

Nükleotid Dizilerinin Aminoasit Formatına Dönüştürülmesi ve Dünya Veri Tabanlarındaki Verilerle Karşılaştırılması

Atila Taner KALAYCIOĞLU *, ** 

* Kafkas Üniversitesi, Veteriner Fakültesi, Mikrobiyoloji Anabilim Dalı, TR-3600 Kars - TÜRKİYE

** Türkiye Halk Sağlığı Kurumu, Moleküler Mikrobiyoloji Araştırma ve Uygulama Laboratuvarı, TR-06100 Ankara - TÜRKİYE

Makale Kodu (Article Code): KVFD-2012-7909

Özet

Moleküler biyoloji ve DNA sekans analizi çalışmalarındaki gelişmeler sonucu elde edilen yeni bilgi ve verilerin hızlı ve hassas bir şekilde değerlendirilmesi ihtiyacı Biyoinformatik adlı bir bilim kullanılması gerekli kılmıştır. Mikroorganizmalardan virüslere ait hedef alınan gen bölgesinin tüm veya kısmi sekans analizleri yardımıyla filogenetik analizler ve genotiplendirme çalışmaları yapılabilmektedir. Viral proteinleri kodlayan gen sekansları protein dizilerine çevrilebilmektedir. Bu sayede belli pozisyonlardaki nükleotid farklılıklarının aminoasit farklılığına da sebep olup olmadığı gibi durumlar değerlendirilebilmektedir. Nükleotid ve aminoasit sekans sonuçları veri tabanlarındaki diğer verilerle mukayese edilip, benzerlik ile ilgili değerlendirmeler yapılabilmektedir. Elde edilen nükleotid ve aminoasit sekans verilerinin veri tabanlarına girilmesi bilgilerin diğer araştırmacılar tarafından kullanılması açısından çok önemlidir.

Anahtar sözcükler: *Biyoinformatik, Veri tabanları, Sekans analizi*

Translation of Nucleotide Sequences to Amino Acids and Their Comparison with Data Stored in Databases

Summary

The advances in the field of molecular biology and DNA sequencing have been leading to obtain a large amount of new data and information. The rapid and precise evaluation of these data has made it imperative to use a new interdisciplinary science called Bioinformatics. Of microorganisms, viruses, whole or partial sequence analysis of the target gene can be used for genotyping and phylogenetic analysis studies. The viral DNA sequences encoding viral protein can be translated to amino acid sequences. In this way, whether or not certain mutation(s) on nucleotide sequences result in amino acid substitutions can be investigated. In addition, the obtained nucleic and amino acid sequence data can be compared with those stored in the databases by means of similarity researches. Thus, it is very important to store sequence results to databases.

Keywords: *Bioinformatics, Databases, Sequence analysis*

GİRİŞ

Moleküler biyoloji ve DNA sekanslama teknolojilerindeki hızlı gelişim sonucu araştırma merkezlerinde devamlı ve büyük miktarda yeni veriler elde edilmektedir. Bu verilerin hızlıca değerlendirilmesi, diğer araştırmacıların erişebilecekleri şekilde depolanması amacıyla uluslararası veri tabanları oluşturulmuştur¹. Veri tabanları, elde edilen bilgi ve verilerin depolandığı, bünyelerinde bulunan yazılımlar sayesinde çeşitli analizlerin yapılabilirdiği, araştırmacıların kullanımına açık programlardır. Günümüzde Biyolojik sistemler ve özellikle sekans analizleri sonucu elde edilen kapsamlı

verileri sağlıklı, doğru biçimde değerlendirme ihtiyacı; biyoloji, biyokimya, tıbbi bilimler, bilişim bilimleri, matematik ve istatistik gibi alanların işbirliği Biyoinformatik adı verilen yeni ve uygulamalı bir bilimin doğmasına sebep olmuştur^{2,3}.

Biyoinformatik, Amerikan Ulusal Biyoteknoloji Bilgi Merkezi (NCBI) tarafından biyoloji, bilgisayar bilimi ve bilgi teknolojilerinin birlikte kullanıldığı bir bilim dalı olarak tanımlanmaktadır⁴.



