



### INVITED REVIEW

#### SARS CoV-2'nin (COVID-19) biyoinformatik destekli identifikasyonu

Mehmet Özkan Timurkan<sup>1</sup>, Hakan Aydın<sup>1</sup>

<sup>1</sup>Atatürk Üniversitesi Veteriner Fakültesi Viroloji Anabilim Dalı, Erzurum, Türkiye

Geliş:13.08.2020, Kabul: 20.10.2020

\*timurkan@gmail.com

#### Bioinformatics-aided identification of SARS CoV-2 (COVID-19)

Eurasian J Vet Sci, 2020, Covid-19 Special Issue, 50-58

DOI: 10.15312/EurasianJVetSci.2020.291

#### Öz

Keşfedilen coronavirusların sayısındaki büyük artış ve sekanslanan coronavirus genomları, bu virus ailesi üzerinde genomik ve biyoinformatik analizler yapmak için bize eşi görülmemiş bir fırsat vermiştir. Coronaviruslar, bilinen tüm RNA virusları arasında en büyük genoma sahip (yaklaşık 30 kb) ailelerden biridir. Virusun çeşitli genleri (ORF1ab, spike, zarf, membran ve nükleokapsid) bulunmaktadır. Ayrıca filogenetik olarak A, B, C ve D alt gruplarından oluşan Betacoronavirus ile birlikte Coronaviridae ailesi dört cinsten oluşmaktadır (Alfacoronavirus, Beta-, Gamma- ve Delta-). Coronavirusların çeşitli gen lokusları kullanılarak yapılan moleküler yapısal analizde, 2003 SARS coronavirusun atası olan virüsle  $4 \times 10^{-4}$  ile  $2 \times 10^{-2}$  arasında bir oranla her yıl değişim geçirdiğini göstermiştir. Coronaviruslar ayrıca rekombinasyona açık viruslardır. Farklı murine hepatitis viruslar (MHV) arasında, farklı infectious bronchitis viruslar arasında, MHV ve bovine coronaviruslar arasında, feline coronavirus (FCoV) tip 1 ve canine coronavirus arasında ve insan coronavirusları arasında bu mutasyon tipi bildirilmiştir. Dolayısıyla mutasyonlara ve değişimlere çok açık olan bir organizma olan virusların sürekli değişen etkileşimlerini incelemek ve anlamak için biyoinformatik analizlere ve sonunda çıkan raporlarla virusların tanımlanmalarına ihtiyaç vardır. Bu derlemede, coronavirus biyoinformatiğinin az bilinen ama araştırmaya muhtaç mevcut ilerlemeleri vurguladık ve gelecekteki gelişmeler için potansiyel yolları tartıştık. Çok kullanılan ve mevcut olan teknolojilere genel bir bakış sunulmuş, SARS CoV-2 özelinde biyoinformatiğin viroloji alanına getireceği bazı önemli avantajlar ve dezavantajlar özetlenmiştir.

**Anahtar kelimeler:** Coronavirus, Covid-19, pandemi

#### Abstract

Current increase in discovery of coronaviruses and their sequenced genomes gave us a unique opportunity to make genomic and bioinformatic analyses of this virus family. Coronaviruses have one of the largest genomes (approx. 30 kb) among known RNA viruses. These viruses have various genes (Open Reading Frame 1ab, spike, envelope, membrane and nucleocapsid). Furthermore Coronaviridae family has 4 genera (Alphacoronavirus, Beta-, Delta- and Gamma-) with Betacoronaviruses consisting of 4 subgroups phylogenetically, namely A, B, C and D. 2003 SARS coronavirus is shown to have  $4 \times 10^{-4}$  ile  $2 \times 10^{-2}$  difference with its common ancestor annually in a molecular structural analysis using different gene loci of coronaviruses. Coronaviruses are also known to have genetic recombination. This type of genetic interaction is reported among different murine hepatitis viruses (MHV), among different infectious bronchitis viruses, between MHV and bovine coronaviruses, between feline coronavirus type 1 and canine coronavirus and among human coronaviruses. Thus we need reports on characterization of viruses by analyzing bioinformatically to investigate and to understand the ever changing virus interactions which are shaped by mutations and genetic changes. In this review we explained and discussed the current status and potential future possibilities of lesser known but much needed coronavirus bioinformatics. We provided a general perspective of present and frequently used techniques and summarized the main advantages and disadvantages of the use of bioinformatic analyses of SARS CoV-2 for virology field.

**Keywords:** Coronavirus, Covid-19, pandemic



## Giriş

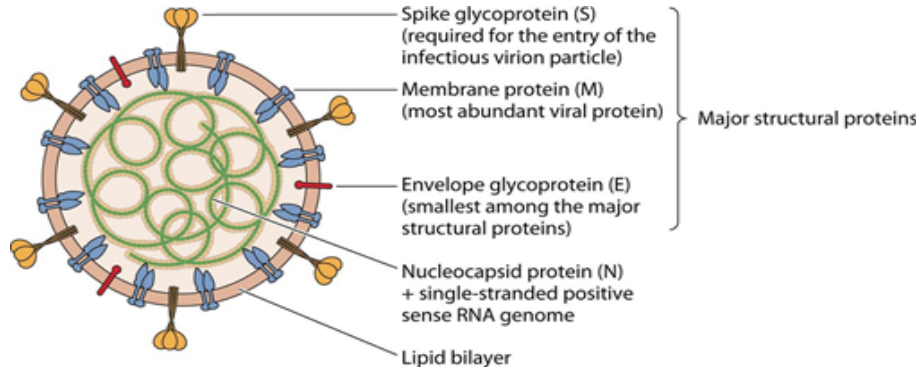
### Biyoinformasyon ve viroloji

Milyenler ile başlayan ve gittikçe ivme kazanan biyolojik gelişmeler ve enfeksiyon durumu viroloji bilimini de etkilemiştir. Viruslar çok basit canlılardır (Şekil 1). Ancak bu basitlikleri onlarla mücadelemizi kolaylaştırmak yerine zorlaştırmaktadır. Viruslarla mücadele ve savaşma noktasında, insanoğlunu zorlayan durumların başında; virusların sürekli değişen (mutasyona uğrayabilen) canlılar olması ve virusların biyoinformasyonunun tam olarak bilinmiyor olması gelmektedir (Kirk ve ark 2015). Viruslar insan ve hayvan sağlığı üzerinde önemli bir yük oluşturmaktadır. Son yıllarda, hem yeni (emerging disease) viral hastalıkların (MERS, COVID-19 vb.) ortaya çıkmasına hem de farklı coğrafi alanlarda bilinen hastalıkların (Zika, West Nile, Crimean Congo Hemorrhagic Fever, Bovine Viral Diarrhea Virus-3, Feline immunodeficiency Virus, Bovine Parainfluenza Virus-3 vb.) yeniden (re-emerging disease) ortaya çıkmasına (Şekil 2) tanık olduk (Oguzoglu ve ark 2010, Timurkan ve Aydın 2019, Timurkan ve ark 2019). Viral hastalıkların ortaya çıkma riskinin artma-

