BİYOİNFORMATİKTE KULLANILAN ÖNEMLİ TERİMLER

Accession number (GenBank): Bir dizi GenBank'a kaydedildiği zaman bu kayıt için verilen yada kayda özel kimlik numarasıdır. Bir büyük harf ve ardından gelen 5 rakam veya 2 büyük harf ve 6 rakamdan oluşur.

Accession number (RefSeq): Bütün bir RefSeq dizisine atanmış kimlik numarasıdır. Sırasıyla iki büyük harf, bir alt çizgi (_) ve 6 rakamdan oluşur (NT_123456).

*** NT_123456 birleştirilmiş kontigler**

***NM_123456 mRNA'lar (mRNA'dan oluşturulmuş cDNA'lar)**

*** NP_123456 proteinler**

*** NC_123456 kromozomlar**

Bit score: Kullanılan skorlama sisteminin istatistiksel özelliklerinin hesaba katılmasıyla ham karşılaştırma skoru S'ten türetilmiş S değeridir. Farklı karşılaştırmalar arasında ilişki kurmak için kullanılabilirler.

BLAST: (Basic Local Alignment Search Tool):

Aynı yada farklı organizmalar arasında nukleotid yada protein dizisi karşılaştırılması ve benzer bölgelerin araştırılması için kullanılan yüksek hızda bir bilgisayar programı.

Blosum (Block Substitution Matrix):

Proteinlerin karşılaştırılması ile elde edilen blokların değişim frekansının gözlemlenmesinden türetilmiş değerlerden oluşan bir değişim matrisidir. Her matris özel bir evrimsel uzaklığa dayanır.

CDS: Bir nukleotid dizisinin kodonları oluşturan bölgesi yada kodlayan dizi.

Conserved Sequence: Bir DNA molekülünde (bir proteindeki Aa dizisinde) evrim süresince değişmeden kalmış olan baz dizisi.

Contig: Bir kromozomun üst üste çakışma gösteren, klonlanmış farklı DNA parçaları grubu.

Domain: Bir proteinin bağımsız olarak katlanabildiği ve çalışılabildiği kabul edilen parçası.

E value (Expectation value): Beklenti değeri, Veritabanı taramasında şans eseri çıkması beklenen, S değerine denk yada daha büyük skorlara sahip benzer dizilerin sayısı. Düşük E değeri büyük skora işaret eder.

EST (Expressed Sequence Tag): Bir cDNA molekülünün, bir genin kimliği olarak kullanılabilecek kısa bir parçası. Genlerin konumlanmasından ve haritalanmasında kullanılır.

Homologue: Dizisi büyük oranda başka bir gene benzeyen gen. Bu genlerin ortak bir ataya sahip olduğu ve benzer fonksiyonlar taşıdığı düşünülür.

Motif: Protein dizisi içinde kısa, korunmuş bir bölge. Motifler genellikle domainlerin yüksek derecede korunmuş bölgeleridir.

Orthologous: Ortak bir atadan geldikleri düşünülen, benzer bir fonksiyonu olabilen, farklı türlere ait homolog diziler.

Paralogous: Aynı tür içinde, gen dupliklasyonu sonucu oluşmuş homolog diziler.

Query: Veritabanındaki tüm dizilerin karşılaştırılacağı giriş dizisi.

Farklı Dizi Formatları

A. Gen Bank DNA Dizi Formatı: Girilen her dizinin tanımlayıcı bilgileri verilir.

Bu bilgiler her satırda ilk bilgi olarak, her biri bir belirleyici ile birlikte gruplara ayrılmış şekilde yazılır.

Örneğin; referans için RF gibi,

LOCUS lokusun ismi

DEFINITION girişin tanımı

ACCESSION orijinal kaynağın accession nımarası

KEYWORDS bu girişin karşı referanslarının yapılabilmesi için anahtar kelimeler

SOURCE DNA'nın elde edildiği organizma

Avrupa Moleküler Biyoloji Laboratuvarı Veri Kütüphanesi Formatı (EMBL)

ID veritabanındaki dizi için kimlik numarası

AC dizinin başlangıcını gösteren accession nımber

DT girişin ve modifikasyonların tarihi

KW anahtar kelimeler

OS, OC kaynak organizma