İletişim (Correspondence)



+90 505 7513858



atakal61@hotmail.com

Bu derlemede, veri tabanları, nükleik asit sekanslarının aminoasit formatına çevrilmesi, dünya veri tabanlarına girilmesi ve veri tabanlarındakilerle karşılaştırılması uygulamaları açıklanacaktır.

VERİ TABANLARI

Dünyada çeşitli canlı ve organizmalara ait nükleotid dizi bilgilerinin organizasyon ve depolama işlemini üstlenen üç veri tabanı vardır. Bunları Amerika Birleşik Devletleri (ABD) merkezli 'NCBI-GenBank' (Amerikan Ulusal Biyoteknoloji Bilgi Merkezi (Gen Bankası; ABD-Maryland) ^{5,6}, Avrupa orijinli 'EMBL' (Avrupa Moleküler Biyoloji Laboratuvarı; İngiltere-Hinxton) ⁷ ve Japonya merkezli 'DDJB' (Japonya veri tabanı; Japonya-Mishima) ⁸ oluşturmaktadır. Bu üç veri tabanı 'International Sequence Database Collaboration (INSDC) olarak işbirliği halinde olup, günlük veri alış verişi yoluyla sekansların üniform bir şekilde depolanıp erişime açık hale getirilmesine olanak sağlarlar ⁶.

GenBank ve Kullanımı

GenBank, nükleotide sekanslarının ve sekanslara ait biyografik ve biyolojik bilgilerin depolandığı en önemli veri tabanı özelliğini taşımaktadır. Bünyesinde 380.000'den fazla genus sınıfında veya daha alt sınıfta organizmaya, insan genom projesi gibi geniş kapsamlı büyük projeler, farklı çalışmalardan ve çevresel örnekleme metotları ile yapılan projeler sonucu elde edilen sekans bilgilerini kapsamaktadır. Nisan 2011 tarihi itibarıyla GenBank üzerinde yaklaşık olarak 135.440.924 sekans kaydında 126.551.501.141 nükleotide ve "Whole genome shotgun sequences" (WGS) bölümünde de 62.715.288 sekans kaydında 191.401.393.188 nükleotide ait bilgi bulunmaktadır ⁶.

GenBank Veri Erişimi

GenBank üzerindeki sekans kayıtlarına "NCBI Entrez" erişim sistemi yoluyla erişilebilir ⁹. Bir guruba veya filogenetik analize konu olan sekanslar topluca "Entrez Popset" ve kavramsal olarak translate edilen protein kodlayan sekanslar (CDS) "Entrez Protein" üzerinde depolanır. GenBank üzerine kayıtlı olan her bir veri bilimsel literatürlere "PubMed" veya "PubMed central" vasıtası ile bağlanır. Ayrıca, NCBI tarafından sunulan bir başka hizmet olan OMIM ile genler ve genetik rahatsızlıklar ile ilgili günümüze kadar belirlenmiş olan mutasyon ve ilgili bilgilere de erişilebilir ³.

Veritabanı Organizasyonu

Sekans kayıtları GenBank'da kaynak taksonomisi veya uygulanan sekans stratejisine göre gruplandırılır. Bu amaçla 12 adet taksonomik bölüm (BCT, ENV, INV, MAM, PHG, PLN, PRI, ROD, SYN, UNA, VRL, VRT) ve altı yüksek verimlilikte bölümler (EST, GSS, HTC, HTG, STS, TSA) bulunmaktadır. Bunlara ilave olarak, PAT bölümü patent ofislerince sağlanan verileri ve WGS bölümü de tüm genom projelerine ait sekansları bulundurmaktadır. Komple genomlar (www.ncbi.

nlm.nih.gov/Genomes) halen GenBank'ın hızlı gelişen bölümünü oluşturmaktadır ^{5,6}.

