



COVID-19 ENFEKSİYONLARINDA DİZİ ANALİZİ YÖNTEMLERİNE GENEL BAKIŞ

OVERVIEW OF SEQUENCE ANALYSIS METHODS IN COVID-19 INFECTIONS

Ferhat Gürkan Arslan¹, Elmas Pınar Kahraman Kılbaş^{*2}, Mustafa Altındış³

¹Fatih Sultan Mehmet Eğitim ve Araştırma Hastanesi Tıbbi Mikrobiyoloji Laboratuvarı, İstanbul, Türkiye, ²Fenerbahçe Üniversitesi, Sağlık Hizmetleri Meslek Yüksekokulu, Tıbbi Laboratuvar Teknikleri Programı, İstanbul, Türkiye, ³Sakarya Üniversitesi Tıp Fakültesi, Tıbbi Mikrobiyoloji AD, Tıbbi Viroloji BD, Sakarya, Türkiye

ORCID iD: Ferhat Gürkan Arslan: 0000-0001-8394-1962; Elmas Pınar Kahraman Kılbaş: 0000-0003-1348-625x; Mustafa Altındış: 0000-0003-0411-9669

***Sorumlu Yazar / Corresponding Author:** Elmas Pınar Kahraman Kılbaş **e-posta / e-mail:** elmspnrkk@gmail.com

Geliş Tarihi / Received:02.01.2022

Kabul Tarihi / Accepted:14.02.2022

Yayın Tarihi / Published:21.03.2022

Öz

Şiddetli akut solunum sendromu koronavirüs 2 (SARS-CoV-2), koronavirüs hastalığı 2019'un (COVID-19) etkeni olarak tanımlandı ve genomik veriler ilk olarak 10 Ocak 2020'de Çin tarafından paylaşıldı. O tarihten itibaren, dünya genelinde toplanan örneklerden viral genomu dizilemek için çok büyük çaba harcandı. Yakın geçmişte, kökenleri izlemek ve bulaşıcı ajanların evrimini anlamak, salgınların yayılma zincirlerini araştırmak, hem etkili ve hızlı moleküler tanı testlerinin geliştirilmesini kolaylaştırmak hem de tedavi ve aşılarda araştırılmasına katkıda bulunmak için, yeni nesil dizileme (NGS) stratejileri, başarıyla kullanılmıştır. Teknoloji ve bilimdeki son gelişmeler, COVID-19'un etkeni olan ağır akut solunum sendromu koronavirüs-2'nin (SARS-CoV-2) genomlarının, bir vakanın tanımlanmasından sonraki saatler veya günler içinde dizilenmesine olanak sağlamıştır. Bu sayede, ilk kez, bir pandeminin halk sağlığı ve epidemiyoloji boyutu gerçek zamanlı olarak izlenebilmektedir. SARS-CoV-2 genom dizilerinin erken paylaşımı, moleküler tanı testlerinin hızla geliştirilmesine olanak sağlayarak, küresel hazırlığa ve karşı önlemlerin tasarımına katkıda bulunmuştur. Hızlı, büyük ölçekli virüs genom dizilimi, viral salgınların dinamiklerini anlama ve kontrol önlemlerinin etkinliğini değerlendirmede oldukça önemlidir. SARS-CoV-2 gen dizilimi, gelişmiş tanımlar, karşı önlemlerin geliştirilmesi ve hastalık epidemiyolojisinin araştırılması dahil olmak üzere birçok farklı alanda kullanılabilir. COVID-19'un etiyolojik ajanının genomik dizisini tam olarak tanımlamak için etkili ve hızlı dizileme yöntemlerinin geliştirilmesi, tanısal moleküler testlerin tasarımı ve pandemi yayılımını azaltmada etkili önlemlerin alınması ve stratejilerin belirlenmesinde temel olmuştur. Mevcut dizilerin sayısından anlaşıldığı gibi, SARS-CoV-2 genomlarına, farklı yaklaşımlar ve dizileme yöntemleri uygulanabilir. Bununla birlikte, her teknoloji ve dizileme yaklaşımının kendi avantajları ve sınırlamaları vardır. Bu derlemede, SARS-CoV-2 genomlarının dizilenmesi için şu andaki mevcut platformlar ve metodolojik yaklaşımlardan bahsedilecektir.

Anahtar Kelimeler: COVID-19, pandemi, SARS-CoV-2, sekans.

Abstract

Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) was identified as the agent of COVID-19, and genomic data was first shared by China on January 10, 2020. Since then, tremendous effort has been devoted to sequencing the viral genome from samples collected around the world. In the recent past, next-generation sequencing (NGS) strategies have been used successfully to trace the origins, understand the evolution of infectious agents, to investigate the chains of the spread of epidemics, to facilitate the development of effective and rapid molecular diagnostic tests, and to contribute to the research of treatments and vaccines. Recent advances in technology and science have allowed the genomes of SARS-CoV-2, the agent of COVID-19, to be sequenced within hours or days after a case is identified. In this way, for the first time, the public health and epidemic size of a pandemic can be monitored in real-time. The early sharing of SARS-CoV-2 genome sequences has allowed the rapid development of molecular diagnostic tests, contributing to global preparedness and the design of countermeasures. Rapid, large-scale sequencing of the virus genome is essential in understanding the dynamics of viral outbreaks and assessing the effectiveness of control measures. SARS-CoV-2 gene sequencing can be used in many different areas, including improved diagnosis, development of countermeasures, and investigation of disease epidemiology. The development of effective and rapid sequencing methods to fully identify the genomic sequence of the etiologic agent of COVID-19 has been fundamental to the design of diagnostic molecular tests and the determination of effective measures and strategies to reduce the spread of the pandemic. Different approaches and sequencing methods can be applied to SARS-CoV-2 genomes, as evidenced by the number of sequences available. However, each technology and sequencing approach has its advantages and limitations. In this review, current platforms and methodological approaches for sequencing SARS-CoV-2 genomes will be discussed.

Keywords: COVID-19, pandemic, SARS-CoV-2, sequence.

Giriş

Ocak 2020'de, daha önce bilinmeyen bir koronavirüsü, insanlarda solunum yolu enfeksiyonu ve ölüm nedeni olarak tanımlandı. Yeni beta koronavirüsün ilk tam genomik dizileri, PCR ve Sanger dizilimi ile desteklenen meta-transkriptomik yaklaşımlarla Aralık 2019'un sonlarında elde edildi. Bir referans genomun varlığı, gerçek zamanlı PCR'ye dayalı tanısal testlerin geliştirilmesini kolaylaştırdı. Şiddetli akut solunum sendromu ile ilişkili koronavirüs virüs türlerine ait olan bu virüs, daha sonra SARS-CoV-2 (Severe Acute Respiratory Syndrome-Related Coronavirus-2) ve neden olduğu hastalık ise koronavirüs hastalığı 2019 (COVID-19) olarak adlandırıldı.¹ COVID-19 (Corona Virus Disease 2019), hızla yayıldı ve 11 Mart 2020'de 110 ülkeden 118.000 vaka ile Dünya Sağlık Örgütü (DSÖ) tarafından pandemi olarak ilan edildi. Devam etmekte olan ve modern tarihin en büyük küresel tehditlerinden birini oluşturan COVID-19 pandemisi, ciddi sosyal ve ekonomik maliyetlere neden oldu.²

Kısmi belli genlerin veya tüm genomların dizilenmesinin, viral patojen genomlarını araştırmada, salgın bulaş dinamiklerini ve yayılma olaylarını anlamada, patojenite ve/veya karşı önlemler (örn. teşhis, antiviral ilaçlar ve aşılar) üzerinde etkili olabilecek mutasyonların izlenmesinde güçlü bir yöntem olduğu bilinmektedir. Sonuçlar, halk sağlığında salgın kontrol kararlarını yönlendirmek için çok önemlidir.¹ Dolayısıyla, COVID-19'un etiyolojik ajanı olan SARS-CoV-2'nin genomik dizisini tam olarak tanımlamak için etkili ve hızlı dizileme yöntemlerinin geliştirilmesi, tanısal moleküler testlerin tasarımı ve pandemi yayılımını azaltmada etkili önlemler ve stratejiler planlamak için temeldir.²

COVID-19 pandemisinde, etkili gerçek zamanlı sürveyans stratejilerinin geliştirilmesine yönelik olarak, SARS-CoV-2 genomunun tam veya tama yakın kısmı ile onun hastalıktan sorumlu kısmının dizilenmesine dayanan çalışmalar yapılmakta ve özel veri havuzlarında depolanmaktadır. Bu veriler, virüsün evrimsel dinamikleri ve klinik olarak anlamlı olabilecek varyantlarının tanımlanmasında birçok çalışmayı teşvik etmektedir. Özellikle, tutarlı ve amaca uygun meta verilere erişimi kolaylaştırmak için sekans verilerinin sürekli olarak işlenmesi, etiketlenmesi ve biriktirilmesi kritik bir ihtiyaçtır.²

SARS-CoV-2 sekanslarını tanımlamak, adlandırmak ve rapor etmek için standartlaştırılmış bir veri hattı henüz kurulmamıştır ancak, DSÖ ve Avrupadaki birçok ülke, pandeminin başlangıcından beri SARS-CoV-2 varyantları dizilenmektedir ve sekansları GISAID (Global Initiative on Sharing All Influenza Data) veya diğer açık veritabanlarında yayınlamaktadır. Virüs özelliklerine ilişkin bilgilerin klinik ve epidemiyolojik verilerle birleştirilmesi önemlidir. SARS-CoV-2 aşılı ve antiviraller kullanılmaya başlandıkça, dolaşımdaki varyantların aşı ile eşleşmesini ve olası antiviral direncin ortaya çıkışını izlemek için sekans verileri giderek daha önemli hale gelecektir.¹