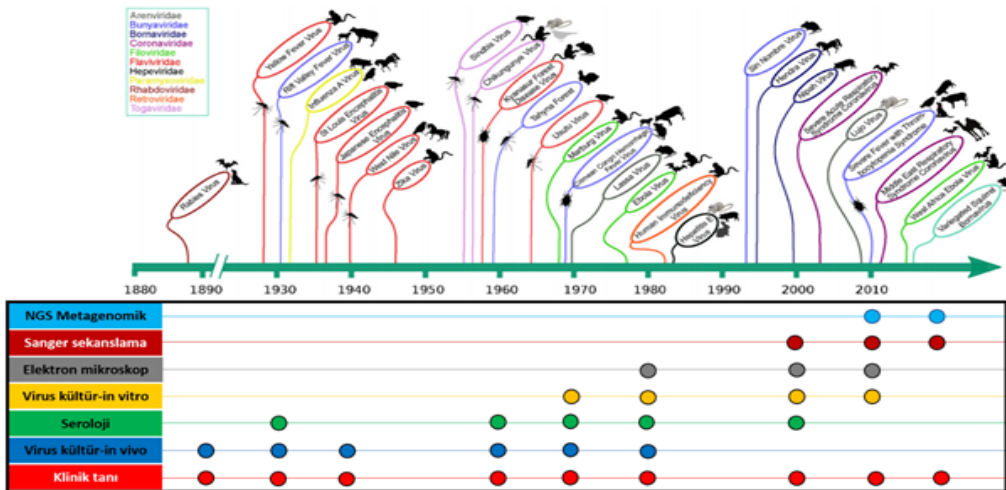
sı, çeşitli sosyal, çevresel ve ekolojik faktörlerden kaynaklanmaktadır. İklim değişikliği, ormansızlaşma, kentleşme ve ticari mal ve malzemelerin, insanların, hayvanların ve hastalık taşıyıcı vektörlerinin benzeri görülmemiş hareketliliği, viral hastalıkların yayılmasını kolaylaştıran ve pandemiler için potansiyel olarak ideal koşullar yaratan unsurlardır (Weiss ve McMichael 2004, WHO 2020).

Viral hastalıkların ekonomik yükü ülkeler açısından oldukça yüksektir. Tüm küresel afetlerin maliyetlerinin şu anda yılda 150 milyar dolar olduğu tahmin edilmektedir ve bunların 30 milyar doları yalnızca bulaşıcı hastalık salgınlarına atfedilebilir. Viruslar sadece insanlarda değil hekimliğimiz açısından hayvanlarda ve çoğu zaman unutulsa da bitkilerde de hastalıklara neden olabilir. Hayvan hastalıkları başta gıda güvenliğimizi etkilemektedir ve ekonomiye önemli zararlar vermektedir. Örneğin, endemik bölgelerde Şap hastalığından kaynaklanan yıllık kayıplar 6.5 ile 21 milyar dolar arasındadır (Knight-Jones ve Rushton 2013).

Viroloji alanında çalışan uzmanlar geleneksel olarak insanlarda, hayvanlarda veya bitkilerde hastalığa neden olan vi-



Şekil 1. SARS CoV-2'nin şematik olarak önemli proteinleriyle beraber yapısı (Şekil; Dhama ve ark 2020'den alınmıştır)



Şekil 2. Emerging/Re-Emerging bazı viral hastalıkların zaman içindeki süreçleri. (Şekil; Ibrahim ve ark 2018 ve Jerome ve ark 2015 yayınlarından derlenmiştir)



rusları incelemeye odaklanmış durumdayız ancak dünyanın canlı kısmı olan biyosferde şaşırtıcı miktarda virus vardır ve bunların çok büyük bir kısmı maalesef tanımlanmamıştır. Bu nedenle, virusların sadece zorunlu hücre içi parazit olduğu görüşü artık geçerli değildir. Doğada var olan viruslar konakçı olarak girdikleri popülasyonların genetik bilgilerini aktarabilir ve depolayabilir ayrıca biyojeokimyasal döngüleri etkileyebilir. Bu nedenle, virusları (insan, hayvan veya bitki virusları) önemli hastalıklara neden olan patojenler olarak bilesek te bunun dışında kalan büyük çoğunluğu tüm ekosistemlerin düzenlenmesinde önemli roller oynayabilir (Ibrahim ve ark 2018).

Virusların evlasyonu (evrimi) çok hızlıdır ve konakçı bağımsızlık sistemlerini ve/veya terapötik müdahaleler ile uygulanan en karmaşık kontrol önlemlerini yani bu tür baskılara yanıt olarak genomlarını hızla değiştirebilirler. Dolayısıyla bu akıllı canlıların genetik kodunu açığa çıkarma onunla savaşmada önemlidir. Virolojide ele alınması gereken birçok temel soru vardır. Örneğin, farklı konaklarda ve ortamlarda virus ailelerinin tüm çeşitliliğini nasıl araştırabiliriz? Viruslar nasıl gelişir ve viral evrimde mutasyon (rekombinasyon) ne kadar önemlidir? Tek bir ortak viral köken var mı yoksa bağımsız kökenler mi buluyoruz? Virusların çeşitli ekosistemlerde taşıdığı dinamik gen havuzunu nasıl belirleyebiliriz? Virusların immün sistemden kaçış mekanizmaları nelerdir ve biz ne kadarını biliyoruz? Virusların doğada var olma süreçleri ve bunu yaparken kullandıkları metotlar nelerdir? Diğer birçok soru, viral hastalıkları kontrol etmek ve tedavi etmek için stratejiler geliştirmemize ve aynı zamanda virusların daha geniş ekolojik rolünü anlamamıza yardımcı olacaktır (Suttle 2005, Marz ve ark 2014).

