

آزمایشگاه آموزشی
بیست و دومین المپیاد
زیست‌شناسی ایران

بیوانفورماتیک

دیتابیس‌ها ۲.

روز پنجم
۹۸/۵/۱۰

اهداف آزمایش:

۱. ادامه آشنایی با مقدمات بیوانفورماتیک

زمان آزمایش: ۹۰ دقیقه



این فایل به منظور آموزش عملی دانش‌پژوهان المپیاد زیست‌شناسی ایران گردآوری شده است.

ادامه معرفی ابزارهای بیوانفورماتیک

NCBI

هدف شما در این جلسه آشنایی و یادگیری BLAST می باشد.

کاربرد های BLAST :

- تعیین اینکه چه پروتئین ها یا توالی های نوکلئوتیدی پارالوگ یا اورتولوگی برای توالی ما وجود دارد.
- تعیین اینکه پروتئین یا نوکلئیک اسید توالی مد نظر ما در یک گونه وجود دارد یا نه.
- تعیین هویت توالی مد نظر ما
- کشف ژن های جدید
- تعیین انواع واریانت های یک ژن یا پروتئین

انواع BLAST:

- Nucleotide BLAST: ورودی : توالی نوکلئوتید جستجو در پایگاه داده نوکلئوتیدی
- Protein BLAST: ورودی : توالی آمینواسیدی جستجو در پایگاه داده پروتئینی
- Blastx: ورودی : توالی نوکلئوتید جستجو در پایگاه داده پروتئینی (در تمام ۶ فریم ممکن)
- tblastn: ورودی : توالی آمینواسیدی جستجو در پایگاه داده نوکلئوتیدی (در تمام ۶ فریم ممکن)

در این بخش می خواهیم برای یک توالی DNA ناشناخته عملیات BLAST را انجام دهیم تا بفهمیم مربوط به چه ژنی در چه گونه ای می باشد .

روی گزینه BLAST در بخش Popular Resources کلیک کنید.

وارد بخش Nucleotide BLAST شوید.

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

Or, upload file

Job Title

Align two or more sequences

Choose Search Set

Database

Organism

Exclude

Limit to

Entrez Query

Program Selection

Optimize for

BLAST

Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)

Show results in a new window

Algorithm parameters

General Parameters

Max target sequences

Short queries

Expect threshold

Word size

Max matches in a query range

Scoring Parameters

Match/Mismatch Scores

Gap Costs

Filters and Masking

Filter

Mask

BLAST

Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)

Show results in a new window

در این صفحه می توانید جستجوی خود را انجام دهید و تنظیمات مناسب را اعمال کنید.

• Browse : در این قسمت می توانید فایل متنی با فرمت FASTA را آپلود کنید. همچنین می توانید توالی را در کادر بالا کپی کنید.

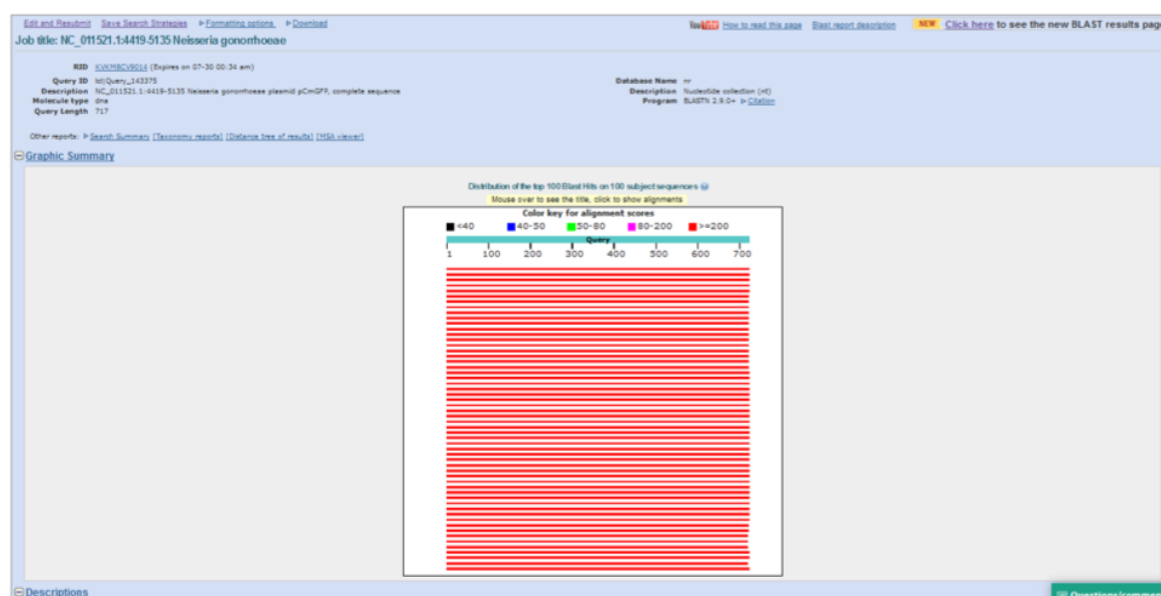
• Query subrange : در این قسمت می توانید تنظیم کنید تنها بخش خاصی از توالی شما در پایگاه داده جستجو شود.

