

Biyoinformatik Temelleri ve Uygulamaları

Editör: Prof. Dr. Mehmet Cengiz BALOĞLU

3. Baskı





Editör: Prof. Dr. Mehmet Cengiz BALOĞLU

BİYOİNFORMATİK TEMELLERİ VE UYGULAMALARI

ISBN 978-625-7052-75-7

Kitap içeriğinin tüm sorumluluğu yazarlarına aittir.

© 2023, PEGEM AKADEMİ

Bu kitabın basım, yayım ve satış hakları Pegem Akademi Yay. Eğt. Dan. Hizm. Tic. A.Ş.ye aittir. Anılan kuruluşun izni alınmadan kitabın tümü ya da bölümleri, kapak tasarımı; mekanik, elektronik, fotokopi, manyetik kayıt ya da başka yöntemlerle çoğaltılamaz, basılamaz ve dağıtılamaz. Bu kitap, T.C. Kültür ve Turizm Bakanlığı bandrolü ile satılmaktadır. Okuyucularımızın bandrolü olmayan kitaplar hakkında yayınevimize bilgi vermesini ve bandrolsüz yayınları satın almamasını diliyoruz.

Pegem Akademi Yayıncılık, 1998 yılından bugüne uluslararası düzeyde düzenli faaliyet yürüten **uluslararası akademik bir yayınevidir**. Yayımladığı kitaplar; Yükseköğretim Kurulunca tanınan yükseköğretim kurumlarının kataloglarında yer almaktadır. Dünyadaki en büyük çevrimiçi kamu erişim kataloğu olan **WorldCat** ve ayrıca Türkiye’de kurulan **Turcademy.com** tarafından yayınları taranmaktadır, indekslenmektedir. Aynı alanda farklı yazarlara ait 1000’in üzerinde yayını bulunmaktadır. Pegem Akademi Yayınları ile ilgili detaylı bilgilere <http://pegem.net> adresinden ulaşılabilir.

3. Baskı: Haziran 2023, Ankara

Yayın-Proje: Şehriban Türüldür
Dizgi-Grafik Tasarım: Tuğba Kaplan
Kapak Tasarım: Pegem Akademi

Baskı: Ay-bay Kırtasiye İnşaat Gıda Pazarlama ve Ticaret Ltd. Şti.
Çetin Emeç Bulvarı 1314. Cadde No: 37A-B Çankaya/ANKARA
Tel: (0312) 472 58 55

Yayıncı Sertifika No: 51818
Matbaa Sertifika No: 46661

İletişim

Macun Mah. 204. Cad. No: 141/A-33 Yenimahalle/ANKARA
Yayınevi: 0312 430 67 50
Dağıtım: 0312 434 54 24
Hazırlık Kursları: 0312 419 05 60
İnternet: www.pegem.net
E-ileti: pegem@pegem.net
WhatsApp Hattı: 0538 594 92 40



Bu kitap
lkemizin Nobel dl sahibi
ilk bilim insanı olan
Prof. Dr. Aziz SANCAR'a
ithaf edilmiřtir.

ÖN SÖZ

Türkiye’de Genetik, Islah ve İstatistik ilmin babası olarak bildiğimiz Prof. Dr. Orhan DÜZGÜNEŞ başta olmak üzere bu ilim sahasında emeği geçen ve ebedi âleme göçmüş olanları rahmetle ve akademik camia adına minnetle yâd ederek söze başlamak istiyoruz. Akademik camiamıza yeni Orhan DÜZGÜNEŞ’lerin, Şaban KARATAŞ’ların katılmaları dileğimizi de tekrarlıyoruz.

Biyoteknolojik metotların ilmi araştırmalara dâhil olması Türkiye’de diğer birçok ülkeye göre daha geç başlamıştır. Ancak çalışmalarda süratle mesafe alınması, çalışmalara erken başlayan ülkelerle aramızdaki farkı birçok alanda kapatmıştır.

Genetik ilminin baş döndürücü hızla devam eden ilerleyişin; yüksek yapılı canlıların tanınması, vasıflarının daha iyi anlaşılması, islahı ve verimlerinin artırılması çalışmalarına takibi zor bir sürat kattığı inkâr edilemez. Bunun yanı sıra prokaryot canlıların fenotipe bağlı taksonomik sınıflandırma ile yapılan tasnifleri de genetik vasıfları çözüldükçe değişikliğe uğrayabilmektedir. Mesela bazı mikroorganizmaların buldukları tür, cins ve hatta familya isimleri değişebilmiştir.

Hızla gelişen, yeni metot ve araştırmalarla ilerleyen biyolojik bilimler ve onun dalları analiz edilirken; “bu bilim kolları bilgilerinin konuşturulmasını, yorumlanmasını ve hatta taçlandırılmasını” temin eden istatistik bilimi de her alana girmiş ve vazgeçilmez olduğunu hissettirmiştir.

İstatistik ilmi, her bilim dalında olduğu gibi zamana ve çağın gereklerine uygun olarak bilginin teknoloji ile daha iyi kullanılabilmesine imkân sağlayan sistemle ilerleme yoluna girmiştir. Bilgisayar ve onun ağları istatistiğin verimine ve kıymetine çok yeni ilaveler yapmıştır.

Biyolojik araştırma ve geliştirme neticelerinin analizi, değerlendirilmesi ve bilgi-veri bankalarının teşkili ile bunların enformatik ağ vasıtası ile verimli kullanılmasının geçmişi 1990’lı yıllarda başlamıştır. Fakat bu hususta Türkiye olarak bizim çalışmalarımız Dünyadan çok daha yavaş ilerlemektedir. Keşke “bilgiyi anında alabileceğimiz ve eşleştireceğimiz veri tabanlarımız tam teşekküllü ve kullanılabilir olsaydı!”