### SARS CoV-2 ve tanımlanması (identifikasyon)

Çin'de binlerce insanı enfekte eden ve dünya çapında hızla yayılan yeni bir coronavirusun küresel bir salgını ile karşı karşıyayız. Ocak 2020'nin sonunda, Dünya Sağlık Örgütü (DSÖ) -World Health Organization (WHO) bunu küresel bir acil durum ilan etti ve pandemi olarak adlandırdı. İlk olarak Çin'in Wuhan kentinde izole edilen virus, yeni coronavirus (new ya da novel CoV - nCoV) adıyla dünyaya duyuruldu. Ancak yapılan biyoinformatik analizlerle 2003 yılında ortaya çıkan şiddetli akut solunum sendromu hastalığı virüsü (Severe Acute Respiratory Syndrome - SARS) ile çok benzer olan (%79-80 genomik benzerlik (Zhou ve ark 2020) ve bu yüzden de onun ikinci nesil mutasyonu anlamında SARS-related coronavirus (SARSr-CoV) ya da SARS CoV-2 adı verildi (Lu ve ark 2020). WHO bu enfeksiyonu 2019'un son zamanlarında ortaya çıkması ve Corona Virus Disease kısaltması ile COVID-19 olarak adlandırmıştır (Huang ve ark 2020). Bu salgın daha önce görülmemiş bir olaydı, bilim camiasının buna tepki verme şekli de öyle oldu. Hastalığın görüldüğü ve görülmediği tüm ülkeler bu duruma bir cevap aradı. Ülkelerin ekonomisi ve bilim

gücü ile hastalığın tanımlanması, yeni kitlerin geliştirilmesi, yeni ilaçların bulunması ve denemesi, tedavi protokolleri gibi birçok şey denendi ve denenmeye de devam etmektedir. Ancak savaşılan düşmanın bilinmesi en önemli durumdur. Virus gibi gözle görünmeyen bir canlının da bu ve bundan sonraki süreçte ne yapacağının öngörülmesi için yapısının, proteinlerinin ve özellikle genomunun açığa çıkarılması yani biyoinformasyonunun yapılması şarttır. SARS-CoV-2, *Nidovirales* takımının, *Coronaviridae* ailesinin, *Orthocoronavirinae* alt ailesinin bir üyesidir ve bu virus dört cinse (genus) *Alphacoronavirus*, *Beta-*, *Gamma-* ve *Delta-* ayrılmıştır (Şekil 3). Alphacoronavirus ve Betacoronavirus cinsleri yarasalardan köken almıştır, Gammacoronavirus ve Deltacoronavirus kuş ve domuz gen havuzlarından gelişmiştir (Dhama ve ark 2020)

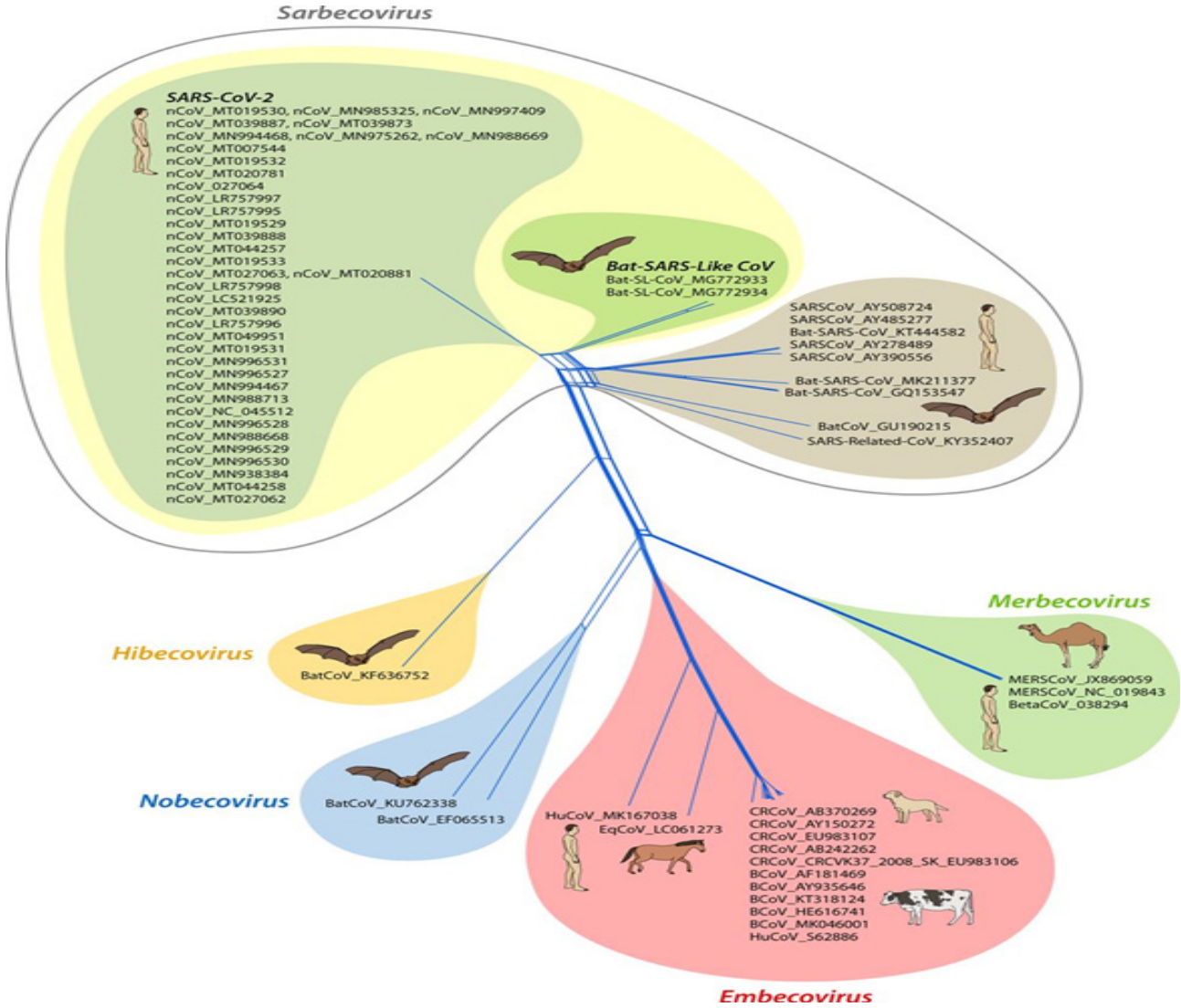
### Viruslar ve biyoinformatik

COVID-19 pandemisinde dünya ve üzerindeki canlılar olarak bu duruma çok hızlı bir şekilde tepki verildi. Bu tepki iyi yöndeydi. Araştırma sonuçlarını neredeyse anında yayınlanmasına izin verilerek, daha önce hiç görülmemiş şekilde virusun genomik yapıları ve hastalarda klinik verileri açık ve hızlı bir şekilde paylaşıldı. Bu, hastalığın ve virusun dinamiklerinin anlaşılmasına, hızlı teşhisin geliştirilmesine yardımcı oldu ve halkın bilgilendirilmesini sağladı.

