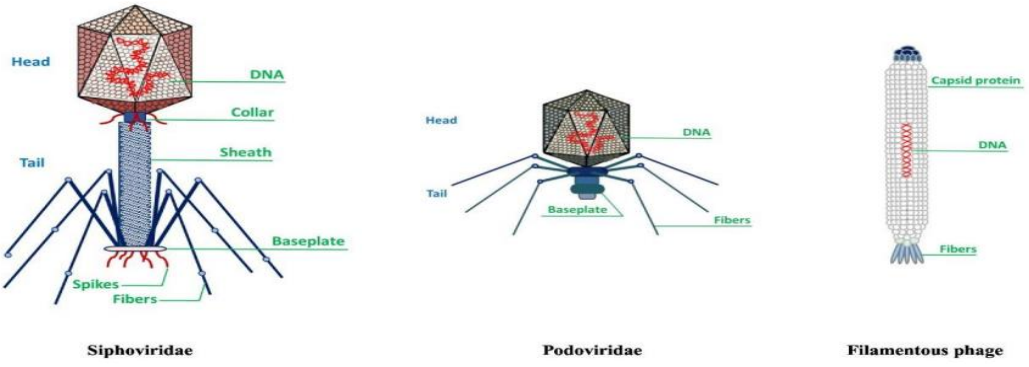
Virom, bir ekosistemde bulunan tüm virüslerin toplamını ifade eder ve mikrobiyomun önemli bir bileşeni olarak kabul edilir. İnsan bağırsağına ait virom, bakteriyofajlar (bakterileri ve arkeleri enfekte eden virüsler), ökaryot virüsleri (insan hücrelerini enfekte eden virüsler) ve endojen viral elementlerden oluşan bileşenlerden oluşur (Bhagchandani ve ark., 2023). Bağırsaktaki diğer mikrobiyom elamanlarıyla karmaşık bir etkileşim içinde olan virom farklı yollarla sağlık ve hastalık durumlarını etkileyebilmektedir. Örneğin bakteriyofajlar, bağırsaktaki bakteriyel popülasyonları düzenleyerek mikrobiyomun kompozisyonunu ve dengesini etkilemektedir. Virom bileşenleri, inflamatuar bağırsak hastalıkları veya otoimmün hastalıklarla ilişkilendirilebilecek konak bağışıklık tepkilerini tetiklemeye ya da baskılamaya neden olabilmektedir. Viromda yer alan bazı virüsler bağırsak bariyer fonksiyonlarını değiştirebilir, inflamasyona yol açabilir veya kanserojen etkiler gösterebilmektedirler. Bunun yanında virom elamanları bakterilerle simbiyotik bir şekilde çalışarak bağırsak homeostazını koruyabilir ve enfeksiyonlara karşı savunma sağlayabilir. Bu dinamik etkileşimler, viromun yalnızca patojen bir yapı olmadığını, aynı zamanda bağırsak ekosisteminde kritik bir düzenleyici rol oynadığını göstermektedir (Wu ve ark., 2024). İnsan viromunu şekillendiren temel faktörler arasında diyet, anne sütü, ilaçlar (örneğin, antibiyotikler ve immünosüpresanlar), konak genetiği, birlikte yaşam, coğrafi konum, altta yatan hastalıklar ve yaşlanma yer almaktadır (Liang ve ark., 2021). İnsan viromu kan, gastrointestinal sistem, solunum sistemi, ürogenital sistem ve deri gibi çeşitli anatomik bölgeler arasında bölgeye özgü çeşitlilik sergiler. Bu bölgeler arasında en yüksek virüs yoğunluğu gastrointestinal sistemde bulunur (ortalama bir yetişkinin bağırsağında gram başına yaklaşık 10^9-10^{10} virüs benzeri partikül (VLP)) ve bir insanda yaklaşık 10^{12} VLP olduğu tahmin edilmektedir (Shkoporov ve ark., 2019). Bunlar, DNA virüsleri (tek sarmallı DNA ve çift sarmallı DNA virüsleri) veya RNA (Ribonükleik Asit) virüslerinden (tek sarmallı RNA ve çift sarmallı RNA virüsleri) oluşmaktadır.

2.1.1.1. Bakteriyofajlar

2.1.1.1.1. Bakteriyofajların Keşfi

Ernest Hankin, 1896 yılında Ganj Nehri sularını analiz ederek bakteriyofajların varlığını ilk kez fark etmiş ve "Yuma ve Ganj suyunun kolera mikrobu üzerindeki bakterisidal etkisi" başlıklı bir çalışma yayımlamıştır (Hankin, 1896). Ancak, bakteriyofajlara ismini veren kişi, onları "bakteri yiyen" olarak tanımlayan Felix d'Herelle'dir (d'Herelle, 1917). O yıllarda enfeksiyon hastalıkları, ölüm oranlarının en önemli nedenlerden biri olması sebebiyle d'Herelle'in keşfi tedavi ve koruyucu hekimlik alanında önemli bir dönüm noktası olmuştur. Ancak antibiyotiklerin keşfiyle birlikte bakteriyofajlara olan ilgi azalmıştır.

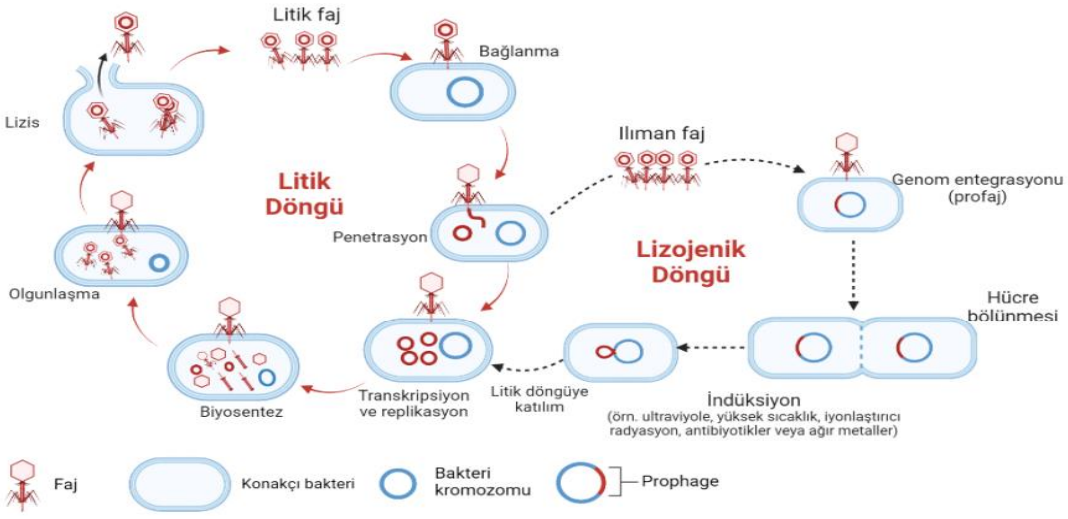
Uluslararası Virüs Taksonomisi Komitesi (ICTV), bakteriyofajları ilk kez 1971 yılında farklı özelliklerine göre sınıflandırmayı denemiştir. Günümüzde Ackermann sınıflandırması yaygın olarak kullanılmaktadır. Bu sınıflandırmaya göre bakteriyofajların %95'ten fazlası, kuyruklu ve çift zincirli DNA'ya (dsDNA) sahip olan Caudovirales takımında yer almaktadır. Bu takım, üç aileye ayrılmaktadır: *Siphoviridae*, *Myoviridae* ve *Podoviridae* (Şekil 2.1). Siphoviridae: Uzun ve esnek kuyruklu olup, Caudovirales takımının %60'ını oluşturur. Myoviridae: Kontraktil kuyruklu olup, %25'lik bir oranla ikinci en yaygın gruptur. Podoviridae: Kısa ve kalın kuyruklu olup, %15'lik bir orana sahiptir (Ackermann, 2007).



Şekil 2.1. Bakteriyofaj yapısının çizimi

2.1.1.1.2. Bakteriyofajların Yapıları ve Yaşam Döngüleri

Tipik bir faj içerisinde genetik materyali bulundurduğu bir baş, boyun, kuyruk ve kuyruk liflerinden meydana gelir. Baş kısmında nükleik asit moleküllerinin birleşerek yumak şeklini oluşturdukları yapı, protein kılıfıyla kaplı bir şekilde bulunur. Bakteriyofajlar genetik materyali içeren nükleik asit ve kapsiti oluşturan nükleokapsid ve DNA'nın bakteriye enjeksiyonunu ve tutunmayı sağlayan kuyruk bölümlerinden oluşur (Ackermann, 2007). Bakteriyofajların litik veya lizojenik hayat döngüleri Şekil 2.2.'de şematize edilmiştir. Litik hayat döngüsü; adsorpsiyon, penetrasyon, biyosentez ve lizis evrelerinden oluşmaktadır (Kutter ve ark., 2005). Son aşama olan liziste ise olgunlaşan yeni fajlar konakçı bakteri hücresini parçalamakta ya da hücre zarında bulunan porlardan dışarı çıkmaktadır (Dowah ve ark., 2018). Lizojenik döngüde ise bakteriyofaj, bakteri hücresi içine girdikten sonra genetik materyalini bakteri DNA'sı ile entegre hale getirmekte ve "profaj" olarak hücre genomunda bulunmaktadır. Profajlar, konakçı bakterinin bölünmesi sırasında bakteri DNA'sı ile birlikte replike olmakta ve yeni bakteri hücrelerine aktarılmaktadır. Bakterinin içinde bulunduğu çevresel koşullar kötüleşmediği sürece bakteriyofaj etkisiz halde varlığını sürdürmektedir. Ancak bakterinin çevresel koşulları bozulduğunda, 5 örneğin besin kaynakları tükendiğinde, bakteri içindeki bakteriyofaj aktif hale gelip çoğalarak bakteriyi parçalamaktadır (Hobbs, 2016).



Şekil 2.2. Litik ve lizojenik döngüde faj

Bakteriyofajlar, genetik çeşitlilikleri ve çoklu döngüler aracılığıyla bakteriyel toplulukları yönlendirebilir ve bunun sonucunda mikrobiyomun dengesini değiştirebilir. Ayrıca, bakteriyofajların yatay gen transferi aracılığıyla bakteriler arasında belli genleri taşıyabileceği ve bu sayede mikroorganizmaların genetik çeşitliliğini artırabileceği de gösterilmiştir. Son yıllarda yapılan çalışmalar, bakteriyofajların inflamatuar başta olmak üzere çeşitli kronik hastalıkların patogenezindeki rolünü vurgulamıştır (Gogokhia ve ark., 2021).

2.1.1.1.3. Bakteriyofajların Uygulama Alanları

Nicastro ve arkadaşlarının 2016 yılında yaptığı çalışma, bakteriyofajların çok yönlü kullanım alanlarını beş ana başlıkta ele almıştır (Nicastro ve ark., 2016). Bu başlıklar şu şekildedir:

i. **Moleküler Biyolojide Kullanımı**:
   - Fajlar, gen vektörleri olarak rekombinant DNA teknolojisinde kullanılır.

ii. **Faj Tiplendirme Uygulamaları**:
   - Fajlar, bakterilerin tanımlanması ve gıda ürünlerinde patojenlerin varlığının tespit edilmesi amacıyla kullanılmaktadır.

iii. **Enfeksiyonların Tedavisinde Kullanımı**:
   - Fajlar, antibiyotiklere alternatif olarak bakteriyel enfeksiyonların tedavisinde kullanılabilir. Yapılan araştırmalar, *Salmonella*, *Shigella*, *E. coli*, *Vibrio cholerae* ve *Staphylococcus* gibi patojenlere karşı etkili olduğunu göstermiştir.

iv. **Biyokoruma Alanında Kullanımı**:
   - Fajlar, gıda kaynaklı patojenlere ve bozulmaya neden olan bakterilere karşı koruyucu bir yöntem olarak uygulanır. Ayrıca, gıda üretim ekipmanlarının yüzey temizliğinde ve biyofilm oluşumunun önlenmesinde kullanılmaktadır.

v. **Faj Gösterim (Phage Display) Teknolojisi**:
   - Faj gösterim teknonolojisi protein-protein, protein-peptit ve protein-DNA etkileşimlerini incelemek için bakteriyofajların kullanıldığı bir yöntemdir.

2.1.1.2. Ökaryotik Hücreleri Enfekte Eden Virüsler (Ökaryotik Virüsler)

İnsan viromunun ökaryotik hücreleri enfekte eden kısmı genellikle insan hücrelerini, daha küçük ökaryotları (örn. protozoalar ve mantarlar) ve diyetle alınabilecek bitki ve hayvan hücre virüslerini içerir (Lecuit ve ark., 2013). İnsan ökaryotik viromu genellikle hastalıklarla ilişkilidir ve tespiti için kültür ve/veya PZR tabanlı yöntemler kullanılmaktadır. Günümüzde yeni nesil dizileme tekniklerinin kullanımıyla, söz konusu virüslerin tespitinde ve tanımlanmasında yeni bir döneme girilmiştir (Ajami ve ark., 2016). Yüksek verimli yeni nesil dizileme, yeni ökaryotik virüs keşfi ve diğer suşlarla karşılaştırma avantajı sunar. Yüksek verimli dizileme, mevcut ve yeni suşların tanımlanmasını sağlamıştır. Örneğin, Avustralya, Virjinya, Nijerya, Pakistan ve Nepal'de çocukların dışkısında yeni astrovirüs suşları keşfedilmiştir (Bosch ve ark., 2014). Bu tür yaklaşımlar, çevresel ve insan örneklerinde yeni suşların hızlı tespiti için moleküler araçların geliştirilmesine, yayılmanın önlenmesine, evrimin anlaşılmasına ve yeni tanı testlerinin geliştirilmesine yardımcı olabilir. *Phycodnaviridae, Herpesviridae, Poxviridae, Mimiviridae* ve *Iridoviridae* ailelerine ait birçok insan ökaryotik virüsü, bebek dışkısı örneklerinde tanımlanmıştır, ancak bakteriyofajlara kıyasla çok daha düşük bağıl bolluklarda bulunmuştur (Pannaraj ve ark., 2018) . Belirli insan ökaryotik virüsleri ayrıca kontamine gıdalar, su ve doğrudan temas yoluyla da bulaşabilir. Bu enfeksiyonların bazıları yaygınken, diğerleri nadirdir. Ayrıca, bazı bireyler semptom geliştirirken, diğerleri asemptomatik kalabilir ve uzun süre virüs yayabilir (Boone ve ark., 2007).

Örneğin, akut hemorajik ateşi olan bireylerde Bas-Congo virüsünün yüksek verimli dizileme ile tanımlanması önemlidir (Grard ve ark., 2012). Akut hemorajik ateş genellikle dengue virüsü, Ebola-Marburg viral hastalıkları, Lassa ateşi, sarı humma, Rift Vadisi ateşi, hantavirüs enfeksiyonları ve Kırım-Kongo hemorajik ateşi gibi hastalıklarla ilişkilendirilir (Wiwanitkit, 2015). Benzer şekilde, yüksek verimli dizileme, etiyolojisi bilinmeyen hastalıklarla ilişkili virüslerin keşfini sağlayabilir. Örneğin, Kawasaki Hastalığı (KD) ilerleyişinde genetik bileşenin önemli olduğu bilinen bir hastalıktır. Ancak, KD ile ilişkili alellerin Doğu Asya ve Kafkas kökenli bireylerde eşit şekilde bulunmasına rağmen, hastalık Doğu Asya popülasyonlarında daha sık görülmektedir. Bu durum sadece genetik yatkınlığın hastalığı açıklamak için

yeterli olmadığına işaret etmektedir. Yapılan çalışmalarda KD ile viral bir etiyolojinin ilişkili olabileceği öne sürülmüş olup yeni nesil dizileme üzerinden kontrol grubunda olmayan torque teno virüs 7'ye (TTV7) ait bir varyant KD hastalarında tespit edilmiştir (Thissen ve ark., 2018).

Diğer örnekler arasında, ensefalit, meningoensefalit ve menenjit gibi hastalıklara neden olan yeni insan astrovirüsleri gibi yeni veya beklenmedik virüsler bulunmaktadır (Zanella ve ark., 2019). İnsan viromunun daha geniş kapsamlı karakterizasyonu, gelecekteki olası salgınları izlemek ve enfeksiyonun altında yatan mekanizmaları anlamak için yardımcı olabilir.

Rotavirüs, norovirüs, enterik adenovirüsler gibi virüsler, bağırsağın inflamasyonunu artırarak bağırsak hastalıklarına yol açabilir (Focà ve ark., 2015) . Bu virüsler, bağışıklık sistemini etkileyerek mikrobiyomun dengesini bozabilirler. Genel olarak, ökaryotik virüsler bir mikrobiyal topluluğun küçük bir kısmını oluşturur, ancak etkileri açıkça gözlemlenebilir. Ökaryotik virüslerin yoğunlukları, semptomatik bireylerden alınan örneklerde asemptomatik bireylere göre genellikle daha yüksektir. İnsan gastrointestinal sistemine yönelik viral metagenomik çalışmaların bir kısmı, ishal şikayeti olan hastaların dışkı örneklerini değerlendirmiştir (Gallardo ve ark., 2024). Bu çalışmalar kapsamında incelenen örneklerde, insan enterovirüsleri, adenovirüsler, kalisivirüsler ve parvovirüsler gibi çeşitli DNA ve RNA virüsleri tespit edilmiştir. Yeni virüslerin tanımlanması, viromun karakterizasyonunun yeni nesil dizileme teknolojileriyle hız kazanmış yönlerinden biridir. Ancak, sekans verileri ortaya konulmuş viral dizilerin çoğu halen tanımlanmamıştır. Bu durum keşfedilmeyi bekleyen büyük bir viral çeşitlilik olduğunu göstermektedir.

Viromdaki bilinen virüslerin varlığı ve dinamikleri üzerine yapılan keşifler, bu virüslerin insan sağlığı üzerindeki etkilerini değerlendirme şeklimizi de değiştirebilir. Örneğin, insan genomuna entegre olan bazı virüsler, kanserle ilişkilendirilmiştir (örneğin, insan papillomavirüs 16, Epstein-Barr virüsü ve daha yakın zamanda keşfedilen Merkel hücre poliomavirüsü). İnsan viromunu karakterize ederken, episomal virüsler ile genom içine entegre olmuş virüsleri ayırt etmek, hastalıkları anlamada önemli bir hedef olabilir. Ayrıca, virom analizleri bilinen virüsleri

beklenmedik dokularda tespit edebilir ve bu durum, hastalıkların yeni mekanizmalarını öne sürebilir (Wylie ve ark., 2012).

2.1.1.3. Endojen Retrovirüsler

İnsan genomunun yaklaşık %8'ini oluşturan ve genoma entegre olabilen endojen retrovirüsler, retroviral genlerin kalıntılarıdır (Santiago-Rodriguez ve ark., 2019). Genellikle defektif olan bu virüslerin transkripsiyonu işlevsel proteinlerin üretimine yol açabilir. Çoğunlukla patojenik olmadıkları düşünülmekle birlikte, bazı otoimmün hastalıklar, tip 1 diyabet, HIV enfeksiyonu, nörodejeneratif hastalıklar, kronik inflamasyon ve bazı kanser türleri (örneğin melanom, meme ve over kanseri) ile belirli endojen retrovirüslerin ilişkilendirilebileceği öne sürülmüştür (Gonzalez-Cao ve ark., 2016; Küry ve ark., 2018; L. Li ve ark., 2013; W. Li ve ark., 2015) .

2.2. Hastalık Mikrobiyom İlişkisi

Günümüzde mikrobiyomun, yalnızca sağlıklı bireylerde yaşamın sürdürülmesindeki temel rollerle sınırlı kalmayıp, aynı zamanda hastalıkların patogenezinde kritik roller oynadığı bilinmektedir. Mikrobiyom ile hastalıklar arasındaki bu karmaşık ilişki, bilimsel araştırmaların odağında yer almakta ve elde edilen bulgular mevcut mevcut tedavi yaklaşımlarını yeniden şekillendirme ve kişiselleştirilmiş tıbbi stratejilerin geliştirilmesine olanak sağlama potansiyeli taşımaktadır.

Çeşitli çevresel, genetik ve yaşam tarzı faktörleri, mikrobiyomda dengesizliklere yol açabilir. Bu dengesizlikler (disbiyozis), kronik hastalıklar başta olmak üzere bir çok hastalığın ortaya çıkmasına, hastalıkların daha hızlı ilerlemesine ve tedavi yanıtının zayıflamasına neden olabilir (Mulder ve ark., 2024). Mikrobiyomda meydana gelen disbiyozisin, bağışıklık yanıtlarının düzensizleşmesine, inflamatuar süreçlerin artmasına ve metabolik bozuklukların gelişmesine katkı sağladığı düşünülmektedir Bağırsak mikrobiyomu, özellikle bağışıklık sistemi ile olan yakın etkileşimi nedeniyle, kronik hastalıkların gelişiminde kritik bir öneme sahiptir. Sağlıklı bir bağırsak mikrobiyotası, bağışıklık sisteminin homeostazını korur ve patojenlere karşı savunma mekanizmalarını destekler. Ancak, mikrobiyomda meydana gelen dengesizlikler, bağışıklık sisteminin aşırı tepki vermesine neden olabilir ve durum inflamasyonun artmasına neden olur. Uzun süreli inflamasyonun ise obezite,

diyabet, kardiyovasküler hastalıklar (KVH), KK, İBH gibi kronik hastalıkların patogenezinde önemli bir rol oynadığı gösterilmiştir (McDonald ve ark., 2024). Örneğin, aşırı vücut yağı birikimiyle karakterize edilen ve kronik inflamasyonla ilişkili bir metabolik hastalık olan obezitenin gelişiminde mikrobiyomun ve viromun önemli rol oynadığı bir çok çalışma ile ortaya konulmuştur (Hu ve ark., 2024). Bu çalışmalarda, bağırsak mikrobiyomunun enerji metabolizmasını, lipid metabolizmasını ve inflamasyonu etkileyerek obezite riskini artırabildiği ve buna ek olarak bakteriyofajların bakteri dengesini etkileyerek ve inflamasyonu artırabileceği ortaya konulmuştur (Cani ve ark., 2007; Karlsson ve ark., 2013). Bir diğer örnek olarak, kan şekeri seviyelerinin kronik olarak yüksek seyretmesiyle karakterize edilen metabolik bir hastalık olan diyabette de bağırsak mikrobiyomundaki disbiyozisin, bağırsak geçirgenliğini artırarak inflamasyonu tetikleyebileceği ve insülin direncine katkıda bulunabileceği gösterilmiştir (Larsen ve ark., 2010; Qin ve ark., 2012). Benzer şekilde, KVH ve mikrobiyom arasındaki ilişki, son yıllarda yapılan araştırmalarla giderek daha fazla ortaya konmaya başlanmıştır (Li ve ark., 2023). Mikrobiyomun metabolik yan ürünleri, örneğin trimetilamin N-oksit (TMAO), ateroskleroz riskini artırabileceği (Koeth ve ark., 2013) ve özellikle bazı virüslerin (örneğin, sitomegalovirüs) inflamasyon ve damar fonksiyon bozukluklarına katkıda bulunduğu düşünülmektedir (Dietrich ve ark., 2013). Otoimmün bir hastalık olan Romatoid artrit (RA)'de ise bağırsak disbiyozunun inflamasyonu tetikleyerek otoimmün yanıtları güçlendirebileceği ve bu durumun eklemdeki hasarı artırabileceği gösterilmiştir (Scher ve ark., 2013).

2.2.1. İnflamatuar Bağırsak Hastalığı ve Mikrobiyom ilişkisi

İBH, bağırsaklarda kronik inflamasyon ile karakterize edilen ve genetik ile çevresel faktörlerin etkileşimi sonucu gelişen bir hastalık grubudur. İBH, iki ana alt gruba ayrılmaktadır; Crohn hastalığı (CH) ve ülseratif kolit (ÜK) 2019 yılı itibarıyla yaklaşık 4,9 milyon kişiyi etkilediği bir tahmin edilen İBH, önemli bir kronik inflamatuar bozukluk olarak değerlendirilmektedir (Villanacci ve ark., 2021).

CH ve ÜK, proinflamatuar sinyaller ve klinik semptomlar açısından örtüşmeler göstermesine rağmen, inflamasyonun şekli, kapsamı, hastalık konumu ve bağırsak mikrobiyotasının bileşimi gibi açılardan farklılık gösterirler (Nishida ve ark., 2018).

Özellikle, bağırsak kommensallerinin değişen kompozisyonu olan disbiyozis, İBH patogenezi ile güçlü bir şekilde ilişkilendirilmiştir (Nishida ve ark., 2018). *Firmicutes* ve *Bacteroidetes* gibi baskın bakteri gruplarında azalma, *Gammaproteobacteria* gibi daha az temsil edilen gruplarda ise artış ile ilişkilendirilmiştir (Kostic ve ark., 2014; VanDussen ve ark., 2015). Örneğin, CH ve ÜK hastalarında bağırsak mikrobiyotasındaki biyoçeşitliliğin azalması sıklıkla bildirilmiştir (Joossens ve ark., 2011). Bununla birlikte, ev içi temasların, aynı ev ortamında yaşayan bireylerin mikrobiyotaları arasında benzerlikler oluşturabileceği ve bağırsak mikrobiyotasının şekillenmesinde çevresel faktörlerin etkili olduğu gösterilmiştir (Lax ve ark., 2014). Bununla birlikte, disbiyozun hastalığın bir sonucu mu yoksa doğrudan nedeni mi olduğu halen tartışma konusudur. Mikroorganizmalar, konak hücrelerle doğrudan etkileşim kurarak, belirli bağışıklık yollarını düzenleyerek ve bağışıklık ile bağışıklık dışı hücreleri aktive ederek İBH patogenezine katkıda bulunabilir. Öte yandan, bağırsak inflamasyonu sonucu bağırsak ortamında meydana gelen değişiklikler de mikrobiyal kompozisyonu etkileyerek disbiyozise yol açabilir (Massimino ve ark., 2023).