• Organism : می توانید گونه هایی را که می خواهید در نتایج جستجو ظاهر شوند را تعیین کنید یا بعضی گونه ها را از جستجو خارج کنید.

- Optimize : در این بخش الگوریتم های مختلف BLAST مانند MegaBlast یا blastn را می توانید انتخاب کنید.
- o MegaBlast برای پیدا کردن توالی های بسیار مشابه ، مناسب برای query های بلند ، سرعت بالا
- o Blastn برای توالی های با تشابه کمتر، دقت بیشتر در پیدا کردن توالی های مشابه ، نامناسب برای توالی های بسیار بلند، سرعت کمتر

- Max target sequences : تعیین تعداد hit های نمایش داده شده در خروجی.
- Expect threshold : حذف نتایجی که دارای E-Value بیشتر از حد تعیین شده هستند.
- Word Size : اندازه کلمات جستجو را تغییر می دهد، عدد بزرگ تر زمان بیشتر و جستجوی بهتری به همراه دارد.
- Gap Cost : تعیین امتیاز منفی شروع و ادامه گپ

در نهایت روی BLAST کلیک کنید تا نتایج ظاهر شود. با توجه به طول توالی و پایگاه داده مورد استفاده و تنظیمات دیگر، زمان انتظار متفاوت خواهد بود.



هرکدام از خطوط قرمز مربوط به یک hit می باشند که بر اساس کم بودن Value-E از بالا به پایین مرتب شده اند. این ترتیب در لیست پایین هم موجود است.

در این جلسه فعالیت های را که می توانیم بر روی نتیجه الایمنت چندگانه انجام دهیم می آموزیم .
 پس از پایان کار صفحه ی زیر را مشاهده خواهید کرد. با کلیک کردن روی تب های مختلف بالای صفحه می توانید اطلاعات بیشتری راجع به این هم راستایی پیدا کنید یا آن را به عنوان ورودی به برنامه های دیگر بدهید.

The screenshot displays the EMBL-EBI Clustal Omega web interface. At the top, there is a navigation bar with links for EMBL-EBI, Services, Research, Training, Industry, and About us. Below this, the Clustal Omega logo is prominently displayed. The main content area shows the results for a job titled 'clustalo-l20180824-163858-0479-29210239-p2m'. The interface includes tabs for 'Alignments', 'Result Summary', 'Phylogenetic Tree', and 'Submission Details'. Under the 'Alignments' tab, there are options to 'Download Alignment File', 'Show Colors', 'View result with Jalview', 'Send to Simple Phylogeny', and 'Send to MView'. The alignment itself is shown as a text-based multiple sequence alignment (MSA) using CLUSTAL O (1.2.4). It lists three sequences: 'HSA_HUMAN' (P69905), 'HSA_MOUSE' (P01942), and 'HSA_CAPRI' (P13786). The alignment shows conserved regions across these species, with gaps indicated by dashes. A note at the bottom states: 'PLEASE NOTE: Showing colors on large alignments is slow.' The footer contains EMBL-EBI contact information, a copyright notice for 2018, and a link to the Terms of use.