Teknik olarak, viral genomların küçük boyutu, birçok virusun tüm genom sekansının belirlenmesini mümkün kılar. Bu diğer herhangi bir canlı sistem için genellikle mevcut olmayan bir avantajdır. Çünkü bilinen en büyük genoma sahip pandoravirusun (Andrade ve ark 2019) genomu (yaklaşık 2.5 Mb) bile bir E.coli genomunun (yaklaşık 5 Mb) ancak yarısı kadardır. Dolayısıyla viruslarda komple genom analizleri daha kolay olabilmektedir. Ancak bu avantaj ile virusların biyoinformasyonu kolay gibi görünse de özellikle RNA genoma sahip virus ailelerinde olmak üzere tüm viruslarda mutasyonlar ve değişimler oldukça sık olmaktadır. Bu yüzden bu dinamik süreçte informasyonun süreklilik arz etmesi esastır (Woo ve ark 2010).

Genoma sahip canlıların (virus, bakteri, parazit vb) genomik sekansının elde edilmesi yalnız başına yeterli olmamaktadır. Bu veri full genom tarzı bile olsa mutlaka bir analize ihtiyacı olacaktır. Viral anlamda; virusun hangi gen bölgesi önemlidir? Virusu oluşturan proteinlerin bağlanma yerleri nelerdir? Konakta virusun tutunduğu reseptörün bile çoğu zaman biyoinformatiği yapılarak patogenezi hakkında bilgi sahibi olunmaya çalışılır. Bununla birlikte, viral genomlar için mevcut sekanslama teknolojilerinde zorluklar ortaya çıkarmaktadır. Çünkü çoğu analiz adımı insan yardımına ihtiyaç duymaktadır ve her yöntemin kendine özgü teknik sınırlamaları vardır (Marz ve ark 2014). Ancak günümüzde SARS CoV-2 genomunun analizinde şimdilerde sekans verilerinin elde edilmesi konusunda oldukça yol kat edilmiş durumdadır. Çünkü hasta bireylerden alınan uygun numuneler, firmalara





Şekil 3. 39 farklı SARS CoV-2 şusunun S (spike) geni bazlı filogenetik analizi (Splits Tree (SplitsTree 4.0) analizi). SARS CoV-2 şuşları, MERS CoV, sığır CoV, köpek CoV, yarasa CoV, yarasa-SL-SARS-CoV ve at CoV dahil olmak üzere geçmiş insan salgınlarından ve hayvan kaynaklı CoV'lar. Analiz, Betacoronavirusların beş tanımlanmış alt türünü, yani Sarbecovirus, Embecovirus, Merbecovirus, Nobecovirus ve Hibecovirusları içerir. Gri alandaki şuşlar, dünyadaki mevcut SARS CoV-2 salgınından ve sarı alan ise SARS CoV-2'nin filogeni olarak en yakın komşuları olan yarasa-SL-CoV'lardan oluşmaktadır (Şekil; Dhama ve ark 2020'den alınmıştır).

verilerek bir PCR analizi dahi yapmadan kısa sürede komple genom sekansı şeklinde elimize ulaşmaktadır. Ancak bilginin yorumlanması ve analizler noktasında viroloji bilimiyle uğraşmayan kişiler bu konuda zorluk çekmektedir. Viroloji alanındaki uzmanlar olarak bu konuda gerekli analizleri yaparak, yan dal bilim insanlarına yardımcı olmaktadır.

Bununla birlikte, biyoinformatik yöntemleri entegre ederek, gelecekte, bir bireyin sınırlı genetik varyasyona sahip bir virus içerip içermediği, bireysel olarak virus özelliklerine dayalı, hastalarda viral evrimi tahmin etmek mümkün olabilir. Burada nihai amaç, bir virus enfeksiyonunun seyrini (prognoz veya taşıyıcılık) tahmin etmek ve buna göre terapötik tedavileri ayarlamak olacaktır.

### SARS CoV-2'nin biyoinformatiğinde kullanılan sistemler ve metotlar

COVID-19 enfeksiyonu için birçok durum bildirilse de yapılması gereken bu virusun bilimsel ve genetik durumunun bir an önce açığa kavuşturulmasıydı. Virusun genomik yapısının, gen bölgelerinin, protein bağlanma noktalarının, ataçman ve nötralizan yapılarının vb., irdelenmesi ve açığa çıkarılması gerekmektedir. Bunun içinde öncelikle virusun izole edilmesi, basit düzeyde gen bölgeleri ve ileri düzeyde tüm genomun sekansı ve sonrasında biyoinformatik analizlerinin yapılması gerekiyordu. Dolayısıyla birçok ülkede sağlık kuruluşları, üniversiteler veya özel sektörler bu virolojik ve epidemiyolojik bilginin açığa çıkarılması için çaba sarf etmiştir.



Tablo 1. Bazı virus biyoinformatik veri tabanlarının ve araçlarının listesi

Virusa Özgü Veritabanları (Database)	Sekans Verileri ve Gen Bankası	Filogenetik Analiz	Epidemiyolojik Veriler ve Virus Konak İlişkileri	Tüm Genom Kurgulama
Nextstrain (Hadfield ve ark 2018)	GenBank/NCBI (Benson ve ark 2012)	MEGA (Kumar ve ark 2018)	COVID-19 Disease Map (Ostaszewski ve ark 2020)	Velvet (Zerbino ve Birney 2008)
NCBI Virus (Brister ve ark 2015)	ENA/EMBL-EBI (Kulikova ve ark 2004)	Armadillo (Lord ve ark 2012)	CoV-Hipathia (Rian ve ark 2020)	ABYSS (Simpson ve ark 2009)
GISAID (Shu ve McCauley, 2017)	DDBJ (Ogasawara ve ark 2020)	Geneious (Kearse ve ark 2012)	Breaking-Cas (Oliveros ve ark 2016)	SPAdes (Bankevich ve ark 2012)
ViPR (Pickett ve ark 2012)	INSDC (Kasch-Mizrachi ve ark 2018)	MrBayes (Huelsenbeck ve Ronquist 2001)	CDC (Center for Disease Control Prevention, <a href="https://www.cdc.gov/">https://www.cdc.gov/</a> )	IDBA-UD (Peng ve ark 2012)
ViralZone (Hulo ve ark 2011)	CoV2ID (Carneiro ve ark 2020)	T-REX (Boc ve ar 2012)	ELIXIR (Blomberg ve Lauer 2020)	VICUNA (Yang ve ark 2012)
VVR (Hatcher ve ark 2017)	GDCTT (Cleemput ve ark 2020)	Phangorn (Schliep 2011)		VrAP (Hölzer ve Marz 2017)
ICTV (Lefkowitz ve ark 2018)		RAxML-NG (Kozlov ve ark 2019)		

Tablo 1’de SARS CoV-2 özelinde tüm virolojik çalışmalar kullanılabilir biyoinformatik veritabanları ve araçlardan bazıları verilmiştir.