Elimizdeki kitap bu sahada mühim bir açığı kapatacak ve yeni bir ufuk açacak, “Biyoinformatik” alanında eğitim gören ve araştırma yapanlara ilham kaynağı olacaktır. Büyük bir emeğin mahsulüdür. İnşallah yeni emek mahsulü bilgi, araştırma ve geliştirme çalışmalarının kılavuzu olarak çok daha ileri seviyede kıymet bulacaktır. Talebe, akademisyen ve araştırmacıların istifade ettiğini görmek kitabın yazarları kadar bizi de ziyadesi ile mutlu kılacaktır. Bu vesile ile başta Muhterem

Hocamız Prof. Dr. Orhan KAVUNCU Beyefendi olmak üzere, kitabın tüm yazarlarına ve kitabın müellifi Prof. Dr. Mehmet Cengiz BALOĞLU' na hizmetlerine yenisini ekledikleri ve bu müstesna eseri vücuda getirdikleri için teşekkür ediyoruz. Kendilerine yardım eden Kıymetli mesai arkadaşlarımıza da müteşekkirimiz ve onları da ayrıca tebrik ediyoruz. Eserden gereği gibi faydalanabilecek olan talebe, akademisyen ve araştırmacılarımıza da bu vesile ile selam, sevgi ve saygılar sunarız.

Prof. Dr. Seyit Aydın
Kastamonu Üniversitesi
Genetik ve Biyomühendislik Bölümü

ÖN SÖZ

Biyolojik olayların tabiatındaki belirsizlik (tesadüfîlik), diğer bilim dallarının ilgilendiği olaylardakinden daha fazladır. Biyolojik bilgilerin sayısal olarak ifadesi, başka yararları yanında bu belirsizliği ölçmek için de Mendel'in çalışmalarından itibaren başvurulan bir yol olmuştur.

Bir olayın nasıl sonuçlanacağına ilişkin belirsizliğin ölçüsü, o sonucun ihtimal sayısıdır. Bu sayı bizim olayın o şekilde sonuçlanacağına dair kanaatimizin derecesini ifade eder. O yüzden Mendel'den itibaren ihtimaller teorisi ve buna bağlı olarak istatistik yöntemlerin biyolojide kullanılması yaygınlaşmış ve gelişmiştir.

Biyolojiyle ilgili bilim dallarında kullanılan İstatistik yöntemlerinin önemli bir bölümü, İngiltere'de Sir Ronald Aylmer Fisher'in (1890-1962) Rothamsted Zirai Araştırma İstasyonunda başlattığı çalışmalarla geliştirilmiştir. Fisher, faktöriyel deneme düzeninin (Factorial Design of Experiments) adını, genlerden aldığı ifade etmektedir. Yirminci yüzyılın başlarına kadar genler, Mendel'in kullandığı tabirle, "**Kalıtım Faktörleri**" olarak ifade ediliyordu. Fisher de birçok lokusa dağılmış olan kalıtım faktörlerinin, mesela verim üzerine etkisini modellemek üzere faktöriyel (çok faktörlü) deneme düzeni modelini geliştirmiştir.

1950'li yılların başından itibaren bir yandan bugünkü genetik bilgilerimizin merkezinde yer alan DNA'nın yapısı ve replikasyonu ile ilgili bilgilerin (Watson & Crick modeli) elde edilmesi, diğer yandan Sibernetik alanındaki gelişmeler sayesinde hesaplamaların bilgisayarda çok daha hızlı ve güvenilir olarak yapılabile hale gelmesi iki önemli dönüm noktasıdır. Bundan sonraki, bilhassa yaklaşık 30 yıl içinde ve sonrasındaki gelişmelerin hızı gerçekten baş döndürücüdür. İnsan Genomu projesinin başlaması ve sonuçlandırılması, sonra birçok Prokaryot ve Ökaryot canlının genom dizisinin bulunması 1980'den bu tarafa gözlenen gelişmelerdir. Görülüyor ki, canlıların genetik yapılarının bilinmesi ve iyileştirilmesi çalışmalarında, temel genetik ve istatistik bilgilere dayanan geleneksel yöntemlerle birlikte, bilişim (informatik) teknolojilerinin çok önemli bir yeri vardır. Bu üç bilgi havuzunun birlikte kullanılmasıyla yeni bir bilim dalı ortaya çıkmıştır: **Biyoinformatik**. Bugün artık DNA, RNA, enzim ve diğer yapı proteinleri düzeyinde elde edilen verilerin bilgisayar ortamında saklanması, analizi, değerlendirilmesi, karşılaştırılması ve erişilmesinde biyoinformatik yöntemlerin kullanılması kaçınılmaz hale gelmiştir.

Ülkemizde Biyoinformatik alanında çalışmalar göreceli olarak yenidir. DNA, RNA ve protein dizilerine ilişkin Biyolojik verilerin elde edilmesi, saklanması, işlenmesi, analiz edilmesi ve birbirleriyle mukayese edilmesi için kat etmemiz gereken daha çok mesafe vardır. Ülkemize özgü canlı materyale ilişkin veri tabanlarının oluşturulması, başka ülkelerdeki veri tabanlarıyla karşılıklı erişim imkânlarının geliştirilmesi için, "**Biyoinformatik**" kelimesinin ilk defa kullanılmaya başlandığı 1970'li yıllardan bu tarafa ABD; İngiltere, Almanya ve Japonya'da gözlenen gelişmeler ve oluşturulan veri tabanlarının benzerleri henüz bizde yoktur.