Bağırsak mikrobiyotasının, bakteriler, arkeler, mantarlar, protistler ve virüsler dahil olmak üzere trilyonlarca farklı mikroorganizmayı içeren karmaşıklığı, bu süreçlerin mekanizmalarını anlamayı zorlaştırmaktadır. Son yıllarda bakterilerin sağlık ve hastalıktaki rolleri giderek daha iyi anlaşılmaya başlanmış olsa da özellikle virüslerin rolü uzun süre göz ardı edilmiştir. Oysa virüsler, bağırsakta prokaryotlardan yaklaşık 10 kat daha fazla bulunur ve hem bakterilerle (bakteriyofajlar) hem de konak hücrelerle (ökaryotik hedefli virüsler) etkileşim yetenekleri nedeniyle bağırsak homeostazını doğrudan etkileyebilirler (Tun ve ark., 2024). İBH hastalarında yapılan virom analizleri, bağırsak mikrobiyotasında bakteriyofaj çeşitliliğinde anlamlı artışlar olduğunu ve bu artışın azalan bakteriyel çeşitlilik ile ilişkili olduğunu ortaya koymuştur (Norman ve ark., 2015). Bu durum, bakteriyofajların, İBH'de bağırsak mikrobiyotasının dinamiklerini etkileyen kritik bir bileşen olduğunu göstermektedir. Ayrıca, bu virom değişikliklerinin, coğrafi olarak farklı popülasyonlardaki İBH hastalarında da tutarlı bir şekilde gözlenmesi, bakteriyofajların hastalık mekanizmasındaki potansiyel rolünü desteklemektedir (Kostic ve ark., 2014). Bu bulgular, bakteriyofajların yalnızca mikrobiyom dengesindeki değişikliklere aracılık

etmekle kalmayıp, aynı zamanda İBH patogenezinde aktif bir rol oynayabileceği yönündeki hipotezleri güçlendirmektedir.

Bu bağlamda, İBH'nın, bağırsak mikroflorasına karşı bağışıklık sisteminin aşırı tepkisi ile geliştiği ve sindirim sisteminde inflamasyon, doku hasarı ve işlev bozukluklarına yol açtığı bilinmektedir. Mikrobiyomda meydana gelen disbiyoz ve virom dinamiklerindeki değişiklikler, İBH patogenezine katkıda bulunan kritik faktörler arasında yer almaktadır. Özellikle bakteriyofajlar, bağırsak mikrobiyotasını doğrudan etkileyerek bu hastalıkların ilerlemesinde önemli bir role sahip olabilir.

2.2.2. Kolorektal Kanser ve Mikrobiyom ilişkisi

KK, kolonda ve/veya rektumda gelişen malign bir hastalık olup dünya genelinde en sık görülen kanser türlerinden biri olarak öne çıkmaktadır (Da Cunha ve ark., 2024). KK, gelişiminde beslenme alışkanlıkları, inflamasyon, genetik faktörler ve bağırsak mikrobiyotası gibi bir dizi faktörün rol oynadığı bilinmektedir (Bürtin ve ark., 2020). İlginç bir şekilde, KK vakalarının yaklaşık %95'i, genetik yatkınlığı olmayan bireylerde ortaya çıkmaktadır (Heijstek ve ark., 2005). Bu durum, çevresel faktörler ve özellikle bağırsak mikrobiyotasının hastalık üzerindeki etkisine işaret etmektedir. KK'in moleküler nedenleri henüz tam olarak anlaşılamamış olmasına rağmen (Taketo ve ark., 2009), bu konuda yapılan araştırmalar, mikrobiyota-metabolit ilişkilerinin önemli bir role sahip olduğunu ortaya koymaktadır.

KK, dünya genelinde tüm kanser vakalarının yaklaşık %10'unu ve kanserle ilişkili ölümlerin %9'unu oluşturmaktadır (Kersten ve ark., 2017). Vakaların çoğunluğu (%90'ı), 50 yaş üzerindeki bireylerde görülmektedir. Tanı anındaki evreye bağlı olarak değişmekle birlikte, KK için küresel 5 yıllık sağ kalım oranı, tüm evreler için ortalama %65 civarındadır (Heijstek ve ark., 2005). Gelişmiş ülkelerde, gelişmekte olan ülkelere kıyasla daha yüksek insidans ve ölüm oranları görülmekte, bu da çevresel faktörlerin ve yaşam tarzının hastalık üzerindeki etkisine işaret etmektedir. Ayrıca, KK tanı ve tedavi süreçleri ile ilişkili yüksek maliyetler nedeniyle sağlık sistemlerine ciddi ekonomik bir yük oluşturmaktadır.

Bağırsak mikrobiyotası, KK'nın gelişiminde hem zararlı hem de yararlı etkileri olabilen metabolitlerin üretimi ile önemli bir rol oynar. Kısa zincirli yağ asitleri (SCFA) ve amino asit metabolizmasıyla ilişkili poliaminler gibi moleküllerin artan seviyeleri, kanser gelişimi ve metastazıyla ilişkilendirilmiştir (Bürtin ve ark., 2020). Otoimmün hastalıklar da KK gelişimine katkıda bulunan diğer faktörler olarak değerlendirilmiştir (Rourke ve ark., 2017). Mikrobiyom bileşimindeki değişikliklerin, inflamatuar süreçleri artırarak kanser gelişimini tetikleyebileceği düşünülmektedir. O'Rourke ve arkadaşlarının çalışmaları, bağırsak mikrobiyotasının kanser patogenezindeki olası rolünü, diyet faktörlerinin bu mikrobiyota ve KK riski üzerindeki etkisini, ayrıca maligniteyi tedavi etmek için kullanılabilecek mevcut stratejileri detaylı bir şekilde incelemiştir. Bu bulgular, mikrobiyomun, KK'le ilişkili yeni prognostik belirteçlerin ve terapötik hedeflerin geliştirilmesinde potansiyel bir kaynak olduğunu göstermektedir.

Sağlıklı bireylerin bağırsak mikrobiyotası, genellikle daha çeşitli ve dengelidir. Buna karşılık, KK hastalarında bağırsak mikrobiyotasında azalan çeşitlilik gözlemlenir (Wilkinson ve ark., 2021). Araştırmalar, belirli bakteriler ile KK arasında bir korelasyon tanımlamıştır (Wilkinson ve ark., 2021). Son yıllarda yapılan metagenom çalışmaları gösteriyor ki, bağırsak mikrobiyotasındaki bölgesel farklılıklara rağmen KK'de sürekli olarak bağıl bolluğu artan varlık yedi bakteriyi tespit edilmiştir. (Dai ve ark., 2018). Vücudun farklı bölgeleri, bağırsak mikrobiyotasının farklı bileşimlerini sergiler (He ve ark., 2018). Dahası, mukozal bakteriyel biyofilmlerin distal KK ile ilişkili olmayan tümör gelişimini etkilediği bulunmuştur; bu da bağırsak mikrobiyotasının anatomiye göre farklı KK tiplerinde farklı roller üstlendiğini göstermektedir (Peters ve ark., 2016). Ameliyat veya hastalıklardan sonra olduğu gibi yerel çevre değiştiğinde bakterilerin kolonun en uzak kısmında daha fazla miktarda toplanma eğiliminde olmuşlardır (Flemer ve ark., 2017). Ek olarak, distal KK'li kişilerin dışkılarında proksimal KK'lı kişilere kıyasla kanser teşvik edici özelliklerle bağlantılı bakteri genlerinin daha yüksek konsantrasyonlarda olduğu tespit edilmiştir (Feng ve ark., 2015a). Kolonla ilişkili hastalıkların erken teşhisi için gelişmiş, doğru ve oldukça hassas moleküler teknikler geliştirmek için, KK'le ilişkili spesifik bakteri türlerini bilmek esastır. Genetik tümör profillerine benzer şekilde, tümörler içindeki çeşitli mikroorganizmaların tespiti sonunda hasta prognozu ve

tedavisine yardımcı olmak için standart bir test haline gelebilir. KK'le ilgili bakteri türleri *Fusobacterium nucleatum, Bacteroides fragilis, Escherichia coli, Streptococcus gallolyticus, H. pylori, E. faecalis, Parvimonas micra, Prevotella intermedia, Alistipes finegoldii* olduğu bilinmektedir (Kamath ve ark., 2024).

KK'li hastaların mikrobiyotasına yönelik araştırmalar ağırlıklı olarak bakteriyom üzerine odaklanmıştır. KK ile ilişkili bakteriyomun anlaşılması hastalık sınıflandırmasına ve altta yatan patogenez ve hastalık etiyolojisinin daha iyi anlaşılmasına katkıda bulunmuş olsa da (Novielli ve ark., 2024) bakteriler kolon mikrobiyomunun yalnızca bir alt kümesini temsil eder. Kolon mikrobiyomunun bir diğer önemli bileşeni olan virüsler giderek daha fazla tanınan KK'i patogenezine katkıda bulunabilecektir (Massimino ve ark., 2020). İnsana özgü virüslerin, mutajenik ve işlevsel olarak manipülatif yetenekleri nedeniyle bakterilerden daha fazla kanserle ilişkili olduğu bilinmektedir (Chang ve ark., 1994; Feng ve ark., 2008; Schiller ve ark., 2012; Shuda ve ark., 2011). Örneğin, insan papilloma virüsü rahim ağzı kanserine, hepatit C karaciğer kanserine ve Hodgkin dışı lenfomaya yol açabilir. Virüsler, varsayıldığından daha fazla KK ilerlemesinde veya bağışıklık bozukluğunda rol oynayabilir (Nakatsu ve ark., 2018a).

Virüsler, özellikle bakteriyofajlar, bağırsak mikrobiyotasını düzenleyen kritik unsurlardır ve KK patogenezinde rol oynayabilirler (Massimino ve ark., 2020). İnsan mikrobiyomunda bakteriyofajlar, bağırsak viromunun %90'ını oluşturur ve Gram negatif bakteriyel konakçıları, örneğin *Fusobacterium nucleatum*, enfekte etme yetenekleri nedeniyle ilgi odağıdır. *Fusobacterium nucleatum*'un, KK ile ilişkili patogenezde önemli bir rol oynadığı bilinmektedir (Hannigan ve ark., 2018).

Kanser viromunun aracılık ettiği bu tür mikrobiyal etkileşim örnekleri, bakteriyofajların bakteriyel mikrobiyomu şekillendirmede mekanik ve farklı işlevlere sahip olduğunu ve KK karsinogenezinde de potansiyel olarak önemli bir rol oynayabileceğini düşündürmektedir (Ho ve ark., 2024). Virom ve KK arasındaki ilişki, mevcut aşamada karmaşık bir yapıya sahiptir, çünkü KK ile doğrudan ilişkilendirilebilecek tek bir insana özgü patojenik virüsün varlığı olası görünmemektedir. Bununla birlikte, insan papilloma virüsü (HPV) ve serviks kanseri

arasında doğrudan nedensel bir ilişki kurulmuştur (Muñoz ve ark., 2003). Bazı çalışmalar, HPV, insan poliomavirüsleri, Simian virüs 40 ve insan herpes virüsleri gibi somatik virüslerin KK karsinogenezindeki rollerini değerlendirmeye çalışmış, ancak bu çalışmaların çoğu kesinlikten uzak sonuçlar ortaya koymuştur. (Dalianis ve ark., 2013; Khabaz ve ark., 2016; Vilchez ve ark., 2004). Mevcut literatürde öne çıkan ortak bir bulgu, KK ile ilişkili bağırsak viromunun disbiyoz durumudur (Hannigan ve ark., 2018; Nakatsu ve ark., 2018). Bu durum, KK hastalarının, kanser olmayan bireylerle kıyaslandığında benzersiz bir bağırsak mikrobiyal profiline sahip olduğunu göstermektedir. KK'nın gelişiminin, tanımlanmış belirli bir patojenik mikroorganizma tarafından doğrudan tetiklenmekten ziyade, mikrobiyal disbiyoz ve mikrobiyal etkileşimlerin kolektif bir örüntüsüyle yönlendirildiği düşünülmektedir. Ayrıca, KK ile ilişkili bağırsak viromunun disbiyoz durumunun, küratif cerrahinin ardından devam edip etmediği de belirsizliğini korumaktadır. Mikrobiyal disbiyozun KK gelişimini yönlendirdiği varsayılırsa, birincil tümörün rezeksiyonunun KK ile ilişkili disbiyoz modellerini tersine çevirmesi ve kalan kanser dışı kolon mukozasının sağlıklı bireylerin mikrobiyomuna benzer bir profil sergilemesi beklenir. Öte yandan, cerrahiden sonra sağlıklı bireylerin kolon mukozasında devam eden disbiyoz, kolonda önceden var olan ve ilk etapta KK'ye yatkınlık yaratmış olabilecek mikrobiyal bir ortam değişikliğini işaret edebilir. Bakteriyofajların, kanserojen bakteri profillerinin temel itici gücü olduğu ve KK patogenezi ile ilişkili bağırsak bakteri topluluklarının düzenlenmesinde kilit rol oynadığı varsayılmaktadır (Liping ve ark., 2024).

Nakatsu ve arkadaşları, KK hastalarının bağırsak bakteriyofaj topluluğunda çeşitlilikte anlamlı artışlar bulmuş ve bu hastalarda zenginleşmiş 22 viral takson tanımlamıştır. Özellikle, *Inovirus* ve *Tunalikevirus* üyelerinin KK hastalarında zenginleştiği ve bu bakteriyofajların KK gelişiminde etkili olabilecek bakteriyel konakçıları düzenlediği belirlenmiştir (Nakatsu ve ark., 2018b).

2.3. Metagenomik Yaklaşımlar

İnsan vücudu virüsler, bakteriler, arkeler, mantarlar ve protozoalar gibi çok çeşitli mikroorganizmalara ev sahipliği yapar. Bu mikroorganizma topluluklarının hem

birbirleriyle hem de konakla olan etkileşimleri, insan sağlığı ve hastalıkları üzerinde büyük bir etkiye sahiptir (Chatterjee ve ark., 2018).

Mikrobiyom çalışmaların kullanılan teknikler hedeflenmiş gen (amplikon) dizileme ve metagenomik dizileme olarak ikiye ayrılırlar. Hedeflenmiş gen dizileme belirli mikroorganizma gruplarının taksonomik profillemesini, bu gruplar arasında evrimsel olarak korunmuş gen bölgelerinin dizilenmesi yoluyla gerçekleştiren bir tekniktir. Bu yöntemde, bakteriler ve arkealar için 16S rRNA, mantarlar için ise 18S rRNA ve ITS bölgeleri, en yaygın kullanılan hedef diziler olarak öne çıkmaktadır. Metagenom dizileme ise bir ortamda bulunan mikroorganizmaların tamamının genomlarını (metagenomlar) dizilemek için kullanılan bir tekniktir.

Metagenomik yaklaşımlar, mikrobiyota ve mikrobiyomun derinlemesine incelenmesini mümkün kılan devrim niteliğinde yöntemler sunmaktadır. Bu yaklaşımlar, mikroorganizmaların laboratuvar ortamında kültüre edilmesi zorunluluğunu ortadan kaldırarak, özellikle mevcut koşullarda kültürlenemeyen mikroorganizmaların analiz edilmesine olanak tanır. Gün geçtikçe yaygınlaşan bu moleküler yöntemler, mikrobiyolojide köklü bir değişimi beraberinde getirmiştir (Rinke ve ark., 2014; Tringe ve ark., 2005). Metagenom çalışmalarının temelinde, çevresel örneklerden doğrudan nükleik asitlerin izole edilmesi ve izole edilen nükleik asitlerin yeni nesil dizileme yöntemleriyle (Next Generation Sequencing, NGS) dizilenmesi yer almaktadır. Bu süreçler sayesinde, ilgili çevrenin mikrobiyal ekosistemi ayrıntılı bir şekilde ortaya çıkarılabilir ve farklı ortamlar karşılaştırılabilir. (Zhang ve ark., 2021).

DNA dizileme teknolojileri, 1977 yılında Frederick Sanger tarafından geliştirilen Sanger dizileme yöntemi ile başlamıştır (Sanger ve ark., 1975). Birinci nesil dizileme teknolojisi olarak kabul edilen bu yöntem, zincir sonlandırma tekniğine dayanarak DNA dizilerinin belirlenmesini sağlamış ve genomik araştırmalarda uzun yıllar boyunca altın standart olarak kabul edilmiştir. Sonraki yıllarda floresan tespit teknolojilerinin ve kapiler elektroforez tekniklerinin geliştirilmesi, DNA dizilemesinde otomasyon çağını başlatmıştır. 20. yüzyılın sonunda, otomatik plaka elektroforezi ve kapiler elektroforez temelli dizileme yöntemleri tanıtılmış ve özellikle İnsan Genomu Projesi (HGP) kapsamında yürütülen çalışmalarda kritik bir

rol oynamıştır. Birinci nesil dizileme teknolojileri, %99,999 doğruluk oranıyla yaklaşık 1000 baz çifti uzunluğundaki DNA dizilerini okuyabilme kapasitesine sahiptir ve bu özelliğiyle altın standart yöntem olarak kabul edilmektedir. Bununla birlikte, yüksek maliyet (1000 baz başına yaklaşık yarım dolar) ve düşük verimlilik (dizileme turu başına yalnızca 96–384 örnek) gibi dezavantajları, geniş ölçekli uygulamalarını kısıtlamaktadır. Sanger dizileme yöntemi, modern dizileme tekniklerinin gelişmesine rağmen, günümüzde hala belirli uygulamalarda yaygın olarak kullanılmaktadır.

Sanger dizilemenin düşük verimliliği, yüksek maliyeti ve zaman alıcı olması nedeniyle büyük ölçekli genom projelerinde sınırlı kalması sebebiyle, 2000'li yılların başında, İnsan Genom Projesi'nin tamamlanmasının ardından, dizileme teknolojilerinde büyük bir dönüşüm gerçekleşmiş ve İkinci Nesil Dizileme (Next-Generation Sequencing, NGS) teknolojileri geliştirilmiştir. Bu dönemde Illumina (Solexa), Roche 454, SOLiD (Applied Biosystems) ve Helicos gibi platformlar, paralel dizileme yetenekleri sayesinde DNA dizileme hızını ve kapasitesini önemli ölçüde artırmıştır. NGS yöntemleri, klonal amplifikasyon ve yüksek derecede paralelizasyon ilkelerine dayanarak aynı anda milyonlarca DNA fragmanının sekanslanmasına olanak tanımış ve dizileme maliyetlerinde çarpıcı bir düşüş sağlamıştır. Özellikle Illumina platformları, kısa okuma (short-read) dizileme stratejisiyle günümüzde en yaygın kullanılan NGS yöntemi haline gelmiştir. 2010'lu yıllarda, Üçüncü Nesil Dizileme teknolojileri geliştirilmiş ve DNA dizileme sürecinde yeni bir aşamaya geçilmiştir. Pacific Biosciences (PacBio) ve Oxford Nanopore Technologies (ONT) tarafından geliştirilen bu yöntemler, tek molekül dizileme (Single-Molecule Real-Time, SMRT) ve nanopor tabanlı dizileme gibi yenilikçi yaklaşımlar sunarak uzun okuma (long-read) dizileme yeteneklerini mümkün kılmıştır. Uzun okuma teknolojileri, yapısal varyant tespiti, genom montajı ve tekrarlayan bölgelerin analizinde büyük avantaj sağlamış ancak yüksek hata oranları nedeniyle halen iyileştirilmesi gereken bazı teknik sınırlamalara sahiptir. Günümüzde, NGS teknolojileri, klinik tanı, kanser genomikleri, mikrobiyom araştırmaları, farmakogenomik ve kişiselleştirilmiş tıp gibi birçok alanda yaygın olarak kullanılmaktadır. Dördüncü nesil dizileme olarak adlandırılan yeni teknolojiler, daha hızlı, taşınabilir, düşük maliyetli ve yüksek doğruluk oranına sahip

sekanslama yöntemleri geliştirmeyi hedeflemektedir. Gelecekte, yapay zekâ destekli veri analizi, doğrudan RNA dizileme ve entegre laboratuvar sistemleri gibi yeniliklerle NGS'nin klinik kullanımı daha yaygın hale gelecek ve genom temelli tıbbi yaklaşımlar önemli ölçüde ilerleyecektir.

2.3.1. Viral Metagenomik

Günümüzde bakteri topluluklarını incelemek için gelişmiş tekniklerin bulunması nedeniyle mikrobiyom çalışmaları genellikle bakterilere odaklanmaktadır. Ancak, yüksek verimli nükleik asit dizileme teknolojilerinin ilerlemesi, insan viromundaki virüslerin tespitini, tanımlanmasını ve karakterize edilmesini kolaylaştırmıştır. 2002 yılında çevresel viral popülasyonların analizi için ilk kez kullanılan shotgun metagenom dizileme yöntemi, bu alanda önemli bir adım olarak kabul edilmektedir (Breitbart ve ark., 2002). Artan sayıda viral metagenomik çalışmada, daha önce bilinmeyen virüsleri, yani "karanlık maddeyi (dark-matter)" tespit etmek için shotgun metagenom dizileme yöntemleri kullanılmaya başlanmıştır (Cantalupo ve ark., 2019). Yeni keşfedilen birçok virüs, mevcut veri tabanlarına eklenmiştir. Ocak 2024 itibarıyla Ulusal Biyoteknoloji Bilgi Merkezi Genom veri tabanı, 11.703 tam viral genom dizisi içermektedir (Tan ve ark., 2024). Bu sayı, tüm viral genomların yalnızca küçük bir bölümünü temsil etse de kamuya açık veri tabanlarının sürekli güncellenmesi, virüs tanımlama platformlarının hassasiyetini artırmaktadır.

Son zamanlardaki patojen tespiti için metagenomiğin kullanımına ilişkin incelemeler de dahil olmak üzere insan viromuyla ilgili çalışmalar artması sebebiyle (Gu ve ark., 2019; Miller ve ark., 2013),viral metagenomik boru hatları üzerine çalışmalarda artış olmuştur (Rose ve ark., 2016).

Viral metagenomik çalışma süreçleri örnek toplama, örnek işleme, dizileme ve biyoinformatik analizleri içerir. Örnek toplama ve muhafaza yöntemleri kan, dışkı, idrar veya toprak gibi örnek türlerine bağlı olarak farklılık gösterir. Viral nükleik asitleri ve viromu etkileyen ana faktörler örneklerin saklama koşullarıdır. Saklama sıcaklığı, saklama tampon içerikleri ve örneklerin dondurulup çözülme döngüsüne girip girmediği gibi koşullar viral nükleik asitlerin korunmasında önemlidir (Santiago-Rodriguez ve ark., 2020). Bunlara ek olarak, viral metagenomik

çalışmalarına ait ıslak laboratuvar süreçleri her adımda kontaminasyon risklerini içermektedir. (Cantalupo ve ark., 2019). Bu nedenle, örnek işleme viral metagenomik analizlerde kritik bir adımdır. Viroma ait nükleik asit saflaştırılma protokolleri başlangıç materyal miktarı, kullanılan tamponlar ve filtre gözenek boyutları açısından farklılık gösterir. Viral nükleik asit ekstraksiyonu ve amplifikasyonu sonrasında NGS teknolojileri kullanılarak dizileme işlemleri yapılır.

Genom birleştirme bir organizmaya ait kısa dizilimlerin daha uzun ve bitişik dizilere dönüştürülmesi sürecini ifade etmektedir. Bu süreçte, birleştirilen diziler daha kapsamlı genomik analizler yapılmasına ve biyolojik anlam çıkarılmasına olanak sağlamaktadır. Viral metagenomik çalışmalarda kısa okuma dizilimlerinin daha uzun ve bitişik viral genomlara dönüştürülmesi amacıyla birleştirme (assembly) işlemi kritik bir rol oynamaktadır. Bu süreç hem viral genomların kapsamlı şekilde karakterize edilmesini hem de metagenomik verilerden yeni viral türlerin tanımlanmasını sağlamaktadır Viral metagenomikte birleştirme genellikle referans tabanlı birleştirme ya da *de novo* birleştirme olmak üzere iki şekilde yapılabilmektedir. Referans tabanlı birleştirmede metagenomik okumalar mevcut viral referans genomlarına hizalanarak birleştirilir ve bu yöntem bilinen viral türler için oldukça verimlidir (Sutton ve ark., 2019). Ancak, düşük çeşitlilikte veya eksik referans genomlara sahip viral topluluklar için sınırlı kalabilmektedir. *de novo* birleştirmede ise, referans genomu olmaksızın okumalar arasındaki bağlantılar ve örtüşmeler kullanılarak birleştirme yapılmaktadır. Bu yöntem karmaşık viral metagenomlarda yeni viral türlerin veya genomik segmentlerin keşfi için kritik bir araçtır. Bununla birlikte, kısa okuma uzunlukları veya düşük veri kalitesi durumunda montaj zorlukları ile karşılaşabilmektedir. Daha uzun viral genom montajları elde etmek, viral metagenomik analizlerin kalitesini ve kapsamını artırmaktadır (Sutton ve ark., 2019) Bunun birkaç önemli avantajı vardır. İlk olarak daha uzun montajlarviral türlerin veya suşların doğru bir şekilde tanımlanma olasılığını artırmaktadır ve bu durum özellikle kısa okuma dizilerinin hizalanması sırasında oluşabilecek hataları minimize etmektedir. *de novo* birleştirme metagenomik veri kümelerinde yeni ve karakterize edilmemiş viral genomları keşfetmeyi mümkün kılmaktadır. Daha uzun montajlar, viral genlerin ve işlevsel bölgelerin eksiksiz bir şekilde anotasyonunu kolaylaştırmaktadır.

Viral genomlar genellikle kısa, mozaik yapıda ve segmentli olabildiğinden metagenomik birleştirme işlemi bakteriyel veya ökaryotik genomlara göre daha zordur. Buna ek olarak, düşük viral bolluk, birleştirme sırasında eksik montajlara neden olabilmekte, viral genomların büyük genetik çeşitliliği, ortak dizilim motiflerinin azlığı nedeniyle birleştirmeyi zorlaştırabilmekte ve çift iplikli ve tek iplikli DNA/RNA virüslerinin farklı yapısal özellikleri, birleştirme stratejilerini optimize etmeyi gerektirmektedir.