SARS-CoV-2 Dizilenmesi İçin Hedefler

COVID-19 salgını, tüm genom dizileme yetkisinin en başından beri halk sağlığı biriminde bulunduğu ilk salgındır. SARS-CoV-2 dizilemesinin mevcut hedefleri iki başlık altında toplanabilir (Çizelge 1). Herhangi bir patojenin dizilemeye dayalı sürveyansının önemli bir bileşeni, verilerin, dizilerin genetik ilişkisine dayalı olarak anlamlı şekilde sınıflandırılmasıdır. Bu, moleküler epidemiyoloji alanındaki

farklı aktörler arasındaki iletişimi kolaylaştırır ve standart epidemiyolojik analiz ile entegrasyon için genomik verilerin basitleştirilmiş tablo haline getirilmesini sağlar. SARS-CoV-2 için birkaç sınıflandırma uygulanmıştır¹:

- Virüs için hCoV-19 terimini kullanan GISAID terminolojisi (www.gisaid.org).³

- Nextstrain isimlendirme (<https://nextstrain.org/ncov>).⁴

- Andrew Rambaut ve diğerleri tarafından soy isimlendirmesi. (<https://cov-lineages.org>).⁵

Nextstrain ve GISAID sınıflandırmaları, küresel olarak dolaşan çeşitliliğin geniş kapsamlı bir kategorizasyonunu sağlamayı hedeflerken, soy isimlendirmesi belirli coğrafi bölgelerdeki salgınlara karşılık gelmektedir.¹

SARS-CoV-2 İçin Dizileme Uygulaması

Uygun Meta Veriler

SARS-CoV-2 genomik verilerinin faydalı olması için bunlara, uygun meta veriler eşlik etmelidir. Meta verileri düzenlemek ve bunları yerel veya herkese açık olarak paylaşmak zaman alabilir, ancak her ikisi de herhangi bir dizileme hattının ayrılmaz bir parçasıdır.⁶

Meta veriler, numune toplama tarihini ve yerini mutlaka içermelidir. Bununla birlikte, ek meta verilerin yayınlanması, bir genomik dizinin potansiyel kullanımını büyük ölçüde artırır. Bu nedenle, mümkün olduğunda, numune tipi ve dizilemenin laboratuvarında nasıl elde edildiğine ilişkin bilgiler dahil edilmelidir. Aynı bireyden çift örnekler veya aynı örnekten çift dizileme açıkça tanımlanmalıdır. Yaş, cinsiyet, eşlik eden hastalıkların varlığı, hastalık şiddeti ve sonucu gibi demografik ve klinik bilgiler ve veri tabanındaki diğer dizilere bağlantılar, bu tür bilgilerin hastanın kimliğini belirleme riskini taşımadığı durumlarda paylaşılması önerilir.⁶

Meta veriler (tarih gibi) için belirli formatlardaki küresel fikir birliği, birçok farklı laboratuvarından alınan genomik dizi verilerinin hızla daha büyük veri kümelerinde derlenmesini ve belirsizliği azaltmasını sağlar. GISAID gibi bazı konsensüs genom depoları, belirli alanlara zaten format kısıtlamaları getirmektedir. Veri havuzları belirli bir formata sahip değilse, WHO klavuzunda belirtilen format kısıtlamaları kullanılabilir.⁶

Hasta gizliliğini riske atmadan tüm Meta verileri paylaşmanın mümkün olmadığı durumlarda, tercihen küresel çalışmalar için en yararlı olan veriler paylaşılmalıdır. Örneğin, örnekleme yeri, tarihi ve seyahat geçmişi filodinamik çalışmalar için hastanın yaşı veya cinsiyetinden daha faydalıdır.⁶

Bazı laboratuvarlar, hastaların tespit edilme şansını azaltmak için verilen tarihleri doğru olarak açıklamazlar. Bu, örneğin numune alma tarihinin her iki tarafında 5 gün içinde yanlış bir tarih seçerek veya numune tarihi olarak dizileme tarihini kullanarak yapılabilir. Bu tür uygulamalar moleküler saate dayalı filogenetik çıkarımı olumsuz etkiler ve ideal olarak bundan kaçınılmalıdır. Buna rağmen bu uygulamaya uyulursa, yeni tarihin tam olarak nasıl seçildiğine dair bilgi not olarak verilmelidir.⁶

Biyogüvenlik ve Biyo-Emniyet

Biyogüvenlik ve biyo-emniyeti değerlendirmek için her zaman risk değerlendirmeleri yapılmalıdır. Bu tür risk değerlendirmelerinin sonuçları, ilgili süreçlerde yer alan çalışanlara iletilmelidir.⁶

Bireysel laboratuvarlar, SARS-CoV-2 protokollerindeki her adım için her zaman yerel risk değerlendirmeleri yapmalıdır.

SARS-CoV-2 materyalinin güvenli bir şekilde taşınmasını sağlamak için uluslararası, ulusal ve yerel mevzuata danışılmalıdır. Bununla birlikte, Dünya Sağlık Örgütü (DSÖ) geniş biyogüvenlik kılavuzları yayınlamıştır. Örnekler, RNA kalitesini koruyan kimyasal yöntemler kullanılarak mümkün olan en erken aşamada (genellikle RNA ekstraksiyonundan önce) inaktive edilmelidir. Tanısal NAAT'lerden önce RNA'yı ekstrakte etmek için kullanılan yöntemler genellikle dizileme için uygundur. Çoğu NAAT'de olduğu gibi, RNA bütünlüğüne zarar verme riski taşıdığı için numune ekstraksiyonundan önce ısıyla inaktivasyon önerilmez.⁶

Etik Hususlar

Hastaların numunelerin toplanması ve dizilenmesi için uygun onay vermesini sağlamak ve verilerin daha sonraki kullanımını, saklanması ve yayınlanmasını değerlendirmek için etik incelemeler yapılmalıdır.⁶ Metagenomik gibi bazı dizileme yaklaşımları, insan genomik verileri üretecektir. İnsan genetik verilerini işlemek için etik onay ve açık hasta onayı alınmadıkça, herhangi bir insan genomik dizisi, mümkün olan en erken aşamada viral veri setinden çıkarılmalıdır. Kişisel veya insan verilerinin saklanması gerekiyorsa, bu tür tüm dosyaların uygun şekilde şifrelenmesi oldukça önemlidir.⁶ Etik incelemeler, hasta gizliliğini riske atmadan paylaşılacak maksimum olası ilgili Meta verileri belirlemelidir.⁶

Örnekleme Stratejisi ve Örnek Seçimi

Yeterli, yüksek kaliteli SARS-CoV-2 RNA'sının edinilmesi, dizileme verimini ve genom dizisi verilerinin nihai kalitesini en üst düzeye çıkarmaya yardımcı olur. Bir RNA örneğinin miktarı ve kalitesini etkileyen durumlar; klinik örnek seçimi, klinik numunenin işlenmesi, viral RNA izolasyon yöntemi ve personelin teknik yeterliliğidir.⁶

Mevcut SARS-CoV-2 dizi verileri, esas olarak, viral genomların tam veya tama yakın dizilenmesi ve yeniden yapılandırılması için yeterli RNA'nın ekstraksiyonuna izin veren yüksek viral yüklerle sahip klinik tanı örneklerinden elde edilir.⁷ NAAT teşhisi için toplanan örnekler tipik olarak dizileme için de uygundur.⁶ Bazı kaynaklar, alt solunum yolundan alınan örneklerin, üst solunum yolundan alınanlardan daha yüksek viral yük içerebileceğini bildirmektedir.⁷ Bununla birlikte, enfeksiyonun seyri sırasında viral yük, solunum ve solunum dışı dokular arasında olduğu gibi farklı solunum bölgeleri arasında da aktif olarak değişir.² Nazal sürüntülerin, boğaz sürüntülerinin ve tükürüğün semptom başlangıcından kısa bir süre sonra ve 25 gün sonrasına kadar yüksek viral yüklerle sahip olduğu bulunmuştur. Numunelerdeki SARS-CoV-2 viral yükü ve viral RNA bolluğu, normalde hastalık başlangıcından sonraki ilk haftada en yüksektir.⁶ Viral genetik materyal, enfekte hücre dizilerinin süpernatantından da izole edilebilir ancak, hücre dizilerinde çoğaltılan viral popülasyonlar laboratuvar pasajı sırasında sıklıkla yeni genetik varyantlar biriktirir.² Dizileme için izolatların, bir moleküler tanı laboratuvarı tarafından önceden işlenmiş pozitif numunelerden seçilmesi,

numune işleme ve nükleik asit ekstraksiyonundaki işlerin tekrarlanmasını önler. Böylece, kaynaklar ve maliyetten açısından tasarruf sağlayabilir. Bazı ticari moleküler tanı kitleri, girdi olarak virallizatları kullanır ve ekstrakte edilmiş RNA'nın depolanmasına izin vermez. Ticari lizis tamponunun bileşenlerinin açıklanmadığı bu gibi durumlarda, hazırlanmış lizatları diğer ticari ekstraksiyon kitleriyle yeniden kullanmak son derece zor olabilir ve doğrudan orijinal klinik numuneden taze inaktivasyon ve ekstraksiyon yapmak gerekli olabilir.⁶