Bununla birlikte kavram olarak kullanılmasa bile biyoinformatik disiplini için gerekli çalışmalar ülkemizde de 1960'lı yılların ikinci yarısından itibaren yapılmaya başlanmıştır. Ankara Üniversitesi Ziraat Fakültesinde o yıllarda rahmetli Prof. Dr. Orhan Düzgüneş'in şahsi gayretleriyle Ziraat Genetik ve İstatistik Kürsüsü (şimdiki adı Biyometri ve Genetik Anabilim Dalı) kurulmuştur. Kürsü, bitki ve hayvan materyalinin genetik yapılarının ıslahı için gerekli olan temel Genetik ve İstatistik bilgileri yanında, biyolojik bilimlerde bilgisayar kullanımının gerekliliğini de fark ederek 1970'lerden itibaren Bilgisayar Programlama dersleri de vermeye başlamıştı. 1977 yılında tamamladığım doktora tezim "**Populasyonların Genetik Yapılarında Seleksiyonla Meydana Gelecek Değişmelerin Çeşitli Faktörlere Bağlılığının Bilgisayarda Simülasyon Yöntemi ile Araştırılması**" başlığını taşıyordu. Bana bu çalışmamda yol gösteren, fikir veren hocam tez danışmanım Prof. Dr. Tahsin Kesici'ydı. Daha sonraki yıllarda uygulamalı istatistik alanında, analitik olarak irdelenmesi mümkün olsa bile deneysel olarak da sınanması yararlı olan birçok ön şart ve varsayımla ilgili araştırma bilgisayarda benzetim (simülasyon) yöntemi ile yapıldı. Halen Biyometri ve Genetik Anabilim Dalında yapılan Genetik çalışmalarda da DNA, RNA dizileri seviyesindeki veriler, çeşitli yazılımlar ve ara yüzler yoluyla bilgisayar ortamında bir veri tabanı olarak kullanılır hale getirilmekte ve evrensel bilgi birikimine katkı sağlanması amacıyla saklanmaktadır.

Benzer çalışmaların yapıldığı akademik birimlerin sayısı bugün oldukça artmış olmakla birlikte henüz yeterli sayı ve seviyede değildir. Halen üniversitemizde Ziraat Fakültelerinde Biyometri ve Genetik yanında Tarımsal Biyoteknoloji Anabilim Dalları, Tıp Fakültelerinde Biyoistatistik, Moleküler Biyoloji ve Tıbbi Genetik birimleri, Fen Edebiyat Fakültelerinin Biyoloji Bölümlerinde Biyoteknoloji, Hesaplamalı Biyoloji Anabilim Dalları, Mühendislik Fakültelerinde Genetik ve Biyomühendislik Bölümleri ve benzerleri bulunmaktadır. Bütün bunlar arasında, biyolojik veri tabanları oluşturulması, başka ülkelerdeki veri tabanlarıyla karşılıklı erişiminin sağlanması ve bu verilerin analizi için gerekli biyoinformatik çalışmaların yapıldığı müstakil birimlere ihtiyaç vardır.

Elinizdeki kitap, öyle ümit ediyorum ki, bu ihtiyacın daha iyi anlaşılması ve akabinde harekete geçilmesi için öncü bir rol oynayacaktır. Kitabın hazırlanmasında emeği geçen herkese, kitabın on bölümüne bilgi birikimlerini interaktif bir şekilde ve ciddi gayretlerle aktaran genç bilim adamlarına, editör M. Cengiz Baloğlu'nun şahsında teşekkür ediyorum.

Prof. Dr. Orhan Kavuncu

Kastamonu Üniversitesi

Genetik ve Biyomühendislik Bölümü

Bölümler ve Yazarları

Editör: Prof. Dr. Mehmet Cengiz BALOĞLU

1. Bölüm: Biyoinformatiğe Giriş

Öğr. Gör. Necdet Mehmet ÜNEL - Kastamonu Üniversitesi

ORCID No: 0000-0002-7522-9278

İlteriş Kutluğ ÖNAL - Kastamonu Üniversitesi

Öğr. Gör. Fadime ÇETİN - Manisa Celal Bayar Üniversitesi

ORCID No: 0000-0002-5511-3008

Merve KELEŞ - Kastamonu Üniversitesi

Prof. Dr. Murat PEKTAŞ - Kastamonu Üniversitesi

ORCID No: 0000-0002-7205-6279

2. Bölüm: Biyolojik Veri Tabanları

Doç. Dr. Yasemin ÇELİK ALTUNOĞLU - Kastamonu Üniversitesi

ORCID No: 0000-0003-2940-7464

Dr. Öğr. Üyesi Kevser Betül CEYLAN - Bartın Üniversitesi

ORCID No: 0000-0002-6006-727X

Dr. Öğr. Üyesi Yusuf CEYLAN - Bartın Üniversitesi

ORCID No: 0000-0001-81867-252

Dr. Öğr. Üyesi Şerife YERLİKAYA - İstanbul Medipol Üniversitesi

ORCID No: 0000-0002-2146-9495

Prof. Dr. Mehmet Cengiz BALOĞLU - Kastamonu Üniversitesi

ORCID No: 0000-0003-2976-7224

3. Bölüm: Veri Tabanları Kullanılarak Dizi Karşılaştırılması

Dr. Öğr. Üyesi Abdulhamit BATTAL - Van Yüzyüncü Yıl Üniversitesi

ORCID No: 0000-0001-6098-3908

4. Bölüm: Gen ve Promotör Bölgelerinin Analizi

Prof. Dr. Gökhan SADİ - Karamanoğlu Mehmetbey Üniversitesi

ORCID No: 0000-0002-6422-1203

Prof. Dr. Özlem DARCANSOY İŞERİ - Başkent Üniversitesi

5. Bölüm: Protein Yapı Analizleri

Prof. Dr. Nursel AÇAR SELÇUKİ - Ege Üniversitesi

ORCID No: 0000-0001-9292-0637

Dr. Ersin GÜNDEĞER - Ege Üniversitesi

Prof. Dr. Cenk SELÇUKİ - Ege Üniversitesi

ORCID No: 0000-0003-3092-2797

6. Bölüm: Moleküler Filogenetik Analizler

Doç. Dr. Aslıhan KURT KIZILDOĞAN - Ondokuz Mayıs Üniversitesi
ORCID No: 0000-0002-9323-0993