2.3.1.1. Viral Metagenomikte Taksonomik Tanımlama ve Zorluklar

Viral metagenomik bakteriyofajlar ve diğer virüs popülasyonları dahil olmak üzere insan viromunun çeşitliliğini ve karmaşıklığını anlamamızı sağlamaktadır. Bu teknoloji, insan sağlığıyla ilişkili yeni viral türlerin keşfi ve viromun konakçı organizma ile etkileşimlerinin anlaşılması için benzersiz bir olanak sunmaktadır.

Taksonomik tanımlama viral metagenomik çalışmaların temel bileşenlerinden biridir ve bu amaçla çeşitli yöntemler geliştirilmiştir. Kullanılan yöntemler hız ve hassasiyet açısından farklı avantaj ve dezavantajlara sahiptir. BLAST, Bowtie ve Diamond gibi araçlar hizalamaya dayalı yöntemler olarak dizi veri tabanlarındaki nükleotit hizalamalarına dayanır. Yüksek hassasiyet gösteren ve doğruluk açısından oldukça hassas olan bu yöntemler için yavaş çalışma süreleri ve kaynak yoğunluğu dezavantaj olarak değerlendirilmektedir. Kompozisyona dayalı sınıflandırma yöntemlerinde ise k-mer eşleşmesi ve nükleotit frekanslarına dayanarak hızlı sınıflandırma sağlanmaktadır. Hızlı ve hesaplama açısından daha verimli olmasına karşın hassasiyetin düşmesi, özellikle düşük benzerlikteki dizilerde doğruluk eksikliği ortaya çıkmaktadır.

Açık okuma çerçevelerini (ORF) kullanarak yeni viral genomları tanımlamayı sağlayan protein yapısına dayalı yaklaşımlarda ise bilinmeyen viral dizileri daha fazla tanımlama potansiyeli bakımından avantajlı ancak analiz sürecinde yüksek bilgi işlem kapasitesi gerekliliği sebebiyle dezavantajlı olarak görülmektedir. Son olarak makine öğrenimi tabanlı yöntemler, DeepVirFinder gibi araçlar, derin öğrenme (Deep Learning, DL) algoritmaları kullanarak metagenomik veri setlerindeki viral dizileri tespit etmektedir (Ren ve ark., 2020). Bu yöntemler referans ve hizalamaya

bağımlı olmayan sınıflandırma sunarken, eğitim verilerindeki önyargılar ve doğruluk standardizasyonunun eksikliği yöntemin dezavantajı olarak değerlendirilmektedir.

Viral metagenomik analizlerde birkaç teknik ve metodolojik zorluk bulunmaktadır. Bunların başında kontaminasyon gelmektedir. Özellikle örnek toplama ve/veya ıslak laboratuvar süreçlerinden kaynaklanan kontaminasyon, analizlerde yanlış sonuçlara yol açabilmektedir. Düşük bağıl bolluk bir diğer zorluk olarak karşımıza çıkmaktadır. Viral genom parçaları genellikle konakçı veya bakteri DNA'sına göre daha az miktarda bulunmaktadır ve bu durum viral dizilerin yeniden yapılandırılmasında belirsizliklere neden olabilmektedir. Yazılım uygulamalarında kullanılan test veri kümelerinin standardizasyon eksikliği, sonuçların karşılaştırılabilirliğini zorlaştıran bir diğer faktördür (Rose ve ark., 2016). Birçok virüsün, laboratuvar ortamında henüz kültüre edilememiş olması virom üyelerinin biyolojik işlevlerinin ve replikasyon mekanizmalarının anlaşılmasını sınırlaması bakımından zorluk olarak değerlendirilmektedir. Son olarak, metagenomik çalışmalar genellikle DNA dizilemesine odaklanmaktadır. Bu durum, hem RNA virüslerinin daha az çalışılmasına sebep olmaktadır. Buna ek olarak, RNA'nın düşük stabilitesi ve RNA virüslerinin tespitindeki teknik zorluklar RNA virüslerinin gerçek bileşimini ortaya çıkarmayı sınırlamaktadır (Garmaeva ve ark., 2019). Tüm bu zorluklar, viral metagenomik çalışmalarının belli ölçüde kısıtlanmasına sebep olmaktadır.

2.4. Mikrobiyom ve Makine Öğrenimi

Mikrobiyom araştırmalarında üretilen verinin miktarı ve çok sayıda öznitelik üretim ihtiyacı bu araştırmalarda Makine Öğrenmesi (Machine Learning, ML) tekniklerinin kullanılmaya başlanmasına a neden olmuştur. Özellikle hasta fenotiplerinin tahmini gibi çalışmalar, ML algoritmaları aracılığıyla başarıyla gerçekleştirilmiştir. Bu algoritmalar, artan veri setlerinden anlamlı desenler öğrenerek, örnekleri çeşitli özelliklerine göre sınıflandırma yeteneği göstermektedir.

ML, çeşitli disiplinlerden elde edilen karmaşık veri örüntülerini tanımlamada etkili bir yöntem olarak kabul edilmektedir. ML'in temel terminolojisi ve gereklilikleri konusunda şu kavramlar öne çıkmaktadır: (1) veri seti (eğitim ve test); (2) örnek; (3)

öznitelik; (4) sınıf etiketi ve (5) çapraz doğrulama. Söz konusu kavramlar aşağıdaki gibi özetlenebilir:

1. **Veri Seti**: Özniteliklere karşı örneklerin matriksel bir yapıda sunulmasıdır. Eğitim veri seti, model oluşturmak ve eğitmek için kullanılırken; test veri seti, modeli değerlendirmek ve hataların tahmini için kullanılır.
2. **Örnek**: Vaka, kayıt veya numune olarak da adlandırılan örnekler, modelin eğitim ve test aşamalarında kullanılan birimlerdir. Örneğin, kanser çalışmaları için toplanan örnekler, hasta ve sağlıklı kontrol grupları.
3. **Öznitelik**: Değişken ya da özellik olarak da bilinen öznitelikler, bir örneği tanımlayan niceliksel ya da kategorik değişkenlerdir. Öznitelik vektörleri bir öznitelik uzayı oluşturur. Kanser biyolojisinde bu, genomik veriler, metagenomik veriler, transkriptomik veriler, proteomikler, metabolomik veriler veya klinik parametreler gibi çeşitli biyolojik ve klinik verilere dayalı olabilir.
4. **Sınıf Etiketi**: Bir örneğin hangi kategoriye ait olduğunu belirten etikettir. Bu etiketler, hasta veya sağlıklı kontrol gibi temel sınıflar ya da hastalığın farklı evrelerini temsil edebilir.
5. **Çapraz Doğrulama**: Bir modelin hata oranlarını ve genelleme kapasitesini ölçmek için kullanılan bir yöntemdir. K-fold, birini dışarıda bırakma (LOO), holdout ve bootstrapping gibi farklı yaklaşımlar mevcuttur. K-fold çapraz doğrulamada, veri seti k alt küme olarak bölünür ve her alt küme bir test veri seti olarak kullanılırken, geri kalan k-1 küme eğitim için kullanılır. Alt küme doğruluklarının ortalaması, modelin genel başarısının bir tahmini olarak kabul edilir (Jagga ve ark., 2015).

Mikrobiyom alanında ML algoritmaları fenotiplerin tahmini, mikrobiyal özelliklerin sınıflandırılması, karmaşık mikrobiyal etkileşimlerin analiz edilmesi ve mikrobiyom bileşimindeki değişikliklerin izlenmesi gibi çeşitli problemlerin çözümü için kullanılmaya çalışmaktadır. ML algoritmaları; denetimli öğrenme, denetimsiz öğrenme ve pekiştirilmeli öğrenme olmak üzere üç ana kategoriye ayrılır.

- **Denetimli Öğrenme**: Sınıf etiketi mevcut olan veri setlerinde, sınıflandırma veya regresyon gibi problemler için uygulanır.
- **Denetimsiz Öğrenme**: Sınıf etiketi mevcut olmadığı durumlarda, kümeleme, yoğunluk tahmini veya boyut indirgeme gibi görevlerde kullanılır.
- **Pekiştirilmeli Öğrenme**: Modelin, bulunduğu ortamla etkileşim kurarak maksimum ödül elde etmek üzere çeşitli aksiyonları belirlemesine dayalı bir yaklaşımdır.

ML ile geliştirilen modeller yeni örneklerin veya hastaların klinik fenotiplerinin tahmininde kullanılabilir. Mikrobiyomdan hastalık teşhisi için bir model oluşturulması aşamalarında ilk adım, çeşitli çevresel ya da konak ile ilişkili mikrobiyal topluluklardan dizileme verilerinin elde edilmesidir.

Klasik ML yöntemleri arasında en popüler yöntemlerden olan lojistik regresyon (LR), rastgele orman algoritması (RF) ve destek vektör makineleri (SVM) yöntemlerinin mikrobiyom verilerinde iyi performans gösterdiği bilinmektedir. Ancak, son zamanlardaki çalışmalarda, yeni yöntemlerle kıyaslama aracı olarak evrilip kullanımları azalmıştır (Hernández Medina ve ark., 2022) . SVM'ler, giriş verilerini iki ayrı sınıfa ayırabilmek için, vektör uzayında iki sınıfın veri noktaları arasında mümkün olan en büyük marja sahip bir hiper düzlem hesaplayarak çalışır (Cortes ve ark., 1995). Hiper düzlem, içinde bulunduğu vektör uzayından bir boyutu daha az olan bir alt uzay olarak tanımlanır. Böyle bir hiper düzlem varsa, veri doğrusal olarak ayrılabilir. Gerçek dünya verilerinin çoğu, doğrusal bir şekilde ifade edilemeyen boyutlar, aykırı değerler veya karmaşık bağımlılıklar nedeniyle tamamen doğrusal olarak ayrılabilir değildir. Bu nedenle, çoğu SVM uygulaması, bu tür hatalara izin verir ve aykırı değerlerin getirdiği hatayı en aza indirmeye çalışırken aynı zamanda doğrusal olarak ayrılabilir veri noktaları arasındaki marjı en üst düzeye çıkarmaya çalışır. Karar ağaçları, her düğümün bir öngörüye göre veri kümesini iki alt kümeye böldüğü bir ağaç yapısıdır. Bu işlem, veri kümesini daha küçük alt kümelere bölen alt düğümlerle tekrarlanabilir. Bir karar ağacı, ağacın yapraklarına sınıf etiketleri atayarak bir sınıflandırıcı olarak kullanılabilir. Düğümler, ağacın her seviyesindeki tahmin değerini en üst düzeye çıkaracak şekilde oluşturulmalıdır; yani boyutlar ve değerler, veri noktalarını hedef sınıflara en iyi şekilde ayıracak şekilde

seçilmelidir. Sınıflandırıcı olarak bir karar ağacı kullanmak, ağaç verimli bir şekilde oluşturulabildiğinden sınıflandırıcı oluşturmanın en hızlı yollarından biridir. Ancak, karar ağaçları birçok gerçek dünya problemi için çok basittir ve aşırı öğrenme (overfitting) eğilimlidir, bu da ağacın eğitim verilerini sınıflandırabilmesine rağmen yeni verileri doğru bir şekilde sınıflandıramayacağı anlamına gelir (Krause ve ark., 2021).

RF yöntemi, karar ağacı sınıflandırıcısının eksikliklerini gidermeye çalışır. Bir değil birkaç (genellikle yüzlerce veya binlerce) karar ağacı kullanır ve ardından nihai sınıflandırma kararını çoğunluk oyuna dayandırır. Aynı boyutları kullanarak, aynı veri kümesinden birden fazla karar ağacı oluşturmak, sınıflandırma performansını iyileştirmeden aynı veya çok benzer ağaçlarla sonuçlanacağından, ağaçlar için özel bir yapılandırma mantığı kullanılmalıdır. Bu yapılandırma mantığı, her ağaç için toplam verinin yalnızca bir alt kümesini kullanmaktan ve tahminlerin seçilebileceği boyutları, oluşturulan ağaçların her seviyesindeki toplam boyutların rastgele bir alt kümesiyle sınırlamaktan oluşur. Ağaçların bu şekilde oluşturulmasının, çok sayıda ağaç kullanılsa bile aşırı uyum riskini önlediği söylenebilir.

Gradyan artırma (Gradient boosting, GB), varyansı azaltmak için kullanılan yaklaşımlardan biridir. RF'nin aksine, GB ağaçlar modeli sırayla oluşturulur ve her ağaç, önceki ağacın hatasını azaltmayı amaçlar. Model/tahmin, sabit bir değerle başlatılır ve her bir yinelemede bir ağaç eğitilir. Her yineleme sonunda, yanlış sınıflandırılan örneklerin ağırlıkları artırılır ve doğru sınıflandırılan örneklerin ağırlıkları azaltılır. Dolayısıyla, her bir ardışık sınıflandırıcı, önceki adımlarda sınıflandırılması zor olan örneklere odaklanır. Yinelemeler sonunda sınıflandırıcı tahminleri, ağırlıklı çoğunluk oyu ile nihai bir tahminde birleştirilir (Teixeira ve ark., 2024). GB yöntemlerinin çeşitli hastalıklar üzerinde karar ağaçlarına göre daha iyi performans gösterdiği ve RF ile yakın sonuçlar verdiği gösterilmiştir (Topçuoğlu ve ark., 2020). LR giriş vektörünün ağırlıklı bir toplamını hesaplar ve daha sonra bu toplama "lojistik fonksiyon" olarak da bilinen sigmoid fonksiyonunu uygular. Sigmoid fonksiyonunun sonucu, bir veri noktasının belirli bir sınıfa ait olma olasılığı olarak yorumlanabilir. Bir LR sınıflandırıcısının eğitimi, giriş vektörü ağırlıklarını, çıktı fonksiyonunun eğitim veri kümesiyle en iyi şekilde eşleşmesi için ayarlamakla

gerçekleşir. Sınıflandırıcıdaki lojistik fonksiyonuna tek giriş, girdi değerlerinin doğrusal bir birleşimi (ağırlıklandırılmış toplam) olduğundan, sınıflandırıcının iyi performans göstermesi için giriş verilerinin çoğunlukla doğrusal olarak ayrılabilir olması gerekir. LR sınıflandırıcısı ile gerçekleştirilen hesaplamalar, aktivasyon fonksiyonu olarak sigmoid fonksiyonunu kullanan bir sinir ağının tek bir nöronundaki hesaplamalarla aynıdır (Krause ve ark., 2021). LR, basitliği nedeniyle mikrobiyomu inceleyen çalışmalarda, özellikle öznitelik seçimi ve kıyaslama için yaygın olarak kullanılan doğrusal bir sınıflandırma modelidir. LR sınıflandırıcıları doğrusal karar sınırlarına sahip olmalarına rağmen, mikrobiyomla ilişkili konak tahmininde doğrusal olmayan muadillerine benzer sonuçlar elde edebilir (Arabameri ve ark., 2020). Lasso ve Elastik ağlar (ENet) gibi LR yöntemleri, fenotip gibi bir çıktıyı, girdilerin doğrusal bir kombinasyonu olarak modelleyerek bu yöntemlerin yorumlanmasını basit hale getirir. Bu yöntemler son zamanlarda konak disbiyoz tahmini çalışmalarında kullanılmış olup, RF gibi diğer yöntemlerle karşılaştırılabilir sonuçlar vermiştir (Pasolli ve ark., 2016).

Son yıllarda, yeni nesil dizileme teknolojilerindeki ilerlemeler, mikrobiyal toplulukların kapsamlı bir şekilde incelenmesine olanak sağlamaktadır. Metagenomik araştırmaların artan uygulama alanları, dizileme maliyetlerinin düşmesi ve mikrobiyomun işlevsel özelliklerinin daha iyi anlaşılmasına yönelik artan ihtiyaç, genetik veri üretiminde hızlı bir yükselişi tetiklemiştir. Nitekim tek bir mikrobiyom çalışması bile yüzlerce gigabayt ham dizileme verisi içerebilir. Ayrıca, analiz sürecinde farklı formatlarda ara sonuçlar üretilmesi gerektiğinden, veri miktarı katlanarak artabilir. Dolayısıyla, bu büyüklükteki verileri işleyebilecek algoritmaların ve mimarilerin geliştirilmesi kritik bir ihtiyaçtır (Krause ve ark., 2021).

Metagenomik verileri miktarının giderek artması ve ilgili araştırmaların durum analizinden çok tanı/tedavi odaklı olmaya başlaması, geleneksel analiz yöntemlerinin sınırlarına ulaşmasına neden olmuştur. Bu noktada yapay zekâ temelli yöntemler meteganom analizlerinde yerini almaya başlamıştır. Yapay zekânın bir alt dalı olan derin öğrenme tabanlı yaklaşımlar, mevcut metotların zayıflıklarını gidermekle kalmayıp, mikrobiyom araştırmalarında yeni perspektifler sunmaktadır. Özellikle bağırsak mikrobiyomu temelli çalışmalarda hastalık tanısı, prognozu ve tedavi

hedefleri için önemli bilgiler elde edilmektedir. Ancak referans kataloglarının sınırlılığı, veri heterojenliği ve özellikle belli mikroorganizmaların düşük bağıl bollukta örneklenmiş olması gibi nedenlerle metagenomik veri analizinde önemli zorluklar yaşanmaktadır. Bu noktada, derin öğrenme yöntemleri yeni yaklaşımlar sunarak bu zorlukların üstesinden gelme potansiyeli barındırması bakımından değerlidir. Son zamanlarda yapılan çalışmalar yeni patojen tespiti, hastalık sınıflandırması ve hastalık tahmini gibi çeşitli alanlarda yapay zeka tabanlı yöntemlerın alana önemli katkı sağlayabileceği gösterilmiştir (Çiftcioğlu ve ark., 2024; Çiftcioğlu ve ark., 2025; Karagoz ve ark., 2024).

3. GEREÇ VE YÖNTEM

3.1. Gereç

Bu tez çalışması, Erciyes Üniversitesi Genom ve Kök Hücre Merkezi (GENKÖK) bünyesinde bulunan Metagenom ve Biyoinformatik Laboratuvarı'nda gerçekleştirilmiştir. Çalışmada yer alan biyoinformatik analizler yüksek hesaplama kapasitesine sahip, 88 işlemcili 512 GB bellekli makine (sunucu) kullanılarak gerçekleştirilmiştir.

3.2. Yöntem

3.2.1.Veri Setleri

Bu tez çalışması, KK ve İBH tanısı alan bireyler ile sağlıklı kontrollerden elde edilen mikrobiyom örneklerine ait metagenomik verileri kullanılmıştır. Araştırma kapsamında bilimsel erişime açık Ulusal Biyoteknoloji Bilgi Merkezi (NCBI) veri tabanından derlenen veri setleri, hastalık gruplarına özgü farklılıkların belirlenmesi ve bu farklılıkların sağlıklı bireylerle karşılaştırılması amacıyla analiz edilmiştir.

KK ve İBH gruplarına ait metagenomik veriler, ilgili hastalık gruplarını temsil eden bilimsel çalışmalardan alınmıştır. Tüm veri setleri, uluslararası veri tabanlarından (NCBI) temin edilmiş, tek bir çalışma dizini altında düzenlenmiş ve analiz için işlenmiştir. Detaylar aşağıda sunulmaktadır:

3.2.1.1. Kolorektal Kanser Veri Setleri

KK hastaları ve sağlıklı kontrollerden alınan mikrobiyom verileri aşağıdaki projeler aracılığıyla elde edilmiştir:

I. PRJEB7774 (NCBI) (46 hasta / 110 Kontrol-366 Gbp) (Q. Feng ve ark., 2015),
II. PRJEB27928 (NCBI) (70 Hasta / 60 Kontol- 364 Gbp) (Wirbel ve ark., 2019),
III. PRJEB12449 (NCBI) (203 Hasta / 223 Kontrol- 862 Gbp) (Ma ve ark., 2021)
IV. PRJEB10878 (NCBI) (64 Hasta / 53 Kontrol- 283 Gbp) (Yu ve ark., 2017)

3.2.1.2. İnflamatuar Bağırsak Hastalığı Veri Setleri

İBH tanısı alan bireylerden ve sağlıklı kontrollerden alınan metagenomik veriler şu projelerden temin edilmiştir:

I. PRJEB1220 (NCBI) (311 Hasta / 211 Kontrol - 1000 Gbp) (Nielsen ve ark., 2014)
II. PRJNA400072 (NCBI) 157 Hasta / 56 Kontrol - 460 Gbp (Franzosa ve ark., 2019)

Bu veri setleri, KK ve İBH'nin mikrobiyal topluluklar üzerindeki etkilerini ve sağlıklı bireylerle arasındaki mikrobiyom farklılıklarını incelemek için kullanılmıştır.

3.2.2. Biyoinformatik Analizler

3.2.2.1. Ham Verilerin Kalite Kontrolü

Ham shotgun metagenom sekans verileri, hatalı nükleotidleri temizlemek, düşük kaliteli okumaları filtrelemek ve genel veri kalitesini artırmak amacıyla FastQC (Babraham ve ark., t.y.) ve Trimmomatic (Bolger ve ark., 2014) gibi kalite kontrol yazılımları kullanılarak analiz edildi. Bu işlem, veri setlerinin daha güvenilir sonuçlar verecek şekilde işlenmesi sağlandı.

3.2.2.2. Verilerin Birleştirilmesi (Assembly)

Mikrobiyom örneklerinden elde edilen ham metagenomik verilerin yüksek kaliteli ve anlamlı bir formatta bir araya getirilmesi amacıyla Iterative De Bruijn graph Assembler for metagenomic data using an Ultra-large De Bruijn graph (IDBA_UD)

(Peng ve ark., 2012) algoritması kullanılarak birleştirme (assembly) yapıldı. IDBA_UD programı, özellikle düşük bollukta bulunan bakteriyel genomları doğru şekilde birleştirmek üzere geliştirilmiştir ve mikrobiyom örneklerinde genetik çeşitliliği yüksek seviyede koruyarak etkin bir şekilde çalışan bir araçtır. Bu yazılım, özellikle uzun okuma uzunluklarına sahip metagenomik verilerle uyumlu olup, "de Bruijn graph" adı verilen bir graf teorisi yöntemini kullanmaktadır. Bu yöntem, sekans verilerindeki örtüşmeleri analiz ederek veri parçalarını birleştirir ve karmaşık mikrobiyal topluluk yapılarının genom dizilerini daha doğru bir şekilde yeniden oluşturmaktadır.

3.2.2.3. Bakteriyel Taksonomik Atama

Veri setlerindeki bakteriyel taksonomik atamalar Metagenomic Phylogenetic Analysis (MetaPhlAn2) (Truong ve ark., 2015) aracı kullanılarak gerçekleştirildi. MetaPhlAn2, metagenomik sekans verilerinden mikrobiyal topluluk yapısını belirlemek için yaygın olarak kullanılan, yüksek hassasiyete sahip bir yazılımdır. Bu araç, belirli bir mikroorganizma grubu için karakteristik olan marker genleri kullanarak metagenomik verilerdeki mikroorganizmaların taksonomik düzeyde tanımlanmasını sağlamaktadır.

Bu analiz süreci, çalışmanın temel hedeflerinden biri olan bakteriyofaj komünitesinin potansiyel konak ilişkilerinin belirlenmesi açısından kritik öneme sahiptir. Bakteriyofajların konak özgüllüğü göstermesi nedeniyle, öncelikle veri setlerinde bulunan bakteriyel topluluğun doğru bir şekilde tanımlanması gerekmektedir. MetaPhlAn2 aracı, bakteriyel taksonomiyi tür düzeyine kadar hassas bir şekilde tanımlayarak, bakteriyofajların konaklarının belirlenmesi için gerekli olan temel veriyi sağlamıştır. Bu analiz, bakteriyel çeşitliliğin yanı sıra hastalık gruplarındaki ve sağlıklı kontrollerdeki mikrobiyal farklılıkları da karşılaştırmalı olarak incelemeye olanak tanıdı.

3.2.2.4. Bakteriyofaj İdentifikasyonu

Bakteriyofajların tespiti, metagenomik verilerin işlenmesinin ardından, yüksek doğruluk sağlayan Transformer (Shang ve ark., 2022) iş akışı ile gerçekleştirildi.

Tanımlanan fajların daha hassas sınıflandırılması ve analizine yönelik olarak, PhaMer adlı derin öğrenme tabanlı bir model kullanıldı. PhaMer, faj adaylarının sınıflandırılmasında yüksek güvenilirlik sağladı ve çalışmada, 0.9 eşik değerinin üzerinde skora sahip fajlar tanımlandı.

3.2.2.5. Bakteriyofajların Taksonomik Sınıflandırılması ve Yaşam Döngüsü

Tanımlanan fajların taksonomik kimliği ve yaşam döngüsü (litik veya lizojenik) hakkında bilgi edinmek amacıyla PhaGCN (GCN classifier for phage taxonomy classification) (Shang ve ark., 2022) ve PhaTYP (Phage TYPe prediction tool) (Shang ve ark., 2022) araçları kullanıldı. PhaGCN, fajların taksonomik sınıflandırılmasını gerçekleştirmek için graf tabanlı bir öğrenme modeline dayanır. PhaTYP (Shang ve ark., 2022) (Predicting lifestyle for bacteriophages using BERT) ise, bakteriyofajların yaşam döngülerini belirleyerek, lizojenik (temperate) ya da litik (virulent) tipte olup olmadıklarını tespit edilmesini sağladı.

Tez çalışmasında ılımlı fajlar filtrelendikten sonra her bir kohorttaki litik bakteriyofaj profillerinin hastalıklarla ilişkisi, makine öğrenme ve transfer öğrenme yöntemleri kullanılarak sınıflandırma performansı ölçüldü. Çalışma kapsamında litik bakteriyofajların üzerine yoğunlaşılmasının temel sebebi bu tip fajların yaşam döngülerinden kaynaklı bağırsak mikrobiyotasındaki bağıl bolluk dengesini değiştirme potansiyellerinden dolayıdır. Ayrıca litik fajların neden olduğu bakteriyel lizis sonucu ortaya çıkan metabolitlerin ve diğer mikrobiyal bileşiklerin bağırsak ekosisteminde kademeli etkiler yaratmasıdır.