Viral RNA'nın korunması, yüksek kaliteli dizi verilerinin üretimi için önemlidir. Bu durum, numune toplama ve analizi arasındaki sürede soğuk zincir sağlanarak, RNA veya numunelerin dondurulup çözülme sayısı azaltılarak ve numune toplanması ile dizilenmesi arasındaki süreyi en aza indirerek başarılabılır. Birkaç günden daha uzun süre 4 °C'de saklanan veya nakledilen RNA'nın, başlangıçta bir RNA stabilizasyon solüsyonunda korunmadığı sürece, dizileme için yeterince yüksek kalitede olması olası değildir. RNA -20 °C'de veya tercihen -80 °C'de saklanabiliyorsa kalite önemli ölçüde daha yüksek olacaktır. Virallizatlar, tipik olarak, ekstrakte edilmiş RNA kadar uzun süre 4 °C'de saklanamaz. Birçok dizileme protokolü, RNA'nın cDNA'ya ters transkripsiyonu veya ikinci zincir sentezi/çift sarmallı DNA PCR amplikonlarının oluşturulması dahil olmak üzere bir numunenin depolama kapasitesini artıran adımları içerir. PCR amplikonları, dizileme kalitesinde azalma olmaksızın 4°C'de aylarca saklanabilir. Bu nedenle, bazı bağlamlarda, materyalin kitaplık hazırlığından önce daha az sıcaklık kısıtlaması ile saklanabilmesi veya nakledilebilmesi için tanısal PCR'lerden sonra bu adımları hızlı bir şekilde gerçekleştirmek uygun olabilir.⁶

Örnek seçimi, amaca ve mevcut kaynaklara uygun olarak yapılmalıdır. Sürveyans yapılacağına, dizileme için temsili virüs suşları, farklı zamanlarda ve farklı coğrafi bölgelerden, buna ilave olarak, farklı demografik özelliklere ve hastalık şiddet spektrumuna sahip hastalardan seçilmelidir. Hedefe yönelik izlemde (örn. salgınlar), virüs evrimini ve virüs genomundaki değişiklikleri daha etkin bir şekilde izlemek için, örnekleme stratejileri, özellikle diğer epidemiyolojik faktörlerle açıklanmadığında, SARS-CoV-2 aşı kaçış varyantlarının, re-enfeksiyona neden olan virüslerin, hayvan popülasyonlarında ortaya çıkan varyantların veya daha bulaşıcı varyantların seçimini içermelidir. Kaynakların kısıtlı olduğu bölgelerde olay/risk tabanlı bir yaklaşım, dizileme ihtiyacını ve maliyetini azaltabilir.¹

Virüs dağılımının belirlenmesi için, toplumda ve hastane ortamlarında SARS-CoV-2 virüslerinin sürveyansına devam edilmesinin yanı sıra, toplumdaki varyant virüslerin prevalansının belirlenmesi ve varyant virüslerin erken saptanması amacıyla COVID-19 vakalarının rastgele örnekleminin kullanılması, kritik öneme sahiptir. Toplum temsil eden örnekleme ek olarak, endişe verici varyantların saptanması için, hedeflenmiş örneklem uygulanmalıdır.¹

Bu tür hedeflenen örneklemede, DSÖ'nün halk sağlığı hedefleri için SARS-CoV-2 genomik dizilimine ilişkin geçici kılavuzu, dizileme için Çizelge 2'deki grupların önceliklendirilmesini önermektedir.

Çizelge 1. SARS-CoV-2 dizilemesi için hedefler¹

Dizilemede Ana Hedefler	Dizilemede Spesifik Hedefler
Ortaya çıkan virüs türlerinin, özellikle endişe verici olup olmadıklarını belirlemek amacıyla,erken tespiti ve tanımlanması	Virüs bulaşma dinamiklerinin araştırılması ve yeni genetik varyantların tanımlanması
	Aşı kaçığı riskini değerlendirmek için virüsün antijenik özelliklerinin modellenmesi
	Aşı bileşimi için virüslerin seçilmesi
	Mutasyonların moleküler tanı, antijen özellikleri ve serolojik yöntemlerin performansı üzerindeki etkisinin değerlendirilmesi
	Sınıflar/soylar ile bulaşıcılık ve hastalık şiddeti veya risk grupları gibi epidemiyolojik veriler arasındaki ilişkinin araştırılması
	Alınan önlemlerin virüs popülasyonu üzerindeki etkisinin anlaşılması
Genetik ve antijenik varyant virüslerin, pandemi üzerindeki etkisinin değerlendirilmesi ve halk sağlığı çalışmasına rehberlik etmek için zaman içinde izlenmesi	Epidemiyolojik kümeler içindeki viralsuşların ilişkisinin değerlendirilmesi ve temas takibinin ve diğer halk sağlığı girişimlerinin desteklenmesi
	Re-enfeksiyonların değerlendirilmesi ve onaylanması
	İnsan sağlığını etkileyebilecek yabani/evcil /çiftlik hayvan popülasyonları içinde ortaya çıkan soyların izlenmesi
	Hastalığın patogeneğinde gözlemlenen mutasyonların ilişkisinin doğrulanması için daha fazla temel araştırma çalışmasının teşvik edilmesi (örneğin enfeksiyon, reseptör bağlanması)
	Mutasyonların antiviral ilaçların performansı üzerindeki etkisinin değerlendirilmesi
	Canlı SARS-CoV-2 aşılı kullanılabılır hale geldiğinde, aşı kaynaklı virüs enfeksiyonları ve bulaşların potansiyel insidansının değerlendirilmesi

Çizelge 2. Dizileme için önceliklendirilmesi önerilen grup¹

SARS-CoV-2 için aşılınmış ve aşıya karşı uygun bir immün yanıt göstermesine rağmen daha sonra SARS-CoV-2 ile enfekte olan bireyler
SARS-CoV-2 enfeksiyonuna duyarlı çok sayıda hayvanla yakın insan-hayvan etkileşimi veya özellikle SARS-CoV-2'ye karşı antikor tedavisi alırken viralsaçılımın uzun süredir devam ettiği bağışıklığı baskılanmış hastalar gibi riskli durumlar
SARS-CoV-2 bulaşıcılığı ve/veya virülansında beklenmedik bir artış veya değişiklik olması
Tanı yöntemlerinin (antikor, antijen, moleküler testler) veya tedavilerin performansında bir değişiklik şüphesi varlığı
Dizilemenin, bulaşma olaylarının anlaşılmasını destekleyebildiği ve/veya enfeksiyon kontrol prosedürlerinin etkinliğini değerlendirebildiği küme araştırmaları sırasında

Yukarıdaki öncelik listesine ek olarak, varyant virüslerin ortaya çıkması için seyahatle ilgili virüs tespitleri izlenmelidir.¹

Rastgele ve Temsili Örneklem için Örneklem Büyüklüğü

Dizilemesi yapılacak örneklerin sayısı, saptama ve izleme olanaklarının düzeyine ve hangi kaynakların mevcut

olduğuna bağlı olmalıdır. Çizelge 3, istenen olanaklara bağlı olarak, temsili veya rastgele örnekleme stratejisi kullanılarak sekanslanması gereken numune sayısını açıklamaktadır. Tüm virüs popülasyonu içindeki belirli varyantları, hedeflenen farklı düşük varyant sıklıklarında, farklı hassasiyet seviyeleri ile saptayabilmek ve miktarını belirleyebilmek için gerekli dizileme sayısını göstermektedir.

Çizelge 3. Toplam 50.000 vaka için zaman birimi ve coğrafi birim başına gerekli dizi sayısı¹

İstenen ayırım gücünün zaman ve coğrafi birim başına gereken sekans sayısı				
Dolaşan tüm virüsler arasında varyantın beklenen prevalansı	Varyant virüslerin varlığını sadece %95 güvenle saptama (kesinlik yok)	Varyant oranının belirlenmesi		
		Düşük hassasiyet (%95 bağıl GA ± %50)	Orta hassasiyet (%95 bağıl GA ± %25)	Yüksek hassasiyet (%95 bağıl GA ± %10)
%25.00	12	46	183	1.114
%10.00	30	138	544	3.133
%5.00	60	289	1.128	5.988
%2.50	119	589	2.236	10.336
%1.00	297	1.455	5.146	17.764
%0.50	588	2.801	8.948	23.212
%0.25	1.156	5.179	14.129	27.379
%0.10	2.748	10.511	21.605	30.669

Örnek 1: Haftalık COVID-19 vaka sayısı yaklaşık 50.000 olan bir ülkede, dolaşımdaki varyantların düzeyi haftalık olarak izlenmek isteniyor. Çizelge 1, 50.000 vaka için, orta düzeyde duyarlılık sütununda ve %1 varyant oranı satırında, istenen performansa ulaşmak için haftalık olarak 5.146 örneğin dizilenmesi gerektiğini göstermektedir.

Örnek 2: Haftalık yaklaşık 12.500 vakası olan bir ülkede, toplumda aylık bazda %0,1'in üzerinde yeni varyantların girişini, herhangi bir varyant düzeyi değerlendirmesi yapmadan tespit etmek istiyor. Bu haftalık 12.500 vaka, değerlendirme aylık veriler üzerinden gerçekleştirileceğinden 50.000 vaka tablosuna karşılık gelmektedir. Çizelge 1, 50.000 vaka için, saptama sütununda ve %0,1 varyant oranı satırında, bu durumda aylık olarak 2.748 örneğin sekanslanacağını göstermektedir.