7. Bölüm: Biyolojik Ağlar

Dr. Öğr. Üyesi Ayten Kübra YAĞIZ - Niğde Ömer Halisdemir Üniversitesi
ORCID No: 0000-0001-9870-2702
Arş. Gör. Caner YAVUZ - Niğde Ömer Halisdemir Üniversitesi
ORCID No: 0000-0003-4427-1227
Dr. Öğr. Üyesi Emre AKSOY - Orta Doğu Teknik Üniversitesi
ORCID No: 0000-0002-9410-2715

8. Bölüm: Genom Projeleri ve Gen Ailelerinin Biyoinformatik Analizleri

Öğr. Gör. Necdet Mehmet ÜNEL - Kastamonu Üniversitesi
ORCID No: 0000-0002-7522-9278
Dr. Öğr. Üyesi Aslı UĞURLU - Kastamonu Üniversitesi
ORCID No: 0000-0003-2131-2823
Dr. Öğr. Üyesi Esra Nurten YER - Kastamonu Üniversitesi
ORCID No: 0000-0002-6368-3916
Dr. Öğr. Üyesi Ferhat ULU - Kastamonu Üniversitesi
ORCID No: 0000-0002-9407-9084
Öğr. Gör. Tevfik Hasan CAN - Manisa Celal Bayar Üniversitesi
ORCID No: 0000-0001-8125-4093
Doç. Dr. Yasemin ÇELİK ALTUNOĞLU - Kastamonu Üniversitesi
ORCID No: 0000-0003-2940-7464
Prof. Dr. Mehmet Cengiz BALOĞLU - Kastamonu Üniversitesi
ORCID No: 0000-0003-2976-7224

9. Bölüm: Gen İfade Analizlerinde Biyoinformatik Yöntemler

Doç. Dr. Musa KAVAS - Ondokuz Mayıs Üniversitesi
ORCID No: 0000-0001-5903-2873

10. Bölüm: Biyoinformatiğin Gelecekteki Yeri ve Eğitimi

Dr. Öğr. Üyesi Bahattin Deniz ALTUNOĞLU - Kastamonu Üniversitesi
ORCID No: 0000-0001-7609-1883
Uzm. Biyolog Pınar BALOĞLU - Kastamonu Üniversitesi
Prof. Dr. Orhan KAVUNCU - Kastamonu Üniversitesi
ORCID No: 0000-0003-4391-9087

İÇİNDEKİLER

Ön Söz.....	iv
Ön Söz.....	vi
Bölmeler ve Yazarları.....	viii

1. BÖLÜM

BİYOİNFORMATİĞE GİRİŞ

1.1. Biyoinformatiğin Tanımı ve Amacı	2
1.2. Biyoinformatiğin Tarihsel Gelişimi	5
1.3. Biyoinformatiğin Uygulama Alanları.....	9
1.3.1. Sağlık Alanında Biyoinformatiğin Uygulama Alanları.....	10
1.3.2. Mikrobiyal Genomların Biyoinformatik Uygulamaları.....	12
1.3.3. Tarımsal Alanda Biyoinformatik Uygulama Alanları	14
Kaynaklar.....	16

2. BÖLÜM

BİYOLOJİK VERİ TABANLARI

2.1. Birincil Veri Tabanları	21
2.2. İkincil Veri Tabanları.....	35
2.3. Özelleşmiş Veri Tabanları	37
2.3.1. TAIR	38
2.3.2. Yeast Genome.....	40
2.3.3. FlyBase.....	41
2.3.4. EcoCyc.....	42
2.3.5. Genome CRISPR ve WGE.....	42
Kaynaklar.....	43

3. BÖLÜM

VERİ TABANLARI KULLANILARAK DİZİ KARŞILAŞTIRILMASI

3.1. İkili Dizi Karşılaştırma	46
3.1.1. BLAST Programı Ana Unsurları.....	46
3.1.2. BLAST Sonuç Sayfasına Genel Bakış	47
3.1.3. Nükleotit BLAST (BLASTn).....	53
3.1.4. Protein BLAST (BLASTp).....	55
3.1.5. Nükleotit Dizisinin Protein Dizisiyle Hizalanması (BLASTx)	58

3.1.6. Amino Asit Dizisinin Kullanılarak Nükleotit Dizilerinin Hizalanması (tBLASTn).....	58
3.1.7. Çoklu Dizi Karşılaştırma	58
3.2. SANSparalel Arayüzü	59
3.3. Sonuç.....	63
Kaynaklar.....	63

4. BÖLÜM

GEN VE PROMOTÖR BÖLGELERİNİN ANALİZİ

4.1. Genomik Dizilerin Analizleri.....	66
4.1.1. Genom Anotasyonu.....	68
4.1.2. Gen Tahmininde Kullanılan Yöntemler.....	69
4.2. Promotör ve Düzenleyici Bölgelerin Tahminlemesi	83
4.2.1. Prokaryotik Promotör ve Düzenleyici Bölgeler	85
4.2.2. Ökaryotik Promotör ve Düzenleyici Bölgeler	86
4.2.3. Tahminleme Algoritmaları ve Yaklaşımlar.....	87
Kaynaklar.....	96