3.2.2.6. Faj-Konak İlişkilendirmesi

Bakteriyofajların konaklarının tespit edilmesi, çalışmanın temel hedeflerinden biri olarak, derin öğrenme tabanlı CHERRY (Host prediction for phages) (Shang ve ark., 2022) modeli kullanılarak gerçekleştirildi. CHERRY, faj-konak ilişkilendirmesini bir bağlantı tahmin (link prediction) problemi olarak ele alır ve faj-prokaryot ilişkilerini çoklu model (multi modal) olarak bir grafik üzerine inşa eder.

- *in situ* Konak Tahmini: Bu aşamada, 628 (non-redundant metagenome-assembled genome)MAG) veri seti kullanılarak örneklerdeki mevcut

konaklar tespit edildi. Bu MAG'ler, GTDB-Tk (Genome Taxonomy Database)(v1.5.1) (Chaumeil ve ark., 2020) kullanılarak taksonomik atamaları gerçekleştirildi.

- *ex situ* Konak Tahmini: CHERRY, ayrıca 62.493 bakteriyel ve arkea genomunu analiz ederek potansiyel dış (*ex situ*) konakları belirlendi.

3.2.3. Makine Öğrenmesi, Derin Öğrenme ve Transfer Öğrenme Yöntemleriyle Sınıflandırma

Bu tez çalışmasında, insan mikrobiyomunda bulunan bakteriyofajların sınıflandırılmasında farklı ML, derin öğrenme ve transfer öğrenme yöntemleri kullanıldı. Metagenomik verilerden elde edilen bakteriyofaj sekansları üzerinde özellik çıkarımı gerçekleştirildi ve sınıflandırma modellerinin eğitimi için uygun veri setleri oluşturuldu.

3.2.3.1. Makine öğrenmesi

Bu çalışmada, kullanılan ML algoritmalar arasında Rastgele Orman (RF) (Y. Liu ve ark., 2012), Yapay Sinir Ağları (Multi-Layer Perceptron, MLP) (Wilson ve ark., 1994) ve Gradyan Güçlendirilmiş Karar Ağaçlar Algoritması (XGBoost) (Chen ve ark., 2016) bulunmaktadır.

Her bir algoritma, veri setlerindeki özelliklerin sınıflandırmadaki ayırt edici gücünü değerlendirmek ve anlamlı özellikleri ön plana çıkarmak amacıyla kullanıldı. Özellikle RF ve XGBoost algoritmaları, değişken önem düzeyi analizi yaparak biyobelirteç olarak değerlendirilebilecek özelliklerin seçilmesini sağladı. MLP ise doğrusal olmayan ilişkilerin modellenmesi ve verinin daha karmaşık yapılarının öğrenilmesinde etkin bir şekilde kullanıldı.

Farklı algoritmaların güçlü yönlerinden faydalanarak daha kapsamlı bir analiz sunmuş ve elde edilen bulguların güvenilirliğini artırmak amacıyla bu yöntemler bir arada kullanıldı. Algoritmaların performansları, doğruluk, F1 skoru, ROC eğrisi gibi metrikler ile karşılaştırılarak değerlendirildi.

3.2.3.2. Transfer Öğrenme

Modelin sınıflandırma performansını artırmak amacıyla derin öğrenme yöntemlerinden transfer öğrenme yöntemi kullanıldı. Veri seti, eğitim ve test aşamalarında rastgele bölünerek %80 eğitim ve %20 test oranında ayrıldı. SMOTE (Synthetic Minority Over-sampling Technique**)** (Chawla ve ark., 2002) kütüphanesi kullanılarak veri dengesizliği giderildi. Aktivasyon fonksiyonu olarak "LeakyRELU" kullanıldı. Kodlayıcı-Kod çözücü mimarisi olarak tam bağlı (Dense) mimari kullanıldı. 2 katmanlı kod çözücü (decoder) kullanıldı. Optimizasyon için "optimizer=adam" kullanıldı. Kodlayıcı ham veriden sıkıştırılmış ve anlamlı bir temsil (latenst space) çıkarırken, kod çözücü ise bu temsili yeniden oluşturmayı öğrenerek veri kaybını minimize etmek için kullanıldı. Buradaki otokodlayıcı, transfer öğrenme sürecinde modelin performansını artırmak için bir özellik çıkarıcı olarak çalıştırıldı veveri setindeki önemli ve anlamlı özellikleri (feature) öğrenip bu bilgileri sınıflandırma işlemi için kullanıma sundu. Modellerin performansı doğruluk (accuracy), duyarlılık (sensitivity), özgüllük (specificity), F1 skoru ve ROC eğrisi altında kalan alan (AUC) gibi metriklerle değerlendirildi. Makine öğrenme algoritmalarının sonuçları karşılaştırılarak sınıflandırma performansları analiz edildi.

3.2.3.3. Öznitelik Seçimi (Feature Selection)

Bu tez çalışmasında, bakteriyofajların biyobelirteç potansiyelini değerlendirmek amacıyla, metagenomik veri setleri üzerinde Rastgele Orman Sınıflandırıcı (Random Forest Classifier) kullanılarak öznitelik seçimi (feature selection) gerçekleştirildi.

3.2.4. İstatistiksel Analiz

Veriler, gruplar arasındaki farkların istatistiksel olarak anlamlı olup olmadığını değerlendirmek amacıyla t-testi ile analiz edilmiştir. Analiz sonucunda, elde edilen p-değerleri ($p < 0.05$ anlamlılık düzeyi) üzerinden gruplar arasındaki farkın anlamlılığı yorumlanmıştır. t-testi sonuçları, ortalama değerler ve standart sapmalar ile birlikte raporlanmıştır

4. BULGULAR

Çalışmaya toplam 1564 bireyden elde edilen 7365 Gbp'lik metagenomik veri dahil edilmiştir. Kalite kontrol aşamasında, düşük kaliteli ve analiz için uygun olmayan diziler belirlenmiş ve 320 Gbp'lik veri çalışma dışı bırakılmış olup toplamda yaklaşık 7.0 Terabyte yüksek kaliteli metagenom verisi, ileri analizler için kullanıma hazır hale getirilmiştir (Tablo 4.1 ve 4.2).

Tablo 4.1. Çalışmaya dahil edilen KK veri setleri

Veri Setleri	Kohort	Ham veri	İşlenen Veri
PRJEB7774	46 Hasta / 110 Kontrol	366 GB	393 GB
PRJEB27928	70 Hasta / 60 Kontrol	364 GB	469 GB
PRJEB12449	203 Hasta / 223 Kontrol	862 GB	1.3 TB
PRJEB10878	64 Hasta / 53 Kontrol	283 GB	1.4 TB
Toplam	383 Hasta / 446 Kontrol	1.8 TB	3.56 TB

Tablo 4.2. Çalışmaya dahil edilen İBH veri setleri

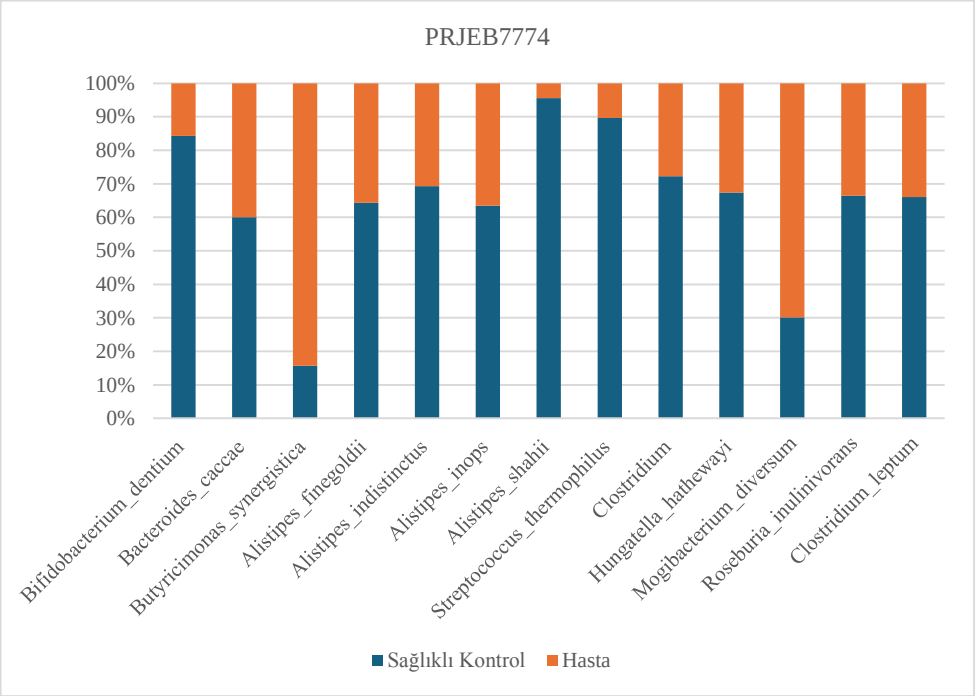
Veri Setleri	Kohort	Ham veri	İşlenen Veri
PRJEB1220	311 Hasta / 211 Kontrol	1 TB	1.3 TB
PRJNA400072	157 Hasta / 56 Kontrol	460 GB	2.6 TB
TOPLAM	468 Hasta / 267 Kontrol	1.46 TB	3.9 TB

4.1. Biyoinformatik Analizler

4.1.1. Kolorektal Kanser Veri Setleri İçin Bakteriyel Çeşitlilik Karşılaştırması

4.1.1.1. Veri Seti 1 (PRJEB7774)

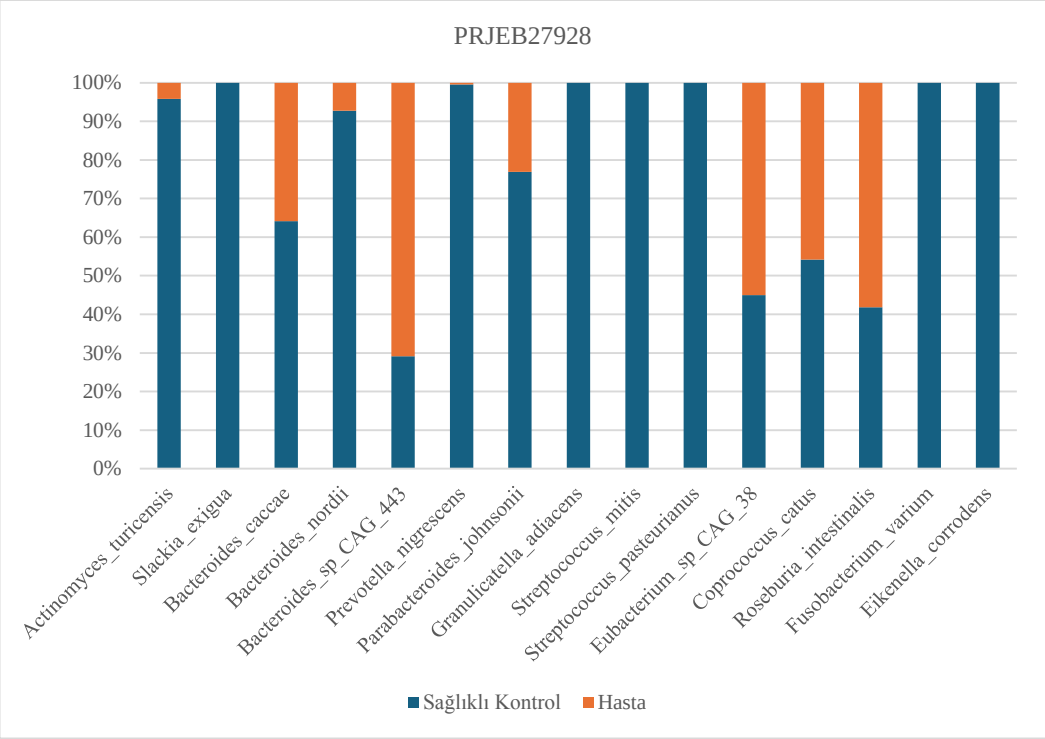
Bu veri seti için gerçekleştirilen analiz sonucunda *Butyricimonas synergistica* ve *Mogibacterium diversum* türlerine ait okumaların istatistiki olarak anlamlı oranda ($p < 0.05$) hasta grubunda fazla olduğu, *Bifidobacterium dentium*, *Bacteroides caccae*, *Alistipes finegoldii*, *Alistipes indistinctus*, *Alistipes inops*, *Alistipes shahii*, *Streptococcus thermophilus*, *Clostridium*, *Hungatella hathewayi*, *Roseburia inulinivorans*, *Clostridium leptum* türlerine ait okumaların ise istatistiki olarak anlamlı oranda ($p < 0.05$) sağlıklı kontrol grubunda fazla bulunduğu tespit edilmiştir. Bu türlere ait bağıl bolluk dağılımları Şekil 4.1'de sunulmuştur.



Şekil 4.1. KK 1 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyel taksonlar.

4.1.1.2. Veri Seti 2 (PRJEB27928)

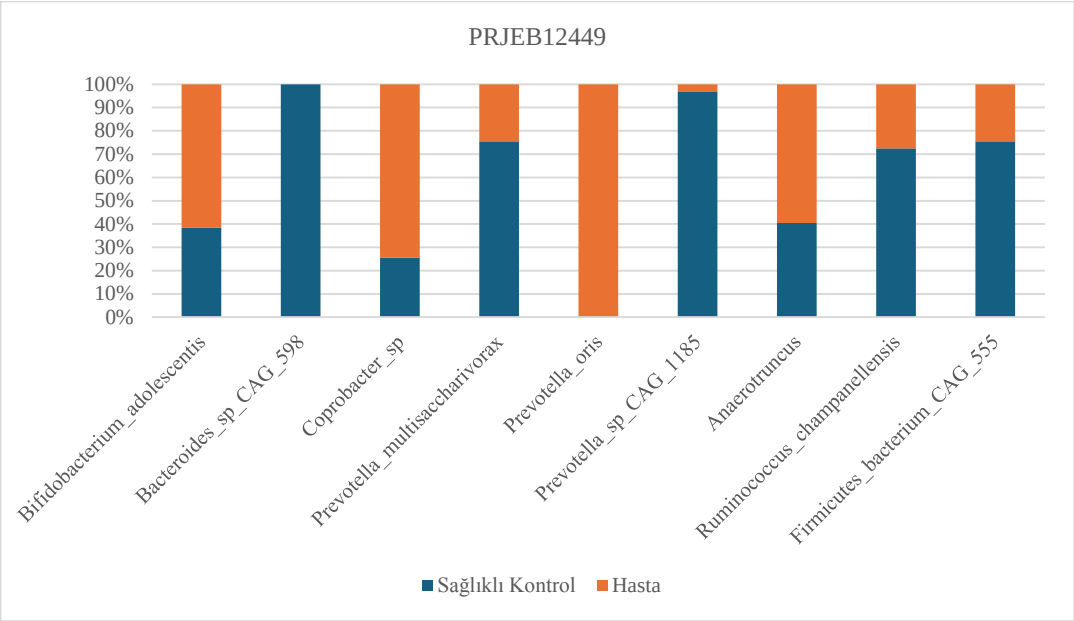
Bu veri seti için gerçekleştirilen analiz sonucunda *Bacteriodes sp CAG 443*, *Eubacterium sp CAG 38* ve *Roseburia intestinalis* türlerine ait okumaların istatistiki olarak anlamlı oranda ($p < 0.05$) hasta grubunda fazla olduğu, *Actinomyces turicensis*, *Slackia exigua*, *Bacteroides caccae*, *Bacteroides nordii*, *Prevotella nigrescens*, *Parabacteroides johnsonii*, *Granulicatella adiacens*, *Streptococcus mitis*, *Streptococcus pasteurianus*, *Coprococcus catus*, *Fusobacterium varium*, *Eikenella corrodens* türlerine ait okumaların ise istatistiki olarak anlamlı oranda ($p < 0.05$) sağlıklı kontrol grubunda fazla bulunduğu tespit edilmiştir. Bu türlere ait bağıl bolluk dağılımları Şekil 4.2'de sunulmuştur.



Şekil 4.2. KK 2 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyel taksonlar.

4.1.1.3. Veri Seti 3 (PRJEB12449)

Bu veri seti için gerçekleştirilen analiz sonucunda *Bacteriodes adolescentis, Coprobacter sp., Prevotella oris* türlerine ait okumaların istatistiki olarak anlamlı oranda ($p < 0.05$) hasta grubunda fazla olduğu, *Bacteriodes sp CAG 598, Prevotella multisaccharivorax, Prevotella sp. CAG 1185, Ruminococcus champanellensis* ve *Firmicutes bacterium CAG 555* türlerine ait okumaların ise istatistiki olarak anlamlı oranda ($p < 0.05$) sağlıklı kontrol grubunda fazla bulunduğu tespit edilmiştir. Bu türlere ait bağıl bolluk dağılımları Şekil 4.3'te sunulmuştur.

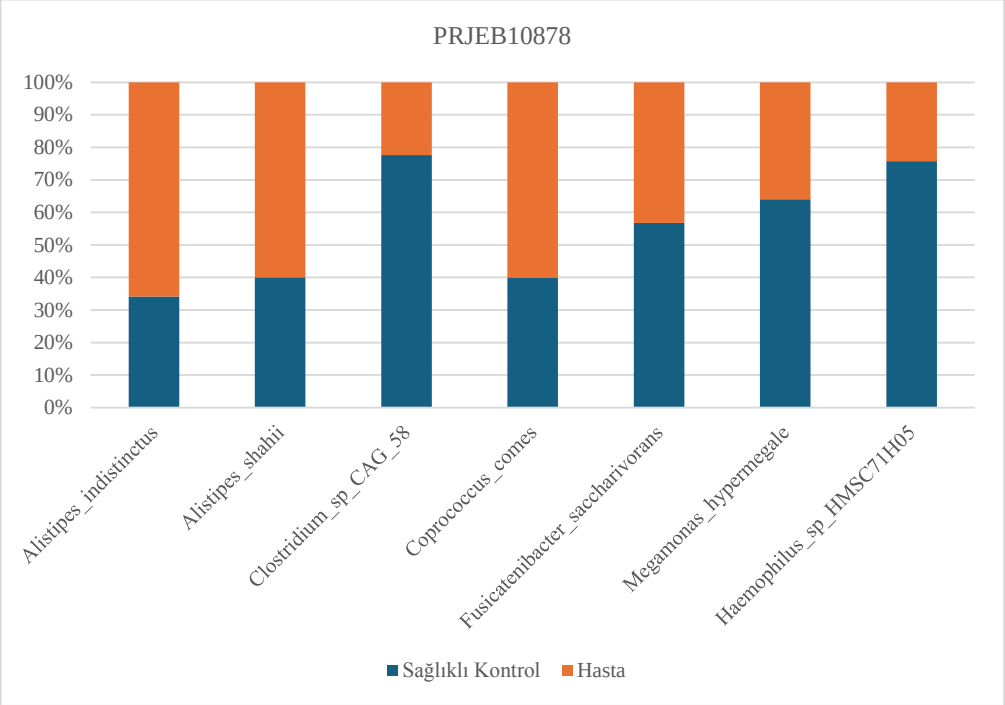


Şekil 4.3. KK 3 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyel taksonlar.

4.1.1.4. Veri Seti 4 (PRJEB10878)

Bu veri seti için gerçekleştirilen analiz sonucunda *Alistipes indistinctus, Alistipes shahii, Coprococcus comes* türlerine ait okumaların istatistiki olarak anlamlı oranda ($p < 0.05$) hasta grubunda fazla olduğu, *Clostridium sp CAG 58, Fusicatenibacter saccharivorans, Megamonas hypermegale, Haemophilus sp HMSC71H05* türlerine ait okumaların ise istatistiki olarak anlamlı oranda ($p < 0.05$) sağlıklı kontrol

grubunda fazla bulunduğu tespit edilmiştir. Bu türlere ait bağıl bolluk dağılımları Şekil 4.4'te sunulmuştur.



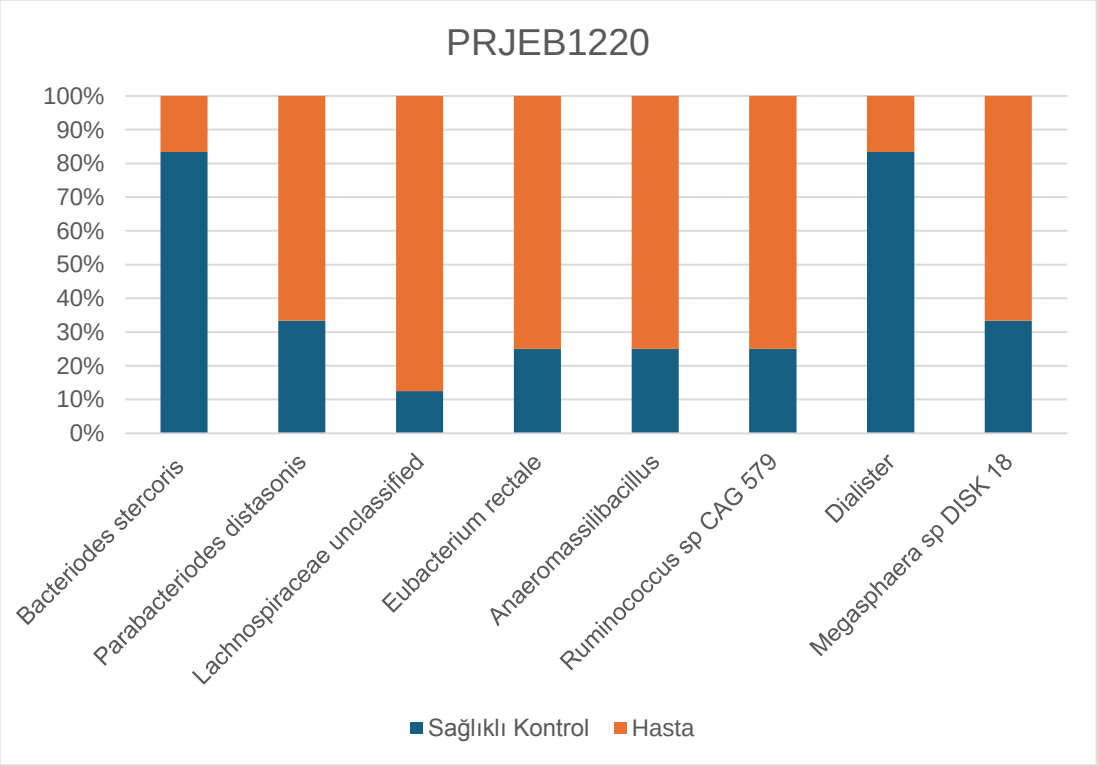
Şekil 4.4. KK 4 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyel taksonlar.

4.1.2. İnflamatuar Bağırsak Hastalığı Veri Setleri İçin Bakteriyel Çeşitlilik Karşılaştırması

Bu çalışmada, İBH ile ilişkili mikrobiyom yapısını anlamak amacıyla iki farklı metagenomik veri seti incelenmiştir. Bu veri setleri, İBH hastalarına ve sağlıklı bireylere ait bağırsak mikrobiyomu örneklerini içermektedir (PRJEB1220, PRJNA400072). Çalışma kapsamında her bir veri seti, genel mikrobiyom profillerinin yanı sıra, bakteriyel türlerin İBH ile ilişkisini değerlendirmek amacıyla analiz edilmiştir.

4.1.2.1. Veri Seti 1 (PRJEB1220)

Bu veri seti için gerçekleştirilen analiz sonucunda *Parabacteriodes distasonis*, *Lachnospiracecae unclassified*, *Eubacterium rectale*, *Anaeromassilibacillus*, *Ruminococcus sp CAG 579*, ve *Megasphaera sp DISK 18* türlerine ait okumaların istatistiki olarak anlamlı oranda ($p < 0.05$) hasta grubunda fazla olduğu, *Bacteriodes stercoris* ve *Dialister* türlerine ait okumaların ise istatistiki olarak anlamlı oranda ($p < 0.05$) sağlıklı kontrol grubunda fazla bulunduğu tespit edilmiştir. Bu türlere ait bağıl bolluk dağılımları Şekil 4.5'te sunulmuştur.

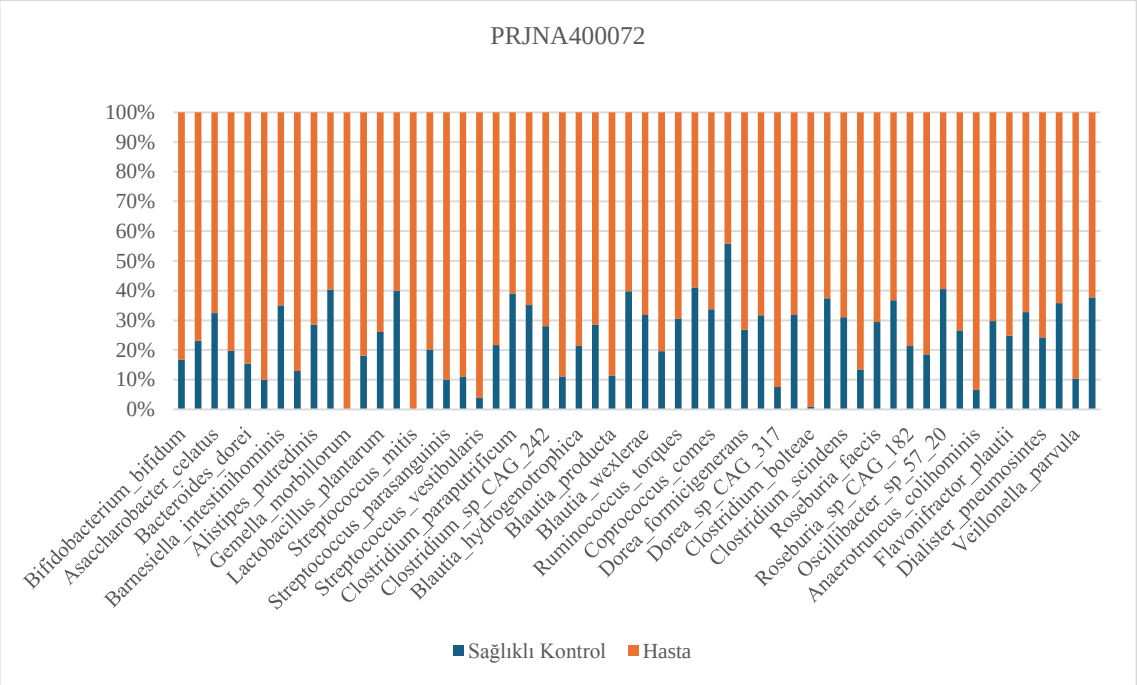


Şekil 4.5. İBH 1 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyel taksonlar.

