



Introducing Bioinformatics in the Biology Olympiad

OPPORTUNITIES
AND CHALLENGES

2024 IBO Educational Conference
12/07/2024

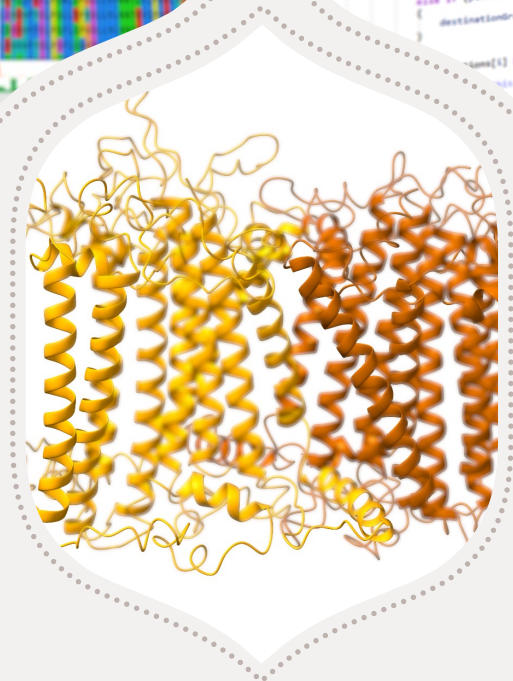
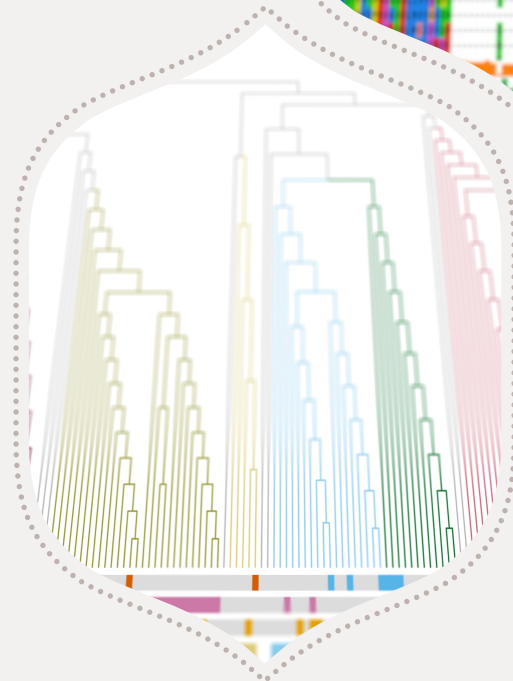
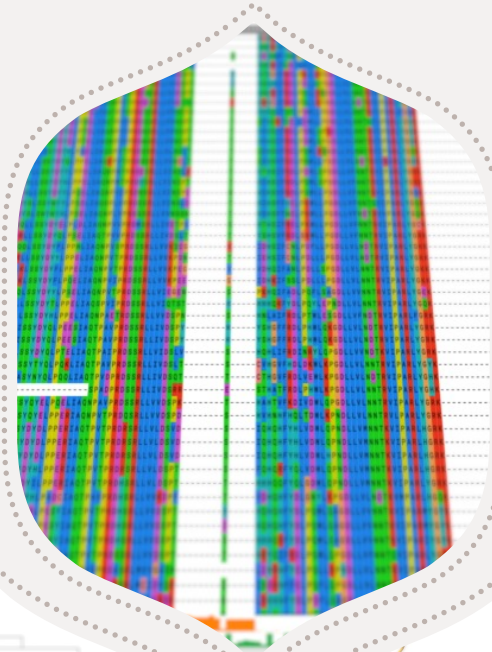


Giorgio Bianchini

Bioinformatics

Bioinformatics [...] is a scientific subdiscipline that involves using computer technology to collect, store, analyze and disseminate biological data and information

– *National Human Genome Research Institute*



Bioinformatics at the IBO

- A practical task explicitly called “Bioinformatics” has appeared in **every IBO since 2019** (except for IBO Challenge II 2022).
- **Even before then**, many practicals have included bioinformatics-adjacent questions.
- Since **2019**, we have included Bioinformatics in the training programme for the Italian team.
- Since **2022**, we have also included a Bioinformatics task in the selection process.

IBO 2018 Iran

On your laptop, there is an application which shows the expected results from a Wright-Fisher model (On your desktop, go to IBO2018Task2 and double-click on “WFM_model.py” and wait for the application to start. There are 4 variables you can play with. The sliders allow you to change the values for each variable. After you have chosen the desired values for each parameter, simply click on “Simulate” to see how heterogeneity (H) changes over time in the Wright-Fisher model, given the chosen parameters. (See the formal definition of H in Part A)



Figure 3: The interface of the Wright-Fisher model app.

IBO 2016 Vietnam

Q.3.1. CALCULATE DISTANCE MATRIX

Calculate the distance matrix based on the character matrix provided in Table 1. The distance between two specimens is defined as the number of characters at which the two specimens show different character states (present: “1”; absent: “0”). Write the numerical results in the **ANSWER SHEET** (8.4 points).

Reconstructing Phylogenetic Relationship using UPGMA

UPGMA (Unweighted Pair Group Method with Arithmetic Mean) is considered the simplest method for reconstructing phylogenetic trees with the assumption that the data provided have constant rates of evolution. In the method, the pair of clusters with the shortest distance is combined into a cluster of higher level at each iteration. To illustrate this concept, consider the numbers of character differences between the taxa (specimens) M, N, O, P, and Q.

Taxa	M	N	O	P	Q
M	0				
N	2	0			
O	6	8	0		
P	4	5	7	0	
Q	7	8	9	7	0

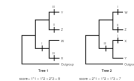
TASK 5: DRAWING A MAXIMUM PARSIMONY TREE

Maximum parsimony is a phylogenetic approach to reconstruct phylogenetic relationships. In this approach, each tree will be evaluated with the number of changes and their weights. The best tree is the one with minimum evolutionary change. Two trees based on the table below for putative taxa W-Z are shown in Figure 4. In this example, Tree 2 is better than Tree 1, due to its lower score.