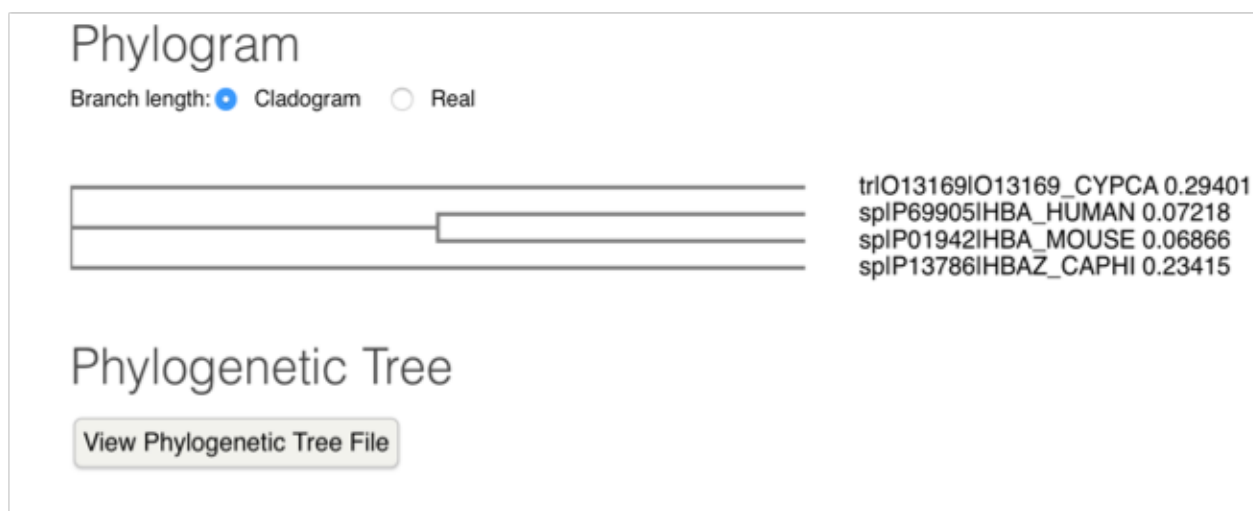
با کلیک بر روی تب Result summary با صفحه‌ی زیر مواجه خواهید شد. هر کدام از لینک های موجود در صفحه اطلاعاتی در مورد همراهی انجام شده به شما می دهد.

با کلیک کردن بر روی لینک Percent Identity Matrix میتوانید درصد تشابه هر جفت توالی را ببینید.

```
#
# Percent Identity Matrix - created by Clustal2.1
#
#
1: tr|O13169|O13169_CYPCA 100.00 50.00 50.70 47.18
2: sp|P69905|HBA_HUMAN 50.00 100.00 85.92 56.34
3: sp|P01942|HBA_MOUSE 50.70 85.92 100.00 56.34
4: sp|P13786|HBAZ_CAPHI 47.18 56.34 56.34 100.00
```

برای رسم درخت فیلوژنتیک باید بر روی تب Send to simple phylogeny کلیک کنید. این برنامه از نتیجه هم راستایی برای رسم درخت استفاده می کند. در صفحه حاصل می توانید تنظیمات کشیدن درخت را تغییر دهید. از مهم ترین این تنظیمات می توان به الگوریتم رسم درخت اشاره کرد. با کلیک کردن روی دکمه submit رسم درخت آغاز می شود.

خروجی این ابزار به شکل روبرو است. نام هر شاخه اولین خط در توالی FASTA ای است که به نرم افزار Multiple Aignment داده شده است. شما میتوانید طول شاخه ها تغییر دهید تا نسبت های واقعی آنها را مشاهده کنید و درخت حاصل را ذخیره کنید.



سؤالات عملی

تنها ابزار های مجاز برای استفاده در این بخش سایت های **EBI , NCBI , Uniport** می باشند و استفاده از اینترنت (غیر از موارد ذکر شده) و یا ابزار های آفلاین (مانند نوت پد یا اکسل) ممنوع می باشد و با فرد خاطی برخورد خواهد شد .

Subpart 1: Your scientist friend invited you to her office in Tehran. She was working on the phylogenic relations of some organisms based on a specific gene. you find her murdered in her office and the only evidence left is a paper found on her desk

This sequence is the key to find my killer!

5 - GCTTCCGTCAGGTGTGAGCCATACAGGTGTTGTTTCAAAAGGCTGCGGC - 3

Find the gene that I was building the phylogenic tree based on it and draw the tree for my favorite organisms!

Ptilocolobus tephrosceles

Macaca nemestrina

Gorilla gorilla gorilla

Papio Anubis

Pan troglodytes

Sus scrofa

Bos Taurus

Human Kind :)

The killer First name is the genus of the organism that have the biggest distance to the nearest node on tree!

Question 1: What was the specific Gene? Find the specific gene based on the sequence. **4 point**

Question 2: Write down some functions of the product of this gene in humans based on your knowledge. **1 point**

Question 3: Draw the UPGMA tree. **9 points**

Save it in Q3 folder.

Question ۴: What is the first name of killer? **۱ point**

زیر بخش ۲ : در زیر چند توالی پروتئینی آورده شده است :

Amino acid	Three letter symbol	One letter symbol*
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

الف (CCCCCPCCCCCPCCCCCCCCCPCCCCCCCCPCCCCC)

ب (HHHHHHHHHHHHHH)

پ (EEQEQQQQEEQEQQEEQEQQQQEEE)

ج (YKCTICGKAFLLESLLKNHVAAH)

سوال ۵) توالی های بالا را به ترتیب فراوانی یافت شدن در پروتئین های انسانی مرتب کنید . ۸ نمره (هر مورد ۲ نمره)

سوال ۶) نتایج سوال ۵ را مورد به مورد توجیه کنید . ۷ نمره