4.1.2.2. Veri Seti 2 (PRJNA400072)

Bu veri seti için gerçekleştirilen analiz sonucunda *Bifidobacterium bifidum*, *Bacteriodes dorei*, *Alistipes putredinis*, *Streptococcus mitis*, *Veillonella parvula*, *Roseburia faecis*, *Blautia* ve Dorea türlerine ait okumaların istatistiki olarak anlamlı oranda ($p < 0.05$) hasta grubunda fazla olduğu, *Coprococcus comes* türüne ait

okumaların ise istatistiki olarak anlamlı oranda ($p < 0.05$) sağlıklı kontrol grubunda fazla bulunduğu tespit edilmiştir. Bu türlere ait bağıl bolluk dağılımları Şekil 4.6'da sunulmuştur.



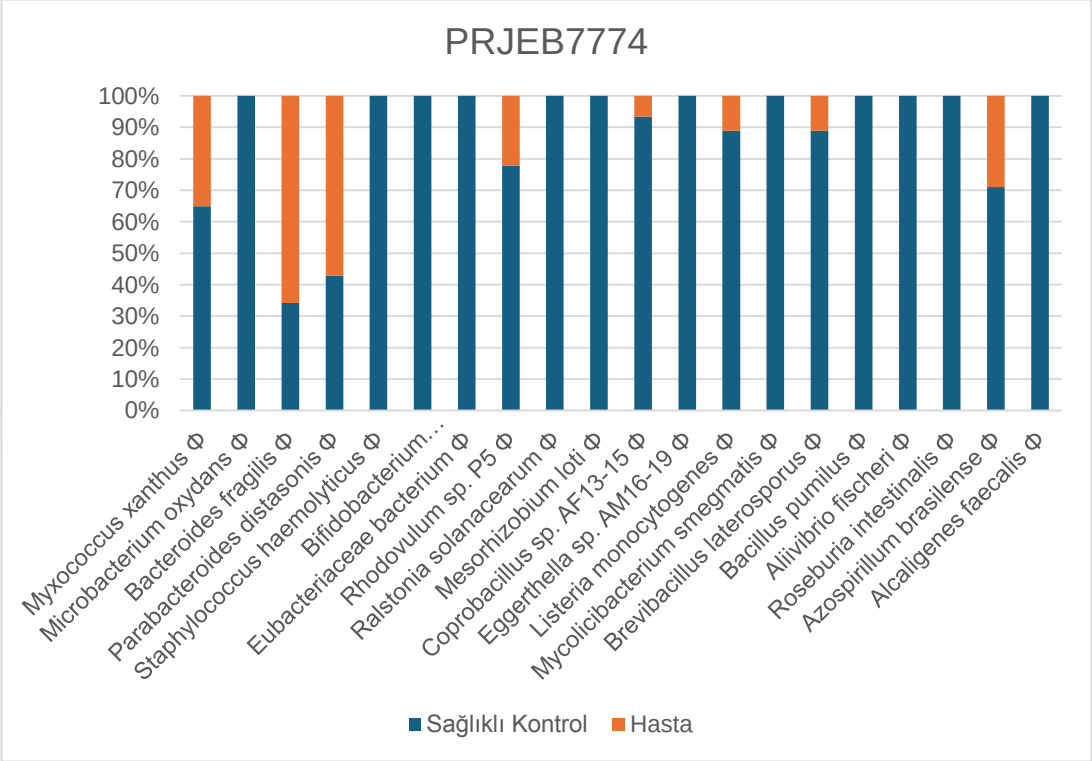
Şekil 4.6. İBH 2 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyel taksonlar.

4.1.3. Kolorektal Kanser Veri Setleri İçin Litik Bakteriyofaj Karşılaştırması

4.1.3.1. Veri Seti 1 (PRJEB7774)

Bu veri seti için gerçekleştirilen analiz sonucunda *Parabacteriodes distasonis* 'in konağı olduğu bakteriyofaja ait okumaların istatistiki olarak anlamlı oranda ($p < 0.05$) hasta grubunda fazla olduğu, *Myxococcus xanthus*, *Microbacterium oxydans*, *Staphylococcus haemolyticus*, *Bifidobacterium pseudocatenulatum*, *Eubacteriaceae bacterium*, *Rhodovulum sp. P5*, *Ralstonia solanacearum*, *Mesorhizobium loti*, *Coprobacillus sp. AF13-15*, *Eggerthella sp. AM16-19*, *Listeria monocytogenes*, *Mycolicibacterium smegmatis*, *Brevibacillus laterosporus*, *Bacillus pumilus*, *Aliivibrio fischeri*, *Roseburia intestinalis*, *Azospirillum brasilense*, *Alcaligenes faecalis* 'in konağı olduğu bakteriyofajlara ait okumaların ise istatistiki olarak

anlamlı oranda (p <0.05) sağlıklı kontrol grubunda fazla bulunduğu tespit edilmiştir. Bu türlere ait bağıl bolluk dağılımları Şekil 4.7'de sunulmuştur.

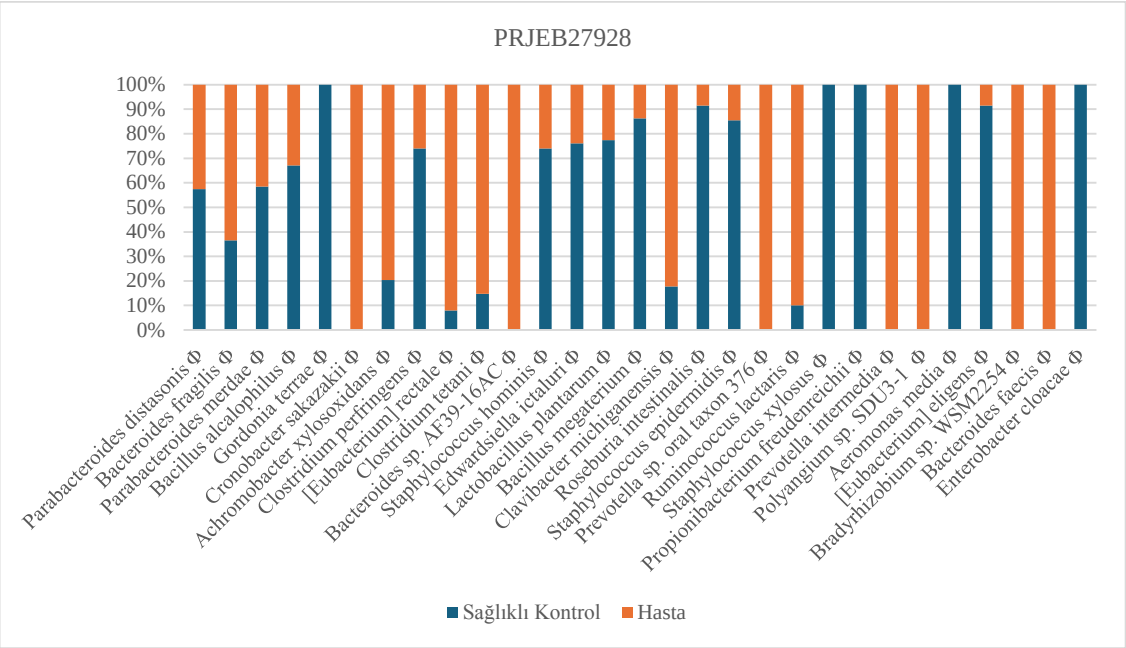


Şekil 4.7. KK 1 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyofajlar.

4.1.3.2. Veri Seti 2 (PRJEB27928)

Bu veri seti için gerçekleştirilen analiz sonucunda *Bacteroides fragilis, Cronobacter sakazakii, Achromobacter xylosoxidans, [Eubacterium] rectale, Clostridium tetani, Bacteroides sp. AF39-16AC, Bacillus megaterium, Staphylococcus epidermidis, Prevotella sp. oral taxon 376, Ruminococcus lactaris, Prevotella intermedia, Polyangium sp. SDU3-1, Bradyrhizobium sp. WSM2254, Bacteroides faecis'in* konağı olduğu bakteriyofajlara ait okumaların istatistiki olarak anlamlı oranda (p <0.05) hasta grubunda fazla olduğu, *Parabacteroides merdae, Gordonia terrae, Clostridium perfringens, Staphylococcus hominis, Lactobacillus plantarum, Roseburia intestinalis, Staphylococcus xylosus, Propionibacterium freudenreichii, Aeromonas media, [Eubacterium] eligens, Enterobacter cloacae* 'nın konağı olduğu

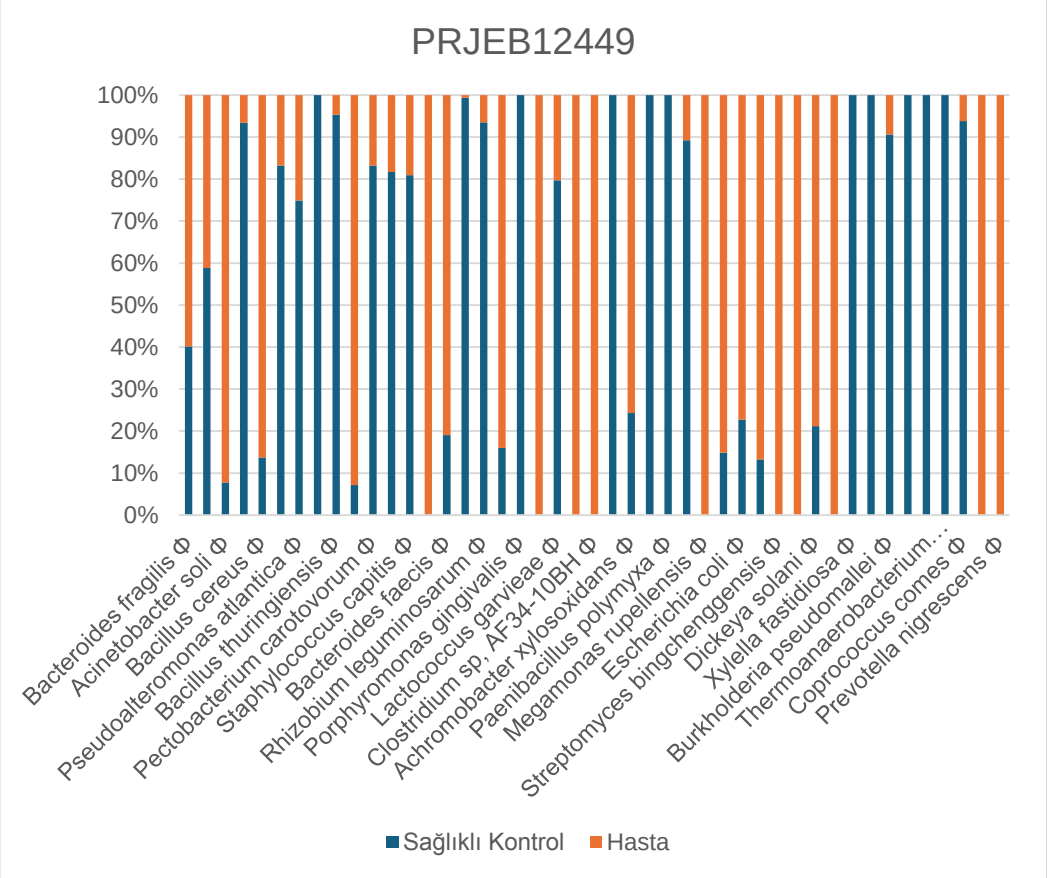
bakteriyofajlara ait okumaların ise istatistiki olarak anlamlı oranda (p <0.05) sağlıklı kontrol grubunda fazla bulunduğu tespit edilmiştir. Bu türlere ait bağıl bolluk dağılımları Şekil 4.8'de sunulmuştur.



Şekil 4.8. KK 2 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyofajlar.

4.1.3.3. Veri Seti 3 (PRJEB12449)

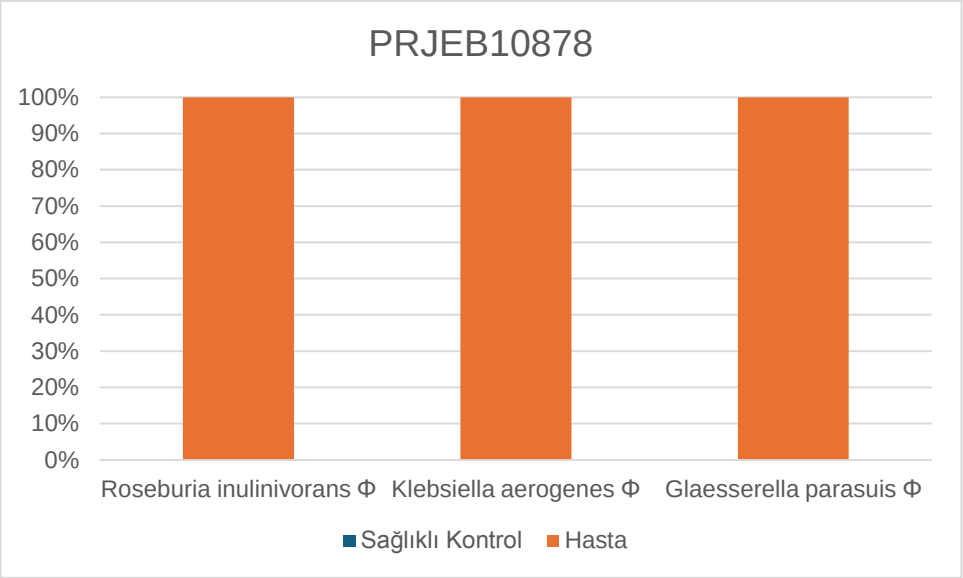
Bu veri seti için gerçekleştirilen analiz sonucunda *Bacteroides fragilis, Acinetobacter soli, Bacillus cereus, Bacteroides cellulosilyticus, Bacteroides faecis, Coprobacter secundus, Achromobacter xylosoxidans, Escherichia coli, Dickeya solani, Bacillus tropicus*'un konağı olduğu bakteriyofajlara ait okumaların istatistiki olarak anlamlı oranda (p <0.05) hasta grubunda fazla olduğu, *Lactobacillus fermentum, Parabacteroides distasonis, Pseudoalteromonas atlantica, Bacillus thuringiensis, Pectobacterium carotovorum, Ruminococcus bromii, Staphylococcus capitis, Gordonia terrae, Rhizobium leguminosarum, Prevotella maculosa*'nın konağı olduğu bakteriyofajlara ait okumaların ise istatistiki olarak anlamlı oranda (p <0.05) sağlıklı kontrol grubunda fazla bulunduğu tespit edilmiştir. Bu türlere ait bağıl bolluk dağılımları Şekil 4.9'de sunulmuştur.



Şekil 4.9. KK 3 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyofajlar.

4.1.3.4. Veri Seti 4 (PRJEB10878)

Bu veri seti için gerçekleştirilen analiz sonucunda *Rosereburia inulinivorans, Klebsiella aerogenes, Glasserella parasuis* konağı olduğu bakteriyofajlara ait okumaların istatistiki olarak anlamlı oranda ($p < 0.05$) hasta grubunda fazla olduğu, sağlıklı kontrol grubunda istatistiki olarak anlamlı oranda ($p < 0.05$) bir takson gözlemlenmemiştir. Bu türlere ait bağıl bolluk dağılımları Şekil 4.10'da sunulmuştur.

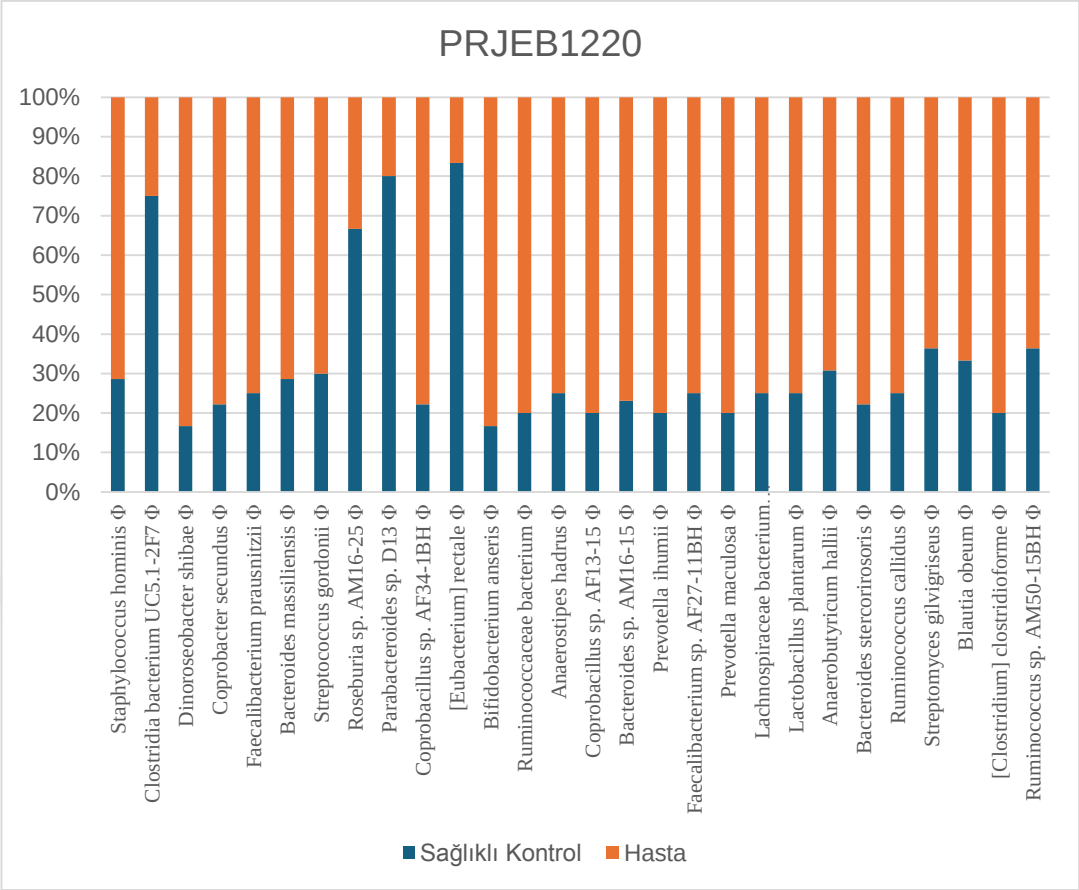


Şekil 4.10. KK 4 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyofajlar.

4.1.4. İnflamatuar Bağırsak Hastalığı Veri Setleri İçin Litik Bakteriyofaj Karşılaştırması

4.1.4.1. Veri Seti 1 (PRJEB1220)

Bu veri seti için gerçekleştirilen analiz sonucunda *Staphylococcus hominis*, *Dinoroseobacter shibae*, *Coprobacter secundus*, *Faecalibacterium prausnitzii*, *Bacteroides massiliensis*, *Streptococcus gordonii*, *Coprobacillus sp. AF34-1BH* , *Bifidobacterium anseris*, *Ruminococcaceae bacterium*, *Anaerostipes hadrus*, *Coprobacillus sp. AF13-15*, *Bacteroides sp. AM16-15*, *Prevotella ihumii*, *Faecalibacterium sp. AF27-11BH*, *Prevotella maculosa*, *Lachnospiraceae bacterium 5_1_63FAA*, *Lactobacillus plantarum*, *Anaerobutyricum hallii*, *Bacteroides stercorirosoris*, *Ruminococcus callidus*, *Streptomyces gilvigriseus*, *Blautia obeum*, *Clostridium clostridioforme*, *Ruminococcus sp. AM50-15BH* konağı olduğu bakteriyofajlara ait okumaların istatistiki olarak anlamlı oranda ($p < 0.05$) hasta grubunda fazla olduğu, *Clostridia bacterium UC5.1-2F7*, *Roseburia sp. AM16-25*, *Parabacteroides sp. D13*, *Eubacterium rectale*'in konağı olduğu bakteriyofajlara ait okumaların ise istatistiki olarak anlamlı oranda ($p < 0.05$) sağlıklı kontrol grubunda fazla bulunduğu tespit edilmiştir. Bu türlere ait bağıl bolluk dağılımları Şekil 4.11'de sunulmuştur.

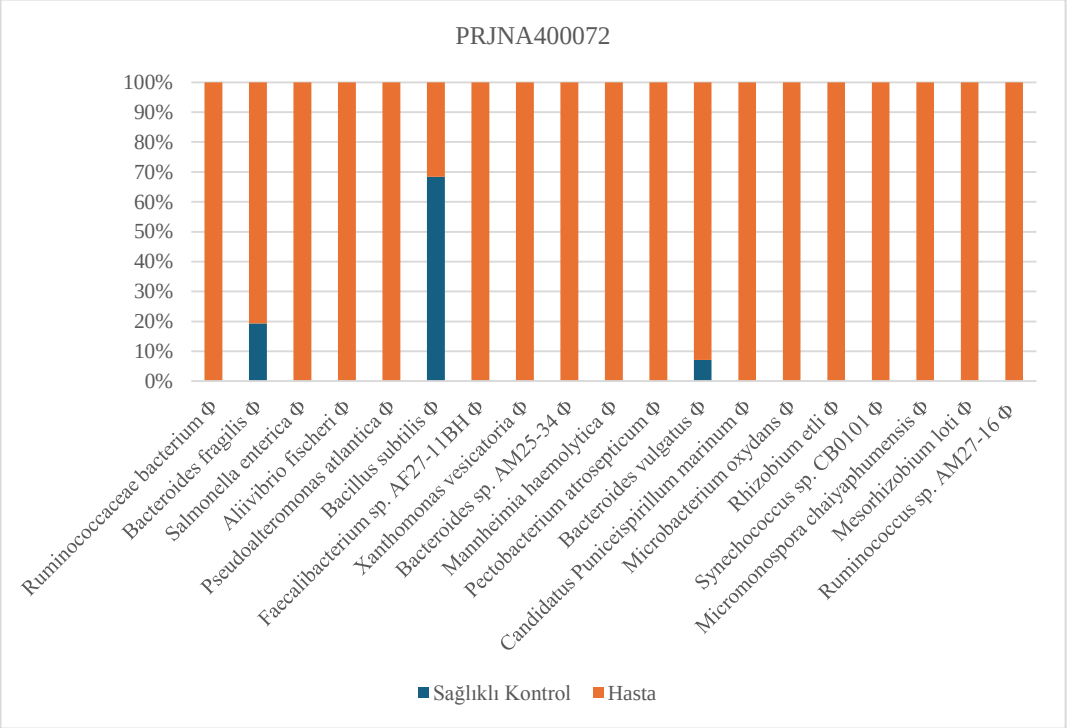


Şekil 4.11. İBH 1 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyofajlar**.**

4.1.4.2. Veri Seti 2 (PRJNA400072)

Bu veri seti için gerçekleştirilen analiz sonucunda *Ruminococcaceae bacterium*, *Bacteroides fragilis*, *Salmonella enterica*, *Aliivibrio fischeri*, *Pseudoalteromonas atlantica*, *Aeromonas media*, *Faecalibacterium sp. AF27-11BH*, *Xanthomonas vesicatoria*, *Bacteroides sp. AM25-34*, *Mannheimia haemolytica*, *Pectobacterium atrosepticum*, *Bacteroides vulgatus*, *Candidatus Puniceispirillum marinum*, *Microbacterium oxydans*, *Rhizobium etli*, *Synechococcus sp. CB0101*, *Micromonospora chaiyaphumensis*, *Mesorhizobium loti*, *Ruminococcus sp. AM27-16* ’un konağı olduğu bakteriyofajlara ait okumaların istatistiki olarak anlamlı oranda ($p < 0.05$) hasta grubunda fazla olduğu, *Bacillus subtilis*’in konak olduğu bakteriyofaja ait okumaların ise istatistiki olarak anlamlı oranda ($p < 0.05$) sağlıklı kontrol

grubunda fazla bulunduğu tespit edilmiştir. Bu türlere ait bağıl bolluk dağılımları Şekil 4.12'de sunulmuştur.



Şekil 4.12. İBH 2 nolu veri setinde hasta ve sağlıklılar arasında istatistiki olarak anlamlı oranda farklılık gösteren bakteriyofajlar.

4.1.5. Tüm veri setlerinin her bir hastalık grubunda ortak analizi

KK metagenomlarında gerçekleştirilen bakteriyel çeşitlilik analizleri sonucunda *Alistipes sp.* ve *Caprobacter sp.* türlerinin istatistiki olarak anlamlı oranda ($p < 0.05$) hasta grubunda fazla olduğu ve diğer KK veri setleriyle örtüştüğü, *Bacteriodes caccae* istatistiki olarak anlamlı oranda ($p < 0.05$) sağlıklı grubunda fazla olduğu ve diğer KK veri setleriyle örtüştüğü keşfedilmiştir. İBH metagenomlarında gerçekleştirilen bakteriyel çeşitlilik analizleri sonucunda diğer İBH veri setleriyle örtüşen türler tespit edilememiştir.

KK metagenomlarında gerçekleştirilen bakteriyofaj çeşitlilik analizleri sonucunda *Bacteroides fragilis*, *Bacteroides faecis*, *Achromobacter xylosoxidans* konağına ait bakteriyofajların istatistiki olarak anlamlı oranda ($p < 0.05$) hasta grubunda fazla

olduğu ve diğer KK veri setleriyle örtüştüğü, *Roseburia intestinalis, Gordonia terrae, Parabacteroides merdae* konağına ait bakteriyofajların istatistiki olarak anlamlı oranda ($p < 0.05$) sağlıklı grubunda fazla olduğu ve diğer KK veri setleriyle örtüştüğü keşfedilmiştir. İBH metagenomlarında gerçekleştirilen bakteriyofaj çeşitlilik analizlerinde *Ruminococcaceae bacterium, Coprobacillus sp. AF13-15* konağına ait bakteriyofajların istatistiki olarak anlamlı oranda ($p < 0.05$) hasta grubunda fazla olduğu ve diğer İBH veri setleriyle örtüştüğü keşfedilmiştir.