#### Virusa özgü veri tabanları (database)

Nextstrain (<https://nextstrain.org/>), patojen mikroorganizma genom verilerini bilimsel olarak ve halk sağlığı açısından dünyanın yararlanması için kurulmuş olan açık kaynaklı bir projedir. Ayrıca patojenlerin evrimini ve oluşacak salgınlarda yayılmayı gösteren güçlü analitik ve görsel açıdan zengin, herkese açık, verilerin sürekli güncellendiği bir veri tabanıdır. Bu veri tabanı içinde SARS CoV-2 haricinde İnfluenza, West Nile, Ebola, Zika, Kızamık, Kabakulak, Dengue ve Enterovirus etkenleri içinde detaylı bilgiler bulunmaktadır. SARS CoV-2 sekans verilerinin gen bankalarına yüklenme dönemlerinde bu veri tabanı oldukça hızlı ve kolay bir şekilde verileri tüm dünya ile paylaşmıştır. Nextstrain ekibi sekans verilerini küresel ve kıtasal düzeyde analiz etmiştir. Bu veri tabanında ülke ve kıta bazlı analizler yapılabilir, sisteme yüklenen verilerle filogeni oluşturulabilir ve tespit edilen tipin hangi grup ve alt gruplara yakın hangi gruplara uzak, analizleri gerçekleştirilebilir.

NCBI Virus (<https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/>), RefSeq, GenBank ve diğer NCBI depolarından özellikle viral sekans verileri için bir topluluk portalıdır. Özellikle virüsler için Amerikan gen bankasının bir ara yüzüdür. Bu ara yüzün amacı GenBank ve diğer NCBI veri tabanlarında arşivlenen verilerin kullanılabilirliğini artırmaktır.

GISAID (Global Initiative on Sharing All Influenza Data) (<https://www.gisaid.org/>), bu veri tabanı önceleri sadece influenza virüsleri için epidemiyolojik veriler vermiştir. An-

cak COVID-19 pandemisi sonrası bu virusa dair hızlı veri paylaşımını desteklemiştir. Bu veri tabanı araştırmacıların salgınlar ve pandemiler sırasında virüslerin nasıl geliştiğinin ve yayıldığına anlaşılmasına yardımcı olmuştur. Ayrıca bu veri tabanında bu virüslerin gen sekansları, virüslerin klinik ve epidemiyolojik verileri ve kuş ve diğer hayvan virüsleriyle ilişkili coğrafi ve türe özgü veriler bulunmaktadır.

ViPR (Virus Pathogen Resource) veritabanı, (<https://www.viprbrc.org/brc/>) Arenaviridae, Bunyaviridae, Caliciviridae, Coronaviridae, Flaviviridae, Filoviridae, Hepeviridae, Herpesviridae, Paramyxoviridae, Picornaviridae, Poxviridae, Reoviridae ve Rhabdoviridae gibi emerging ve re-emerging viral hastalık aileleri için sekans verilerinin işlenmesi, analizi, 3D protein yapılarının çıkarılması ve filogenetik analiz gibi bir çok veriyi sunan bir veri tabanıdır. COVID-19 döneminde hem Coronaviridae ailesi hemde SARS CoV-2 için özel data, araçlar ve analizler için özelleştirilmiştir.

ViralZone (<https://viralzone.expasy.org/>), virion ve genom şekilleriyle birlikte genel moleküler ve viral epidemiyolojik bilgileri sağlayan bir veritabanıdır. Her virus veya virus ailesinin bir sayfası, UniProtKB / Swiss-Prot viral protein girişlerine kolay erişim sağlamaktadır.

VVR (virus variation resource) (<https://www.ncbi.nlm.nih.gov/genome/viruses/variation/>), bir NCBI alt ara yüzü olan VVR bazı virüsler için veri kümeleri oluşturmuştur. Bu ara yüzde İnfluenza virus, rotavirus, dengue, west nile, ebola, zika ve MERS-coronavirus için analizler ve görselleştirme araçları bulunmaktadır.

ICTV (International Committee on Taxonomy of Viruses) (<https://talk.ictvonline.org/taxonomy/>), Viroloji için çok





önemli veri tabanlarından biri olan ICTV, ilk kez 1966 yılında kurulmuş ve her üç yılda bir komitesinin başkanı ve yardımcıları ile sekreteryasını güncelleyerek bugünlere gelmiştir. Herhangi bir analiz ve aracı olmayan veri tabanı virusların taksonomisini veri sunmaktadır. Bu komite, yaygın virus adlarının veya hastalık adlarının belirlenmesi yerine virus taksonlarının (yani türler, cinsler, familyalar vb.) belirlenmesi ve adlandırılmasıyla ilgilenir. Yeni bir viral hastalığın salgını için karar verilmesi gereken üç isim vardır: hastalık, virus ve virus türü. Dünya Sağlık Örgütü (WHO) hastalıktan, uzman virologlar virustan, ICTV ise virus türü yani taksondan sorumludur.

### Sekans verileri ve gen bankası

Hiç kuşkusuz ki bu kategorinin en önemli üyesi dünya gen bankası olarak ta bilinen Amerikan National Center for Biotechnology Information (NCBI) olarak bildiğimiz genbankasıdır (<https://www.ncbi.nlm.nih.gov/>). Ağustos 2020 itibariyle Severe acute respiratory syndrome coronavirus 2 anahtar kelimesi ile nucleotid seçeneği seçildiği takdirde 13418 tane bu virusa dair sekans verisi girilmiş durumdadır. Ülkemiz özelinde ise sadece 56 verinin gen bankasında olduğu görülmektedir.