5. BÖLÜM

PROTEİN YAPI ANALİZLERİ

5.1. Proteinlerin 3 Boyutlu Yapısı.....	98
5.2. Proteinlerin 3 Boyutlu Modellenmesi	102
5.2.1. Ab initio Yöntemler	103
5.2.2. Homoloji Modellemesi (Benzetme Yoluyla Modelleme).....	104
5.2.3. Giydirmeye Yoluyla Modelleme (Threading).....	106
5.2.4. Küçük Parçalardan Bütünleyerek Modelleme.....	108
5.2.5. Kuantum Mekaniksel Yöntemler	108
5.3. Proteinlerin 3 Boyutlu Modellenme Örnekleri.....	109
5.3.1. Discovery Studio Visualizer Kullanımı	109
5.3.2. Kuantum Mekaniksel Hesaplamalar	120
Kaynaklar.....	134

6. BÖLÜM

MOLEKÜLER FİLOGENETİK ANALİZLER

6.1. Filogenetik Ağaç: Temel Kavramlar.....	137
6.2. Ağaç Oluşturmak için Uygun molekülün seçimi-Moleküler Markörler	140

6.3. Moleküler Filogenetik Analiz Basamakları	141
6.3.1. Analiz için gerekli dizi verilerini toplama	141
6.3.2. Dizi hizalama.....	142
6.3.3. Filogenetik ağaç oluşturma-uygun metodun seçimi	144
6.3.4. En iyi ağacın belirlenmesi.....	152
6.3.5. Ağacın yorumlanması	154
6.4. MEGA Yardımıyla Filogenetik Ağaç Oluşturma	156
Kaynaklar.....	161

7. BÖLÜM

BİYOLOJİK AĞLAR

7.1. Sinyal İletimi ve Gen Kontrol (DNA-Protein Etkileşim - DPE) Ağları	172
7.2. DNA-Protein Etkileşiminin Belirlenmesi İçin Kullanılan Teknikler	175
7.2.1. In vitro DNA-Protein Etkileşimleri.....	175
7.2.2. In vivo DNA-Protein Etkileşimleri.....	176
7.2.3. DNA-Protein Etkileşimlerini Tahmin Eden Hesaplamalı Yöntemler	179
Kaynaklar.....	211

8. BÖLÜM

GENOM PROJELERİ VE GEN AİLELERİNİN BİYOİNFORMATİK ANALİZLERİ

8.1. İnsan Genom Projesi	226
8.1.1. Türkiye Genom Araştırması.....	228
8.2. Hayvan Genom Projeleri.....	230
8.3. Bitki Genom Projeleri.....	236
8.4. Genomlara Ait Veri Tabanları	239
8.4.1. NCBI Genom Veri Tabanı	240
8.4.2. Phytozome	245
8.5. Genomlarda Gen Ailelerinin Belirlenmesi	249
8.5.1. AREB/ABF sinyal yolağı	251
8.5.2. MYC /MYB sinyal yolağı	252
8.5.3. CBF/DREB sinyal yolağı	252
8.5.4. NAC (NAM, ATAF and CUC) and ZF-HD (zinc-finger homeodomain) sinyal yolaqları.....	253
8.5.5. Diğer sinyal yolaqları.....	254
Kaynaklar.....	264

9. BÖLÜM

GEN İFADE ANALİZLERİNDE BİYİNFORMATİK YÖNTEMLER

9.1. Gen İfadesi Analiz Yöntemleri	285
9.2. Gen İfadesi Analizi İçin Kullanılan Veri Tabanları.....	291
9.3. Gen İfade Seviyelerinin Biyoinformatik Yollarla Belirlenmesi	296
9.3.1. RNA dizileme yöntemiyle elde edilen verilerin biyoinformatik araçlarla incelenmesi	296
9.3.2. Mikroarray ile gen ifade analizleri için kullanılan biyoinformatik araçlar	303
Kaynaklar.....	305

10. BÖLÜM

BİYİNFORMATİĞİN GELECEKTEKİ YERİ VE EĞİTİMİ

10.1. Biyoinformatiğin Gelişimi ve Gelecekteki Yeri	312
10.2. Biyoinformatik Eğitimi.....	317
10.2.1. Biyoinformatik Eğitim Programlarının Kısa Tarihi: Kendi Kendine Öğrenmekten Yapılandırılmış Programlara	318
10.2.2. Biyoinformatik Eğitim Programlarının Güncel Yapısı	319
10.2.3. Biyoinformatik Eğitimini İyileştirmede Bilimsel Çalışmalar.....	324
Kaynaklar.....	326