Her bir veri setinde tespit edilen faj çeşidi nicelik olarak değerlendirildiğinde, KK'ye ait dört veri setinde toplam 21 faj çeşidinin istatistiki olarak hastalarda anlamlı oranda fazla olduğu buna karşın iki İBH veri setinde 42 farklı fajın anlamlı oranda hasta grubunda fazla olduğu tespit edilmiştir.

4.2. Makine Öğrenmesi Yöntemleri ile Veri Setlerinin Sınıflandırma Sonuçları

KK ve İBH veri setlerinde bakteriyofaj bağıl bolluklarına dayalı olarak yapılan analizlerde makine öğrenimi algoritmalarının sınıflandırma performansları değerlendirilme sonuçları aşağıda sunulmuştur.

4.2.1. Kolorektal Kanser Veri Setlerine Ait Sınıflandırma Sonuçları

Veri seti 1 için yapılan analizler sonucunda fajların hastalık ve sağlık durumunu ayırt etme gücü %64 olarak hesaplanmıştır ve en iyi sınıflandırma performansı gösteren algoritmanın MLP olduğu raporlanmıştır. Veri Seti 2 bu değer %71, veri seti 3 için % 74, veri seti 4 için ise % 54 olarak hesaplanmıştır (Tablo 4.3). Sonuç olarak, çalışma kapsamında fajların hastalık ve sağlık durumunu ayırt etme gücü en yüksek %74 olarak hesaplanmış ve MLP algoritmasının sınıflandırma başarısının daha üstün olduğu tespit edilmiştir (Tablo 4.3).

Tablo 4.3. KK veri setleri için sınıflandırma performansı sonuçları

KK					
Veri Set 1 (PRJEB7774)	Doğruluk Ortalaması	Kesinlik Ortalaması	Duyarlılık	f1 skoru ortalaması	Roc Auc skor ortalaması
Rastgele Orman Algoritması (Random Forest)	0.69	0.20	0.04	0.06	0.50
Gradyan Güçlendirilmiş Karar Ağaçlar Algoritması (XGBoost)	0.67	0.37	0.30	0.32	0.56
Yapay Sinir Ağı (MLP)	0.76	0.72	0.34	0.44	0.64
Veri Set 2 (PRJEB27928)	Doğruluk Ortalaması	Kesinlik Ortalaması	Duyarlılık	f1 skoru ortalaması	Roc Auc skor ortalaması
Rastgele Orman Algoritması (Random Forest)	0.71	0.68	0.88	0.76	0.70
Gradyan Güçlendirilmiş Karar Ağaçlar Algoritması (XGBoost)	0.65	0.66	0.75	0.70	0.64
Yapay Sinir Ağı (MLP)	0.71	0.75	0.75	0.72	0.71
Veri Set 3 (PRJEB12449)	Doğruluk Ortalaması	Kesinlik Ortalaması	Duyarlılık	f1 skoru ortalaması	Roc Auc skor ortalaması
Rastgele Orman Algoritması (Random Forest)	0.69	0.71	0.60	0.65	0.68
Gradyan Güçlendirilmiş Karar Ağaçlar Algoritması (XGBoost)	0.60	0.59	0.58	0.58	0.60
Yapay Sinir Ağı (MLP)	0.74	0.72	0.72	0.72	0.74
Veri Set 4 (PRJEB10878)	Doğruluk Ortalaması	Kesinlik Ortalaması	Duyarlılık	f1 skoru ortalaması	Roc Auc skor ortalaması
Rastgele Orman Algoritması (Random Forest)	0.53	0.57	0.79	0.66	0.53
Gradyan Güçlendirilmiş Karar Ağaçlar Algoritması (XGBoost)	0.55	0.58	0.59	0.57	0.54
Yapay Sinir Ağı (MLP)	0.45	0.47	0.40	0.42	0.45

4.2.2. İnflamatuar Bağırsak Hastalığı (İBH) Veri Setlerinin Sınıflandırma Sonuçları

Veri seti 1 için yapılan analizler sonucunda fajların hastalık ve sağlık durumunu ayırt etme gücü %66 olarak hesaplanmıştır. En iyi sınıflandırma performans gösteren algoritma RF olmuştur. Veri Seti 2 için bu değer %61 olarak hesaplanmış olup XGBoost'un en iyi sınıflandırma performans ortaya koyduğu tespit edilmiştir (Tablo 4.4).

Tablo 4.4. İBH veri setleri için sınıflandırma performansı sonuçları

İBH					
Veri Set 1 (PRJEB1220)	Doğruluk Ortalaması	Kesinlik Ortalaması	Duyarlılık	f1 skoru ortalaması	Roc Auc skor ortalaması
Rastgele Orman Algoritması (Random Forest)	0.68	0.71	0.76	0.73	0.66
Gradyan Güçlendirilmiş Karar Ağaçlar Algoritması (XGBoost)	0.65	0.69	0.72	0.70	0.64
Yapay Sinir Ağı (MLP)	0.64	0.66	0.73	0.70	0.63
Veri Set 2 (PRJNA400072)	Doğruluk Ortalaması	Kesinlik Ortalaması	Duyarlılık	f1 skoru ortalaması	Roc Auc skor ortalaması
Rastgele Orman Algoritması (Random Forest)	0.72	0.73	0.97	0.83	0.49
Gradyan Güçlendirilmiş Karar Ağaçlar Algoritması (XGBoost)	0.56	0.72	0.40	0.74	0.61
Yapay Sinir Ağı (MLP)	0.68	0.74	0.86	0.79	0.51

4.3. Transfer Öğrenme Modeli ile Veri Setlerinin Sınıflandırma Sonuçları,

4.3.1. Kolorektal Kanser Veri Setlerinin Sınıflandırma Sonuçları

Veri seti 1 için yapılan analizler sonucunda fajların hastalık ve sağlık durumunu ayırt etme gücü %78 olarak hesaplanmıştır. En iyi sınıflandırma performans gösteren algoritma XGBoost olmuştur. Veri seti 2 için fajların hastalık ve sağlık durumunu ayırt etme gücü %69 olarak hesaplanmış olup, RF ve XGBoost algoritmaların iyi sınıflandırma performansı gösterdiği raporlanmıştır. Veri seti 3 için fajların sınıflandırma gücü %65 olarak hesaplanmıştır. En iyi sınıflandırma performans gösteren algoritma XGBoost olmuştur. Veri seti 4 için fajların hastalık ve sağlık durumunu ayırt etme gücü %54 olarak, en iyi sınıflandırma performans gösteren algoritma ise XGBoost olarak raporlanmıştır (Tablo 4.5)

Tablo 4.5. KK veri setlerinin transfer öğrenme modeli ile gerçekleştirilmiş sınıflandırma performansı sonuçları

KK					
Veri Set 1 (PRJEB7774)	Doğruluk Ortalaması	Kesinlik Ortalaması	Duyarlılık	f1 skoru ortalaması	Roc Auc skor ortalaması
Rastgele Orman Algoritması (Random Forest)	0.66	0.67	0.69	0.69	0.68
Gradyan Güçlendirilmiş Karar Ağaçlar Algoritması (XGBoost)	0.76	0.77	0.78	0.69	0.78
Yapay Sinir Ağı (MLP)	0.71	0.70	0.59	0.62	0.60
Veri Set 2 (PRJEB27928)	Doğruluk Ortalaması	Kesinlik Ortalaması	Duyarlılık	f1 skoru ortalaması	Roc Auc skor ortalaması
Rastgele Orman Algoritması (Random Forest)	0.70	0.71	0.69	0.70	0.69
Gradyan Güçlendirilmiş Karar Ağaçlar Algoritması (XGBoost)	0.70	0.69	0.69	0.69	0.69
Yapay Sinir Ağı (MLP)	0.64	0.63	0.54	0.53	0.53
Veri Set 3 (PRJEB12449)	Doğruluk Ortalaması	Kesinlik Ortalaması	Duyarlılık	f1 skoru ortalaması	Roc Auc skor ortalaması
Rastgele Orman Algoritması (Random Forest)	0.62	0.61	0.60	0.61	0.60
Gradyan Güçlendirilmiş Karar Ağaçlar Algoritması (XGBoost)	0.67	0.66	0.65	0.65	0.65
Yapay Sinir Ağı (MLP)	0.55	0.54	0.53	0.54	0.53
Veri Set 4 (PRJEB10878)	Doğruluk Ortalaması	Kesinlik Ortalaması	Duyarlılık	f1 skoru ortalaması	Roc Auc skor ortalaması
Rastgele Orman Algoritması (Random Forest)	0.48	0.47	0.46	0.42	0.45
Gradyan Güçlendirilmiş Karar Ağaçlar Algoritması (XGBoost)	0.65	0.64	0.54	0.49	0.54
Yapay Sinir Ağı (MLP)	0.48	0.47	0.50	0.45	0.50

4.3.2. İnflamatuar Bağırsak Hastalığı (İBH) Veri Setlerinin Sınıflandırma Sonuçları

Veri seti 1 için yapılan analizler sonucunda fajların hastalık ve sağlık durumunu ayırt etme gücü %60 olarak hesaplanmış olup bu veri seti için en iyi sınıflandırma performans gösteren algoritmalar RF ve XGBoost olarak raporlanmıştır. Veri Seti 2 için fajların sınıflandırma gücü gücünün %74 en iyi sınıflandırma performansı algoritmanın ise RF olduğu tespit edilmiştir (Tablo 4.6).

Tablo 4.6. İBH veri setlerinin transfer öğrenme modeli ile gerçekleştirilmiş sınıflandırma performansı sonuçları

İBH					
Veri Set 1 (PRJEB1220)	Doğruluk Ortalaması	Kesinlik Ortalaması	Duyarlılık	f1 skoru ortalaması	Roc Auc skor ortalaması
Rastgele Orman Algoritması (Random Forest)	0.61	0.60	0.60	0.60	0.60
Gradyan Güçlendirilmiş Karar Ağaçlar Algoritması (XGBoost)	0.62	0.6	0.61	0.61	0.60
Yapay Sinir Ağı (MLP)	0.61	0.60	0.58	0.5	0.58
Veri Set 2 (PRJNA400072)	Doğruluk Ortalaması	Kesinlik Ortalaması	Duyarlılık	f1 skoru ortalaması	Roc Auc skor ortalaması
Rastgele Orman Algoritması (Random Forest)	0.74	0.73	0.74	0.74	0.74
Gradyan Güçlendirilmiş Karar Ağaçlar Algoritması (XGBoost)	0.70	0.66	0.63	0.64	0.62
Yapay Sinir Ağı (MLP)	0.72	0.71	0.63	0.66	0.62

Transfer öğrenme yöntemleri KK ve İBH veri setlerinde, bakteriyofaj bağıl bollukları kullanılarak sınıflandırma performansını iyileştirmek amacıyla uygulanmıştır. Transfer öğrenme ile elde edilen sonuçlar, bakteriyofaj bağıl bolluklar üzerinden yapılan analizler sonucunda sınıflandırma kapasitesi belli bir miktar güçlendirmiştir. Genel olarak, KK veri setinde XGBoost algoritması, transfer öğrenme sonrası daha yüksek doğruluk ve f1 skoru değerlerine ulaşmıştır. İBH veri setinde ise RF algoritması ile doğruluk artışı gözlemlenmiştir.

5. TARTIŞMA

İnsan bağırsak mikrobiyomu, konakçı organizma ile sürekli bir etkileşim içinde olan, bakteriler, arkeler, mantarlar ve virüslerden oluşan karmaşık bir ekosistemdir. Mikrobiyom bağışıklık sisteminin gelişimi, metabolik süreçlerin düzenlenmesi ve patojenlere karşı direnç gibi önemli roller üstlenir (Masheghati ve ark., 2024). Mikrobiyomun viral bileşenleri arasında yer alan bakteriyofajlar bakteriyel popülasyonların düzenlenmesinde ve dolayısıyla mikrobiyal ekosistemin dengesinin korunmasında kritik bir rol oynar (Shen ve ark., 2024). Ancak, fajların insan sağlığı üzerindeki doğrudan etkileri ve hastalıklarla ilişkileri henüz tam olarak aydınlatılamamıştır. Bu tez çalışması kapsamında, kolorektal kanser (KK) ve inflamatuar bağırsak hastalığı (İBH) hastalarının bağırsak mikrobiyomlarına ait, shotgun metagenom dizileme yöntemleriyle elde edilmiş metagenomik veriler analiz edilerek insan mikrobiyomunun kritik bir bileşeni olan bakteriyofaj ekosistemi incelenmiştir.

Tez çalışması kapsamında tüm veri setlerinin her bir hastalık grubunda ortak yapılan bakteriyel taksonomik analizi sonucunda; KK metagenomlarında *Alistipes sp.* ve *Caprobacter sp*. türlerinin istatistiki olarak anlamlı oranda hasta grubunda, *Bacteriodes caccae*'nin istatistiki olarak anlamlı oranda sağlıklı grubunda fazla olduğunu raporlanmıştır. Kolorektal kanserle ilişkili patojenik bir bakteri olan *Alistipes* türü üzerine yapılan çok sayıda çalışma, sağlıklı bireylerle karşılaştırıldığında *Alistipes* cinsinin kolorektal kanser mukozasında belirgin şekilde zenginleştiğini ortaya koymuştur (Dai ve ark., 2018b; Park ve ark., 2021; Li ve ark., 2021). Ayrıca kolorektal kanserin progresyonu ile *Alistipes sp.*'nin pozitif korelasyon gösterdiği raporlanmıştır (Liu ve ark., 2023). Bu tez çalışmasında *Alistipes sp.*'nin KK hastalarında sağlıklılara göre istatistiki olarak daha fazla bulunması önceki çalışmalarla örtüşmektedir.

Bu çalışmada KK hastalarında Bacteroides fragilis, Bacteroides faecis ve Achromobacter xylosoxidans bakteriyofajlarının istatistiksel olarak anlamlı oranda sağlıklılardan fazla olduğu tespit edilmiştir. Bacteroides fragilis, sağlıklı insan bağırsak mikrobiyomunda doğal olarak bulunan bir bakteri olmakla birlikte, enterotoksijenik alt türleriyle (ETBF) patojenik özellikler sergileyebilir. ETBF'nin ürettiği toksinler, bağırsağın epitel bariyer bütünlüğünü bozarak kronik inflamasyona yol açabilmekte ve KK gelişiminde rol oynayabilmektedir (Pandey ve ark., 2023). Bağırsak mikrobiyotası içerisinde önemli bir yer tutan bu bakteri, fırsatçı enfeksiyonlara neden olabilir ve bağışıklık sistemini etkileyebilir. *Bacteroides fragilis*'in genetik çeşitliliği ve yatay gen transferi yeteneği, patojenik özelliklerini artırabilir (Evanovich ve ark., 2022). *Bacteroides fragilis*'in bağırsak enflamasyonu ile ilişkili olduğu ve interlökin-10 üretimini uyardığı gösterilmiştir (Cohen-Poradosu ve ark., 2011). *Bacteriodes faecis* insan bağırsağımda bulunan anaerobik, gram negatif, spor oluşturmayan çubuk şeklinde ve hareketsiz bir bakteri olup kolon kanserine ilerleyebilecek ülseratif kolit ve İBH'de klinik semptomlara katkı sağlayabileceği gösterilmiştir (Mohebali ve ark., 2023). *Achromobacter xylosoxidans* ise genellikle su ve toprak gibi çevresel ortamlarda bulunan, gram-negatif bir bakteri türüdür. Bu bakteri, özellikle bağışıklık sistemi zayıflamış bireylerde fırsatçı enfeksiyonlara neden olabilir. Kanser hastaları *A. xylosoxidans* kaynaklı bakteriyemi açısından artmış bir risk altındadır (Sánchez-Vargas ve ark., 2015). Michire ve ark. tarafından yapılan çalışmaya göre kanser hastalarının bağırsak mikrobiyomunda *A. xylosoxidans'*ın kontrol gruplarına kıyasla daha yüksek oranlarda bulunduğu tespit edilmiştir (Michire ve ark., 2022). Yapılan çalışmalar *Bacteriodes fragilis, Bacteroides faecis* ve *Achromobacter xylosoxidans* türlerinin bağırsak inflamasyonu ile ilişkili olduğunu göstermektedir. Bağırsakta inflamasyon geliştiğinde, bu tür bakteriler sayıca artabilir ve onlara özgü bakteriyofajlar da paralel olarak çoğalabilir (Gogokhia ve ark., 2021). Bu teoriye uygun olarak bu tez çalışmasında söz konusu bakterilere ait fajlar anlamlı oranda fazla bulunmuştur. Ancak bakteriyofajların bağışıklık sistemi ile etkileşim halinde olduğu da bilinmektedir. Filamentöz bakteriyofajlar bağırsak epitel hücrelerinde proinflamatuar yanıtları tetikleyebilir veya bağırsak epitel hücrelerinde IL-8 üretimini artırarak inflamasyonu şiddetlendirebilir (Varadan ve ark., 2024). Buna ek olarak fajlar doğrudan bağırsak

immün sistemini uyarabilir ve inflamatuar hastalıkların şiddetlenmesine neden olabilir. Özellikle *Bacteroides fragilis*'e özgü fajların interferon-gamma (IFN-γ) üretimini artırarak inflamasyonu tetiklediği gösterilmiştir (Gogokhia ve ark., 2019).

Benzer şekilde malignitesi olan, uzun süre hastanede kalan, birçok kez cerrahi girişim uygulanan hastalarda enfeksiyon etkeni olduğunu belirtilmiştir. Altta yatan hastalıkları nedeni ile çeşitli invaziv girişimlere maruz kalan, uzun süre hastanede yatan hastalarda gelişen çeşitli nazokomiyal enfeksiyonların (sepsis, yara enfeksiyonu, otit, menenjit) etkeninin *A. xylosoxidans* olduğu bildirilmiştir. Bu çalışmanın sonuçları tez verileriyle paralellik göstermektedir. Dolayısıyla, bu çalışmada hastalarda anlamlı oran artmış olduğu tespit edilen fajların, KK ile ilişkilendirilmiş bakterilerin fajları olmasından bağımsız olarak, immün sistemi uyarma özellikleri sebebiyle hastalığa katkı sağladığı söylenebilir. Belli fajların farklı konak aralığına sahip olabilmesi bu durumu destekler. Doğada düşünüldüğünden daha yaygın olarak fenetipik ya da genotipik geniş konak aralıklı (broad-host-range) bakteriyofajlar bulunabilmektedir. Fenotipik olarak geniş konak aralığına sahip olanlarda bireysel faj parçacıkları birden fazla konağı enfekte edebilir. Genotipik olanlarda ise bireysel parçacıklar yalnızca tek bir konakçıyı enfekte eder; ancak belirli mutasyonlar hızlı konak değiştirmeye olanak tanıyarak geniş konak aralıklı alt tiplerin oluşumunu sağlar. Ayrıca ekolojik parametreler, örneğin konak yoğunluğu ve konak çeşitliliği, belirli bir ortamda geniş konak aralıklı bakteriyofajların varlık olasılığını ve çeşitliliğini belirleyen faktörlerdir (Jonge ve ark., 2019).

Bu tez çalışmasında KK veri setlerinde *Bacteroides fragilis, Bacteroides faecis* ve *Achromobacter xylosoxidans* fajlarının hastalarda daha fazla bulunması ilgili bakterilerin bağıl bolluklarının da fazla olmasıyla ilişkilendirilebilir. Ancak çalışma kapsamında yapılan bakteriyel takson analizlerinde bu bakteriler için hasta-sağlıklı grupları karşılaştırmasında herhangi bir anlamlı sonuç elde edilememiştir. Bu duruma sebep olarak bu üç fajın belirtilen bakteriyel konaklar dışında farklı konakları da olabileceği hipotezi ortaya atılabilir. Benzer bir şekilde Jonge ve arkadaşlarının çalışmasında tez çalışmasında gözlenen bu durum şu şekilde açıklanmıştır; doğada daha önce düşünüldüğünden daha yaygın olan geniş konak aralıklı (broad-host-range) bakteriyofajlar, çoğu ortamda hem dar konak aralıklı

(narrow-host-range) hem de geniş konak aralıklı bakteriyofajların bir çeşitliliğine sahip olduğu görülmektedir. Fenotipik geniş konak aralıklı bakteriyofajlarda, bireysel faj parçacıkları birden fazla konağı enfekte edebilir. Genotipik geniş konak aralıklı bakteriyofajlarda ise, bireysel parçacıklar yalnızca tek bir konakçıyı enfekte eder; ancak belirli mutasyonlar, nesiller arasında hızlı bir konak değiştirmeye olanak tanıyarak geniş konak aralıklı alt türler topluluğu oluşumunu sağlar. Geniş konak aralıklı bakteriyofajlar, doğal ya da genetik mühendisliği ile üretilmiş olsalar da birden fazla suşun neden olduğu enfeksiyonların tedavisinde faydalıdır. Ayrıca ekolojik parametreler, örneğin konak yoğunluğu ve konak çeşitliliği, belirli bir ortamda geniş konak aralıklı bakteriyofajların varlık olasılığını ve çeşitliliğini belirleyen faktörlerdir (Jonge ve ark., 2019).

Bu tez çalışmasında *Bacteroides fragilis* bakteriyofajlarının İBH veri setlerinin bir tanesinde de hasta grubunda anlamlı oranda fazla tespit edilmiştir. Bu durum KK ve İBH'nin patogenezinde ortak mekanizmalar ve bağırsak mikrobiyotasındaki ortak değişikliklerle ilişkili olabileceği düşünülmektedir. *Bacteroides fragilis* bakteriyofajları, özellikle antibiyotik dirençli *B. fragilis* türlerine karşı potansiyel terapötik ajanlar olarak değerlendirilmektedir (Bakuradze ve ark., 2023). Bakuradze ve arkadaşları tarafından yapılan bir çalışmada, B. fragilis bakteriyofajının yüksek stabiliteye sahip olduğu ve tedavi amaçlı kullanılabileceği gösterilmiştir (Bakuradze ve ark., 2023). *B. fragilis*'in genom analizi yapıldığında, bu bakterinin gastrointestinal sistemde hem simbiyotik hem de patojenik roller üstlendiğini göstermektedir (Patrick, 2022). Özellikle, ülseratif kolit (ÜK) hastalarında yapılan bir çalışmada, *B. fragilis*'in ileal kesecik iltihabı (pouchitis) gelişimi öncesinde ve sırasında anlamlı düzeyde çoğaldığı gösterilmiştir (Vineis ve ark., 2016). Bu durum, *B. Fragilis*'in zorlu inflamatuar koşullara karşı adaptasyon yeteneği nedeniyle bağırsak mikrobiyotasındaki homeostatik dengenin bozulmasında önemli bir aktör olabileceğini göstermektedir. Mevcut tez çalışmasında, *B. fragilis* konağına ait bakteriyofajların İBH hastalarında anlamlı düzeyde artış gösterdiğinin belirlenmesi, *B. fragilis*'in patobiont potansiyeline ilişkin önceki bulguları desteklemektedir. Bu durum, sadece *B. fragilis*'in değil, ona ait bakteriyofajların da bağırsak mikrobiyotasındaki dinamikleri değiştirme potansiyeline sahip olabileceğini düşündürmektedir. Mevcut literatürle karşılaştırıldığında, *B. fragilis*'in mikrobiyota

içindeki patojenik potansiyeli ve immün sistem ile etkileşimi, bu bakteriyofajların hastalıkların tanı ve takibinde kullanılabileceğini göstermektedir.

Tez çalışmasında, İBH veri setlerinde KK veri setlerine göre daha fazla sayıda faj türü hastalıkla ilişkilendirilmiştir (KK için dört veri setinde toplam 21 faj çeşidi, İBH için iki veri setinde toplam 42 faj çeşidi). Bu durum KK'nin sadece kolonu tutmasına karşın, İBH'nin ülseratif kolit ve inflamatuvar bağırsak hastalığı olmak üzere sindirim kanalının farklı bölgelerini ya da tamamını tutabilen doğasıyla ilişki olabilir. Bakteriyofajların bağışıklık sistemi ile etkileşim halinde olması ve bakteriyel popülasyonları modüle ederek mikrobiyota dengesini etkilemesi, daha geniş bir alanı etkileyen İBH'de etkili faj çeşdinin fazla olmasına neden olmuş olabilir. Bağırsaktaki bakteriyofaj çeşitliliğinin artmasının hastalık şiddetiyle ilişkili olduğu bilinmektedir. Özellikle ülseratif kolit hastalarından alınan bakteriyofajlar, sağlıklı bireylere nakledildiğinde farelerde bağırsak inflamasyonunun arttığı gözlemlenmiştir (Sinha ve ark., 2021).

Çalışma kapsamında *Roseburia intestinalis* bakterisinin KK hastalarında, bu bakteriye ait fajın ise sağlıklı kontrol grubunda istatistiksel olarak anlamlı düzeyde daha fazla olduğu tespit edilmiştir. Chattopadhyay ve arkadaşları tarafından yapılan bir çalışmada *R.intestinalis*'in kolorektal kanser hasta gruplarında sıkça raporlandığı ve hastalıklı ilişkili olduğu tespit edilmiştir (Chattopadhyay ve ark., 2021). KK ile ilişkilendirilmiş bir bakterinin fajının sağlıklıda fazla bulunması, bağırsak mikrobiyotasındaki faj-bakteri etkileşimlerinin hastalık patogenezinde önemli bir rol oynayabileceğinin göstergesi olabilir. Benzer şekilde İBH veri setlerinde yapılan analizler sonucunda, *Eubacterium rectale*'in hasta grubunda sağlıklı kontrollere kıyasla istatistiki olarak anlamlı oranda fazla bulunduğu, buna karşın *Eubacterium rectale* bakteriyofajının sağlıklı kontrol grubunda istatistiki olarak anlamlı oranda fazla olduğu tespit edilmiştir. Literatürde, *Eubacterium rectale*'in KK ve İBH hastalarında "sürücü bakteri" olarak rol oynadığı ve bu hastalıkların ilerlemesinde aktif görev alabileceği daha önce rapor edilmiştir (Wang ve ark., 2021).