Avrupanın gen bankası olarak bilinen EMBL-EBI (European Molecular Biology Laboratory (EMBL) - European Bioinformatics Institute (EBI) (<https://www.ebi.ac.uk/>) ve bunların bünyesinde kurulan European Nucleotide Archive (ENA) (<https://www.ebi.ac.uk/ena/browser/home>) bulunmaktadır. Ağustos 2020 itibariyle buraya girilmiş SARS CoV-2 için 13925 sekans verisi bulunmaktadır.

Japonya da Uzak Doğu kökenli üçüncü bir gen bankası veri tabanı daha bulunmaktadır. Japonya'nın gen bankası olan DNA Data Bank of Japan (DDBJ) (<https://www.ddbj.nig.ac.jp/index-e.html>) içinde Ağustos 2020'ye kadar SARS CoV-2 anahtar kelimesiyle 12385 sekans verisi girilmiş durumdadır.

Aslında her üç veri tabanı da ortak bir uluslararası kolaborasiyon olan International Nucleotide Sequence Database Collaboration (INSDC)'ye (<http://www.insdc.org/>) bağlı olarak çalışırlar. Veri tabanlarının ara yüzleri farklı olsa da birbirleri arasında bu konsorsiyum sayesinde bir bağlantı bulunmaktadır. INSDC'ye girilen bilgiler ve deneysel konfigürasyonlarla ilgili bağlamsal bilgilerle zenginleştirilmiş hizalamalar (alignment) ve kurgulamalar (assembly) ile işlevsel açıklamalara kadar, veri ham okuma spektrumunu kapsayacak şekilde çalışır (Karsch-Mizrachi ve ark 2018).

CoV2ID ([http://covid.portugene.com/cgi-bin/COVid\\_home.cgi](http://covid.portugene.com/cgi-bin/COVid_home.cgi)), SARS-CoV-2'nin saptanması ve COVID-19'un tedavisi için moleküler yöntemlerin değerlendirilmesini kolaylaştırmak için oluşturulmuş bir veri tabanıdır. Sadece sekans veri-

leri değil tanı testleri için PCR primerleri de bulunmaktadır. GDCTT (<https://www.genomedetective.com/app/typing-tool/cov/>), bu web aracı sekans sonrası gelen ham datanın nükleotid dizisinin programa yükleme sonrası coronavirus türlerini ve genotiplerini belirlemek için Blast ve filogenetik yöntemleri kullanmak üzere tasarlanmıştır.

### Filogenetik analiz

Filogenetik analizler ve sonrasında oluşturulan ağaçlar (phylogenetic tree), literatürdeki en geleneksel grafik sunum modelidir. Ancak bu yaklaşım, evrimsel hızdaki değişkenlik, virusların fiziksel değişimi ve viruslar ile konakçıları arasındaki evrimsel ilişkileri karıştıran (bir çeşit geçmişe dair kayıtların olmaması) ciddi zorluktur. Ancak yine de literatürlerde çok çeşitli filogenetik analiz yapmaya yarayan araç ve yazılım bulunmaktadır. Bu programlar çeşitli parametreler (Kimura vb.) ve farklı metotlar (maximum likelihood, neighbor-joining, UPGMA, maximum parsimony, bayesian, distance matrix vb.) kullanarak fiogenetik ağaçları oluştururlar. MEGA (Molecular Evolutionary Genetics Analysis) programı, bilindiği kadarıyla dünyada birçok çalışmada (Timurkan ve ark 2018, Timurkan ve Alcigir 2017) kullanılmış bir analiz programıdır. Bunun yanında Armadillo Workflow Platform, Geneious, MrBayes, T-REX, Phangorn ve RAXML-NG vb. daha birçok filogenetik ağaç oluşturabilecek bilgisayar programları bulunmaktadır (Tablo 1).

### Epidemiyolojik veriler

COVID-19 Disease Map, (<https://covid.pages.uni.lu/>) bu veri tabanında popülasyon odaklı geniş bir perspektifte COVID-19'un moleküler mekanizmalarının bir bilgi havuzu oluşturulması hedeflenmiştir. COVID-19 Disease Map aslında literatür kanıtlara dayalı olarak oluşturulmuş moleküler etkileşim diyagramlarının bir derlemesi niteliğindedir. Ayrıca SARS-CoV-2'ye özgü konakçı-patojen etkileşimlerine de odaklanmıştır.

CoV-Hipathia (<http://hipathia.babelomics.org/covid19/>), bu web aracı, SARS CoV-2 tarafından enfeksiyona dahil olduğu bilinen sinyal yolları bağlamında gen ekspresyon seviyelerindeki ve/veya genomik mutasyonlardaki birleşik değişikliklerin sonuçlarının yorumlanması için bilgisayar tabanlı (in-slico) bir insan sinyalleme modeli oluşturur. Dolayısıyla temelde bilgisayar tabanlı olarak hastalığın patogenezi hakkında bilgi vermektedir.

Breaking-Cas (<http://www.clinbioinfospa.es/CovidResources>), günümüzde yeni bir teknoloji olan CRISPR tabanlı genom düzenleme için oligo kılavuz tasarım yapabilen bir veri tabanıdır. Bu sitede SARS CoV-2 genomları ile BioinfoGP CNB'de ENSEMBL genomları için CRISPR-Cas denemeleri için bir çeşit kılavuz RNA'ların etkileşimli tasarımı yapılabilmektedir.





CDC (Centers for Disease Control and Prevention, <https://www.cdc.gov/>), Hastalık Kontrol ve Önleme Merkezi olarak tabiki bu pandemide gerekli önlemleri bildirerek ve anlık açıklamalarla dünya sağlık örgütüyle çalışmışlardır. Tüm bilgileri kendi sitelerinin haricinde özellikle epidemiyolojik veriler ile tanı algoritmaları ve moleküler virolojik bilgileri <https://github.com/CDCgov/SARS-CoV-2-Sequencing> gibi platformlarda da sunmuşlardır.

ELIXIR (<https://elixir-europe.org/services/covid-19>), Avrupa yaşam bilimleri için araştırma altyapısı olan ELIXIR, Avrupa SARS CoV-2 araştırması üzerinde çalışan araştırmacılar ve konsorsiyumlar tarafından kullanılabilir bir dizi hizmet sunmaktadır. Bir epidemiyolojik veri bankası gibi düşüneneğimiz ELIXIR'in üç hedefi bulunmaktadır. Bunlardan birincisi Avrupa ülkelerinde COVID-19 veri bankası oluşturmak için Avrupa genelinde COVID-19 veri platformlarını birbirine bağlamak; ikincisi COVID-19 verilerini açık, doğru ve uzun vadede yeniden kullanılabilir hale getirmek için iyi bir veri yönetimi teşvik etmek ve üçüncüsü de tekrarlanabilir ve iş birliğine dayalı bilimi teşvik etmek için online web araçları, çeşitli programlar ve epidemiyolojik hesaplama kaynaklarını sağlamayı hedeflemiştir (Tablo 1).