ŞEKİLLER LİSTESİ

Şekil 1.1. Disiplinler arası bir bilim olan Biyoinformatik	4
Şekil 1.2. Biyoinformatik alanındaki gelişmeler	8
Şekil 1.3 Biyoinformatiğin uygulandığı alanlar	9
Şekil 2.1 Amerikan Ulusal Biyoteknoloji Bilgi Merkezi internet sayfasının ana sayfa görüntüsü.....	22
Şekil 2.2 GenBank veri tabanında bir gene ait GenBank erişim numarası ve genel bilgileri gösteren internet sayfası.....	23
Şekil 2.3 GenBank'ın çapraz referanslama sonucunda oluşan farklı biyolojik ve literatür bilgileri	24
Şekil 2.4 Fasta formatında yazılmış bir protein dizisi.....	24
Şekil 2.5 BLAST programının ana sayfası ve GenBank erişim numarası girilerek BLAST uygulamasının başlatılması.....	25
Şekil 2.6 BLAST sonrası açılan yeni pencerede aranan proteinin bulunduğu proteinlerin ailesi, veri tabanında bulunan diğer proteinlerle benzeme oranları, açıklamaları ve E değerleri.....	26
Şekil 2.7 Her kaynak başlığı altında açılan alt başlıklar	27
Şekil 2.8 Avrupa Moleküler Biyoloji Laboratuvarı-Avrupa Biyoinformatik Enstitüsü.....	35
Şekil 2.9 Exspasy biyoinformatik kaynak portalı.....	36
Şekil 2.10 TAIR web sayfası.....	38
Şekil 2.11 SGD ana sayfası.....	40
Şekil 2.12 FlyBase veri tabanı ana sayfası.....	41
Şekil 2.13 WGE veri tabanının ana sayfası.....	43
Şekil 3.1 BLAST veri tabanı ara yüz ekran görüntüsü.....	49
Şekil 3.2 BLAST programı ana unsurları	50
Şekil 3.3 BLAST Sonuç sayfası.....	51
Şekil 3.4 Yapılan hizalamaların bulunan skor sonuçlarına göre gösterilmesi	51
Şekil 3.5 Tanımlamalar bölümü.....	52
Şekil 3.6 Hizalama sonuçlarının gösterilmesi.....	53
Şekil 3.7 BLASTn sorgu yapılacak dizinin girilmesi	54
Şekil 3.8 DST mRNA dizisiyle oluşan hizalama.....	55
Şekil 3.9 FZP protein dizisinin BLASTp programına FASTA formatında girilmesi.....	56
Şekil 3.10 BLASTp programı için veri tabanı ve program seçimi	56
Şekil 3.11 FZP proteininin hizalanması sonucu oluşan sonuç sayfası	57
Şekil 3.12 FZP protein dizisiyle anlamlı olarak hizalanan diziler	57
Şekil 3.13 FZP protein dizisiyle hizalanan protein dizisinin eşleşme sonucu.	58
Şekil 3.14 SANSpaarellel ara yüzü görüntüsü	60
Şekil 3.15 SANSpaarellel sonuç sayfası.....	61
Şekil 3.16 SANSpaarellel BLAST-like sonuç sayfası ekranı	61
Şekil 3.17 SANSpaarellel MSA sonuç sayfası.....	61

Şekil 3.18 SANSparalel Mview sonuç sayfası.....	62
Şekil 3.19 SANSparalel LOGO sonuç sayfası	62
Şekil 3.20 SANSparalel Jalview sonuç sayfası	62
Şekil 4.1 Protein kodlayan genlerin sayısı genom boyutuyla doğru orantılıdır.	66
Şekil 4.2 Prokaryotik canlılarda kesintisiz devam eden, başlangıç ve bitiş kodonlarıyla sınırlı kodlama bölgeleri açık okuma çerçevesi olarak tanımlanır.	67
Şekil 4.3 Transkripsiyon sonrasında öncül mRNA'daki intronların farklı kombinasyonlarda çıkartılıp geri kalan kısımlarının tekrar birleştirilmesi alternatif kesme/yapıştırma (splice) olarak adlandırılır.	69
Şekil 4.4 Ökaryotik genlerin tahmininde kullanılan yöntemler ve genetik dizilerde genleri bulmak için kullanılan algılama algoritmaları.	70
Şekil 4.5 Prokaryotik bir canlıda bulunan ve işlevsel bir protein üreten genlerin genel yapısı. Başlangıç ve bitiş noktalarıyla sınırlandırılmış kodlama bölgesi gösterilmiştir.	71
Şekil 4.6 Prokaryotik mRNA üzerindeki ribozom bağlanma (Shine-Dalgarno) ve translasyon başlangıç bölgesinin (AUG) varlığı, genomik dizi içerisinde gen kodlayıcı bölgelerin başlangıcını gösterir.....	72
Şekil 4.7 NCBI internet sitesinin ORF tahmini yapan ara yüzü	73
Şekil 4.8 Araştırılan genomik dizi bilgisinin ORF finder ara yüzüne girilmesiyle farklı ORF bölgeleri ortaya çıkarılır. Ayrıca, bu ORF bölgeleri protein dizilerine de dönüştürülür.	74
Şekil 4.9 SmartBLAST analiziyle ORF tarafından belirlenen protein dizisi için kimliklendirme ve homoloji analizi yapılmaktadır.	74
Şekil 4.10 Kodlama bölgelerinde üçüncü nükleotitlerin G veya C olma olasılığı diğer bazlara oranla çok daha fazladır (A). Test edilen dizilerde üçüncü bazların analizi gerçekleştiğinde G veya C içeriği yoğun olan bölümler kodlama dizileriyle ilişkilendirilir (B).	75
Şekil 4.11 Prokaryotik bir gen için başlangıç/bitiş kodonları ve kodlama bölgesi içerisindeki nükleotitlerin olasılık dağılımlarını gösteren Markov modeli gen tanımlamada yaygınlıkla kullanılmaktadır.	76
Şekil 4.12 Transkripsiyon sonlandırıcı ikincil yapıları oluşturan tamamlayıcı bölgelerinin varlığı, kodlama bölgesinin sonunu işaret eder.	77
Şekil 4.13. GeneMarkS programının ara yüzü. Bu program sayesinde prokaryotik genlerin tahmininde oldukça başarılı sonuçlar elde edilmektedir.	78
Şekil 4.14 İnsan intronlarının kesim noktalarında bulunan bazların bulunma olasılıklarını gösteren dizi logoları. İtronların %99'u GT ile başlayıp AG ile sonlanır. Ancak bazı intronların GC ile başladıkları da gösterilmiştir	79
Şekil 4.15 Kodlama bölgelerinin sonunda bulunan poliadenilasyon sinyali (AATAAA motifi) gen tahmininde kullanılan önemli bir parametredir. .80	
Şekil 4.16 GENSCAN ana sayfası. Bu programla aynı anda 1 milyona yakın baz genomik içerik açısından analiz edebilir.	82