KK'i ve İBH metagenom veri setlerinin makine öğrenme algoritmaları ve transfer öğrenme modeli kullanılarak faj bağıl bollukları üzerinden yapılan analizlerle,

hastalık sağlıklı durumu sınıflandırma becerisi hesaplanmıştır. Bu analizler neticesinde, makine öğrenimi algoritmalarıyla gerçekleştirilen sınıflandırmalarda, KK için %74, İBH için ise %66 başarı oranı elde edilmiştir. MLP algoritması, RF ve XGBoost algoritmalarına kıyasla en yüksek performansı sergileyen yöntem olarak belirlenmiştir. ML'nin yüksek performansı, özellikle veri setindeki doğrusal olmayan ilişkiler ve karmaşık etkileşimleri modelleme konusundaki üstün yeteneğinden kaynaklanmaktadır. Omik çalışmalarında yüksek sınıflandırma başarıları nedeniyle bu algoritma sıklıkla tercih edilmektedir (Licciardi ve ark., 2024). Bu çalışmada kullanılan veri setlerindeki heterojenite, veri dengesizliği ve örneklem sayısındaki yetersizlikler sebebiyle sınıflandırma başarısının olması gerekenden düşük çıkmış olabileceği ihtimaline karşı transfer öğrenme yöntemiyle sınıflandırma başarısının artırılması hedeflenmiştir. Transfer öğrenme modeli, önceden eğitilmiş modellerin bilgi birikiminden faydalanarak, sınırlı veri ve dengesiz veri setlerinde daha yüksek genelleme yeteneği sağlayarak sınıflandırma doğruluğunu artırmayı amaçlamaktadır (Hosna ve ark., 2022). Bu yöntem, omik verilerin karmaşıklığı ve heterojen yapısı göz önünde bulundurulduğunda, sınıflandırma başarılarını iyileştirmek için güçlü bir araç olarak öne çıkmaktadır. Transfer öğrenmesinin, sınırlı sayıda örnekle birden fazla hastalık için mikrobiyal toplulukların sınıflandırılması ve KK'in ardışık evrelerinde bağırsak mikrobiyomundaki değişikliklerin tahmini gibi çeşitli bağlamlarda mikrobiyal topluluk sınıflandırmasını kolaylaştırabileceği gösterilmiştir (Chong ve ark., 2022). Bu çalışmaların ışığında KK ve İBH veri setleri üzerinde gerçekleştirilen transfer öğrenme modeli sonuçlarına göre KK için %78, İBH için %74 sınıflandırma başarısına ulaşılmıştır. KK veri setinde XGBoost algoritması, MLP ve RF algoritmalarına kıyasla en yüksek performansı sergileyen yöntem olarak belirlenmiştir. İBH veri setinde ise RF algoritması, XGBoost algoritması ve MLP kıyasla en yüksek performansı sergileyen yöntem olarak belirlenmiştir.

Bu tez çalışması, bakteriyofajların bağırsak mikrobiyotasındaki dinamikleri ve iki kronik hastalık olan KK ve İBH ile ilişkilerini anlamaya yönelik önemli bulgular sunmaktadır. Ancak, çalışmanın bazı kısıtlamaları bulunmaktadır. Çalışmada kullanılan veri setleri bilimsel kullanıma açık platformlardan edinilmiştir. Bu veriler genellikle total DNA izolasyon basamaklarını içeren ıslak laboratuvar süreçleri sonucunda üretilmiş olup, bakteriyofaj nükleik asitlerini hedefleyen bir filtrasyon,

zenginleştirme ya da viral izolasyon basamağı içermemektedir. Her ne kadar total DNA izolasyonu sonrasında gerçekleştirilen metagenom dizilemede (belirli derinlikte olduğu sürece) yüksek miktarlarda faj DNA'sı da diziliyor olsa da bu çalışmada kullanılan verilerin spesifik olarak bakteriyofaj izolasyonu ve analizine yönelik tasarlanış projelerde üretilmemiş olması faj çeşitliliğinin tam olarak ortaya konulmama potansiyeli açısından kısıtlılık oluşturmaktadır. Bir diğer kısıtlılık olarak, kullanılan veri setlerinde yer alan hastaların metaverilerinin (klinik bilgiler, hastalık alt tipleri gibi) yokluğu, çalışmaya dahil edilen hastalık alt fenotipleri arasındaki faj varyasyonlarının incelenememesine neden olmuştur.

Sonuç olarak, bu tez çalışması ile farklı veri setlerine ait mikrobiyom faj kompozisyonları incelenerek yeni sinyaller keşfedilmiştir. Bu biyobelirteçler KK'i ve İBH'nın özellikle erken teşhisinde non-invaziv yöntemlerin geliştirilmesi için kullanılma potansiyeli taşımaktadır.

6. SONUÇ VE ÖNERİLER

Bu tez çalışmasında, KK'i ve İBH hastalarının bilimsel kullanıma açık veri tabanlarından erişilen shotgun metagenom yöntemleriyle dizilenmiş veriler üzerinde virom analizleri yapılarak, faj bağıl bollukları hesaplanmış, hasta-sağlıklı gruplarda karşılaştırılmış, yapay zekâ algoritmaları ve öznitelik seçimleri yapılarak fajların hastalık ve sağlık durumunu ayırt etme gücü analiz edilmiştir.

Sonuç olarak, bu tez, mikrobiyom sınıflandırma problemlerinde makine öğrenimi algoritmalarının güçlü bir araç olduğunu ve transfer öğrenme yönteminin sınıflandırma başarımlarını daha da ileriye taşıyabildiğini ortaya koymaktadır. Ancak, bu yöntemlerin daha geniş veri setleri ve farklı hastalık durumlarında test edilmesi, virom dünyasının daha iyi anlaşılmasına ve bu alandaki bilgi birikiminin artırılmasına katkı sağlayacaktır.

Bu tez çalışması, insan mikrobiyomunun bakteriyofaj kompozisyonu ile hastalık durumları arasındaki ilişkilerin aydınlatılmasında ML ve transfer öğrenmenin uygulanabilirliğini ortaya koymuştur. Elde edilen sonuçlar, bakteriyofajların potansiyel biyobelirteç olarak kullanılabileceğini ve bu alanda ileri araştırmaların yapılmasının önemini vurgulamaktadır. Ayrıca bu çalışma, mikrobiyom verileriyle çalışan araştırmacılar için bir yol haritası sunarak, ML algoritmalarının seçimi ve optimizasyonunda dikkat edilmesi gereken noktaları ortaya koymuştur.

Transfer öğrenme, yalnızca mevcut veri setlerinde değil, gelecekte yapılacak mikrobiyom araştırmalarında da geniş bir potansiyele sahiptir. Özellikle, bu yöntemin daha büyük veri setleri, farklı omik türleri (örn. metatranskriptomik veya metabolomik) ve diğer hastalık bağlamlarında uygulanması, mikrobiyom araştırmalarını yeni bir boyuta taşıyabilir. Transfer öğrenme bu tez çalışmasında sadece sınıflandırma performansını artırmakla kalmamış, aynı zamanda mikrobiyom verilerinden elde edilen biyolojik bilgilerin daha geniş bir bağlamda

kullanılabilirliğini göstermiştir. Bu yöntem, mikrobiyomun hastalık teşhisinde ve tedavi hedeflerinde daha etkili kullanılmasına yönelik önemli bir adımdır.

Sonuç olarak bu tez kapsamında bakteriofaj komünitelerinin çalışmaya konu iki hastalık ile ilişkilendirilmesi yapılmış, mikrobiyom sınıflandırma problemlerinde makine öğrenimi algoritmalarının güçlü bir araç olabileceği ve transfer öğrenme yönteminin sınıflandırma başarımlarını daha da ileriye taşıyabildiğini ortaya konulmuştur. Ancak, bu yöntemlerin daha geniş veri setleri ve farklı hastalık durumlarında test edilmesi, virom dünyasının daha iyi anlaşılmasına ve bu alandaki bilgi birikiminin artırılmasına katkı sağlayacaktır.

Önemli bulgular;

1- KK'ye veri setlerinde, *Bacteroides fragilis, Bacteroides faecis, Achromobacter xylosoxidans* bakteriyofajları, İBH'ye ait veri setlerinde *Ruminococcaceae bacterium, Coprobacillus sp. AF13-15* bakteriyofajlarına ait okumaların istatistiki olarak anlamlı oranda hasta grubunda fazla olduğu tespit edildi.

2- KK ve İBH veri setlerinde *Bactreriodes fragilis* konağına ait bakteriyofajın hasta gruplarında ortak olduğu tespit edildi.

3- İBH veri setlerinde KK veri setlerine göre daha fazla sayıda faj türü hastalıkla ilişkilendirilmiştir (KK için dört veri setinde toplam 21 faj çeşidi, İBH için iki veri setinde toplam 42 faj çeşidi).

4- *Eubacterium rectale* bakterisinin İBH hasta grubunda sağlıklı kontrollere kıyasla istatistiki olarak anlamlı oranda fazla bulunduğu, buna karşın *Eubacterium rectale* bakteriyofajının sağlıklı kontrol grubunda istatistiki olarak anlamlı oranda fazla olduğu tespit edildi.

5- *Roseburia intestinalis* bakterisinin KK hastalarında, bu bakteriye ait fajın ise sağlıklı kontrol grubunda istatistiki olarak anlamlı oranda daha fazla bulunduğu tespit edilmiştir.

6- Faj bağıl bollukları üzerinden yapay zekâ algoritmalarıyla ve öznitelik seçimleriyle yapılan analizler sonucunda, fajların hastalık ve sağlık durumunu ayırt etme gücü KK için en yüksek %74, İBH için en yüksek %66 olarak hesaplanmıştır.

7- Sınıflandırma performansını artırmak için kullanılan transfer öğrenme modelleriyle fajların hastalık ve sağlık durumunu ayırt etme gücü KK için en yüksek %78, İBH için en yüksek %74 sınıflandırma başarısına ulaşmıştır.

7. KAYNAKÇA

Ackermann HW. 5500 Phages examined in the electron microscope. *Arch Virol*. 2007;152(2):227-243.

Ajami NJ, Petrosino JF. Chapter 5.1 - Enteric Viral Metagenomics. In: Svensson L, Desselberger U, Greenberg HB, Estes MK, eds. *Viral Gastroenteritis*. Academic Press; 2016:523-533.

Arabameri A, Asemani D, Teymourpour P. Detection of Colorectal Carcinoma Based on Microbiota Analysis Using Generalized Regression Neural Networks and Nonlinear Feature Selection. *IEEE/ACM Trans Comput Biol Bioinform*. 2020;17(2):547-557.

Bakuradze N, Merabishvili M, Kusradze I. Characterization of a Bacteriophage GEC_vB_Bfr_UZM3 Active against Bacteroides fragilis. *Viruses*. 2023;15:41-49.

Bedani R, Cucick ACC, Albuquerque MAC de, LeBlanc JG, Saad SMI. B-Group Vitamins as Potential Prebiotic Candidates: Their Effects on the Human Gut Microbiome. *J Nutr*. 2024;154(2):341-353.

Belibasakis GN, Seneviratne CJ, Jayasinghe RD, Vo PTD, Bostanci N, Choi Y. Bacteriome and mycobiome dysbiosis in oral mucosal dysplasia and oral cancer. *Periodontol 2000*. 2024;96(1):95-111.

Bhagchandani T, Nikita, Verma A, Tandon R. Exploring the Human Virome: Composition, Dynamics, and Implications for Health and Disease. *Curr Microbiol*. 2023;81(1):16-25.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2125.

Boone SA, Gerba CP. Significance of Fomites in the Spread of Respiratory and Enteric Viral Disease. *Appl Environ Microbiol*. 2007;73(6):1687-1696.

Bosch A, Pintó RM, Guix S. Human Astroviruses. *Clin Microbiol Rev*. 2014;27(4):1048-1074.

Breitbart M, Salamon P, Andresen B. Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci U S A*. 2002;99(22):14250-14255.

Bürtin F, Mullins CS, Linnebacher M. Mouse models of colorectal cancer: Past, present and future perspectives. *World J Gastroenterol*. 2020;26(13):1394-1426.

Cadwell K. The Virome in Host Health and Disease. *Immunity*. 2015;42(5):805-813.

Cani PD, Amar J, Iglesias MA. Metabolic endotoxemia initiates obesity and insulin resistance. *Diabetes*. 2007;56(7):1761-1772.

Cantalupo PG, Pipas JM. Detecting viral sequences in NGS data. *Curr Opin Virol*. 2019;39:41-48.

Chang Y, Cesarman E, Pessin MS. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science*. 1994;266(5192):1865-1869.

Chatterjee A, Duerkop BA. Beyond Bacteria: Bacteriophage-Eukaryotic Host Interactions Reveal Emerging Paradigms of Health and Disease. *Front Microbiol*. 2018;9:1394-1415.

Chattopadhyay I, Dhar R, Pethusamy K. Exploring the Role of Gut Microbiome in Colon Cancer. *Appl Biochem Biotechnol*. 2021;193(6):1780-1799.

Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*. 2020;36(6):1925-1927.

Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Int Res*. 2002;16(1):321-357.

Chen C, Zhang Y, Yao X. Characterizations of the Gut Bacteriome, Mycobiome, and Virome in Patients with Osteoarthritis. *Microbiol Spectr*. 2022;11(1):e01711-22.

Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery*

and Data Mining. KDD ’16. Association for Computing Machinery; 2016:785-794.

Chong H, Zha Y, Yu Q. EXPERT: transfer learning-enabled context-aware microbial community classification. *Brief Bioinform*. 2022;23(6):bbac396.

Çiftcioğlu UGE, Nalbanoglu OU. DeepGum: Deep feature transfer for gut microbiome analysis using bottleneck models. *Biomed Signal Process Control*. 2024;91:105984.

Çiftcioğlu UGE, Nalbantoglu OU. Disrobiom: A novel approach to discover robust biomarkers from gut microbiome datasets with deep-learning algorithms. *Biomed Signal Process Control*. 2025;100:106935.

Cohen-Poradosu R, McLoughlin R, Lee JC, Kasper D. Bacteroides fragilis-stimulated interleukin-10 contains expanding disease. *J Infect Dis*. 2011;204 3:363-371.

Conceição-Neto N, Zeller M, Lefrère H. Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Sci Rep*. 2015;5(1):16532-16540.

Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273-297.

Da Cunha T, Vaziri H. Interval Colorectal Cancer in Inflammatory Bowel Disease: A Review. *J Clin Gastroenterol*. 2024;58(1):1-15.

Dai Z, Coker OO, Nakatsu G. Multi-cohort analysis of colorectal cancer metagenome identified altered bacteria across populations and universal bacterial markers. *Microbiome*. 2018;6(1):70-85.

Dai Z, Coker OO, Nakatsu G. Multi-cohort analysis of colorectal cancer metagenome identified altered bacteria across populations and universal bacterial markers. *Microbiome*. 2018;6:70-85.

Dalianis T, Hirsch HH. Human polyomaviruses in disease and cancer. *Virology*. 2013;437(2):63-72.

Dietrich T, Sharma P, Walter C, Weston P, Beck J. The epidemiological evidence behind the association between periodontitis and incident atherosclerotic cardiovascular disease. *J Clin Periodontol*. 2013;40 Suppl 14:S70-84.

Dorobisz K, Dorobisz T, Pazdro-Zastawny K. Assessment of Prognostic Factors, Clinical Features Including the Microbiome, and Treatment Outcomes in Patients with Cancer of Unknown Primary Site. *Cancers*. 2024;16(19):3416.

Dowah ASA, Clokie MRJ. Review of the nature, diversity and structure of bacteriophage receptor binding proteins that target Gram-positive bacteria. *Biophys Rev*. 2018;10(2):535-542.

Evanovich E, Mattos PJ de SM, Guerreiro J. Comprehensive Comparative Genomic revels: Bacteroides fragilis is a reservoir of antibiotic resistance genes in the gut microbiota. *bioRxiv*. Published online 2022.

Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science*. 2008;319(5866):1096-1100.

Feng Q, Liang S, Jia H. Gut microbiome development along the colorectal adenoma–carcinoma sequence. *Nat Commun*. 2015;6(1):6528-6537.

Flemer B, Lynch DB, Brown JMR. Tumour-associated and non-tumour-associated microbiota in colorectal cancer. *Gut*. 2017;66(4):633-643.

Franzosa EA, Sirota-Madi A, Avila-Pacheco J. Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat Microbiol*. 2019;4(2):293-305.

Fusco W, Lorenzo MB, Cintoni M. Short-Chain Fatty-Acid-Producing Bacteria: Key Components of the Human Gut Microbiota. *Nutrients*. 2023;15(9):2211.

Gallardo P, Izquierdo M, Viver T. A metagenomic approach to unveil the association between fecal gut microbiota and short-chain fatty acids in diarrhea caused by diarrheagenic Escherichia coli in children. *Microb Cell Graz Austria*. 2024;11:116-127.

Garmaeva S, Sinha T, Kurilshikov A, Fu J, Wijmenga C, Zhernakova A. Studying the gut virome in the metagenomic era: challenges and perspectives. *BMC Biol*. 2019;17(1):84-95.

Gogokhia L, Buhrke K, Bell R. Expansion of Bacteriophages Is Linked to Aggravated Intestinal Inflammation and Colitis. *Cell Host Microbe*. 2019;25 2:285-2998.

Gogokhia L, Round JL. Immune–bacteriophage interactions in inflammatory bowel diseases. *Curr Opin Virol*. 2021;49:30-35.

Gonzalez-Cao M, Iduma P, Karachaliou N, Santarpia M, Blanco J, Rosell R. Human endogenous retroviruses and cancer. *Cancer Biol Med*. 2016;13(4):483-488.

Grard G, Fair JN, Lee D. A Novel Rhabdovirus Associated with Acute Hemorrhagic Fever in Central Africa. *PLOS Pathog*. 2012;8(9):e1002924.

Gu W, Miller S, Chiu CY. Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. *Annu Rev Pathol Mech Dis*. 2019;14(Volume 14, 2019):319-338.

Hannigan GD, Duhaime MB, Ruffin MT, Koumpouras CC, Schloss PD. Diagnostic Potential and Interactive Dynamics of the Colorectal Cancer Virome. *mBio*. 2018;9(6):e02248-18.

Haynes M, Rohwer F. The Human Virome. In: Nelson KE, ed. *Metagenomics of the Human Body*. Springer; 2011:63-77.

He Y, Wu W, Zheng HM. Regional variation limits applications of healthy gut microbiome reference ranges and disease models. *Nat Med*. 2018;24(10):1532-1535.

Heijstek MW, Kranenburg O, Borel Rinkes IHM. Mouse models of colorectal cancer and liver metastases. *Dig Surg*. 2005;22(1-2):16-25.

Hernández Medina R, Kutuzova S, Nielsen KN. Machine learning and deep learning applications in microbiome research. *ISME Commun*. 2022;2(1):98-109.

Ho SX, Law JH, Png CW. Alterations in colorectal cancer virome and its persistence after surgery. *Sci Rep*. 2024;14(1):2819-2835.

Hobbs Z, Abedon ST. Diversity of phage infection types and associated terminology: the problem with "Lytic or lysogenic." *FEMS Microbiol Lett*. 2016;363(7):fnw047.

Hodgkinson K, El Abbar F, Dobranowski P. Butyrate's role in human health and the current progress towards its clinical application to treat gastrointestinal disease. *Clin Nutr*. 2023;42(2):61-75.

Holtappels D, Alfenas-Zerbini P, Koskella B. Drivers and consequences of bacteriophage host range. *FEMS Microbiol Rev*. 2023;47(4):fuad038.

Hosna A, Merry E, Gyalmo J, Alom Z, Aung Z, Azim MA. Transfer learning: a friendly introduction. *J Big Data*. 2022;9(1):102-119.

Hu X, Yu C, He Y. Integrative metagenomic analysis reveals distinct gut microbial signatures related to obesity. *BMC Microbiol*. 2024;24(1):119-130.

Jagga Z, Gupta D. Machine learning for biomarker identification in cancer research - developments toward its clinical application. *Pers Med*. 2015;12(4):371-387.

Jonge PA de, Nobrega FL, Brouns SJJ, Dutilh BE. Molecular and Evolutionary Determinants of Bacteriophage Host Range. *Trends Microbiol*. 2019;27(1):51-63.

Joossens M, Huys G, Cnockaert M. Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut*. 2011;60(5):631-637.

Kamath HS, Shukla R, Shah U. Role of Gut Microbiota in Predisposition toColon Cancer: A Narrative Review. *Indian J Microbiol*. 2024;64(3):1-13.

Karagoz MA, Nalbantoglu OU. A self-supervised learning model based on variational autoencoder for limited-sample mammogram classification. *Appl Intell*. 2024;54(4):3448-3463.

Karlsson FH, Tremaroli V, Nookaew I. Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature*. 2013;498(7452):99-103.

Kersten K, de Visser KE, van Miltenburg MH, Jonkers J. Genetically engineered mouse models in oncology research and cancer medicine. *EMBO Mol Med*. 2017;9(2):137-153.

Khabaz MN, Nedjadi T, Gari MA. Simian Virus 40 is Not Likely Involved in the Development of Colorectal Adenocarcinoma. *Future Virol*. 2016;11(3):175-180.

Koeth RA, Wang Z, Levison BS. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat Med*. 2013;19(5):576-585.

Kostic AD, Xavier RJ, Gevers D. The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology*. 2014;146(6):1489-1499.

Krause T, Wassan JT, Mc Kevitt P, Wang H, Zheng H, Hemmje M. Analyzing Large Microbiome Datasets Using Machine Learning and Big Data. *BioMedInformatics*. 2021;1(3):138-165.

Krause T, Wassan JT, Mc Kevitt P, Wang H, Zheng H, Hemmje M. Analyzing Large Microbiome Datasets Using Machine Learning and Big Data. *BioMedInformatics*. 2021;1(3):138-165.

Küry P, Nath A, Créange A. Human Endogenous Retroviruses in Neurological Diseases. *Trends Mol Med*. 2018;24(4):379-394.

Kutter E, Sulakvelidze A, eds. *Bacteriophages: Biology and Applications*. CRC Press; 2005.

Lax S, Smith DP, Hampton-Marcell J. Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science*. 2014;345(6200):1048-1052.

Lazarevic V, Whiteson K, Huse S. Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. *J Microbiol Methods*. 2009;79(3):266-271.

Lecuit M, Eloit M. The human virome: new tools and concepts. *Trends Microbiol*. 2013;21(10):510-515.

Leon-Derecho CMP de. Chapter 17 - Microbiome-based therapeutics. In: Dable-Tupas G, Karunakaran R, Lim PPC, Otero MCB, eds. *Human Microbiome Drug Targets*. Elsevier; 2025:233-244.

Li L, Deng X, Linsuwanon P. AIDS Alters the Commensal Plasma Virome. *J Virol*. 2013;87(19):10912-10915.

Li N, Bai C, Zhao L, Sun Z, Ge Y, Li X. The Relationship Between Gut Microbiome Features and Chemotherapy Response in Gastrointestinal Cancer. *Front Oncol*. 2021;11:781697.

Li W, Lee MH, Henderson L. Human endogenous retrovirus-K contributes to motor neuron disease. *Sci Transl Med*. 2015;7(307):307ra153.

Li Y, Ma J, Meng J. Structural changes in the gut virome of patients with atherosclerotic cardiovascular disease. *Microbiol Spectr*. 2023;12(1):e01050-

Liang G, Bushman FD. The human virome: assembly, composition and host interactions. *Nat Rev Microbiol*. 2021;19(8):514-527.

Licciardi A, Fiannaca A, La Rosa M, Urso MA, La Paglia L. A Deep Learning Multi-omics Framework to Combine Microbiome and Metabolome Profiles for Disease Classification. In: Wand M, Malinovská K, Schmidhuber J, Tetko

IV, eds. *Artificial Neural Networks and Machine Learning – ICANN 2024*. Springer Nature Switzerland; 2024:3:14-25.

Light SH, Nagler CR. Regulation of immune responses to food by commensal microbes. *Immunol Rev*. 2024;326(1):203-218.

Lin Z, Chen Q, Ruan HB. To die or not to die: Gasdermins in intestinal health and disease. *Semin Immunol*. 2024;71:101865.

Liping Z, Sheng Y, Yinhang W. Comprehensive retrospect and future perspective on bacteriophage and cancer. *Virol J*. 2024;21(1):278-290.

Liu J, Huang X, Chen C. Identification of colorectal cancer progression-associated intestinal microbiome and predictive signature construction. *J Transl Med*. 2023;21(1):373.

Liu Y, Wang Y, Zhang J. New Machine Learning Algorithm: Random Forest. In: Liu B, Ma M, Chang J, eds. *Information Computing and Applications*. Springer; 2012:246-252.

Lopetuso LR, Ianiro G, Scaldaferri F, Cammarota G, Gasbarrini A. Gut Virome and Inflammatory Bowel Disease. *Inflamm Bowel Dis*. 2016;22(7):1708-1712.

Ma Y, Zhang Y, Xiang J. Metagenome Analysis of Intestinal Bacteria in Healthy People, Patients With Inflammatory Bowel Disease and Colorectal Cancer. *Front Cell Infect Microbiol*. 2021;11:599734.

Malongane F, Berejena T. Exploring the microbiome present in fermented indigenous African foods and their potential impact on human health. *J Agric Food Res*. 2024;16:101101.

Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. *Microbiome*. 2015;3(1):31-45.

Masheghati F, Asgharzadeh MR, Jafari A, Masoudi N, Maleki-Kakelar H. The role of gut microbiota and probiotics in preventing, treating, and boosting the immune system in colorectal cancer. *Life Sci*. 2024;344:122529.

Massimino L, Lovisa S, Antonio Lamparelli L, Danese S, Ungaro F. Gut eukaryotic virome in colorectal carcinogenesis: Is that a trigger? *Comput Struct Biotechnol J*. 2020;19:16-28.

Massimino L, Palmieri O, Facoetti A. Gut virome-colonising Orthohepadnavirus genus is associated with ulcerative colitis pathogenesis and induces intestinal inflammation in vivo. *Gut*. 2023;72(10):1838-1847.

McDonald E, Bombei K, Weber M. The Gut Microbiome in Chronic Diseases. *Celebr Res*. Published online April 11, 2024.