RNA Ekstraksiyonu

SARS-CoV-2 RNA ekstraksiyonu, biyogüvenlik düzeyi (BGD) 2 laboratuvarları gerektirir. RNA, tüm RNA ekstraksiyonu veya viral RNA'nın zenginleştirilmesi için ticari olarak mevcut çok çeşitli kitlerden herhangi biri kullanılarak klinik örneklerden, kültürlenmiş izolatlardan veya çevresel örneklerden ekstrakte edilebilir ve saflaştırılabilir. Standart metodolojiler, nükleazlarıninhibe eden, viral RNA'nın bozulmamasını sağlayan Guanidin tuzunun ve proteini denatüre etmek ve eritmek için fenolün, virüsü etkin bir şekilde etkisiz hale getirmesini içerir. Viral RNA ekstraksiyon protokolleri, RNA geri kazanımını artırmak için genellikle poli-A RNA gibi taşıyıcı RNA'nın eklenmesini önerir. Taşıyıcı RNA'nın varlığı, amplikonveyaHybridcapture'a dayalı SARS-CoV-2 genom dizileme yöntemlerini etkilemese de özellikle metatranskriptomik yöntemleri olumsuz etkileyebilir. Bu nedenle taşıyıcı RNA'nın kullanımı dikkatle değerlendirilmelidir. Alternatif olarak, viral RNA ekstraksiyonları için lizis tamponuna lineer poliakrilamid eklenmesi önerilmiştir.⁸ RNA ekstraksiyonu sırasında veya sonrasında, özellikle metatranskriptomik kütüphane hazırlıkları için, bir DNase işlemi de önerilir. NanoDropspektrofotometreler veya QubitFluorometre ile ölçülmüş ve kullanıma kadar -80 °C'de saklanmış RNA, yüksek hassasiyetli bir RNA testi kullanılarak kalitatif olarak analiz edilebilir. Dizilemeden önce, SARS-CoV-2

RNA'nın varlığı ve miktarı, her hedef için Ct (eşik döngü) değerleri sağlayan bir veya daha fazla viral geni (yani RdRp, orf1ab, E ve N) hedefleyen qRT-PCR kullanılarak değerlendirilebilir.⁹

Kontrol Numuneleri

Tampon veya su gibi negatif kontrol numuneleri birden fazla numune içeren herhangi bir dizileme çalışmasına, mümkün olan en erken aşamada dahil edilmeli ve dizilemenin ardışık tüm aşamalarında örneklerle ilerlemelidirler. Laboratuvarında veya biyoinformatik işlem sırasında meydana gelen kontaminasyonu dışlamak için bu son derece önemlidir.⁶

Bilinen genetik dizilere sahip pozitif kontrol örnekleri, konsensus çağırması için yeni kabul edilen veya uyarlanmış biyoinformatik veri hatlarını doğrulamak için faydalı olabilir, ancak her dizileme çalışmasına dahil edilmesi gerekmez.⁶

Dizileme Stratejileri

Klinik örneklerden SARS-CoV-2 dizilimi için çeşitli yöntemler mevcuttur. Dizileme stratejileri, genomik dizi hakkında önceden bilgi gerektirmeyen metagenomik yaklaşımları ve genom bilgisine dayanan hedefli yaklaşımları içerir. Her iki yaklaşım da tipik olarak dizilemeden önce SARS-CoV-2 genetik materyalini ortamda bulunan diğer RNA/DNA'lara göre zenginleştirmeye çalışır. Yeterli rezidüel RNA mevcutsa ve uygun şekilde saklandıysa, tanısal testler için ekstrakte edilen RNA kullanılarak çoğu yaklaşım gerçekleştirilebilir.⁶ Farklı uygulamalar için yöntemle ilgili önemli olan ana parametreler, seçilen kütüphane oluşturma yaklaşımını (hedeflenmemiş, yakalama tabanlı, amplikon tabanlı), oluşturulan okuma uzunluğunu, hata oranını ve hata profilini, sekanslama derinliğini ve genom veya kısmen sekanslanmış genom boyunca tek tip kapsama alanını içerir (Çizelge 4).

Birçok genomik surveyans amacı için genomun tam veya tama yakın bir konsensüs dizisi yeterlidir. Bu, açık kaynaklı ARTIC protokolü¹⁰, ticari kitler (Illumina ve IonTorrent platformları için mevcuttur) veya in-house protokoller gibi multipleksamplikon testleri kullanılarak uygun maliyetli bir şekilde yapılabilir. Bulaş ve/veya re-enfeksiyonun

doğrulanmasında, buna ilişkin kanıtlara önemli ölçüde katkıda bulunabilecek minör varyantların belirlenmesinde daha kapsamlı dizileme önerilir.¹

Kısa ampikon yöntemleri, ampikon uzunluğundan daha büyük veya benzer boyuttaki genom değişikliklerinin doğru tespitine ve haplotipleri belirlemeye/yeniden yapılandırmaya izin vermez. Bu nedenlerle, bu uygulamalar için, uzun okuma dizileme teknolojileriyle kombine daha uzun ampikonlar, yakalama tabanlı veya hedeflenmemiş kütüphaneler kullanılmalıdır. Tercih edilen yöntem Sangerdizilemesi ise, S geninin tamamı dizilenmelidir.¹

Yeni nesil dizileme (NextGenerationSequencing) teknolojileri, metagenomik örneklerden yeni virüslerin tanımlanması, tam veya tama yakın viral genom dizilerinin yeniden yapılandırılması ve viral evrimin ve yarı türlerin analizi dahil olmak üzere, virolojideki çeşitli uygulamalar için tercih edilen yöntem haline gelmiştir. NGS (NextGenerationSequencing) tabanlı yaklaşımların amaca en uygun avantajlarından biri, tam uzunluktaki viral genomların, bilinmeyen veya yeterince tanımlanmamış

virüsler için dahi, kültürle zenginleştirilmiş viral preparatlardan veya doğrudan klinik örneklerden başlayarak yeniden oluşturulabilmesidir. SARS-CoV-2 pandemisinde, hem ikinci hem de üçüncü nesil NGS teknolojileri başarıyla uygulanmış ve farklı üreticiler tarafından bağımsız olarak birkaç özel kütüphane hazırlama protokolü geliştirilmiştir.² Projenin nihai hedefleri ve eldeki biyolojik numunenin türü, en uygun dizileme stratejisinin seçimini belirleyen ana faktördür. Numunenin türü (örneğin klinik numuneler, çevresel numuneler, hücre kültürleri), viral yük (genellikle numune kaynağı ile ilgilidir), RNA ekstraksiyon prosedürü, RNA kalitesi, eş zamanlı yapma/otomasyon gereksinimleri ve diğer hususların tümü deneysel hedeflerle bağdaştırılmalıdır(viral genomun örnekler arası veya örnek içi varyasyonlarının araştırılması, viral ve konak transkriptomunun ve epitranskriptomunun incelenmesi, tek hücre çalışmaları, vb.).² Bugüne kadar, kavramsal olarak farklı dört yaklaşım uygulanmıştır (Çizelge 5).

Çizelge 4. SARS-CoV-2 genom dizileme uygulamaları ve önerilen teknolojiler¹

Uygulama	Önerilen sekanslama platformları	Önerilen kütüphane oluşturma yaklaşımları	Önerilen okuma uzunluğu	Önerilen yerel (yaklaşık) minimum kapsam
Bulaş modelleri, tür/soy belirlemesi, re-enfeksiyonun teyidi, fenotipik olarak ilgili mutasyonlar, veri raporlaması	MiSeq/NextSeq/iSeq/NovaSeq (Illumina), IonTorrent (ThermoFisher), MinION (Oxford Nanopore), Sequel Sistemi (PacBio)	Amplikon tabanlı (ARTIC, ticari, in-house)	>100 bp	>10x genomun >%95'inden fazla
Re-enfeksiyon ve/veya doğrudan bulaşma onayı (azınlık varyantlarının gerekli olduğu durumlarda)	MiSeq/NextSeq/iSeq/NovaSeq (Illumina), Ion S5 serisi/Genexus (ThermoFisher)	Amplikon tabanlı (ARTIC, ticari, in-house)	>100 bp	>500x genomun >%95'inden fazla
Derinlemesine genom analizi (büyük insersiyon veya delesyonlar, rekombinasyon, yeniden düzenlemeler, yarı tür haplotipleri)	Minion (Oxford Nanopore), Sequel Sistemi (PacBio)	Amplikon tabanlı (>1000 bp parçacıklar), yakalama tabanlı, hedeflenmemiş	>100 bp	>500x genomun >%95'inden fazla
Bilinmeyen patojenlerin veya oldukça farklı suşların tespiti	MiSeq/NextSeq/iSeq/NovaSeq (Illumina), Ion S5 serisi/Genexus (ThermoFisher), MinION (Oxford Nanopore)	Hedeflenmemiş RNA dizilimi, β -CoV'ye özgü RT-PCR	>100 bp	Örnek başına > 5 Gbp veri

Çizelge 5. SARS-CoV-2 sekanslama yaklaşımlarının özellikleri^{2,11}

	Shotgun metatranskriptomik	Amplikon-tabanlı	Hybridcapture-tabanlı	Doğrudan RNA sekanslama*
Hedefler	SARS-CoV-2, konak mikrobiyotası ve enfeksiyona konak yanıtı	SARS-CoV-2 genomu	SARS-CoV-2 genomu	SARS-CoV-2 ve konak transkriptom ve epitranskriptome
Ko-enfeksiyon saptama	Evet	Hayır	Hayır/Evet (gen paneline bağlı)	Evet
Minimum okuma sayısı	20-50 milyon	5-20 milyon	5-20 milyon	0,5 milyon
Genom kaplama	≥99%	≥95-99%	≥95-99%	≥99%
SNV tanımlamasında doğruluk	Yüksek	Yüksek	Orta düzey	Düşük
İstenen örnek viral yük (Ct)	<24-28	≥24-28	≥24-28	<24-28
Örnek RNA girişi (ng)	10-200	1-50	10-50	≥1000
Örnek tip	Hasta örnekleri	Hasta numuneleri, çevresel numuneler	Hasta numuneleri, çevresel numuneler	Viral hücre kültürleri
Maliyet	Yüksek	Düşük	Orta düzey	Yüksek
NGS dizileme platformları	Yüksek veya ultra yüksek verimli platformlar	Orta verimli platformlar	Orta veya yüksek verimli platformlar	ONT

*Doğrudan RNA dizilemesinden sadece 1 veri seti şu anda halka açık depolarda mevcuttur.