		I	II	III
Character	W	1	0	0
	X	1	0	1
	Y	0	0	1
	Z	0	1	0
Weight		1	2	3

$$\text{score} = \sum N_i \cdot w_i$$

N_i is number of changes in trait i in tree and w_i is weight of trait i



IBO 2018 Iran

In this next step, we map character evolution on a given phylogenetic tree such as the one shown in Fig. 4.2. The tree represents a worldwide consensus between different hypotheses about the evolution of sand plants.

We will use a character mapping procedure called delimited transformation (DELTRANS). Proceeded according to the following protocol:

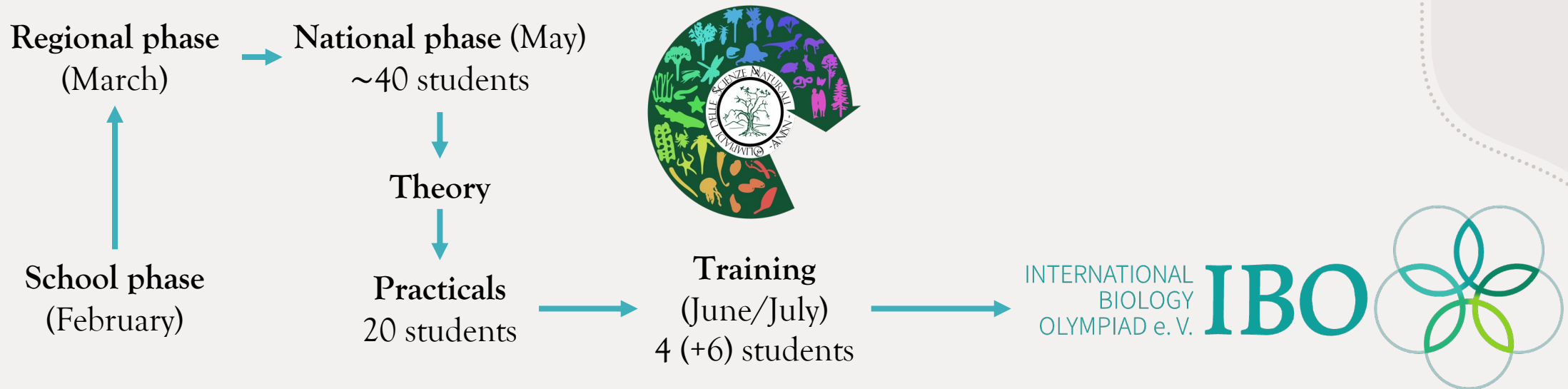
1. Characters are only allowed to change forward from 0 to 1.
2. Minimize the number of times that a character changes on the tree (principle of parsimony).
3. If it is impossible to select a change in a given character to just a single branch, then let the character change more than once (parallel evolution).
4. Indicate on the tree on which branch a given character changes state. Notice that changes of characters evolving in parallel (i.e. changing state several times) is indicated with two vertical bars (||), whereas a unique character state change (i.e. changing a single time) is indicated with a single vertical bar (|). Finally, indicate the number of the character that changes its state from 0 to 1 such as shown on the tree below.



IBO 2015 Denmark

The Italian Natural Sciences Championships

- Note: used to be called Olympiad until 2022, but were required by the Ministry of Education to change name because of the 2026 Winter Olympic Games (which are clearly more important).
- Three local and national selection phases, followed by training:



The Italian Natural Sciences Championships

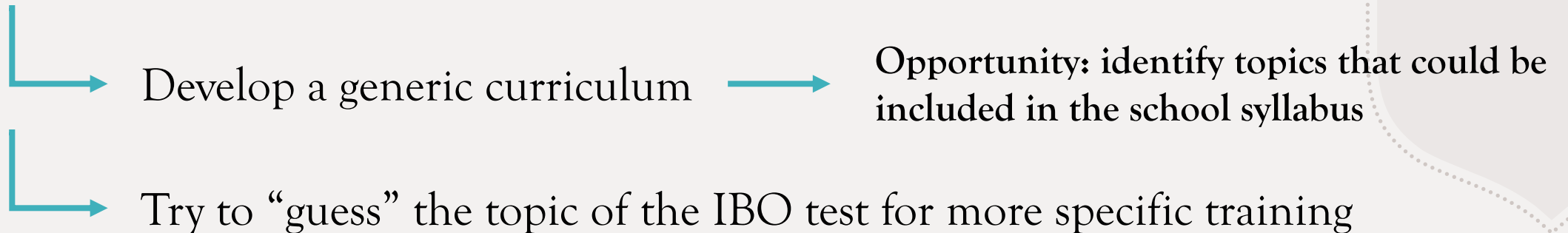
- ♦ Established in 2002, Italy participates in IBO since 2008.
- ♦ **Goals:**
 - To create an opportunity for students to test their inclination for studying and understanding natural phenomena.
 - To compare school settings across the Italian regions.
 - To identify a reference curriculum for Natural Sciences, shared between the varied realities of Italian schools.
 - To compare teaching in Italian schools with other nations.
 - To stimulate thoughts about curriculum adjustments, also thanks to the comparison with foreign schools.
- ♦ Test difficulty and topics must be appropriate for Italian schools.
 - We are selecting the best students according to the Italian school system.
 - Hopefully this correlates with the students that would perform best in the IBO, but not necessarily.
 - There are certainly distortions, exacerbated by the lack of funds.

The Italian Natural Sciences Championships

- The **theoretical task** consists of multiple-choice questions and lasts 80 minutes.
- **Practical tasks** were introduced in 2009, to help select the best students for IBO.
 - Until 2022, two topics: **Biochemistry** and **Ecosystematics**, each lasting ~1 hour.
 - Despite lasting longer, the **practical tasks weigh 50%** as much as the theoretical task (33% of the total).
- In 2022, in-person practicals could not be held. Instead, online “replacement practical tests” were developed, which did include a **bioinformatics** task.
- A true **Bioinformatics** practical was introduced in 2023, first as part of Biochemistry (extended to 95 minutes) and in 2024 as a separate task lasting 45 minutes.
- Bioinformatics has featured in the **training programme** since 2019.