Michire A, Anghel R, Draghia PM. The Microbiota and the Relationship with Colorectal Cancer: Surgical Complications—A Review. *Gastrointest Disord*. 2022;4(2):66-76.

Miller RR, Montoya V, Gardy JL, Patrick DM, Tang P. Metagenomics for pathogen detection in public health. *Genome Med*. 2013;5(9):81-95.

Miyauchi E, Shimokawa C, Steimle A, Desai MS, Ohno H. The impact of the gut microbiome on extra-intestinal autoimmune diseases. *Nat Rev Immunol*. 2023;23(1):9-23.

Mohebali N, Weigel M, Hain T. *Faecalibacterium prausnitzii*, *Bacteroides faecis* and *Roseburia intestinalis* attenuate clinical symptoms of experimental colitis by regulating Treg/Th17 cell balance and intestinal barrier integrity. *Biomed Pharmacother*. 2023;167:115568.

Mulder RH, Kraaij R, Schuurmans IK. Early-life stress and the gut microbiome: A comprehensive population-based investigation. *Brain Behav Immun*. 2024;118:117-127.

Nakatsu G, Zhou H, Wu WKK. Alterations in Enteric Virome Are Associated With Colorectal Cancer and Survival Outcomes. *Gastroenterology*. 2018;155(2):529-541.e5.

Nicastro J, Wong S, Khazaei Z, Lam P, Blay J, Slavcev RA. *Bacteriophage Applications - Historical Perspective and Future Potential*. Springer International Publishing; 2016.

Nielsen HB, Almeida M, Juncker AS. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat Biotechnol*. 2014;32(8):822-828.

Nishida A, Inoue R, Inatomi O, Bamba S, Naito Y, Andoh A. Gut microbiota in the pathogenesis of inflammatory bowel disease. *Clin J Gastroenterol*. 2018;11(1):1-10.

Norman JM, Handley SA, Baldridge MT. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell*. 2015;160(3):447-460.

de Noronha SISR, de Moraes LAG, Hassell JE. High-fat diet, microbiome-gut-brain axis signaling, and anxiety-like behavior in male rats. *Biol Res*. 2024;57(1):23-45.

Novielli P, Romano D, Magarelli M. Explainable artificial intelligence for microbiome data analysis in colorectal cancer biomarker identification. *Front Microbiol*. 2024;15.

O'Rourke KP, Loizou E, Livshits G. Transplantation of engineered organoids enables rapid generation of metastatic mouse models of colorectal cancer. *Nat Biotechnol*. 2017;35(6):577-582.

Pandey H, Tang DWT, Wong SH, Lal D. Gut Microbiota in Colorectal Cancer: Biological Role and Therapeutic Opportunities. *Cancers*. 2023;15(3):866.

Park J, Kim NE, Yoon H. Fecal Microbiota and Gut Microbe-Derived Extracellular Vesicles in Colorectal Cancer. *Front Oncol*. 2021;11:650026.

Pasolli E, Truong DT, Malik F, Waldron L, Segata N. Machine Learning Meta-analysis of Large Metagenomic Datasets: Tools and Biological Insights. *PLOS Comput Biol*. 2016;12(7):e1004977.

Patrick S. A tale of two habitats: Bacteroides fragilis, a lethal pathogen and resident in the human gastrointestinal microbiome. *Microbiol Read Engl*. 2022;168(4):45-65.

Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*. 2012;28(11):1420-1428.

Peters BA, Dominianni C, Shapiro JA. The gut microbiota in conventional and serrated precursors of colorectal cancer. *Microbiome*. 2016;4(1):69.

Qin J, Li Y, Cai Z. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*. 2012;490(7418):55-60.

Ren J, Song K, Deng C. Identifying viruses from metagenomic data using deep learning. *Quant Biol*. 2020;8(1):64-77.

Rinke C, Lee J, Nath N. Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat Protoc*. 2014;9(5):1038-1048.

Rose R, Constantinides B, Tapinos A, Robertson DL, Prosperi M. Challenges in the analysis of viral metagenomes. *Virus Evol*. 2016;2(2):vew022.

Saedi S, Derakhshan S, Hasani A. Recent Advances in Gut Microbiome Modulation: Effect of Probiotics, Prebiotics, Synbiotics, and Postbiotics in Inflammatory Bowel Disease Prevention and Treatment. *Curr Microbiol*. 2024;82(1):12.

Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol*. 1975;94(3):441-448.

Santiago-Rodriguez TM, Hollister EB. Human Virome and Disease: High-Throughput Sequencing for Virus Discovery, Identification of Phage-Bacteria Dysbiosis and Development of Therapeutic Approaches with Emphasis on the Human Gut. *Viruses*. 2019;11(7):656-670.

Scher JU, Sczesnak A, Longman RS. Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. *eLife*. 2013;2:e01202.

Schiller JT, Castellsagué X, Garland SM. A review of clinical trials of human papillomavirus prophylactic vaccines. *Vaccine*. 2012;30 Suppl 5(0 5):F123-138.

Schoggins JW, MacDuff DA, Imanaka N. Pan-viral specificity of IFN-induced genes reveals new roles for cGAS in innate immunity. *Nature*. 2014;505(7485):691-695.

Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol*. 2016;14(8):e1002533.

Shang J, Tang X, Guo R, Sun Y. Accurate identification of bacteriophages from metagenomic data using Transformer. *Brief Bioinform*. 2022;23(4):bbac258.

Shay E, Sangwan N, Padmanabhan R, Lundy S, Burkey B, Eng C. Bacteriome and mycobiome and bacteriome-mycobiome interactions in head and neck squamous cell carcinoma. *Oncotarget*. 2020;11(25):2375-2386.

Shen Y, Qian Q, Ding L. High-throughput single-microbe RNA sequencing reveals adaptive state heterogeneity and host-phage activity associations in human gut microbiome. *Protein Cell*. Published online May 23, 2024:pwae027.

Shkoporov AN, Hill C. Bacteriophages of the Human Gut: The “Known Unknown” of the Microbiome. *Cell Host Microbe*. 2019;25(2):195-209.

Shuda M, Kwun HJ, Feng H, Chang Y, Moore PS. Human Merkel cell polyomavirus small T antigen is an oncoprotein targeting the 4E-BP1 translation regulator. *J Clin Invest*. 2011;121(9):3623-3634.

Sinha A, Li YH, Mirzaei M. Transplantation of bacteriophages from ulcerative colitis patients shifts the gut bacteriome and exacerbates the severity of DSS colitis. *Microbiome*. 2021;10.

Sung J, Rajendraprasad SS, Philbrick KL. The human gut microbiome in critical illness: disruptions, consequences, and therapeutic frontiers. *J Crit Care*. 2024;79:154436.

Sutton TDS, Clooney AG, Ryan FJ, Ross RP, Hill C. Choice of assembly software has a critical impact on virome characterisation. *Microbiome*. 2019;7(1):12.

Taketo MM, Edelmann W. Mouse models of colon cancer. *Gastroenterology*. 2009;136(3):780-798.

Tan L, Guo Z, Wang X, Kim DY, Li R. Utilization of nanopore direct RNA sequencing to analyze viral RNA modifications. *mSystems*. 2024;9(2):e01163-23.

Teixeira M, Silva F, Ferreira RM, Pereira T, Figueiredo C, Oliveira HP. A review of machine learning methods for cancer characterization from microbiome data. *NPJ Precis Oncol*. 2024;8(1):123-135.

Thissen JB, Isshiki M, Jaing C. A novel variant of torque teno virus 7 identified in patients with Kawasaki disease. *PLOS ONE*. 2018;13(12):e0209683.

Topçuoğlu BD, Lesniak NA, Ruffin MT, Wiens J, Schloss PD. A Framework for Effective Application of Machine Learning to Microbiome-Based Classification Problems. *mBio*. 2020;11(3):e00434-450.

Tringe SG, von Mering C, Kobayashi A. Comparative metagenomics of microbial communities. *Science*. 2005;308(5721):554-557.

Truong DT, Franzosa EA, Tickle TL. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods*. 2015;12(10):902-903.

Tun HM, Peng Y, Massimino L. Gut virome in inflammatory bowel disease and beyond. *Gut*. 2024;73(2):350-360.

VanDussen KL, Marinshaw JM, Shaikh N. Development of an enhanced human gastrointestinal epithelial culture system to facilitate patient-based assays. *Gut*. 2015;64(6):911-920.

Varadan AC, Grasis JA. Filamentous bacteriophages induce proinflammatory responses in intestinal epithelial cells. Published online June 24, 2024.

Vilchez RA, Butel JS. Emergent Human Pathogen Simian Virus 40 and Its Role in Cancer. *Clin Microbiol Rev*. 2004;17(3):495-508.

Villanacci V, Reggiani-Bonetti L, Salviato T. Histopathology of IBD Colitis. A practical approach from the pathologists of the Italian Group for the study of the gastrointestinal tract (GIPAD). *Pathologica*. 2021;113(1):39-53.

Vineis JH, Ringus DL, Morrison HG. Patient-Specific Bacteroides Genome Variants in Pouchitis. *mBio*. 2016;7(6):e01713-1716.

Wang Y, Wan X, Wu X, Zhang C, Liu J, Hou S. Eubacterium rectale contributes to colorectal cancer initiation via promoting colitis. *Gut Pathog*. 2021;13(1):2.

Wilkinson JE, Franzosa EA, Everett C. A framework for microbiome science in public health. *Nat Med*. 2021;27(5):766-774.

Wilson E, Tufts DW. Multilayer perceptron design algorithm. In: *Proceedings of IEEE Workshop on Neural Networks for Signal Processing*. ; 1994:61-68.

Wirbel J, Pyl PT, Kartal E. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat Med*. 2019;25(4):679-689.

Wiwanitkit S, Wiwanitkit V. Acute viral hemorrhage disease: A summary on new viruses. *J Acute Dis*. 2015;4(4):277-279.

Wu Y, Peng Y. Ten computational challenges in human virome studies. *Virol Sin*. 2024;39(6):845-850.

Wylie KM, Truty RM, Sharpton TJ. Novel Bacterial Taxa in the Human Microbiome. *PLOS ONE*. 2012;7(6):e35294.

Yachida S, Mizutani S, Shiroma H. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat Med*. 2019;25(6):968-976.

Yu J, Feng Q, Wong SH. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut*. 2017;66(1):70-78.

Zanella MC, Lenggenhager L, Schrenzel J, Cordey S, Kaiser L. High-throughput sequencing for the aetiologic identification of viral encephalitis, meningoencephalitis, and meningitis. A narrative review and clinical appraisal. *Clin Microbiol Infect*. 2019;25(4):422-430.

Zhang L, Chen F, Zeng Z. Advances in Metagenomics and Its Application in Environmental Microorganisms. *Front Microbiol*. 2021;12.

Zhu X, Li B, Lou P. The Relationship Between the Gut Microbiome and Neurodegenerative Diseases. *Neurosci Bull*. 2021;37(10):1510-1522.

Mehmet Hora Doktora Tezi İntihal Raporu

ORİJİNALLİK RAPORU

%5	%4	%2	%2
BENZERLİK ENDEKSİ	İNTERNET KAYNAKLARI	YAYINLAR	ÖĞRENCİ ÖDEVLERİ

BİRİNCİL KAYNAKLAR

1	Submitted to The Scientific & Technological Research Council of Turkey (TUBITAK) Öğrenci Ödevi	%1
2	Erentürk, Şaduman. "Birincil liflere atık kağıt lifi ve kuru sağlamlık maddesi ilavesinin kağıdın özelliklerine etkisi", Bartin University (Turkey), 2024 Yayın	%1
3	acikbilim.yok.gov.tr İnternet Kaynağı	%1
4	www.researchgate.net İnternet Kaynağı	<%1
5	adudspace.adu.edu.tr:8080 İnternet Kaynağı	<%1
6	tmc.dergisi.org İnternet Kaynağı	<%1
7	dergipark.org.tr İnternet Kaynağı	<%1
8	Submitted to Erciyes Ãniversitesi Öğrenci Ödevi	<%1
9	Uncu, Beste. "Veri Madenciliği ile Yeni Ilaç Hedeflerinin Saptanmasına Yönelik Interaktomikler: in Silico bir Senaryo", Dokuz Eylul Universitesi (Turkey), 2024 Yayın	<%1
10	Submitted to Abant İzzet Baysal Universitesi Öğrenci Ödevi	<%1

ÖZGEÇMİŞ VE ESERLER LİSTESİ

ÖZGEÇMİŞ

1. Kişisel Bilgiler

- **Adı ve Soyadı:** Mehmet HORA

2. Eğitim Bilgileri

- Erciyes Üniversitesi, Sağlık Bilimleri Enstitüsü, Biyoinformatik Sistemler Biyolojisi A.D. (2025)
- Erciyes Üniversitesi, Sağlık Bilimleri Enstitüsü, Tıbbi Mikrobiyoloji A.D. (2019)
- Erciyes Üniversitesi, Fen Fakültesi, Biyoloji Bölümü (2015)

3. Akademik Deneyimler ve Çalışmalar

3.1 Araştırma projeleri

- **Yürütücü:** Tübitak 1002-B Acil Destek Projesi, Proje no : 323S130, "İnsan Mikrobiyomunda Bulunan Bakteriyofaj Komünitesinin Konakçı Hastalıklarıyla Olan İlişkisinin Araştırılması" ,Proje Başlangıç/Bitiş Tarihleri: 01.02.2024 - 01.08.2024.
- **Bursiyer:** Tüseb B Grubu Ar-Ge Projesi, Proje no : 32774 no'lu "Çoklu İlaca Dirençli Gram Negatiflerin Etken Olduğu Üriner Sistem Enfeksiyonlarında Bakteriyofaj Tedavisinin Etkinliğinin Arastırılması", Proje Başlangıç/Bitiş Tarihleri: 14.06.2024 – Devam ediyor
- **Bursiyer:** Tübitak 1001 Projesi, Proje no : 123R030, "Akciğer Kanserinin Girişimsel Olmayan Tanısına Yönelik Oral Mikrobiyoma Dayalı Yapay Zeka Temelli Bir Tanı Yöntemi Geliştirilmesi", Proje Başlangıç/Bitiş Tarihleri: 15.11.2023 – Devam ediyor

- **Proje Personeli:** Tübitak 1507- Kobi Ar-Ge Projeleri 7201241, "Post-Prandiyal Glisemik Tepkileri Sağlıklı Seviyede Tutacak Mikrobiyom Güdümlü Kişiselleştirilmiş Beslenme Öneri Sisteminin Geliştirilmesi", Proje Başlangıç/Bitiş Tarihleri: 01.09.2020 - 28.02.2022.
- **Bursiyer:** Tübitak 1001 Projesi, Proje no : 120E092, "Mikrobiyom Analizinde Yeni Problem Tanımları Ve Çözümlerine Dair Yapay Öğrenme Temelli Yöntemlerin Araştırılması", Proje Başlangıç/Bitiş Tarihleri: 01.08.2020 - 01.08.2023.
- **Bursiyer:** Tübitak 1001 Projesi, Proje no :116S290, "Ortopedik Protez Enfeksiyon Mikrobiyotasının Metagenomik Yöntemlerle De Novo Incelenmesi", Proje Başlangıç/Bitiş Tarihleri: 01.07.2017 - 01.04.2020.
- **Bursiyer**: Tübitak 3001 Projesi, Proje no: 315S263, "Kolistin Dirençli Gram-Negatif Patojenlerde Direnç Kazanım Paternlerinin Tüm Genom Dizi Analizleriyle Ortaya Çıkarılması",Proje Başlangıç/Bitiş Tarihleri: 01.01.2017 - 01.07.2018.

3.2. Uluslararası hakemli dergilerde yayımlanan eserler (SCI ve SCI-Expended)

- Gundogdu A., Nalbantoglu O.U., Sarikaya I., Karis G., Erdogan M.N., **Hora M.,** Tunali, V. 2023. "Culture-independent assessment of probiotic supplement consistency in commercially available probiotic supplements". FOOD BIOSCIENCE, 53, 102709. **(Q1, IF:5.2)**
- Karakan T., Gundogdu A., Alagözlü H., Ekmen N., Ozgul S., Tunali V., **Hora M.,** Beyazgul D., Nalbantoglu O. U. 2022. "Artificial intelligence-based personalized diet: A pilot clinical study for irritable bowel syndrome." GUT MICROBES, 14(1), 2138672. **(Q1, IF:12.2)**
- Ciloglu F.U., **Hora M.,** Gundogdu A., Kahraman M., Tokmakci M., Aydin, O. 2022. "SERS-based sensor with a machine learning based effective feature extraction technique for fast detection of colistin-resistant *Klebsiella pneumoniae*." ANALYTICA CHIMICA ACTA, 1221, 340094.**(Q1, IF:6.2)**

- Sahin S., Gundogdu A., Nalbantoglu U., Kadioglu P., Karaca Z., Hacioglu A., Urhan M. E., Unluhizarci K., Demir A.N., **Hora M.,** Durcan E., Elbüken G., Dokmetas H.S., Zuhur S.S., Kelestimur, F. 2022. "Acromegaly is associated with a distinct oral and gut microbiota." PITUITARY, *25*(3), 520-530.**(Q1, IF:3.3)**
- Eken, A., Erdem, S., Haliloglu, Y., Zehra Okus, F., Cakir, M., Fatih Yetkin, M., Akcakoyunlu, M., Karayigit, M. O., Azizoglu, Z. B., Bicer, A., Gur, T. N., Aslan, K., **Hora, M**., Oukka, M., Altuntas, H. D., Ufuk Nalbantoglu, O., Gundogdu, A., Mirza, M., & Canatan, H. (2021). "Temporal overexpression of IL-22 and Reg3γ differentially impacts the severity of experimental autoimmune encephalomyelitis". IMMUNOLOGY, 164(1), 73-89. **(Q1, IF:7.3)**
- Tunali, V., Arslan, N. Ç., Ermiş, B. H., Derviş Hakim, G., Gündoğdu, A., **Hora, M.**, & Nalbantoğlu, Ö. U. (2024). "A Multicenter Randomized Controlled Trial of Microbiome-Based Artificial Intelligence-Assisted Personalized Diet vs Low-Fermentable Oligosaccharides, Disaccharides, Monosaccharides, and Polyols Diet: A Novel Approach for the Management of Irritable Bowel Syndrome". THE AMERİCAN JOURNAL OF GASTROENTEROLOGY, 119(9), 1901-1912 **(Q1, IF:8.0)**
- Gundogdu, A., Nalbantoglu, O. U., Karis, G., Sarikaya, I., Erdogan, M. N., **Hora, M.**, & Aslan, H. (2024). "Comparing microbial communities in mucilage and seawater samples: Metagenomic insights into mucilage formation in the Marmara Sea." ENVIRONMENTAL SCIENCE AND POLLUTION RESEARCH, 31(48), 58363-58374. **(Q1, IF:13.3)**
- Cetiner, O. F., **Hora, M.**, Kafadar, I. H., Kilic, A. U., & Gundogdu, A. (2024). "16S rRNA sequencing reveals microbiota differences in orthopedic implants between aseptic loosening and prosthetic joint infection cases." ACTA MICROBIOLOGICA ET IMMUNOLOGICA HUNGARICA **(Q4, IF:1.3)**

- Saglam, M. F., Gündogdu, A., **Hora, M.**, Kandemir, H., & Sengul, I. F. (2021). "Synthesis of pyrrolo[3,2-c]carbazole-2-carbohydrazides and pyrrolo[3,2-c]carbazol-2-yl-1,3,4-oxadiazoles and their in vitro antibacterial evaluation." SYNTHETIC COMMUNICATIONS. **(Q3, IF:1.9)**
- Gizem Karış, Herdem Aslan, **Mehmet Hora**, Aycan Gundogdu, Uncovering Antibiotic Resistance Genes in Marmara Sea Mucilage via Shotgun Metagenome Sequencing, JOURNAL OF GLOBAL ANTIMICROBIAL RESISTANCE. **(Q2, IF:3.7)**
- **Mehmet Hora**, Ozkan Ufuk Nalbantoglu, Aycan Gundogdu, Mehmet Hora, "Unveiling the Origins of Antimicrobial Resistance Genes in Wastewater: A Shotgun Metagenomic Study." JOURNAL OF GLOBAL ANTIMICROBIAL RESISTANCE. **(Q2, IF:3.7)**

3.3. Uluslararası hakemli dergilerde yayımlanan makaleler (diğer indeksler)

- Astley D.J., Spang L., Parnian F., Vollmerhausen T., Kilic H., **Hora M.,** Gundogdu A., Katouli, M. 2022. "A comparative study of the clonal diversity and virulence characteristics of uropathogenic Escherichia coli isolated from Australian and Turkish (Turkey) children and adults with urinary tract infections." GERMS,*12*(2), 214.**(Q3, IF:1.64)**

3.4. Bilimsel toplantılarda sunulan bildiriler

3.4.1.Uluslararası

- Gündoğdu A., Nalbantoglu O.U., Ozkan E., **Hora M.,** Onuk E. 2022. "Koroner arter hastalığının tanısında sıvı biyopsi amaçlı kan mikrobiyotası biyobelirteçlerinin keşfi." Uluslararası XL. TMC Kongresi, ANTALYA, TÜRKİYE, 16-20 Kasım. (Sözlü Sunum)
- Üçkayabaşı A., Nağıyev T., Gündoğdu A., **Hora M.,** Kuleci S., Köksal F. 2022. "Akciğer kanserli hastalarda yeni nesil dizileme ile akciğer mikrobiyom analizi" Uluslararası XL. TMC Kongresi, ANTALYA, TÜRKİYE, 16-20 Kasım. (Sözlü Sunum)

- Gündoğdu A., Aslan H., Nalbantoglu O.U., Karis G., Bali A., Sarikaya I., Erdogan M.N., **Hora M.** 2022. "Marmara Denizi'nin farklı su katmanlarından alınan su örnekleri için metagenom temelli bir profilleme çalışması." Uluslararası XL. TMC Kongresi, ANTALYA, TÜRKİYE, 16-20 Kasım. (Sözlü Sunum)
- Uysal F., **Hora M.,** Gündoğdu A., Tokçakçı M., Aydın Ö. 2021. "Detection of colistin resistant Klebsiella pneumoniae bacteria using surfaceenhanced Raman spectroscopy combined with machine learning". BIO Türkiye Organizasyonu Uluslararası Biyoteknoloji Kongresi, İSTANBUL, TÜRKİYE, 9 - 11 Eylül. (Sözlü Sunum).
- Nalbantoğlu Ö.U., Kiliç H., Ulu Kiliç A., **Hora M.,** Gündoğdu A. 2018. "Whole-genome sequencing reveals recently acquired virulence islands in *Ralstonia mannitolilytica* isolated from high-risk patients", 28th European Congress of Clinical Microbiology and Infectious Diseases, MADRID, ISPANYA, 21-21 Nisan 2018.
- **Hora M.,** Yalçın K., Gündoğdu A. 2018. "Turkish Microbes: Previously non-catalogued genes were detected in the gut metagenomes of Turkish individuals", 2nd International Human Microbiota: In health and diseases congress, ANTALYA, TÜRKİYE, 18-22 April. (Sözlü Sunum)

3.4.2. Ulusal

- **Hora M**., Yalçın K., Saka S., Nalbantoğlu Ö.U., Gündoğdu A. 2018. "Parkinson Hastalığında Bağırsak Mikrobiyomu Değişikliklerinin *in silico* incelenmesi", 3. Yaşam Bilimleri Kongresi, KAYSERİ, TÜRKİYE, 23-24 Şubat.
- Saka S., **Hora M.,** Yalçın K., Gündoğdu A., Nalbantoğlu Ö.U. 2018. "Nanogözenek DNA dizilemede kullanılan baz çağırma yazılımlarının performans analizi", 3. Yaşam Bilimleri Kongresi, KAYSERİ, TÜRKIYE, 23-24 Şubat.

4. İş Deneyimleri

- **Öğrenci Personel:** Erciyes Üniversitesi Gülser-Mustafa Gündoğdu Merkez Laboratuvarı, Bakteriyoloji Laboratuvarı (2012-2019)
- **Biyoinformatik Uzmanı:** Enbiosis Biyoteknoloji (2019 – Devam ediyor.)

5. Ödüller ve Burslar

- **En iyi bilimsel çalışma ödülü-AGU 3. Yaşam Bilimleri Kongresi- 2018.** *"Parkinson Hastalığında Bağırsak Mikrobiyomu Değişikliklerinin in silico incelenmesi"*

- **Uluslararası XXXVIII. Türk Mikrobiyoloji Cemiyeti Kongresi, Kongre Katılım ve Konaklama Bursu- 2018**
- **Uluslararası XL. Türk Mikrobiyoloji Cemiyeti Kongresi, Kongre Katılım ve Konaklama Bursu-2022**
- **33rd International Congress of Antimicrobial Chemotherapy (ICC), Young Investıgator Travel Awardee-2024**

6. Yabancı Dil Bilgisi

- **İngilizce** (72,50/YÖKDİL,2019)
- **İletişim düzeyi:** İngilizce, C1 İleri

Fransızca, A1 Başlangıç

7. Teknik ve Kişisel Beceriler

- **Laboratuvar Becerileri**
    - Genel Moleküler Biyoloji Teknikleri
        - Nükleik Asit İzolasyonu (DNA/RNA)
        - PCR teknikleri
        - Agaroz Jel Elektroforezi
        - Klonlama Teknikleri
        - Protein Analiz Teknikleri
        - Sekanslama Sistemleri (Sanger sekanslama,Yeni Nesil Sekanslama)

- Genel Mikrobiyoloji Teknikleri
  - Kültürleme ve İzolasyon
  - Mikrobiyal İdentifikasyon
  - Antimikrobiyal Duyarlılık Testleri
  - Rutin Mikrobiyoloji Laboratuvar İşlemleri
- Biyoinformatik
  - Metagenom analizleri
  - Genom Boyu İlişkilendirme Çalışmaları (GWAS)
  - Mikrobiyom analizleri
  - Virom analizleri
  - Makine Öğrenimi, Derin Öğrenme ve Yapay Zeka modelleri geliştirme ve uygulama
  - İstatistik, Biyoistatistik

- **Programlama Dilleri**
  - Python
  - Pythorch
  - Tensor Flow
  - R