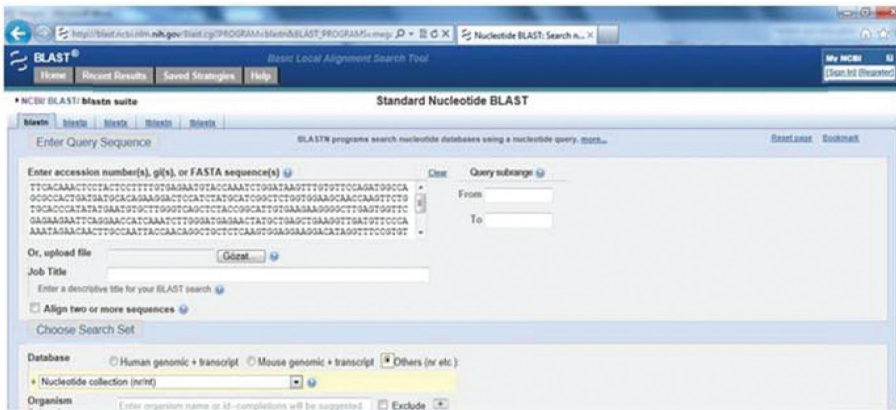
NÜKLEOTİD DİZİLERİNİN AMİNOASİT FORMATINA ÇEVİLMESİ

DNA dizilerinin protein kodlayan bölümleri ilk olarak Messenger RNA (mRNA) dönüştürüldükten sonra proteini oluşturan aminoasit dizisine çevrilir. Bu aşamada translasyonun başlatılıp, sonlandırıldığı açık okuma alanının 'Open Reading frame' (ORF) belirlenmesi önemlidir. ORF, DNA molekülünden bir parça olup, kodladığı proteine ait aminoasitlere dönüştürüldüğünde stop kodon içermez ¹⁰. Bir viral gen dizisine ait ORF'in belirlenmesi o genin kodladığı protein esas alınarak yapılacak antiviral veya aşı geliştirme çalışmaları açısından önemlidir ¹¹.

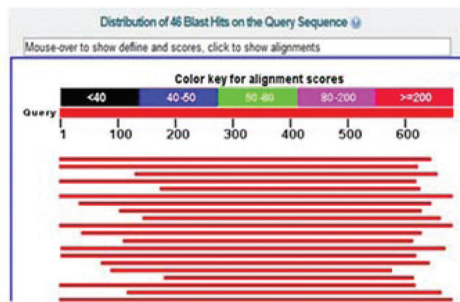
Uygulamada bir virüse ait herhangi bir genin kısmi veya tümü için sekans analizi yapıldığında doğru ORF belirlenmelidir. Bu amaçla çevrimiçi olarak "Expasy Translate tool" (<http://www.expasy.org/tools/dna.html>) veya "NCBI ORF finder" (<http://www.ncbi.nlm.nih.gov/gorf.html>) programları kullanılabilir. Bir gene ait tüm veya kısmi bir DNA dizisi 5'ileri (forward) ve 3' ters (reverse) yönde üçer adet olmak üzere altı potansiyel ORF'e sahiptir. Bir ORF ATG nükleotidleri ve kodladıkları Metionin (M) ile başlar ve genellikle TAA, TAG veya TGA nükleotidlerinden oluşan bir STOP kodon ile sonlanır.

Örnek olarak Ülkemizde 2002 yılından itibaren özellikle Kelkit vadisi ve çevresinde Kırım Kongo Kanamalı Ateşi (KKKA) hastalığı görülmektedir. Bu hastalığa sebep olan Kırım Kongo Kanamalı Ateş virüsü (KKKAV) izolatlarının S ve/veya M, L gen segmentlerinin sekans analizleri yapılarak, ülkemizde ve dünyada daha önce elde edilmiş sekans verileri ile kıyaslanıp, virüslerin filogenetik analizleri yapılabilmektedir. Elde edilen nükleotid sekansları aminoasit formatına çevrilerek nükleotid dizilerindeki farklılıkların aminoasit farklılıklarına (mutasyon varyasyon) sebep olup olmadığı belirlenebilmektedir ¹². Bu durum virüsün antijenik yapısı ve izolatlar arasındaki antijenik farklılıkların araştırılması açısından önemlidir.