Bioinformatics training for the Italian team

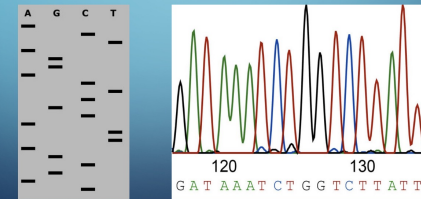
- **Challenge:** what does “Bioinformatics” even mean?
- We have no more than a couple of days to spend on this topic, so we cannot cover everything.



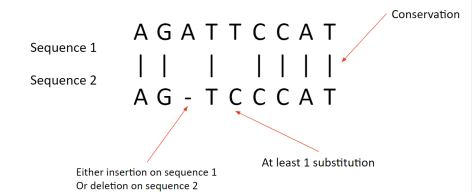
Bioinformatics training for the Italian team

Offenere sequenze

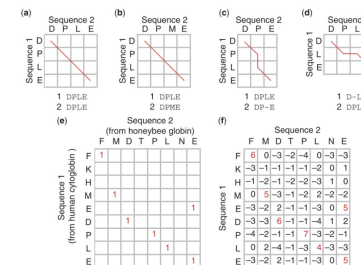
- Sequenziamento
 - Sanger



Sequence alignments



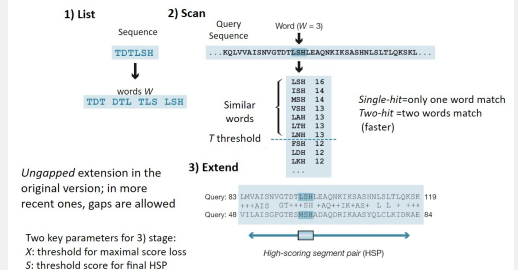
Setting up a matrix



Two proteins are arranged in a matrix and an optimal path along a diagonal is sought

Basic local alignment search tool (BLAST)

It's a *heuristic* technique, considered the standard for similarity searches in sequence databases as it offers both speed and sensitivity



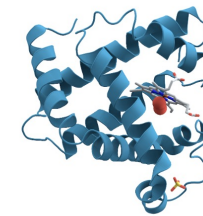
Bioinformatics training for the Italian team

La struttura delle proteine

Oltre alla propria sequenza di amminoacidi (struttura primaria) una proteina è caratterizzata anche dalla propria struttura tridimensionale.

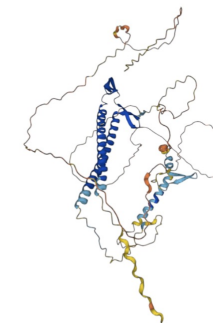
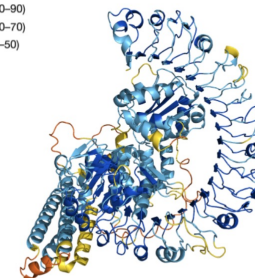


Le prime strutture 3D di proteine (emoglobina e mioglobina) sono state risolte nel 1959 da Max Perutz e John Kendrew con la cristallografia a raggi X (Premio Nobel per la Chimica nel 1962)



Valutare l'accuratezza delle predizioni

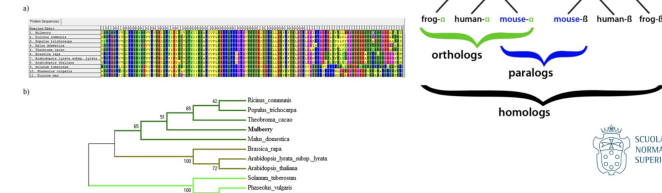
pLDDT e [90-100]
pLDDT e [70-90]
pLDDT e [50-70]
pLDDT e [0-50]



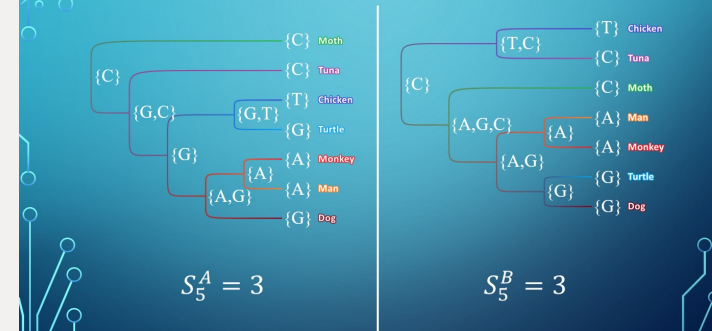
Bioinformatics training for the Italian team

Sfruttare l'informazione evolutiva

I geni si modificano nel corso dell'evoluzione e la loro storia si può ricostruire dalla loro sequenza



MAXIMUM PARSIMONY



UPGMA



Bioinformatics training for the Italian team

- Try to “guess” the topic of the IBO test for more specific training

2019



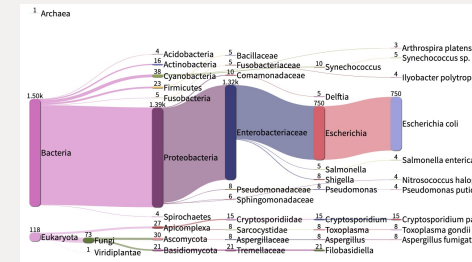
(announced by the organisers)

- Challenge 1: commercial software (but free trial)
- Challenge 2: not all students have laptops
- Challenge 3: laptops with different OSs

└─ USB stick with read-only custom Linux

2024

Pavian



- Free and browser-based ⇒ easier
- Different venue with a computer lab

Developing bioinformatics practicals

- ♦ **Opportunity: bioinformatics can be related to almost every biological subject.**
 - ♦ We can create engaging linked tasks, where bioinformatics is integrated with the wet-lab part of the practical.
- ♦ **Challenge: bioinformatics is not (yet) part of the school curriculum.**
 - ♦ We cannot assume that the students know **anything**.
 - ♦ We cannot give too much weight to this task.
 - ♦ Even more than usual, the test assesses the students' ability for data mining and to follow instructions.