### Tüm genom kurgulama

Tüm genom okuma ve kurgulama için birçok araç geliştirilmiştir. Örneğin; Velvet, (Zerbino ve Birney 2008) ve ABySS (Simpson ve ark 2009) vb. Ancak bu araçlar, viral UTR (untranslate region) bölgelerindeki tekrarlar ve az okuma veya düzensiz okuma bölgeleri içinde tüm genomlar için kullanılamaz. Bunların yerine SPAdes, IDBA-UD, VICUNA ve VrAP vb. programlar kullanılabilir. Bu konu komple genom analizlerinin artmasıyla daha popüler hale daha ileri yıllarda gelecektir.

### Öneriler

Biyoinformatik konusu çağın konusu olarak karşımıza çıkmaktadır. Burada SARS CoV-2 özelinde virusların biyoinformatiği ile ilgili güncel bilgiler paylaşıldı. Ancak viruslardaki sekonder yapıların analizleri, virus evrimi vb. daha birçok konu da bu bağlamda bu derlemenin konularına eklenebilir niteliktedir. Ancak SARS CoV-2 özelinde yeni bir hastalık olan COVID19 için ilerleyen zamanlarda yeni algoritmalarla bu konu daha iyi aydınlatılabilir.

Virus biyoinformatiğinin geleceği, hızlı ve spesifik biyoinformatik yazılım geliştirmeye, bazen virusa özgü veri tabanlarının ve araçlarının kurulmasına ve ortak disiplinler arası (genetik, moleküler biyoloji, enformatik vb.) araştırma projelerinin oluşturulmasına bağlıdır. Bunların yanında bu web ara yüzleri ve programların çalıştırılması ve kullanılmasında araştırmacıların tek bir programdan ziyade farklı program-

larla bu durumu irdelemesi daha doğru olacaktır. Çünkü biyolojik bir durumun mekanik diyebileceğimiz bir bilgisayar alt yapısında tek başına doğru sonuç veremediği çoğu kez bildirilmiştir. Nitekim en çok biyolojinin bilgisayara yansımaları olarak kullandığımız istatistik programlarda ve analizlerde bunu görmekteyiz. Bu noktada tek sorumluluk araştırmacılara düşmemektedir. Araştırmacılar birleşerek toplulukları ve konsorsiyumları ve bu birlikteliklerde ülkeleri bu paydaya dahil edip güçlü biyoinformasyon birimleri kurulabilir.

SARS CoV-2 özelinde hali hazırda var olan birçok veri tabanı ve biyoinformatik araçlar, pandemi ve sonrasındaki süreçte çok hızlı bir şekilde bu duruma adapte olarak araştırmacılara katkı sağlamışlardır. Nitekim insanları etkileyen bir enfeksiyon olması sebebiyle de hem tanı algoritmaları hızlı bir şekilde geliştirilmiş hem de sonuçların yorumlanması kolaylıkla olmuştur. Ancak yine de unutulmamalıdır ki etken viral bir ajan olduğundan virusları da en iyi tanıyan, bilen ve onlarla direkt temasta olan viroloji alanındaki uzmanlar olduğundan virus ile çalışmak isteyen her araştırmacının viroloji uzmanından yardım alması gerekmektedir.

### Çıkar Çatışması

Yazarlar herhangi bir çıkar çatışması bildirmemiştir.

### Finansal Kaynak

Bu çalışma sırasında, yapılan araştırma konusu ile ilgili doğrudan bağlantısı bulunan herhangi bir ilaç firmasından, tıbbi alet, gereç ve malzeme sağlayan ve/veya üreten bir firma veya herhangi bir ticari firmadan, çalışmanın değerlendirme sürecinde, çalışma ile ilgili verilecek kararı olumsuz etkileyecek maddi ve/veya manevi herhangi bir destek alınmamıştır.

### Kaynaklar

- Andrade ACDSP, de Miranda Boratto PV, Rodrigues RAL, Bastos TM, et al., 2019. New isolates of pandoraviruses: contribution to the study of replication cycle steps. *J Virol*, 93(5), e01942-18.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, et al., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput Biol*, 19 (5), 455-477.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi, I, et al., 2012. GenBank. *Nucleic acids research*, 41(1), 36-42.
- Blomberg N, Lauer KB, 2020. Connecting data, tools and people across Europe: ELIXIR's response to the COVID-19 pandemic. *Eur J Hum Genet*, 28, 719-723
- Boc A, Diallo AB, Makarenkov V, 2012. T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic acids res*, 40(1), 573-579.