Shotgun Metatranskriptomik

Shotgun metagenomik dizileme, bir numunedeki tüm DNA'yı sorgulayabilen ve genom dizileri hakkında herhangi bir ön bilgi olmaksızın karmaşık mikroorganizma topluluklarının tanımlanmasına izin veren, kültürden bağımsız bir tekniktir. Metagenomik dizileme, daha önce tanımlanmamış patojenlerin tanımlanması için son derece güçlü bir araçtır. Bu yaklaşım, mikrobiyal toplulukların bileşimi hakkında ayrıntılı ve kantitatif bilgiler sunarak, klinik mikrobiyolojide tedavi stratejilerini yönlendirmek için kullanılabilmesi ilave katkı sağlar.^{2,6}

Shotgun metatranskriptomikleri çeşitli klinik numune türlerinden SARS-CoV-2 genomunun tamamını veya tama yakın bölümünü elde etmek için başarıyla uygulanmıştır. Metagenomik/metatranskriptomik yöntemler, diğer viral ve bakteriyel DNA/RNA'ları da tanımlayabildiğinden, ikincil enfeksiyonlar ve tedavi kararları hakkında fikir verme ve hasta sonuçlarını tahmin etme potansiyeline sahiptir. Ayrıca, metatranskriptomikler, enfekteepitelyal hücreler ve aktive immün hücrelerden konakçı transkriptlerini sağlayabildiğinden, hastalarda immün sistem reaksiyonunun anlık olarak tam bir görüntüsünü verebilir, virüs-konak etkileşimi çalışmalarına katkıda bulunabilir ve hatta hastaların sınırlı genotiplendirilmesini kolaylaştırabilir.²

Konakçının veya diğer SARS-CoV-2 olmayan genetik materyalin ortamdan uzaklaştırılması, oluşturulan dizi verilerinde daha yüksek oranda SARS-CoV-2 okumasını ve dolayısıyla tam bir genomu kurtarma şansının daha yüksek olmasını sağlar. Bu nedenle SARS-CoV-2 metagenomik

yaklaşımları, tipik olarak, RNA ekstraksiyonundan önce santrifüjleme, filtrasyon veya istenmeyen DNA/RNA'nın kimyasal veya enzimatik olarak çıkarılması yoluyla konakçı ve bakteriyel hücrelerini uzaklaştırma adımlarını içerir. Bu, bronkoalveolar lavaj gibi hücrelerin daha kolay ayrılabilmesi için sıvı numuneler için daha uygundur. Ribozomal RNA (rRNA) ve DNA içeriği de virüs RNA dizilimi için kütüphane hazırlığı sırasında azalır ve taşıyıcı RNA genellikle ekstraksiyonlardan çıkarılır veya lineer poliakrilamid ile değiştirilir. Tüm bu önlemlere rağmen, yine de numuneler, dizilenebilen yüksek miktarlarda hedef dışı konakçı DNA/RNA içerebilir. Bu nedenle metagenomik yaklaşımlar genellikle yüksek virüs yüklerine sahip numunelerin kullanılmasında faydalıdır. Alternatif olarak, genellikle çok sayıda okumanın oluşturulması gerekir; bu şekilde, SARS-CoV-2 genetik materyali, okumaların yalnızca küçük bir bölümünü temsil etse bile, virüs genomunun tamamını elde etmek yine de mümkün olacaktır.⁶

Tipik bir iş akışı, RNA'nın parçalara ayrılması, birinci ve ikinci zincir cDNA sentezi ve seçilen NGS teknolojisine göre kütüphane hazırlığından oluşur. Çoğu çalışma Illumina platformunu kullanmış olsa da, Oxford NanoporeTechnology (ONT), klinik örneklerden influenza virüsleri için tasarlanmış bir protokolün modifikasyonu yoluyla, shotgunmetatranskriptomik için de kullanılmıştır.^{12,13} ONT kütüphane hazırlığı için ≥1 µg DNA gereksiniminin karşılanmasında sekans bağımsız tek primeramplifikasyon (SISPA) adımı kullanılır.^{8,14} SISPA'nın

(Sequence Independent Single-Primer Amplification) olası hatalarına rağmen, bu yaklaşım, düşük miktarlarda RNA'dan bile eksiksiz SARS-CoV-2 genom yapılarının hızlı bir şekilde oluşturulmasına izin verir. Bugüne kadar kullanımı sınırlı olmasına rağmen, SARS-CoV-2'nin shotgun metatranskriptomikleri için Pacific Bioscience (PacBio) teknolojisi de uygundur.¹⁵

Shotgun metatranskriptomik yaklaşımı, SARS-CoV-2'nin tanımlanmasında kullanılmıştır ve yeni ortaya çıkan SARS-CoV-2 suşlarını dizilemek için de tercih edilen yöntemdir. Viral dizi hakkında önceden bilgi gerektirmez ve birbirinden farklı bölgelerin "capture" ve "amplikon" yaklaşımları üzerindeki olası etkilerinin önüne geçer. Prensip olarak, viral genom dışında, viralsubgenomik RNA'lar, olası transkripsiyon sonrası modifikasyonlar ve kütüphane hazırlama iş akışına bağlı olarak negatif zincirli ara ürünlerin tümü, shotgunmetagenomik ile incelenebilir. Ayrıca, yeterli düzeyde genom yapısı elde edildiğinde, konakçı gen ekspresyonuna ilişkin bazı bilgilere ek olarak, yarı türlerden veya koenfeksiyonlardan alınan örnek içi virüs varyantlarının doğru bir değerlendirmesini sağlayabilir ve daha önce de belirtildiği gibi, enfeksiyon sırasında konakçı gen ekspresyonuna ilişkin tahminlere izin verebilir.² Hedeflenen zenginleştirmeye dayalı yaklaşımlarla karşılaştırıldığında, önemli ölçüde daha yüksek bir dizileme derinliği (>2 G baz) gereklidir.² Bir SARS-CoV-2 genomu oluşturmak için daha fazla veri üretilmesi gerektiğinden, genellikle hedeflenen yakalama tabanlı veya amplikon tabanlı dizileme yaklaşımlarından daha maliyetlidir. Bununla birlikte, metagenomikler için faydalı olan santrifüjleme gibi ön-işlem aşamalarının moleküler tanı deneylerinde yapılmaması, metagenomik dizileme için ön-işlem aşamalarını içeren yeni ekstraksiyonların gerçekleştirilmesini gerektirebilir. Hedefli dizileme yaklaşımlarının genellikle daha uygun maliyetli olması ve daha az kaynak gerektirmesi, metagenomik yaklaşımların (örneğin patojen keşfi, koenfeksiyonların tespiti) faydalarının gerekli olmadığı durumlarda daha uygun olabilirler. Metagenomik yaklaşımların başarısı yöntemlere göre değişmektedir. Birkaç çalışma, gerçek zamanlı PCR (qPCR) döngü eşikleri (Cts) yaklaşık 25-30'un üzerinde olan örneklerde birkaç metagenomik dizileme analizinin başarısında hızlı bir düşüş olduğunu göstermiştir. Bu tür numuneler için, multipleks ve yakalama tabanlı PCR yöntemleri, genom boyunca metagenomik dizilemeye göre tutarlı bir şekilde daha yüksek içerik sağlar. Tam genomu elde etmek için üretilmesi gereken numune başına dizi okuma sayısı, numune tipine, konakçı materyali çıkarmak için ön tedavi prosedürlerine ve viremi düzeyine bağlı olacaktır.⁶ Viral yük, numune alma tekniğindeki varyasyonun yanı sıra hastalar arasındaki doğal yük farklılıklarından dolayı klinik numunelerde çok büyük farklılıklar gösterir. SARS-CoV-2 okumalarının oranı, viral yüklerin benzer olduğu durumlarda bile numuneler arasında büyük farklılıklar gösterebilir.¹⁶⁻¹⁸

Düşük viral yüke sahip numuneler için, virüsün hücre kültüründe çoğalmasına izin verilerek viral genetik materyalin oranı teorik olarak artırılabilir. Bununla birlikte, virüs kültürüyle ilişkili biyogüvenlik riskleri, kültürlenmemiş klinik numunelerle ilişkili risklerden önemli ölçüde daha yüksektir. Güvenli kullanım ve depolamayı sağlamak için kapsamlı ek prosedürlerle birlikte biyogüvenlik seviyesi 3 tesisleri gereklidir. Ek olarak, hücre kültürüne pasajlama, dizilerde, orijinal klinik örnekte mevcut olmayan yapay mutasyonlara ve sonraki analizler için önemli etkilere neden olabilir. Bu nedenle, özellikle,

dizileme duyarlılığını iyileştirmek için diğer bait-capture (tuzak yakalama) ve amplikon tabanlı yaklaşımlar mevcut olduğundan, hücre kültürünü yalnızca SARS-CoV-2 dizilimi için virüs genetik materyalini çoğaltmak amacıyla kullanılmamalıdır.⁶