Administering bioinformatics practicals

- ♦ **Opportunity: a bioinformatics practical is “cheap”**
 - ♦ It does not require any specialised equipment or reagents (other than a computer).
- ♦ **Challenge 1: we do need computers that support the software used for the task**
 - ♦ While laptops are widespread in Italy, there will be occasional students who do not have a computer and should not be penalised.
 - ♦ Italian schools *usually* have access to a computer lab, but the equipment is often outdated or not well maintained.
- ♦ **Challenge 2: students + computer = cheating**
 - ♦ Allowing students to access a computer (maybe even the Internet, if we are running a browser-based test) could likely lead to cheating issues, which must be considered while designing the task.

Bioinformatics practical 2022: insulin receptor

- This was part of the online “**replacement practical test**” and was linked to a Biochemistry task.
- In the first part, students were given data about a patient with high blood glucose and high blood insulin, thus concluding that the **patient had issues with the insulin receptor**.
- In the bioinformatics part, they were given the amino acid sequence for the patient’s insulin receptor gene and had to **identify the mutation** most likely involved in the disease.
- Inspired by a real case report (Tuhan et al 2017, DOI: 10.4274/jcrpe.4577).

Bioinformatics practical 2022: insulin receptor

Per analizzare la sequenza del recettore per l'insulina, procedi a estrarre l'mRNA da un campione di tessuto del paziente, lo amplifichi utilizzando una PCR con trascrittasi inversa e lo sequenzi con il metodo di Sanger. Una volta ottenuta la sequenza nucleotidica dell'mRNA, lo traduci “*in silico*” (ovvero, utilizzando un programma al PC) per ottenere la sequenza amminoacidica del recettore.

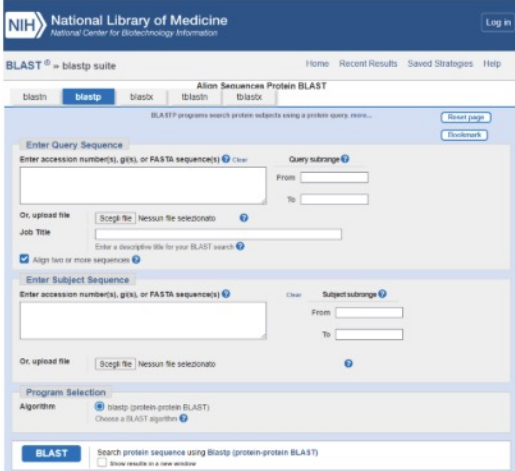
Negli esseri umani (*Homo sapiens*), l'isoforma “lunga” del recettore dell'insulina è composta da 1382 amminoacidi. La sequenza di amminoacidi che hai ottenuto dal paziente è effettivamente lunga 1382 amminoacidi, ed è la seguente:

```
MATGRRGAAAPLLVAVALLGAAGHLYPGEVCPGMDIRNLTRELHELENCVIEGHLQILLMFKTRPEDFRDLSFPKLIMITDYLL
LFRVYGLLESLKDLFPNLTVIRGSRFFNYALVIFEMVHLKELGLYNLMNITRGSVRIEKNNELCYLATIDWSRILDSVEDNYIVLNKDD
NEECGDICPGTAGKTNCPATVINGQFVERCWTHSHCQKVCPTICKSHGCTAEGLCCHSECLGNCSQDDPTKCVACRNFDGRCVET
CPPPYHFDQWRCVNFSCQDLHHCKNSRRQGCHQYVIHNNKCIPECPSGYTMNSNLLCTPCLGCPKPVCHILEGEKTDIVSTSAQE
LRGCTVINGSLIINIRGGNNLAAELEANGLIEEISGYLKIRRSYALVLSFFRKLRIRGETLEIGNYSFYALDNQNLRLQLWDWSKHN
LTITQGKLFHYNPKLCLSEIHKMEEVSGTKGRQERNDIALKTNGDQASCENELLKFSYIRTSFDKILLRWEPTYWPPDFRDLLGFMLFY
KEAPYQNVTEFDGQDAGCSNSWTVDIDPPLRSNDPKSQNHGWLMRGLKPWTQYAI FVKTLVTFSDERRTYGAKSDI IYVQTDATNPS
VPLDPI SVSNSSSQIILKWKPPSDPNGNITHYLVFERQAEDELFELDYCLKGLKLPRTWSPPFSEDSQKHNSQSEYEDSAGECCSC
PKTDSQILKELESSFRKTFEDYLNHVVFVPRKTSSTGAEDPRPSRKRRLSGDVGNVTVVPTVAAPNTSSTSTPTSPPEEHRPFKEV
VNKESLVISGLRHFTGYRIELQACNQDTPPEERCSVAAYVSARTMPEAKADDIVGPTHEIFENNVVHLMQELKEPNGLIVLYEVSYRR
YGDDELHLCVSRKHFALERGCRRLGLSPGNYSVRIRATSLAGNSWTEPTYFYVTDYLDVPSNIAKIIIGPLIFVFLFSVVISIYLF
RKRQPDGPGPLGYASSNPEYLSASDVFPSCSVYVPEDEWEVSREKITLLRELQGSFGMVYEGNARDIVKGEAETRVAVKTVNESASLRER
IEFLNQASVMKGFTHHHVRLGLGVSKGQPTLVVMEMLMAHGDLKSYLRSLRPEAENNPGRPPPTLQEMIQMAAEIADGMAYLNKKFVH
RDLAARNCMVAHDFTVKIGDFGMDTRDIYETDYRKGGKGLLPVRWMAPESLKDGVTTSDDMWSFGVVLWEITSLAEQPYQGLSNEQVL
KFVMDGGYLDQPDNCPERVTDLMRMCWQFNPKMRPTFLEIVNLLKDDLHPSFPEVFSFHSEENKAPSESELEMEFEDMENVPLDRSSHC
QREEAGGRDGGSSLGFKRSYEEHIPYTHMNGGKNGRILTLPRSNPS
```

La sequenza del recettore del paziente, da sola, è di utilità limitata: per poterla interpretare, è necessario confrontarla con una sequenza di riferimento (cioè “normale”), che si può trovare in un database bioinformatico. Nei database bioinformatici, a ciascuna sequenza è associato un “Accession number” (AN), che permette di identificarla univocamente. L'AN della sequenza amminoacidica di riferimento del recettore dell'insulina umano è **NP_000199.2**.