Şekil 4.17 Örnek DNA dizisinin GENSCAN analizi. Sol alt bölgede tahmin edilen proteinin amino asit dizisi, sağ alt köşede ise tahmin edilen kodlama dizisi gösterilmektedir.....	82
Şekil 4.18 Ökaryotik promotörlerin ve bazı motiflerin temsili gösterimi.....	86
Şekil 4.19 Cluster-Buster uygulamasının ara yüzü.	92
Şekil 5.1 Nötral ve zwitteriyonik amino asitlerin genel yapısı.....	98
Şekil 5.2 Polikondenzasyon yoluyla amino asitlerden peptit ve proteinlerin oluşması	100
Şekil 5.3 Katlanma (Levinthal) ikilemi.....	103
Şekil 5.4 Kalıp karşılaştırma metodundaki basamakların temsili gösterimi.....	105
Şekil 5.5 CATH veri tabanına göre protein sınıflandırması	107
Şekil 5.6 SCOP veri tabanına göre protein sınıflandırması	107
Şekil 5.7 Discovery Studio Visualizer 16 için bir PDB dosyasının nasıl açıldığını sırayla adım adım gösterilmesi.	111
Şekil 5.8 Dizi Penceresinin nasıl açıldığını sırayla gösteren ekran görüntüleri.	112
Şekil 5.9 Macromolecules düzeninin ayarlanması.	113
Şekil 5.10 İkincil Yapının nasıl görüneceğini seçme ekranı.....	114
Şekil 5.11 İkincil Yapının nasıl görüneceğini seçmek için ortaya çıkan diyalog penceresi.	114
Şekil 5.12 Residue kimliği bir ipucu olarak görünür.	114
Şekil 5.13 Grup oluşturma	115
Şekil 5.14 Grup adı tanımlama	115
Şekil 5.15 Seçilen bölgenin ortalanıp yakınlaştırılması.....	116
Şekil 5.16 Molekül açma.....	116
Şekil 5.17 Publication Quality' ye ayarının yapılması	117
Şekil 5.18 Display Style penceresi.....	117
Şekil 5.19 Hidrojen bağlarının gösterimi	117
Şekil 5.20 H Bağlarının molekül üzerinde gösterimi.....	118
Şekil 5.21 İki atom arasındaki uzaklığı ölçme.	118
Şekil 5.22 Aradaki mesafeyi gösteren sayıyı düzenleme.	118
Şekil 5.23 Mesafe düzenleme penceresi.....	119
Şekil 5.24 Molekülün DSV formatında kaydedilmesi.	119
Şekil 5.25 Molekülün resim olarak kaydedilmesi.....	119
Şekil 5.26 Pubchem ve PDB ana sayfası.	120
Şekil 5.27 Set torsions düğmesi	121
Şekil 5.28 Set Torsions düğmesine basıldıktan sonra, yazılımın belirlediği torsiyon açılı (Sarı renk ile belirtilmiş)	121
Şekil 5.29 Torsiyon açısının ne kadar döndürüleceğini gösteren yer	122
Şekil 5.30 Setup Calculation seçimi	122
Şekil 5.31 Calculation penceresi.....	122
Şekil 5.32 Local Monitor seçimi	122
Şekil 5.33 Local Monitor penceresi.....	123

Şekil 5.34 Display Spreadsheet seçimi	123
Şekil 5.35 Seçilen konformerin farklı olarak kaydedilmesi.....	123
Şekil 5.36 Bulunan konformer sayısını gösteren yer.....	123
Şekil 5.37 mol2 formatında moleküllerin kaydedilmesi	124
Şekil 5.38 Farklı kaydetme sırasında ortaya çıkan yük seçimi	124
Şekil 5.39 Farklı kaydetme sırasında ortaya çıkan “Select molecules” penceresi.....	124
Şekil 5.40 Molekülün kaydedildiğini gösteren son pencere.....	124
Şekil 5.41 Open düğmesi.....	125
Şekil 5.42 Gaussian hesabının ayarlanması.....	125
Şekil 5.43 İş tipinin belirtilmesi.....	126
Şekil 5.44 Metodun belirtilmesi.....	126
Şekil 5.45 İş ile ilgili istenilen bilgilerin girilmesi.	127
Şekil 5.46 İşlemci, hafıza gibi bilgilerin girilmesi (Link 0).	127
Şekil 5.47 Çözücü model seçimi.....	127
Şekil 5.48 Giriş dosyasının kaydedilmesi	128
Şekil 5.49 Giriş Dosyasının kaydedileceği yerin seçimi.	128
Şekil 5.50 Giriş dosyası	128
Şekil 5.51 Hesap çalıştırma ekranı	128
Şekil 5.52 Çıktının GausView’da açılması.	129
Şekil 5.53 Çıktının Mol2 formatında kaydedilmesi.....	130
Şekil 5.54 Mol2 olarak kaydedilecek konformerin nereye kaydedileceğinin seçimi	130
Şekil 5.55 GausView’da çıktının incelenmesi.....	130
Şekil 5.56 Hesap sonucunun özeti.....	131
Şekil 5.57 Notepad++ programında çıktının incelenmesi.....	131
Şekil 6.1 Filogenetik ağaç	138
Şekil 6.2 Filogenetik ağaç formları.....	138
Şekil 6.3 Köklü filogenetik ağaçlarda dallanma tipleri.....	139
Şekil 6.4 Filogenetik ağaçta OTU’ların çeşitli şekillerde gruplandırılması.	139
Şekil 6.5 Paralog ve ortolog kavramları.....	140
Şekil 6.6 Çoklu hizalama örneği.....	143
Şekil 6.7 Örnek bir filogenetik ağaç	155
Şekil 7.1 Transkripsiyon faktörlerinin gen kontrol mekanizmaları.....	174
Şekil 7.2 Transkripsiyonel kontrol ve sinyal iletimi birbirini takip eden karmaşık gen regülasyonlarını içermektedir. Bir gen kontrol ağı, bir hücredeki farklı genlerin zaman, mekân ve miktar olarak ifade edilme oranını yöneten birbirine oldukça bağlı proseslerin karmaşık bir setidir.....	175
Şekil 7.3 Kromatin immüno çökeltme analizi	178
Şekil 7.4 Maya çift-hibrid (Y2H) yöntemi ile protein-protein etkileşimlerinin keşfedilmesi	188