Amplikon Tabanlı Dizileme

Amplikon dizileme, araştırmacıların analizlerinin kapsamını yalnızca sınırlı sayıda/türde bir seçim dizisiyle kısıtlamalarını sağlar. Bu yaklaşım oldukça spesifik ancak, "hedeflenecek" dizi hakkında önemli bir ön bilgi gerektirir. Virüs genomunun ayrı bölgelerinin amplifikasyonu için çok spesifik primerlere dayanan, klinik örneklerden SARS-CoV-2 nükleik asitlerinin saptanmasına yönelik tanıtılabilir RT-PCR testleri, amplikon dizilemenin özel bir biçimi olarak kabul edilebilir. SARS-CoV-2 dizilemesine yönelik amplikon tabanlı yaklaşımlar, birinci zincir cDNA sentezini takiben multipleksPCR'ler ile genom amplifikasyonundan oluşan bir zenginleştirme iş akışını benimser. Amaç, viral genomun tamamını veya belirli kısımlarını kapsayan amplikon havuzları üretmektir. SARS-CoV-2 için, amplikonların sayısı ve boyutunda farklılık gösteren birkaç farklı multipleks PCR tasarımı önerilmiştir.² Amplikon dizileme, düşük miktarlarda RNA ve bozulmuş numunelere karşı oldukça spesifik ve güvenilirdir. Viral olmayan okumalar nadir olduğu için metatranskriptomik yaklaşıma göre daha az dizileme gereklidir.² Yakalama tabanlı yaklaşımlardan farklı olarak, amplikon tabanlı yaklaşımlar, hedeflenen dizi ve kullanılan primerler arasındaki uyumsuzluğa tolerans göstermez. Bu nedenle, hedeflenen genomik çeşitlilik nispeten düşük olmalıdır ve/veya hedef sekans, primerlerin daha fazla korunmuş genomik bölgeleri hedefleyecek şekilde tasarlanması için yeterince bilinmelidir. SARS-CoV-2'nin insanlarda yakın zamanda ortaya çıktığı ve bu nedenle nispeten düşük küresel genomik çeşitlilik gösterdiği göz önüne alındığında, PCR tabanlı yaklaşımlar şu anda SARS-CoV-2 dizilimi için oldukça uygundur. Ancak, amplikon hatalarının oluşumunun izlenmesi ve primerbağlanma bölgelerindeki değişimlerin bir sonucu olarak bozulma meydana geldiğinde primerlerin değiştirilmesi gerekir.⁶

Optimize edilmiş PCR tabanlı yaklaşımlar PCR Ct değerleri 30'a kadar olan numunelerden rutin olarak tüm SARS-CoV-2 virüs genomlarının oluşturulmasına izin verir. Ct değerleri 30-35 olan numunelerden rutin olarak kısmi genomlar oluşturulabilir. Ancak bu değerler yaklaşık değerlerdir; Ct, farklı tanı yöntemleri arasında değişebileceğinden amplifikasyon başarısının mükemmel bir belirleyicisi değildir ve farklı tip ve kalitedeki numunelerin kullanılması duyarlılığı etkiler. Ek olarak, PCR tanı testlerinde hedeflenen genomik bölgeler, yaygın amplikon bazlı sekanslama yaklaşımlarında kullanılanlardan tipik olarak çok daha kısadır. Bu nedenle RNA degradasyonu tipik olarak PCR bazlı sekanslamayı PCR tanılarından daha fazla etkileyecektir.⁶ Amplikonsekanslama teorik olarak uygun ve ucuz olsa da, dikkate alınması gereken bazı sınırlamaları vardır. İlk olarak, primer verimliliğindeki farklılıklar veya primer bağlanma bölgelerindeki olası varyantlar nedeniyle, genom boyunca amplifikasyon bütünüyle kaçınılarak azalmış uyum ile tam olmayan birleşmeye neden olabilir. Ayrıca, primerler referans SARS-CoV-2 genom sekansı üzerinde tasarlandığından, bu yaklaşım büyük yapısal varyantları tanımlayamayabilir ve yüksek düzeyde genomik sapma varlığında sistematik sınırlamalar gösterebilir.² SARS-CoV-2 amplikon hazırlığı için, bazıları belirli NGS platformlarına uyarlanmış çeşitli ticari kitler ve ticari

olmayan protokoller mevcuttur. Dizileme derinliği marjinal bir düşünce olduğundan, kütüphaneler orta-verimli laboratuvar platformlarında dizilenebilir (yani IlluminaNextSeq ve Miseq; Iontorrent platformları, vb.). Ek olarak, önceki viral salgınlarda başarıyla uygulanmış olan hızlı, amplikon tabanlı bir protokolün SARS-CoV-2'ye uyarlanması yaklaşımıyla, SARS-CoV-2'nin amplikon dizilimi, ONT ve PacBio gibi Tek Molekül Dizilimi (SingleMoleculeSequencing/SMS) teknolojilerinin kısa yanıt süreleri ile birleştirildiğinde, bulaş zincirinin hızlı sürveyansı için kullanılabilir.¹⁹

Hybridcapture-Zenginleştirme Dizileme

Hybridcapture, amplikon tabanlı dizilemeye benzer şekilde, araştırmacıların yalnızca önceden tanımlanmış dizileri veya bir genomun, özellikle ilgilendikleri uygun bölgelerini hedeflemelerini sağlayan bir dizileme stratejisidir. Hybridcapture kullanan hedef zenginleştirme stratejileri, protein kodlayan genlerin ekzonlarının hızlı ve uygun maliyetli dizilimini (ekzom dizilimi) sağlamak için başlangıçta insan genomik çalışmaları için geliştirilmiştir.²⁰ Bir metagenomik dizileme kitaplığının hazırlanmasının ardından, dizilemeden önce SARS-CoV-2 genetik materyalini zenginleştiren yakalama tabanlı yaklaşımlar gerçekleştirilebilir. Bu tür yaklaşımlar, viral RNA'dan tuzak DNA veya RNA'lara ters transkripte olmuş DNA'nın hibritlenmesine dayanır. Bu tuzaklar, SARS-CoV-2 genomunun bölgelerini tamamlayıcı olacak şekilde tasarlanmıştır. Bir tuzığa (örn. konak DNA) başarılı bir şekilde bağlanmamış olan hedef dışı kütüphane materyali, enzimatik veya fiziksel yaklaşımlar kullanılarak çıkarılabilir. Bu, diğer koenfeksiyonları tespit etme şansını azaltır, ancak SARS-CoV-2 genomu ile eşleşecek beklenen dizileme okumalarının sayısını artırarak, multipleks çalışmalarda daha fazla örneğin etkin bir şekilde birlikte dizilenmesine olanak tanır.⁶

Hybridcapture, spesifik biyotinlenmiş problemlara hibridizasyon yoluyla, hedeflenen genetik materyali zenginleştirir ve shotgun metatranskriptomiklerine kıyasla önemli ölçüde azaltılmış bir sekanslama derinliği sağlar. Kütüphaneler laboratuvar platformlarında sekanslanabilir (Illumina NextSeq ve Miseq, Iontorrent, vb.). Genel olarak Hybridcapture-zenginleştirme yöntemleri, amplikon tabanlı yöntemlerden daha fazla sayıda parçacıklara/problemlere dayanır ve hedef dizilerin daha eksiksiz belirlenmesini sağlar. Ayrıca, hedef bölgelerin yakalanması, mükemmel tamamlayıcılığa PCR-amplikon üretimine göre daha az bağımlı olduğundan, hibridizasyon yoluyla yakalama genellikle genomik değişkenliğe karşı daha dayanıklıdır. Bazı çalışmalarda, Hybridcapture dizilemenin, SARS-CoV-2 genomlarının dizilenmesi için amplikon bazlı yöntemlerden daha az duyarlı olduğu tespit edilip düşük viral yüke sahip örnekler için uygulanması önerilmemekten, diğer çalışmalarda hibridizasyonla zenginleştirme çok düşük viral yüke sahip numuneler için bile başarılı olmuştur.^{18,21} Yakalama tabanlı yöntemler ayrıca, örnek içi varyantların tarafsız temsili de sunabilir. Aynı örnek üzerinde shotgun metatranskriptomik ve/veya Hybridcapture ile tahmin edilen alel frekans dağılımları arasında yüksek düzeyde uyum olduğu bildirilmiştir.^{18,21} Illumina tarafından geliştirilen SARS-CoV-2 genom yakalama zenginleştirme iş akışı, SARS-CoV-2 ve diğer solunum yolu virüslerinin eşzamanlı tespiti için problemler içerdiğinden dikkat çekicidir.²

PCR amplikon tabanlı yaklaşıma göre yakalama tabanlı yaklaşım kullanmanın bir avantajı, yakalama tabanlı yaklaşımların, prob dizilerinden %10-20'lik dizi

farklılıklarını tolere edebilmesidir. Bu, PCR tarafından tolere edilen uyumsuzluktan daha yüksektir, primer dizilerinden böyle bir sapma, yüksek amplikon hatası riskine yolaçacaktır. Bu nedenle, nispeten farklı SARS-CoV-2 dizilerini başarılı bir şekilde zenginleştirmek için yakalama tabanlı yaklaşımlar kullanılabilir. Yakalama tabanlı yaklaşımlar, PCR amplikon tabanlı yaklaşımlardan genellikle daha karmaşık ve daha pahalıdır.⁶

Doğrudan RNA Sekanslama

Yukarıda bahsedilen stratejilerin tümü, RNA'nın ters transkripsiyonunu, kütüphane oluşturmadan önce nükleik asitlerin az veya fazla manipülasyonunu gerektirir ve post-transkripsiyonel modifikasyonlar ve transkriptlerin bileşim oranı dahil olmak üzere bilgi kaybına neden olabilir. SMS, dizileme teknolojilerinde nispeten yeni bir gelişme olup, amplifikasyon olmaksızın ve bazı durumlarda (örneğin, ONT ile doğrudan RNA sekanslama) ters transkripsiyon olmaksızın tek nükleik asit moleküllerinin dizisinin doğrudan belirlenmesine izin verir. SMS teknolojileri genellikle "klasik" NGS yöntemlerinden daha uzun okumalar sağlar, ancak daha yüksek hata oranları olduğu bildirilmektedir.²² ONT tarafından doğrudan RNA dizileme protokolü kurulumu, transkripsiyon sonrası değişikliklerin saptanmasına izin verebilir. Ek olarak, uzun okumalar sayesinde, bu teknolojiler tek olgun ve öncü transkriptlerin ve koronavirüs enfeksiyonu sırasında meydana gelenler gibi karmaşık transkripsiyonel modellerin (rekombinasyon, alternatif transkript olgunlaşması, nadir transkripsiyonel izoformlar, vb.) çok doğru rekonstrüksiyonlarını sağlayabilir.