Per confrontare la sequenza ottenuta dal paziente con la sequenza di riferimento, decidi di utilizzare uno strumento bioinformatico chiamato BLAST (Basic Local Alignment Search Tool). Puoi accedere a questo strumento dall'indirizzo <https://blast.ncbi.nlm.nih.gov/Blast.cgi>.

- Fai click sul pulsante “Protein BLAST” a destra, che aprirà l'interfaccia dello strumento.
- Seleziona la casella “Align two or more sequences”. A questo punto, l'interfaccia risulterà composta da tre sezioni (“Enter Query Sequence”, “Enter Subject Sequence”, “Program Selection”):



Bioinformatics practical 2022: insulin receptor

Query	301	CHQYVIHNNKCIPECPSGYTMSSNLLCTPCLGPCPKVCHILEGEKTIDSVTSAQELRGC	360
Sbjct	301L.....	360
Query	361	TVINGSLIINIRGGNNLAAELEANLGLIEEISGYLKIRRSYALVLSFFRKRLIRGETL	420
Sbjct	361	420
Query	421	EIGNYSFYALDNQNLRLQLDWSKHNLITITQGKLFFHYNPKLCLSEIHKMEEVSGTKGRQE	480
Sbjct	421	480
Query	481	RNDIALKTNGDQASCENELKFSYIRTSFDKILLRWEYPWPPDFRDLLGFMLFYKEAPYQ	540
Sbjct	481	540
Query	541	NVTEFDGQDACGSNSWTVDIDPPLRSNDPKSQNHGWLMRGLKPWTQYAI FVKTLVTFS	600
Sbjct	541	600
Query	601	DERRTYGAKSDIIYVQTDATNPSVPLDPIVSNSSSQIILKWKPPSDPNGNITHYLVFE	660
Sbjct	601	660
Query	661	RQAEDSELFELDYLKGLKLPRTWSPPFESQKHQSEYEDSAGECCSCPKTDSQIL	720
Sbjct	661	720
Query	721	KELEESSFRKTFEDYLHNWVFVPRKTSSTGAEDPRPSRKRSLGVDGNVTVVPTVAAF	780
Sbjct	721A.....	780
Query	781	PNTSSTSTPTSPPEHRPFQKVNKESLVISGLRHFTGYRIELQACNQDTPEERCSVAAYV	840
Sbjct	781V.....	840
Query	841	SARTMPEAKADDIVGPVTHEIFENNVLHMLWQELKEPNGLIVLYEVSYYRYGDEELHLCV	900
Sbjct	841P.....	900
Query	901	SRKHFALEGRGLRGLSPGNYSVRIRATSLAGNSWTEPTYFYVTDYLDVPSNIAKIIIG	960
Sbjct	901	960
Query	961	PLIFVFLFSVWIGSIYFLRKRQPDGPLYASSNPEYLSASDVFPSCVYVPDEWEVSR	1020
Sbjct	961	1020
Query	1021	EKITLLRELQGSFGMVYEGNARDIVKGEAETRVAVKTVNESASLRERIEFLNQASVMKG	1080
Sbjct	1021I.....E.....	1080
Query	1081	FTCHHVRLLGVSQKQPTLVVLMELMAHGDLSYLRSLRPEAENNPGRPPPTLQEMIQMA	1140
Sbjct	1081	1140

Domanda 19. Le differenze tra la sequenza del paziente e la sequenza di riferimento sono dovute ad alcune mutazioni presenti nel genoma del paziente. Per ciascuna di esse, indica l'amminoacido nella sequenza di riferimento, l'amminoacido nella sequenza del paziente, e la posizione all'interno della sequenza (nota che con le impostazioni di default BLAST mostra 60 amminoacidi per riga – ad esempio, la seconda riga mostra gli amminoacidi da 61 a 120 inclusi).

(1.5 punti per ogni risposta corretta; 0.75 punti se gli amminoacidi sono corretti ma la posizione è ± 2)

Domanda 20. Qual è il tipo delle mutazioni che hai identificato?

(1 punto se corretta, -0.5 punti se errata)

- a. Mutazioni sinonime
- b. Mutazioni missenso
- c. Mutazioni non senso

Bioinformatics practical 2022: insulin receptor

Questa analisi ti ha permesso di identificare alcune mutazioni nel paziente rispetto alla sequenza di riferimento, che potrebbero essere alla base della patologia di cui soffre. Tuttavia, è anche possibile che le mutazioni non influiscano sulla funzione del recettore. Per ottenere ulteriori informazioni riguardo a queste mutazioni, decidi di confrontare la sequenza del recettore dell'insulina umana con quella di alcune altre specie di mammifero.

Per fare ciò, hai bisogno degli AN delle sequenze di riferimento per queste altre specie. Per ottenerli, puoi utilizzare ancora una volta BLAST.

- Accedi nuovamente a BLAST dall'indirizzo <https://blast.ncbi.nlm.nih.gov/Blast.cgi>, quindi fai click sul pulsante "Protein BLAST".
- Deseleziona la casella "Align two or more sequences". L'interfaccia risulterà composta da tre sezioni ("Enter Query Sequence", "Choose Search Set" e "Program Selection").
- All'interno della grande casella di testo nella sezione "Enter Query Sequence" inserisci l'AN della sequenza di riferimento umana.
- Nel campo "Organism", che si trova nella sottosezione "Standard" della sezione "Choose Search Set", inizia a scrivere "*Felis catus*". Mentre scrivi, si aprirà un menu con dei suggerimenti; quando vedi comparire il suggerimento "*Felis catus* (taxid:9685)", cliccaci sopra.
- Fai click sul pulsante blu "BLAST" in basso. Ancora una volta, la pagina si aggiornerà; attendi fino a quando vengono visualizzati i risultati della ricerca.