- Briester JR, Ako-Adjei D, Bao Y, Blinkova O, 2015. NCBI viral genomes resource. *Nucleic acids res*, 43(1), 571-577.
- Carneiro J, Gomes C, Couto C, Pereira F, 2020. CoV2ID: Detection and Therapeutics Oligo Database for SARS-CoV-2. *bioRxiv*, In press.
- Cleemput S, Dumon W, Fonseca V, Abdool Karim W, et al., 2020. Genome Detective Coronavirus Typing Tool for rapid identification and characterization of novel coronavirus genomes. *Bioinformatics*, 36(11), 3552-3555.
- Dhama K, Khan S, Tiwari R, Sircar S, et al., 2020. Coronavirus disease 2019 – COVID-19. *Clin Microbiol Rev*, 33, e00028-20.
- Hadfield J, Megill C, Bell SM, Huddleston J, et al., 2018. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*, 34(23), 4121-4123.
- Hatcher EL, Zhdanov SA, Bao Y, Blinkova O, et al., 2017. Virus Variation Resource–improved response to emergent viral outbreaks. *Nucleic acids res*, 45(1), 482-490.
- Hölzer M, Marz, M, 2017. Software dedicated to virus sequence analysis “bioinformatics goes viral”. *Adv Virus Res*, 99, 233–257.
- Huang C, Wang Y, Li X, Ren L, et al., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*, 395, 497–506.
- Huelsenbeck JP, Ronquist F, 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8), 754-755.
- Hulo C, de Castro E, Masson P, Bougueleret L, et al., 2011. ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res*. 39, 576–582.
- Ibrahim B, McMahon DP, Hufsky F, Beer M, et al., 2018. A new era of virus bioinformatics. *Virus Res*, 251, 86-90.
- Jerome H, Vattipally SB, Thomson EC, 2015. Can we identify potential viral zoonoses before they cross the species barrier? *Microbiology Today*, 42, 150-153.
- Karsch-Mizrachi I, Takagi T, Cochrane G, 2018. The international nucleotide sequence database collaboration, *Nucleic Acids Res*, 46(1), 48–51.
- Kearse M, Moir R, Wilson A, Stones-Havas S, et al., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647-1649.
- Kirk MD, Pires SM, Black RE, Caipo M, et al., 2015. World Health Organization estimates of the global and regional disease burden of 22 foodborne bacterial, protozoal, and viral diseases, 2010: a data synthesis. *Plos Med*, 12, e1001921.
- Knight-Jones T, Rushton J, 2013. The economic impacts of foot and mouth disease – what are they, how big are they and where do they occur? *Prev Vet Med*, 112(3–4), 161–173.
- Kozlov AM, Darriba D, Flouri T, Morel B, et al., 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, 35(21), 4453-4455.
- Kulikova T, Aldebert P, Althorpe N, Baker W, et al., 2004. The EMBL nucleotide sequence database. *Nucleic acids res*, 32(1), 27-30.
- Kumar S, Stecher G, Li M, Knyaz C, et al., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*, 35(6), 1547-1549.
- Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, et al., 2018. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic acids res*, 46(1), 708-717.
- Lord E, Leclercq M, Boc A, Diallo AB, et al., 2012. Armadillo 1.1: an original workflow platform for designing and conducting phylogenetic analysis and simulations. *Plos one*, 7(1), e29903.
- Lu R, Zhao X, Li J, Niu P, et al., 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*, 395, 565–574.
- Marz M, Beerwinkel N, Drosten C, Fricke M, et al., 2014. Challenges in RNA virus bioinformatics. *Bioinformatics*, 30(13), 1793-1799.
- Ogasawara O, Kodama Y, Mashima J, Kosuge T, et al., 2020. DDBJ Database updates and computational infrastructure enhancement. *Nucleic acids res*, 48(1), 45-50.
- Oguzoglu TC, Timurkan MO, Muz D, Kudu A, et al., 2010. First molecular characterization of feline immunodeficiency virus in Turkey. *Arch Virol*, 155(11), 1877-1881.
- Oliveros JC, Franch M, Tabas-Madrid D, San-León D, et al., 2016. Breaking-Cas - interactive design of guide RNAs for CRISPR - Cas experiments for ENSEMBL genomes. *Nucleic acids res*, 44(1), 267-271.
- Ostaszewski M, Mazein A, Gillespie ME., Kuperstein I, et al., 2020. COVID-19 Disease Map, building a computational repository of SARS-CoV-2 virus-host interaction mechanisms. *Sci data*, 7(1), 1-4.
- Peng Y, Leung, H.C.M., Yiu, S., Chin, F.Y.L., 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28 (11), 1420–1428.
- Pickett BE, Sadat EL, Zhang Y, Noronha JM, et al., 2012. ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic acids res*, 40(1), 593-598.
- Rian K, Esteban-Medina M, Hidalgo MR, Cubuk C, et al., 2020. Mechanistic modeling of the SARS-CoV-2 disease map. *bioRxiv*, In press.
- Schliep KP, 2011. phangorn: phylogenetic analysis in R. *Bioinformatics*, 27(4), 592-593.
- Shu Y, McCauley J, 2017. GISAID: Global initiative on sharing all influenza data—from vision to reality. *Eurosurveillance*, 22(13), 30494.
- Simpson JT, Wong K, Jackman SD, Schein JE, et al., 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res*, 19, 1117–1123.
- Suttle CA, 2005. Viruses in the sea. *Nature*, 437, 356–361.
- Timurkan MO, Alcigir ME, 2017. Phylogenetic analysis of a partial L1 gene from bovine papillomavirus type 1 isolated from naturally occurring papilloma cases in the northwestern region of Turkey. *Onderstepoort J Vet Res*, 84(1), 1-6.







- Timurkan MO, Aydın H, Alkan F, 2018. Detection and molecular characterization of canine adenovirus type 2 (CAV-2) in dogs with respiratory tract symptoms in shelters in Turkey. *Vet arhiv*, 88(4), 467-479.
- Timurkan MO, Aydın H, Sait A, 2019. Identification and molecular characterisation of bovine parainfluenza virus-3 and bovine respiratory syncytial virus-first report from Turkey. *J Vet Res*, 63(2), 167-173.
- Timurkan MÖ, Aydın H, 2019. Increased genetic diversity of BVDV strains circulating in Eastern Anatolia, Turkey: first detection of BVDV-3 in Turkey. *Trop Anim Health Prod*, 51(7), 1953-1961.
- Weiss RA, McMichael AJ, 2004. Social and environmental risk factors in the emergence of infectious diseases. *Nat Med*, 10(12), 70-76.
- Woo PC, Huang Y, Lau SK, Yuen KY, 2010. Coronavirus genomics and bioinformatics analysis. *Viruses*, 2(8), 1804-1820.
- World Health Organization (WHO), 2020. Novel Coronavirus (COVID-19) Situational Reports. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situational-reports/>. Accessed: 10.08.2020.
- Yang X, Charlebois P, Gnerre S, Coole MG, et al., 2012. De novo assembly of highly diverse viral populations. *BMC Genomics*, 13, 475.
- Zerbino DR, Birney E, 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*, 18, 821-829.
- Zhou P, Yang XL, Wang XG, Hu B, et al., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579, 270-273.

#### **Yazar Katkıları**

Fikir/Kavram: Mehmet Özkan Timurkan  
Tasarım: Mehmet Özkan Timurkan, Hakan Aydın  
Denetleme/Danışmanlık: Mehmet Özkan Timurkan, Hakan Aydın  
Veri Toplama ve/veya İşleme: Mehmet Özkan Timurkan  
Analiz ve/veya Yorum: Mehmet Özkan Timurkan, Hakan Aydın  
Kaynak Taraması: Mehmet Özkan Timurkan, Hakan Aydın  
Makalenin Yazımı: Mehmet Özkan Timurkan  
Eleştirel İnceleme: Mehmet Özkan Timurkan, Hakan Aydın

**CITE THIS ARTICLE:** Timurkan MÖ, Aydın H, 2020. SARS CoV-2'nin (COVID-19) biyoinformatik destekli idenfikasyonu. *Eurasian J Vet Sci, Covid-19 Special Issue*, 50-58