Dizileme Teknolojisinin Seçilmesi

SARS-CoV-2 genetik materyalini zenginleştirmek için ilk numune hazırlığından sonra, kütüphaneler tipik olarak herhangi bir virüs için uygun olan standart dizileme protokolleri kullanılarak hazırlanabilir. Protokol, kullanılan cihaza bağlı olacaktır. Dizileme teknolojisini belirlemeden önce, farklı teknolojilerin çalışma süresi, maliyetler, kullanım kolaylığı, sonrasında veri işleme, verim (veri üretim hızı) ve dizileme doğruluğu dikkate alınmalıdır.⁶

Geleneksel sekanslama (Sanger sekanslama), ayrı fragmanları (1000 bp'ye kadar) ayrı reaksiyonlarda dizilemek için kullanılabilir. SARS-CoV-2'nin tüm genom dizilimi, her hasta numunesi için en az 30 farklı amplikonun ayrı ayrı dizilenmesini gerektirir. Bu nedenle, Sanger dizilimi, yeni nesil dizilemeyi takiben düzenekteki boşlukları doldurmak için veya bir tanı testinin başarısızlığını takiben primer bağlanma bölgeleri gibi kısa bölgelerdeki virüs çeşitliliğini araştırmak için, genomların kısa parçalarını dizilemede oldukça yararlıdır.⁶

Yeni nesil dizileme platformları, rutin, tüm genom dizilimi için daha uygundur. SARS-CoV-2 için bugüne kadar yaygın olarak kullanılan dizileme platformları arasında Illumina, IonTorrent ve Oxford Nanopore Technologies yer almaktadır. Tüm yeni nesil dizileme platformları, tek bir çalışmada birden fazla örneğin birlikte dizilenmesine izin verir. Her teknolojinin temel avantajları ve sınırlamaları Çizelge 6'da özetlenmiştir. Tüm platformlar SARS-CoV-2 konsensüs genomları oluşturmak için uygun olsa da, bazıları belirli dizileme programı hedeflerini karşılamak için daha uygun olabilir. Örneğin, hızlı geri dönüş süresi klinik uygulamalar için önemli olabilirken, okuma düzeyi doğruluğu, konak içi çeşitliliğin araştırılması için daha önemli olabilir.⁶

Çizelge 6. SARS-Cov-2 dizi analizi için yaygın olarak kullanılan platformlar ve özellikleri⁶

Cihaz	Avantajlar	Kısıtlamalar	Cihaz çalışma süresi	Dizileme verimi	Göreceli kullanılabilirlik ve maliyet
Sanger dizileme	Erişim kolaylığı Kullanım kolaylığı Uygun maliyetli dizileme (birkaç hedef gerektiğinde)	•Çok düşük verim •Amplikonlar (genellikle 1000 bp'den fazla olmayan) ayrı ayrı amplifiye edilmeli ve dizilenmeli •Tam genomlar için pahalı •Metagenomik için uygun değil	Sıklıkla birkaç saat	Her bir çalışmada 100kB-2 Mb	Yaygın kullanım Birkaç hedef için nispeten düşük maliyet
Illumina (ör. iSeq, MiniSeq, MiSeq, NextSeq, HiSeq, NovaSeq)	Çok yüksek dizileme ürünleri mümkün Çok yüksek doğruluk iSeq taşınabilir Verileri işleme yöntemleri iyi yapılandırılmıştır	•Illumina iSeq dışında, satın alınması ve bakımı diğer bazı platformlara kıyasla pahalı •Maksimum okuma uzunluğu 2x300 bp.	Cihaza bağlı olarak 10-55 saat	Cihaza bağlı olarak 1.2-6000 Gb	Yüksek bakım ve başlangıç maliyetleri Hesaplı işletme maliyetleri
Oxford Nanopore Technologies (Flongle, MinION, GridION, PromethION)	•Taşınabilir, doğrudan dizileme •Gerçek zamanlı veriler •Düşük başlangıç ve bakım maliyetleri •Yeterli veri elde edilmez dizilemeyi durdurabilir •Çok uzun okumalar yapılabilir (SARS-CoV-2 genomunun tam uzunluğunu aşan)	•Homopolimerlerle ilgili zorluklar •Okuma başına hata oranı ~%5'tir, bu nedenle yüksek doğrulukta konsensüs dizileri elde etmek için uygun veri hatlarının kullanılması kritik öneme sahiptir. •Çoğaltma dizilimi kullanılmadığı sürece konak içi varyasyonu belirlemek için şu anda uygun değil	Hemen okunabilir Gerektiğinde birkaç güne kadar izlenebilir ve çalıştırılabilir	Flongleflow hücresi için <2Gb'den PromethION flow hücresi için 220 Gb'ye kadar PromethION'da 48 adede kadar flow hücresi kullanılabilir	Bakım gerektirmez ve başlangıç maliyetleri düşük Hesaplı işletme maliyetleri
IonTorrent	Dizileme başladığında hızlı geri dönüş	•Homopolimerlerle ilgili zorluklar •Satın almak pahalı •400 bp civarında maksimum tipik okuma uzunlukları	Çipe ve cihaza bağlı olarak 2 saat-1 gün	Cihaza ve çiplere bağlı olarak 30 Mb-50Gb	Hesaplı maliyetler

*Çeşitli araçların bu listesi, SARS-CoV-2 genomik dizilimi için en sık kullanılan araçlara genel bir bakış sağlamak içindir ve bu ürünlerin DSÖ tarafından onaylandığı anlamına gelmez.

Veri Depolama ve Erişim için Mevcut Kaynaklar

Bir salgın sırasında, ilgili epidemiyolojik ve klinik Meta verilerle birlikte patojen genom dizisi verilerinin hızlı paylaşımı, halk sağlığı yanıtında genomik dizilemenin etkisini en üst düzeye çıkaracaktır. Özellikle güncel pandemi sırasında, mevcut bilgilerin zamanında toplanması ve açık verilere doğrudan erişim esastır. Bununla birlikte, etkili önleme stratejilerinin uygulanması, ilaç kullanımı ve aşı gelişiminin desteklenmesi ve hastalığın ve etkilerinin anlaşılması için gerekli unsurlardır. SARS-CoV-2 sekans verilerinin ve ilişkili Meta verilerin dikkatli bir şekilde derlenmesi ve biriktirilmesi, hem epidemiyolojik çalışmalar

hem de kapsamlı ilişkilendirme çalışmalarında ve gelecekte takip çalışmalarının etkinleştirilmesinde önemlidir. Paylaşılan Meta veriler her zaman en azından numune toplama tarihini ve yerini içermelidir ancak, ek Meta veriler dizinin potansiyel uygulamalarını büyük ölçüde artıracaktır. Bu nedenle, mümkün olduğunda Meta veriler, numune tipine, dizinin nasıl elde edildiğine, hasta seyahat geçmişine ve demografik veya klinik bilgilere ilişkin verileri içermelidir.^{2,6}

Dizilerin yaygın olarak kullanılan, aranabilir platformlar aracılığıyla paylaşılması, verilerin erişilebilirliğini artırır. Uygun Meta verilere sahip SARS-CoV-2 genetik dizileri, sıklıkla birden fazla platform aracılığıyla paylaşılır.

Konsensüs genomlarını paylaşmak için kamuya açık veri tabanları, Ulusal Biyoteknoloji Bilgi Merkezi (NCBI), Avrupa Moleküler Biyoloji Laboratuvarı'nın Avrupa Biyoinformatik Enstitüsü (EMBL-EBI) ve Japonya DNA Veri Bankası'nı (DDBJ) içerir. Uygun Meta verilere sahip ham okuma verileri, NCBI Dizi Okuma Arşivi (SRA), EMBL-EBI ENA ve DDBJ Dizi Okuma Arşivi'ni içeren Uluslararası Nükleotid Dizisi Veri İşbirliğinin (INSDC) depoları aracılığıyla paylaşılabilir. Konsensüs genomları için halka açık bir veritabanı örneğin GISAID EpiCoV'tur. COVID-19 Veri Portalı, COVID-19 ile ilgili tüm biyomedikal veri kaynaklarının paylaşımını ve erişimini kolaylaştırmaya çalışır.⁶

SARS-CoV-2 genomik verilerinin yaygın olarak kullanılan havuzunu temsil eden GISAID EpiCovportalı, her bir viral genomla ilişkili numune türü, dizileme teknolojisi ve dizileme protokollerini ve hasta durumu gibi (örneğin hastaneye yatırılmış veya taburcu edilmiş) temel klinik notları içermektedir. Diğer önemli olabilecek hasta bilgileri (örn. cinsiyet, yaş) sistematik olarak toplanmaz. EpiCov'daki veriler herkese açık olarak erişilebilir olsa da, kullanıcılar kaydolmalı ve verileri üçüncü taraflara dağıtmamayı kabul etmelidir. Veri kullanımı araştırma amaçlarıyla sınırlıdır, ham dizileme verileri depolanamaz ve program niteliğinde erişim mevcut değildir. Özellikle, ham ve işlenmiş viral dizi verileri, International Nucleotide Sequence Database Collaboration (INSDC) havuzlarından birinde bulunmalıdır.²³ İnsan genetik verilerinin her zaman yürürlükteki yasalara ve düzenlemelere uygun olarak yönetilmesi ve mümkün olduğunda EGA ve dbGAP gibi özel güvenli veri havuzları aracılığıyla kullanıma sunulması gerekmektedir.²⁴ Ayrıca, tüm omik veri türleri için, veri kümelerinin faydasını ve gelecekteki yeniden kullanılabilirliğini en üst düzeye çıkarmak için ilgili meta veri standartlarına dikkatli bir şekilde bağlı kalmanın gerekli olduğu da unutulmamalıdır.⁶