Uygulamada KKKAV izolatlarının S veya M segment bölgeleri polimeraz zincir reaksiyonu (PZR), sekanslanır. Elde edilen ileri (5' forward) ve ters (3' reverse) yönde nükleotide sekans verileri "BioEdit" (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>) veya "Clustal W" (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) gibi programlar kullanılarak karşılaştırılır ^{13,14}. Böylece hatalı olan veya iyi okunmamış bölgeler tespit edilir. Bu bölgeler sekans analiz sonucunu gösteren "Sequencher" (<http://genecodes.com/>) gibi programlar kullanılarak kontrol edilip, hatalar düzeltildikten sonra doğrulanmış sekans verisi elde edilir. Elde edilen sekans verisi 'Expasy translate tool' (<http://web.expasy.org/translate/>) programı ile aminoasit formatına çevrilir.



a



Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
HQ22814.1	Crimean-Congo hemorrhagic fever virus isolate KASTAMONU 229-201	1123	1173	94%	0.0	99%	
HQ22815.1	Crimean-Congo hemorrhagic fever virus isolate SAMSUN 1153-2009 c	1121	1131	91%	0.0	99%	
HQ291206.1	Crimean-Congo hemorrhagic fever virus isolate TOKAT 873-2009 glyc	946	946	77%	0.0	99%	
JX308616.1	Crimean-Congo hemorrhagic fever virus isolate Corum 2020-2011 em	1114	1114	90%	0.0	99%	
JF323543.1	Crimean-Congo hemorrhagic fever virus isolate Kayseri 1761-09 glyci	802	809	66%	0.0	99%	
GQ227034.1	Crimean-Congo hemorrhagic fever virus strain Turkey-Kelkit06 segme	1208	1208	100%	0.0	99%	

b

```
>gb|JX308616.1 Crimean-Congo hemorrhagic fever virus isolate Corum 2020-2011
envelope glycoprotein precursor, gene, partial cds
Length=760
```

```
Score = 976 bits (528), Expect = 0.0
Identities = 536/540 (99%), Gaps = 0/540 (0%)
Strand=Plus/Plus
```

```
Query 1 GGGTGTATACCATCAACAGGGTGAAGTCATTCAAGCTATGCGAAAACAGCGCCACAGGG 60
Sbjct 140 GGGTGTATACCATCAACAGGGTGAAGTCATTCAAGCTATGCGAAAACAGCGCCACAGGG 199

Query 61 AAAACCTGTGAGATAGACAGCACTCCGGTTAAGTGCAGGCAAGGTTTCTGTTAAAAATC 120
Sbjct 200 AAAACCTGTGAGATAGACAGCACTCCGGTTAAGTGCAGGCAAGGTTTCTGTTAAAAATC 259

Query 121 ACCCAAGAGGGGAAGGGGCCAGTAAAATTAATCTAGAGGCTCAGAGGTTGCTCTGGATGCT 180
Sbjct 260 ACCCAAGAGGGGAAGGGGCCAGTAAAATTAATCTAGAGGCTCAGAGGTTGCTCTGGATGCT 319

Query 181 TGGCACTCAAGCTGTGAAGTAATGATACCTAAAGGCAGTGGAGATATCCTAGTGGACTGT 240
Sbjct 320 TGGCACTCAAGCTGTGAAGTAATGATACCTAAAGGCAGTGGAGATATCCTAGTGGACTGT 379

Query 241 TCAGCGGGGAGCAACATTTCTTAAAAGACAACCTAATTGACCTAGGGTGCCCCATATC 300
Sbjct 380 TCAGGTGGGAGCAACATTTCTTAAAAGACAACCTAATTGACCTAGGGTGCCCCATATC 439

Query 301 CCGCTATTGGGTAAGATGGCTATTTACATTTGCAGAATGTCAAATCATCCTAGAACAACC 360
Sbjct 440 CCGCTATTGGGTAAGATGGCTATTTACATTTGCAGAATGTCAAATCATCCTAGAACAACC 499

Query 361 ATGGCTTTTCTCTTCTGGTTCAGCTTTGGCTATGTTATCACTTGCTATTTTGCAGGCT 420
Sbjct 500 ATGGCTTTTCTCTTCTGGTTCAGCTTTGGCTATGTTATCACTTGCTATTTTGCAGGCT 559

Query 421 CTTTTTATTCAATTGATAAATTATGGAACTAGGGAAAAAGATCAAGCAATATAGGGAA 480
Sbjct 560 CTTTTTATTCAATTGATAAATTATGGAACTAGGGAAAAAGATCAAGCAATATAGGGAA 619

Query 481 CTGAACCTCAAACCTGCACCATATGTGAGACAGCTCCTGTCAATGCAATAGATGCTGAA 540
Sbjct 620 CTGAACCTCAAACCTGCACCATATGTGAGACAGCTCCTGTCAATGCAATAGATGCTGAA 679
```