Con questa ricerca, hai cercato tutte le proteine "simili" alla sequenza di riferimento umana, all'interno del genoma di *Felis catus* (ovvero, del gatto domestico). BLAST mostrerà un gran numero di risultati, ordinati in base alla loro significatività statistica. Per ciascun risultato, BLAST mostra una serie di informazioni, incluso l'AN, che si trova nella colonna all'estrema destra.

Information: Your search is limited to records that include: *Felis catus* (taxid:9685)

Job Title: Protein Sequence

RID: 85WD9UDY016 Search expires on 07-03 02:12 am [Download All](#) ▼

Program: BLASTP [Citation](#) ▼

Database: nr [See details](#) ▼

Query ID: lcl|Query_8237276

Description: unnamed protein product

Molecule type: amino acid

Query Length: 1382

Other reports: [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#) ⓘ

Filter Results

Organism: only top 20 will appear ☐ exclude
Type common name, binomial, taxid or group name
[+ Add organism](#)

Percent Identity: to **E value:** to **Query Coverage:** to

[Filter](#) [Reset](#)

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments [Download](#) ▼ [Select columns](#) ▼ Show 100 ▼ ⓘ

☒ select all 100 sequences selected [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#) [MSA Viewer](#)

	Description ▼	Scientific Name ▼	Max Score ▼	Total Score ▼	Query Cover ▼	E value ▼	Per Ident ▼	Acc. Len ▼	Accession
<input checked="" type="checkbox"/>	insulin receptor isoform X1 [Felis catus]	Felis catus	2778	2778	100%	0.0	95.88%	1381	XP_023100560.2
<input checked="" type="checkbox"/>	insulin receptor isoform X2 [Felis catus]	Felis catus	2769	2769	100%	0.0	95.73%	1380	XP_023100561.2
<input checked="" type="checkbox"/>	insulin receptor isoform X3 [Felis catus]	Felis catus	2690	2690	98%	0.0	95.65%	1369	XP_023100562.2

Bioinformatics practical 2022: insulin receptor

Per ottenere le sequenze di riferimento per le altre specie di mammifero, dovresti ripetere questi passaggi. Poiché il tempo a disposizione è limitato, le riportiamo invece qui sotto:

Specie	Nome comune	Accession number
<i>Equus caballus</i>	Cavallo domestico	XP_023500375.1
<i>Sus scrofa</i>	Maiale	XP_020939599.1
<i>Mus musculus</i>	Topolino comune	XP_006508763.1
<i>Canis lupus dingo</i>	Dingo	XP_025313168.2
<i>Rhinolophus ferrumequinum</i>	Pipistrello ferro di cavallo maggiore	XP_032990230.1
<i>Loxodonta africana</i>	Elefante africano	XP_023396235.1
<i>Ornithorhynchus anatinus</i>	Ornitorinco	XP_028906070.1
<i>Gracilinanus agilis</i>	Opossum agile e gracile (un marsupiale sudamericano)	XP_044541320.1

Puoi finalmente confrontare la sequenza del paziente e la sequenza di riferimento umana con le sequenze di questi altri mammiferi, usando nuovamente BLAST.

[Download](#) ▼

Query range 14: 781 to 840

Query	781	PNTSSTSTPTSPPEEHRPFKEKVNKESLVISGLRHFTGYRIELQACNQDTPEERCSVAAYV	840
XP_023500375.1	781S..R.....S.....	840
XP_032990230.1	781	..V.....S.....M.....S.....	840
XP_025313168.2	744K.....S.....	802
XP_020939599.1	781	S.....S.....R.....A.....	840
XP_006508763.1	783	..V...IV...Q.....S.D.....	842
XP_023396235.1	764	..V...RV...A.....V.....	823
XP_028906070.1	777	S.....PA..E.R..QK.....ES.....H...H.AQ.S.....	836
XP_044541320.1	841	F.A...TA.A.T..PK.....KS.....Q.....H.EQ.S..GL....	900

[Download](#) ▼

Query range 15: 841 to 900

Query	841	SARTMPEAKADDIVGPVTHEIFENNVVHLMWELKPNGLIVLYEVSRYRYGDEELHLCV	900
XP_023500375.1	841P.....	900
XP_032990230.1	841P.....	900
XP_025313168.2	803P.....	862
XP_020939599.1	841I.....P.....	900
XP_006508763.1	843P.....	902
XP_023396235.1	824P.....	883
XP_028906070.1	837VSD..I...K...P.....N.W.L.EQ...H...	907

XP_044541320.1	901A...VL.....R...P.H.....I...N.W.V.....	961
--------------------------------	-----	--	-----

NEENMLPFSFL

|

T

Bioinformatics practical 2022: insulin receptor

- ♦ **Challenge 1:** students + web browser = cheating
 - ♦ **Solution 1:** students were at their schools with a supervising teacher and were also remotely supervised by us via video call.
 - ♦ **Solution 2:** a “data mining” test – students had to reason on the sequence and generate data.
- ♦ **Challenge 2:** in a school computer lab, we cannot use custom software
 - ♦ **Solution:** use a browser-based test
- ♦ **Challenge 3:** bioinformatics pre-requirements
 - ♦ **Solution:** we provided very detailed instructions with screenshots.
- ♦ **Opportunity:** we managed to fit a bit of evolutionary biology in the test.

Bioinformatics practical 2023: α -amylase

- This test was part 2 of the Biochemistry task.
- In part 1, students were asked to measure the activity of α -amylase using Lugol's solution, in three conditions: **native enzyme**, enzyme **incubated with EDTA**, enzyme **incubated with EDTA + CaCl₂**. They were also given data about the enzyme activity when **incubated with EDTA** and other divalent ions.
 - The goal was for the students to understand that α -amylase requires a Ca²⁺ ion.
- In the bioinformatics part, the students had to use **UniProt** and **PDB** to gather information about the enzyme.

Bioinformatics practical 2023: α -amylase

Parte 2.1 – UniProt

Accedi al database UniProt digitando l'indirizzo <https://www.uniprot.org/> nel browser. La pagina iniziale del database contiene una casella di testo in cui puoi inserire il nome di una proteina (in inglese) per ottenere informazioni al riguardo. Scrivi in questa casella “**Salivary Amylase**” (senza virgolette) e fai click su “**Search**”. Se la ricerca non produce risultati, controlla di non aver commesso errori di battitura.