Sonuç

Son yıllarda, pandemilere yol açabilecek ortaya çıkan salgınları belirleme, sınırlama, hazırlık ve sağlık kapasitesi oluşturmak için yatırımları genişletme ve sürdürme ihtiyacına ciddi biçimde yoğunlaşmıştır. Bu bağlamda, son deneyimlerde (SARS, MERS, Zika ve Ebola gibi) görüldüğü gibi, yeni bulaşıcı hastalıkların yayılmasını izlemek ve bunlara karşı koymak için, ortaya çıkan patojenlerin genomik dizilerinin yeniden yapılandırılmasında ultra hızlı ve uygun maliyetli yöntemler, önemli araçlardır.²

Genom dizi verileri, sadece birkaç ay içinde, SARS-CoV-2'nin insanlarda olası yayılma süresinin yeniden yapılandırılmasına, virüsün yayılmasını izlemede temel olan viralsuşların sınıflandırılmasına ve çeşitli seçici baskıların etkisi altında olabilecek SARS-CoV-2 genomundaki bölgelerin tanımlanmasına yönelik sistemlerin geliştirilmesine katkı sağlamıştır.²⁵⁻²⁸ Yüksek verimli transkriptomik, SARS-CoV-2 gen ekspresyonu, gen ürünlerinin birleşim oranları ve gen viral gen ekspresyonunun düzenlenmesinin olası moleküler mekanizmaları (post-transkripsiyonel modifikasyonlar dahil) hakkında yeni anlayışlar sağlamıştır.^{11,29,30} SARS-CoV-2 genomundaki, muhtemelen artan/azalan virülans veya insan konakçılara olası adaptasyon ile bağlantılı olabilecek genetik varyantlar vurgulanmıştır.³¹ SARS-CoV-2 dizi verilerine erişim sağlayan çok sayıda veri tabanı ve kaynak mevcuttur. Bununla birlikte, maksimum fayda için, ilişkili ham veriler ve Meta veriler kritik unsurlardır ve

olabildiğince kapsamlı olmalı, standart formatlarda sunulmalı ve erişilebilir olmalıdır.²

Çıkar Çatışması Beyanı

Yazarlar arasında herhangi bir çıkar çatışması bulunmamaktadır.

Yazar Katkıları

FGA, EPKK Fikir; FGA: Tasarım; FGA, MA: Veri toplama; FGA: Kaynak tarama; FGA, EPKK: Analiz ve/veya yorum; FGA, MA: Makale yazımı; MA: Eleştirel İnceleme; EPKK: Yayınlama süreci

Kaynaklar

1. European Centre for Disease Prevention and Control. (2021). Sequencing of SARS-CoV-2: first update (Technical Guidance). ECDC: Stockholm. <https://www.ecdc.europa.eu/en/publications-data/sequencing-sars-cov-2>. Published September 2021. Accessed September 10, 2021.
2. Chiara M, D'Erchia AM, Gissi C, et al. Next generation sequencing of SARS-CoV-2 genomes: challenges, applications and opportunities. *Brief Bioinform.* 2021;22(2):616-630. doi: 10.1093/bib/bbaa297.
3. GISAID. hCoV-19 Reference Sequence. www.gisaid.org. Published October 2021. Accessed October 17, 2021.
4. GISAID. Genomic epidemiology of novel coronavirus - Global subsampling. <https://nextstrain.org/ncov>. Published October 2021. Accessed October 17, 2021.
5. Rambaut A, Holmes EC, O'Toole Á, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol.* 2020;5:1403-1407. doi: 10.1038/s41564-020-0770-5.
6. World Health Organization. (2021). Genomic sequencing of SARS-CoV-2: a guide to implementation for maximum impact on public health, 8 January 2021. World Health Organization. <https://apps.who.int/iris/handle/10665/338480>. Published October 2021. Accessed October 10, 2021.
7. Li C, Zhao C, Bao J, Tang B, Wang Y, Gu B. Laboratory diagnosis of coronavirus disease-2019 (COVID-19). *Clin Chim Acta.* 2020;510:35-46. doi: 10.1016/j.cca.2020.06.045.
8. Greninger AL, Chen EC, Sittler T, et al. A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America. *PLoS One.* 2010;5(10):e13381. doi: 10.1371/journal.pone.0013381.
9. Carter LJ, Garner LV, Smoot JW, et al. Assay Techniques and Test Development for COVID-19 Diagnosis. *ACS Cent Sci.* 2020;6(5):591-605. doi: 10.1021/acscentsci.0c00501.
10. ARTIC NETWORK. (2020). SARS-CoV-2. <https://artic.network/ncov-2019>. Published October 2021. Accessed October 12, 2021.
11. Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H. The Architecture of SARS-CoV-2 Transcriptome. *Cell.* 2020;181(4):914-921. doi: 10.1016/j.cell.2020.04.011.
12. Chan JF, Yuan S, Kok KH, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet.* 2020;395(10223):514-523. doi: 10.1016/S0140-6736(20)30154-9.
13. Lewandowski K, Xu Y, Pullan ST, et al. Metagenomic Nanopore Sequencing of Influenza Virus Direct from Clinical Respiratory Samples. *J Clin Microbiol.* 2019;58(1):e00963-19. doi: 10.1128/JCM.00963-19.
14. Kafetzopoulou LE, Efthymiadis K, Lewandowski K, et al. Assessment of metagenomic Nanopore and Illumina sequencing for recovering whole genome sequences of chikungunya and dengue viruses directly from clinical samples. *Euro Surveill.* 2018(10);23:1800228. doi: 10.2807/1560-7917.ES.2018.23.50.1800228.
15. Lam TTY, Jia N, Zhang YW, et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature.* 2020;583:282-285. doi: 10.1038/s41586-020-2169-0.

16. Zhang H, Ai JW, Yang W, et al. Metatranscriptomic Characterization of Coronavirus Disease 2019 Identified a Host Transcriptional Classifier Associated With Immune Signaling. *Clin Infect Dis.* 2021;73(3):376-385. doi: 10.1093/cid/ciaa663.
17. Butler DJ, Mozsary C, Meydan C, et al. Shotgun Transcriptome and Isothermal Profiling of SARS-CoV-2 Infection Reveals Unique Host Responses, Viral Diversification, and Drug Interactions. *bioRxiv Preprint.* 2020 Update in: *Nat Commun.* 2021;12:1660. doi: 10.1101/2020.04.20.048066.
18. Xiao M, Liu X, Ji J, et al. Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples. *Genome Med.* 2020;12(1):57. doi: 10.1186/s13073-020-00751-4.
19. Meredith LW, Hamilton WL, Warne B, et al. Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect Dis.* 2020;20(11):1263-1271. doi: 10.1016/S1473-3099(20)30562-4.
20. Albert TJ, Molla MN, Muzny DM, et al. Direct selection of human genomic loci by microarray hybridization. *Nat Methods.* 2007;4(11):903-905. doi: 10.1038/nmeth1111.
21. Maurano MT, Ramaswami S, Zappile P, et al. Sequencing identifies multiple early introductions of SARS-CoV-2 to the New York City region. *Genome Res.* 2020;30(12):1781-1788. doi: 10.1101/gr.266676.120.
22. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouli Q. Opportunities and challenges in long-read sequencing data analysis. *Genome biology.* 2020;21(1):1-16. doi: 10.1186/s13059-020-1935-5.
23. Karsch-Mizrachi I, Takagi T, Cochrane G. Sequence Database Collaboration IN. The international nucleotide sequence database collaboration. *Nucleic acids research.* 2016; 44(D1):48-50. doi: 10.1093/nar/gkx1097.
24. Mailman MD, Feolo M, Jin Y, et al. The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet.* 2007;39(10):1181-1186. doi: 10.1038/ng1007-1181.
25. Chiara M, Horner DS, Gissi C, Pesole G. Comparative Genomics Reveals Early Emergence and Biased Spatiotemporal Distribution of SARS-CoV-2. *Mol Biol Evol.* 2021;38(6):2547-2565. doi: 10.1093/molbev/msab049.
26. Boni MF, Lemey P, Jiang X, et al. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat Microbiol.* 2020;5(11):1408-1417. doi: 10.1038/s41564-020-0771-4.
27. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill.* 2017;22(13):30494. doi: 10.2807/1560-7917.ES.2017.22.13.30494.
28. Kosakovsky Pond SL, Poon AFY, Velazquez R, et al. HyPhy 2.5-A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. *Mol Biol Evol.* 2020;37(1):295-299. doi: 10.1093/molbev/msz197.
29. Di Giorgio S, Martignano F, Torcia MG, Mattiuz G, Conticello SG. Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. *Sci Adv.* 2020;6(25):eabb5813. doi: 10.1126/sciadv.abb5813.
30. Picardi E, Mansi L, Pesole G. A-to-I RNA editing in SARS-CoV-2: real or artifact? *BioRxiv: Prepr Serv Biol.* 2020. doi: 10.1101/2020.07.27.223172.
31. Pachetti M, Marini B, Benedetti F, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J Transl Med.* 2020;18(1):179. doi: 10.1186/s12967-020-02344-6.