Şekil 7.5 Afinite saflaştırması - kütle spektrometresi (AP-MS) ile protein-protein etkileşimlerinin belirlenmesi.	190
Şekil 7.6 Bimoleküler floresan tamamlama (BiFC) analizi, canlı hücrelerdeki protein etkileşimlerinin basit ve doğrudan görselleştirilmesini sağlar	193
Şekil 7.7 Saccharomyces cerevisiae metabolik ağı.....	204
Şekil 7.8 Bir metabolik ağın yeniden yapılandırılması iş akışı.	205
Şekil 7.9 Streptococcus pnömonisi'nin yeniden yapılandırılmış metabolik ağlarının farklı çizimleri.	206
Şekil 7.10 Metabolik yollar veya şebekelerle yakından ilişkili birçok ağ vardır. Örneğin, Arabidopsis thaliana'daki sülfür metabolizması farklı metabolizmalarla etkileşim içerisinde.	209
Şekil 8.1 Ortolog genler arasındaki ayrılma oranları.....	233
Şekil 8.2 NCBI tarafından eklenen ökaryotik genomların sayısı	244
Şekil 8.3 Her yıl NCBI tarafından açıklanan ökaryotik genomlar	244
Şekil 8.4 Stres toleransının ve duyarlılığının düzenleyici ağı	251
Şekil 9.1 Santral dogmanın şematik gösterimi ve bilgi akışı.....	284
Şekil 9.2 Gen ifade analizlerinde takip edilen iş akış şeması.....	285
Şekil 9.3 Illumina dizileme sistemindeki genel iş akışı.....	291
Şekil 9.4 Örnek bir GEO veri taraması.....	293
Şekil 9.5 DNA dizileme sonucu elde edilen verilerin analiz basamakları.....	296
Şekil 9.6 Örnek bir FastQC sonucu.....	298
Şekil 10.1 PubMed veri tabanında biyoinformatik anahtar kelimesinin genel araştırılması.	315
Şekil 10.2 PubMed veri tabanında biyoinformatik anahtar kelimesinin özel araştırılması.	315
Şekil 10.3. Lisans düzeyinde Biyoinformatik Programlarının Derslerinin Dağılımı	322

TABLolar LİSTESİ

Tablo 2.1 Tüm NCBI kaynaklarını içeren panel başlıklarının listesi.....	27
Tablo 2.2 TAIRde yer alan veriler ve kaynakları	39
Tablo 2.3 FlyBase veri tabanına ait istatistiki veriler.....	41
Tablo 3.1 BLAST programları.....	48
Tablo 3.2 BLAST analizi için hazırlanan veri tabanları	50
Tablo 3.3 Çoklu hizalama için yaygın olarak kullanılan veri tabanı kullanmayan programlar	59
Tablo 4.1 Gen tanımlamada kullanılan bazı programlar ve internet adresleri.	83
Tablo 5.1 Konformer analiz ve optimizasyon sonucu elde edilen tablo.....	132
Tablo 5.2. Protein yapı ve fonksiyon çalışmalarında 1900-2002 yılları arasında Kimya, Fizyoloji ve Tıp alanlarında verilen Nobel ödülleri	132
Tablo 6.1 MSA, homoloji ve filogenetik analizlerde kullanılan bazı programlar ve özellikleri.....	146
Tablo 6.2 Filogenetik ağaç oluşturmak için kullanılan yöntemler listesi ve güçlü/zayıf yönleri	153
Tablo 7.1 DNA-protein ağlarının belirlenmesi için kullanılan veri tabanları.	183
Tablo 7.2 Muhtemel bağlanma bölgesi tahmin programları.	201
Tablo 7.3 Kenetlenme tabanlı protein küme analiz programları.	201
Tablo 7.4 Şablon tabanlı protein küme analiz programları.....	202
Tablo 7.5 Protein-protein ağlarının belirlenmesi için kullanılan veri bankaları.....	202
Tablo 7.6 Metabolik ağların belirlenmesi için kullanılan veri bankaları.....	210
Tablo 8.1 C. elegans'ın genom bilgisi	235
Tablo 8.2 Phytozome Versiyon 0, 1, 2 ve 12 karşılaştırılması	247
Tablo 8.3 Türler ve Türlerle ait BTF sayıları.....	254
Tablo 8.4 Farklı bitkilerin genomlarında biyoinformatik yöntemler kullanılarak tespit edilen transkripsiyon faktörleri ve gen ailelerinin listesi.....	255
Tablo 9.1 Gen ifade analizlerinde kullanılan yöntemler	286
Tablo 9.2 DNA array çeşitleri ve üreticileri.....	289
Tablo 9.3 Gen ifade analizlerinde kullanılan veri tabanları ve erişim adresleri	295
Tablo 9.4 Genel mikroarray ham veri dosya türleri.....	304
Tablo 10.1 Biyoinformatik iş gücü tiplerine göre eğitim programının sağlaması gereken temel	320
Tablo 10.2. Biyoinformatik eğitimi müfredat yapısı	321
Tablo 10.3. Biyoinformatik lisans ders programlarında yer alan zorunlu dersler.....	323
Tablo 10.4. Biyoinformatik konulu eğitim çalışmalarında öğretimi yapılan kavramlar ve konu alanlarına göre dağılımları	324