La pagina si aggiornerà, mostrando i risultati della ricerca riguardante l'amilasi salivare sotto forma di tabella. Le colonne della tabella includono varie informazioni, tra cui un codice che identifica univocamente ciascuna voce del database (“Entry”), il nome del gene e della proteina, il nome scientifico dell'organismo in cui è stata identificata, e la lunghezza della sequenza amminoacidica.

- Completa la seguente tabella, indicando il codice (Entry) e l'organismo corrispondente per l'amilasi salivare di quattro specie diverse (4 punti).

Entry	Genere	Specie

Ora, fai click sull'Entry del primo risultato (P0DUB6). La finestra si aggiornerà mostrando alcune informazioni riguardo alla proteina.

- Qual è il nome del gene che codifica per questa proteina? (1 punto)

- Da quanti amminoacidi è composta la sequenza della proteina? (1 punto)

P0DUB6 · AMY1A_HUMAN

Protein ¹	Alpha-amylase 1A	Amino acids	511 (go to sequence)
Gene ¹	AMY1A	Protein existence ¹	Evidence at protein level
Status ¹	UniProtKB reviewed (Swiss-Prot)	Annotation score ¹	
Organism ¹	Homo sapiens (Human)		

Entry Variant viewer **53** Feature viewer Genomic coordinates Publications External links History

BLAST Download Add Add a publication Entry feedback

Function¹

Calcium-binding enzyme that initiates starch digestion in the oral cavity (PubMed:12527308). Catalyzes the hydrolysis of internal (1->4)-alpha-D-glucosidic bonds, yielding a mixture of maltose, isomaltose, small amounts of glucose as well as small linear and branched oligosaccharides called dextrins (PubMed:12527308).

Caution
Three distinct genes (AMY1A, AMY1B and AMY1C), located in a gene cluster on 1p21, encode proteins sharing the same peptidic sequence.

Catalytic activity¹

Endohydrolysis of (1->4)-alpha-D-glucosidic linkages in polysaccharides containing three or more (1->4)-alpha-linked D-glucose units.
EC:3.2.1.1 (UniProtKB | ENZYME RHEA)

Cofactor¹

Protein has several cofactor binding sites:

Ca²⁺ (UniProtKB | RHEA | CHEBI:29108)
Note: Binds 1 Ca²⁺ ion per subunit.

chloride (UniProtKB | RHEA | CHEBI:17996)
Note: Binds 1 Cl⁻ ion per subunit.

Features

Showing features for binding site¹, active site¹, site¹.

Download

1 50 100 150 200 250 300 350 400 450 500 511

| | | | |

Bioinformatics practical 2023: α -amylase

Inizia a scorrere questa pagina. Nella sezione “**Function**” sono indicate informazioni riguardo alla funzione dell’enzima, tra cui i suoi cofattori (“cofactor”, piccole molecole che si associano all’enzima e ne rendono possibile la funzione catalitica), i siti di legame per varie molecole (“binding site”), il sito attivo (“active site”), e altri amminoacidi interessanti (“site”).

5. Completa la seguente tabella, indicando per ciascun cofattore dell’enzima:
- Il nome o la formula chimica
 - Il numero di unità (molecole o ioni) di cofattore che si legano a ciascuna molecola di enzima
 - La posizione nella sequenza dell’amminoacido (o degli amminoacidi) con cui interagisce

Nota che la tabella potrebbe avere più righe del necessario (4 punti).

Cofattore	Unità cofattore / enzima	Amminoacidi coinvolti

Continua a scorrere la pagina, individuando le informazioni necessarie a rispondere alle seguenti domande:

6. Su quale cromosoma si trova il gene che codifica per questa proteina? (1 punto)

7. Una volta che l’enzima viene prodotto, dove si localizza nella cellula? (1 punto)

8. Da quante subunità è formato l’enzima completo? (1 punto)

Nella sezione “**PTM/Processing**” sono riportate informazioni riguardo alle modifiche post-traduzionali (PTM = post-translational modifications) subite dall’enzima; queste includono alcuni amminoacidi che vengono modificati in vario modo, la formazione di ponti disolfuro, e un peptide segnale.

9. Indica la posizione degli amminoacidi che formano il peptide segnale. (1 punto)

10. Qual è la funzione di questo peptide segnale? (2 punti)

Nella sezione “**Structure**” sono presenti alcuni link a database che contengono informazioni riguardo alla struttura della proteina in esame. Sono presenti varie strutture diverse, che sono state realizzate in studi differenti con varie metodologie. Un parametro importante per giudicare la qualità di una struttura è la risoluzione (“Resolution”), misurata in ångström ($1 \text{ \AA} = 10^{-10} \text{ m}$); in prima approssimazione, puoi assumere che essa rappresenti la minima distanza tra due atomi che appaiono come entità distinte all’interno della struttura.

11. Indica il numero identificativo (“Identifier”) e la risoluzione della struttura di questa proteina con miglior risoluzione. (1 punto)

Numero identificativo	Risoluzione

Bioinformatics practical 2023: α -amylase

Parte 2.2 – PDB

Ora esamineremo la struttura tridimensionale della proteina utilizzando il database PDB. Puoi chiudere la finestra di UniProt, oppure aprire un'altra scheda. Puoi accedere al database PDB digitando l'indirizzo <https://www.rcsb.org/> nella barra degli indirizzi del browser.

La home page di PDB è più affollata rispetto a quella di UniProt, ma puoi notare in alto a destra una casella di testo in cui inserire il termine da cercare. Scrivi **1C8Q** (questo è uno dei numeri identificativi che erano presenti nella tabella su UniProt) e fai click sull'icona con la lente d'ingrandimento per avviare la ricerca.