c

Şekil 1. KKKAV Kastamonu 241-2012 izolatı 540 nükleotide uzunluğunda kısmi M segment sekansı BLAST analizi

a) BLAST sayfasına 540 nükleotide uzunluğunda KKKAV Kastamonu 241-2012 izolatı kısmi M segment envelope glycoprotein precursor geni (Gn) sekansın kopyalanması, b) Sonuçların grafik formatında gösterilmesi ve GenBank veri tabanında en yüksek oranda altı benzer eşleşmenin verilmesi, c) BLAST analizi yapılan Kastamonu 241-2012 izolatı (Query) ile 99% oranında sekans homolojisi gösteren Corum 2020-2011 izolatı (Subject) eşleşmesinin gösterilmesi

Fig 1. BLAST analysis of 540 nucleotide long partial M segment sequence of the KKKAV Kastamonu 241-2012 isolate

a) Initial BLAST input page with a 540 nucleotide long KKKAV Kastamonu 241-2012 isolate partial M segment envelope glycoprotein precursor gene (Gn) sequence, b) Results of the search are given in graphic format at the top, along with the first six hits from the GenBank, c) BLAST analysis of Kastamonu 241-2012 isolate (Query), which displayed 99% sequence homology with the isolate Corum 2020-2011 (Subject)

BLAST-NÜKLEOTİDE VE PROTEİN DİZİ EŞLEŞTİRME ANALİZİ

“Basic Local Alingment Search Tool” (BLAST) (<http://blast.ncbi.nlm.nih.gov/>) yazılımı NCBI tarafından geliştirilmiş ve biyoinformatik uygulamalarında araştırmacılar tarafından sıklıkla kullanılmaktadır. BLAST analizi ile hedef gen bölgesine ait kısmi veya tüm sekans dizisi (nükleotid- aminoasit) dünyanın değişik merkezlerinde çeşitli araştırmacılar tarafından GenBank veri tabanına girilen sekanslar ile karşılaştırılır. Aynı ve en yüksek benzerliğe sahip olan eşleşmelerdeki benzerlik oranı yüzde olarak verilir ¹⁵⁻¹⁷. BLAST analizi kullanım kolaylığı ve araştırmacıya hızlıca bilgi vermesi açısından önemlidir. Böylece analiz edilen mikroorganizmaya ait sekans verisiyle genetik analiz veya genotiplendirme çalışmaları yapılabilmektedir. Bu durum özellikle ülkemizde dolaşımda olan KKKAV'lerinin genetik olarak takip edilmesi açısından önemlidir. Örnek olarak, ‘Türkiye Halk Sağlığı Kurumu’ (THSK), Moleküler Mikrobiyoloji Araştırma ve Uygulama Laboratuvarında sekansı yapılan KKKAV Kastamonu 241-2012 izolatının (KSTM241/12) 540 nükleotid uzunluğunda kısmi M segment envelope glycoprotein precursor (Gn) geni nükleotid sekansının BLAST analizi uygulaması *Şekil 1 (a,b,c)*'de gösterilmiştir. Bu uygulamada sekansı değerlendirilen KSTM241/12 izolatın ülkemizde daha önce sekans analizi yapılan izolatlarla yüksek oranda benzerlik gösterdiği belirlenmiştir.

NÜKLEOTİD VE AMİNOASİT DİZİ SEKANSLARININ GenBank VERİ TABANINA GİRİLMESİ

GenBank üzerine verilerin girişi elektronik ortamda çalışmayı yapan araştırmacılar tarafından “BankIt” veya “Sequin” programları kullanılarak yapılmaktadır (www.ncbi.nlm.nih.gov/GenBank/). Bilimsel dergilerin birçoğu gönderilen bir yayının yayınlanması için yayın içeriğinde kullanılan sekansların veri tabanına gönderilmesini şart koşmaktadır. Nükleotid ve aminoasit dizi bilgileri veri tabanlarına; ait oldukları organizma adı, kaynağı, izolasyon yeri, izolasyon tarihi, ait olduğu genin adı, veri tabanına giriş yapan araştırmacıların çalıştığı merkezlerinin adı ve adres bilgileri ile birlikte girilebilmektedir. Her bir gönderi için genelde iki iş günü içinde bir takip numarası verilmekte ve gönderilen bilgiler ve sekanslar veri tabanı tarafından kontrol edilmektedir. Eğer her hangi bir tarih belirtilmiş ise o tarihte veya tarih belirtilmemiş ise veri tabanı tarafından belirlenen bir tarihte veri erişime açılmaktadır. Bu süreçte veri tabanı tarafından veriye ait bilgiler mail ortamında yazara kontrol amacı ile gönderilir. Düzeltme ve ilave bilgiler yine mail ortamında veri tabanına iletilir. Kabul edilen ve erişime açılan her bir veri için bir kayıt numarası (accession number) verilmekte ve bu numara diğer veri tabanları (DDJB ve EMBL) ile paylaşılmaktadır. Kayıt numarası GenBank veri tabanının üyelik hattı kısmında yer alıp, sekans verisinde ve/veya diğer

bilgilerde değişiklikler olsa bile kayıt ömrü boyunca sabit kalmaktadır. Kayıt numarası bilimsel yayınlarda veriyi kaynak gösterme amacıyla da kullanılmaktadır. Sonuç olarak, her GenBank kaydı sekans ve sekansa ait bilgileri içermekte olup kayda özgü erişim numarasına sahiptir. Kabul edilen sekans verisi ile ilgili ilave bilgi ve düzeltmeler veriyi giren araştırmacı tarafından update@ncbi.nlm.nih.gov adresine iletilir ^{5,6}.

GenBank'a ve diğer veri tabanlarına kayıt edilen sekans verilerine ait bilgiler iki bölümden oluşur. İlk bölümü; sekans verisini veritabanına kaydeden araştırmacılara ait bilgiler, laboratuvar adresi, sekansı yapılan organizmaya ve kaynağına ait bilgiler (Strain-izolat adı, ülke, izolasyon kaynağı, izole edilen canlı türü, tarih...), sekans dizisinin başlangıç ve bitiş kodonları, ait olduğu gen ve kodladığı protein ismi gibi bilgileri bulduran sayfa oluşturur. İkinci bölümü; gene ait nükleotid ve aminoasit sekanslarını FASTA formatında içeren kısım oluşturur ^{5,6}.

SONUÇ

Bilimsel gelişmelerdeki baş döndürücü gelişim ve araştırmaların sürekliliği bir yandan birçok problemin çözümüne katkıda bulunurken, diğer yandan da yeni araştırmalar için temel oluşturmaktadır. Elde edilen bilgi ve verilerin ileri düzeyde analiz ihtiyacı Biyoinformatik biliminin önemini ve kullanımını gittikçe artırmaktadır. GenBank, içeriğinde çeşitli canlı ve mikroorganizmalara ait sekans verilerinin depolandığı ve bünyesinde benzerlik analizlerinin yapılabildiği BLAST gibi programı bulduran oldukça sık kullanılan kapsamlı bir veri tabanıdır. Bilginin paylaşılması ve diğer araştırmacıların faydalanması açısından elde edilen sekans verilerinin veri tabanlarına kaydedilmesi büyük önem taşımaktadır.

KAYNAKLAR

- 1. Akçalı A:** Genetik kodların uluslararası paylaşımı. *Türk Hijyen ve Deneysel Biyoloji Dergisi*, 65 (3): 109-110, 2008.
- 2. Karabulut E, Karaoğlu E:** Biyoinformatik ve biyoistatistik. *Hacettepe Tıp Derg*, 41, 162-170, 2010.
- 3. Polat M, Karahan AG:** Multidisipliner yeni bir bilim dalı: Biyoinformatik ve tıpta uygulamaları. *S.D.Ü Tıp Fak Derg*, 16 (3): 41-50, 2009.
- 4. Bioinformatics:** <http://www.ncbi.nlm.nih.gov/About/primer/bioinformatics.html>, Accessed:20/10/2012.
- 5. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW:** GenBank. *Nucleic Acids Res*, 39, 32-37, 2011.
- 6. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW:** GenBank. *Nucleic Acids Res*, 40, 48-53, 2012.
- 7. Leinonen R, Akhtar R, Birney E, Bower L, Cerdano-Tárraga A, Cheng Y, Cleland I, Faruque N, Goodgame N, Gibson R, Hoad G, Jang M, Pakseresht N, Plaister S, Radhakrishnan R, Reddy K, Sobhany S, Ten Hoopen P, Vaughan R, Zalunin V, Cochrane G:** The European Nucleotide Archive. *Nucleic Acids Res*, 39, 28-31, 2010.
- 8. Kaminuma E, Mashima J, Kodama Y, Gojobori T, Ogasawara O, Okubo K, Takagi T, Nakamura Y:** DDBJ launches a new archive database with analytical tools for next-generation sequence data. *Nucleic Acids Res*, 38, 33-38, 2010.

9. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Federhen S, Feolo M, Fingerman IM, Geer LY, Helmberg W, Kapustin Y, Krasnov S, Landsman D, Lipman DJ, Lu Z, Madden TL, Madej T, Maglott DR, Marchler-Bauer A, Miller V, Karsch-Mizrachi I, Ostell J, Panchenko A, Phan L, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Slotta D, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Wang Y, Wilbur WJ, Yaschenko E, Ye J: Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*, 40, 13-25, 2012.
10. Translation and Open Reading Frames: http://bioweb.uwlax.edu/genweb/molecular/seq_anal/translation/translation.html, Accessed: 20/10/2012.
11. Yan Q: Bioinformatics databases and tools in virology research: an overview. *In Silico Biol*, 8 (2): 71-85, 2008.
12. Kalaycioglu AT, Durmaz R, Uyar Y, Unaldi O, Aksekili E, Ozkul A, Korukluoglu G, Ertek, M: Lack of genetic diversity in Crimean-Congo Hemorrhagic Fever viruses in Turkey: Assessment of present and future patterns of disease. *J Med Virol*, 84 (3): 471-478, 2012.
13. Hall T: BioEdit: An important software for molecular biology. *GERF Bulletin of Biosciences*, 2 (1): 60-61, 2011.
14. Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucl Acids Res*, 22, 4673-4680, 1994.
15. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res*, 25 (17): 3389-3402, 1997.
16. Zhang Z, Schäffer AA, Miller W, Madden TL, Lipman DJ, Koonin EV, Altschul SF: Protein sequence similarity searches using patterns as seeds. *Nucleic Acids Res*, 26 (17): 3986-3999, 1998.
17. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk J, McGinnis A, Madden TL: NCBI BLAST: A better web interface. *Nucleic Acids Res*, 36, 5-9, 2008.