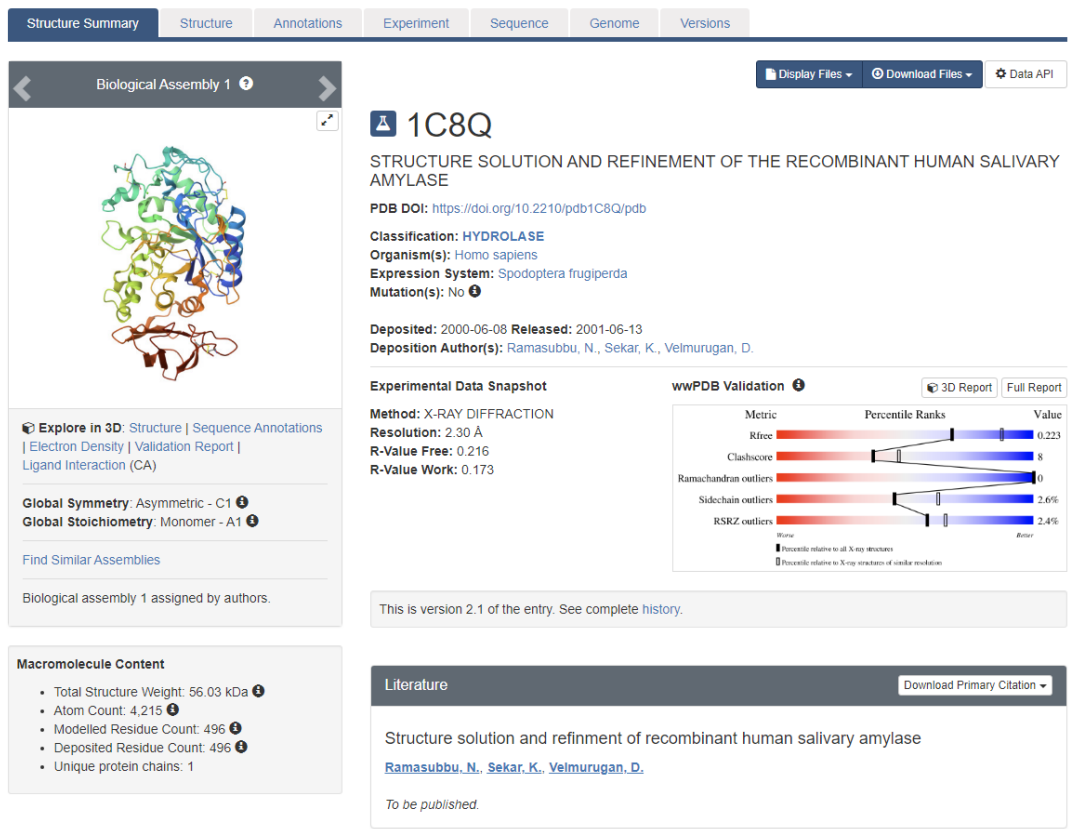
A questo punto, la pagina si aggiornerà, mostrando la struttura con il numero identificativo che hai inserito. La pagina iniziale contiene informazioni generali riguardo a questa struttura.

12. Qual è il peso molecolare di questa struttura? (0.5 punti)

13. Da quanti amminoacidi è composta la struttura? (0.5 punti)

14. Il numero di amminoacidi da cui è composta questa struttura è lo stesso rispetto alla proteina che abbiamo analizzato su UniProt? Se no, perché? (1 punto)

Nella parte sinistra della pagina, puoi notare una rappresentazione tridimensionale della proteina; se fai click sul link “**Structure**” accanto a “3D View”, si aprirà una pagina con una versione ingrandita e interattiva della proteina.



The screenshot displays the PDB entry page for 1C8Q. The top navigation bar includes tabs for Structure Summary, Structure, Annotations, Experiment, Sequence, Genome, and Versions. The main content area is divided into several sections:

- Biological Assembly 1:** A 3D ribbon diagram of the protein structure.
- 1C8Q:** The entry ID and title: "STRUCTURE SOLUTION AND REFINEMENT OF THE RECOMBINANT HUMAN SALIVARY AMYLASE".
- PDB DOI:** <https://doi.org/10.2210/pdb1C8Q/pdb>
- Classification:** HYDROLASE
- Organism(s):** Homo sapiens
- Expression System:** Spodoptera frugiperda
- Mutation(s):** No
- Deposited:** 2000-06-08 **Released:** 2001-06-13
- Deposition Author(s):** Ramasubbu, N., Sekar, K., Velmurugan, D.
- Experimental Data Snapshot:** Method: X-RAY DIFFRACTION, Resolution: 2.30 Å, R-Value Free: 0.216, R-Value Work: 0.173.
- wwPDB Validation:** A bar chart showing percentile ranks for various metrics: Rfree (0.223), Clashscore (8), Ramachandran outliers (0), Sidechain outliers (2.6%), and RSRZ outliers (2.4%).
- Global Symmetry:** Asymmetric - C1
- Global Stoichiometry:** Monomer - A1
- Find Similar Assemblies:** A link to find similar structures.
- Biological assembly 1 assigned by authors.**
- Macromolecule Content:** Total Structure Weight: 56.03 kDa, Atom Count: 4,215, Modelled Residue Count: 496, Deposited Residue Count: 496, Unique protein chains: 1.
- Literature:** A section for the primary citation, titled "Structure solution and refinement of recombinant human salivary amylase" by Ramasubbu, N., Sekar, K., Velmurugan, D.

Bioinformatics practical 2023: α -amylase

La proteina è rappresentata in verde scuro, mentre alcune altre molecole sono rappresentate da delle “palline” di colori vari. Passando il mouse sulla struttura o sulle “palline”, in basso a destra comparirà uno specchietto che dice su quale amminoacido o su quale molecola accessoria è posizionato il mouse. Puoi fare zoom utilizzando la rotella del mouse, ruotare la vista trascinando con il pulsante sinistro del mouse premuto, o spostare la telecamera trascinando con il pulsante destro del mouse premuto.

Il tuo obiettivo è localizzare lo ione calcio (rappresentato da una pallina verde chiaro) che è presente in questa struttura. Per aiutarti, puoi nascondere gli altri elementi della struttura facendo click con i simboli con l'occhio (👁️) nella parte destra dell'interfaccia. Una volta che hai localizzato lo ione calcio, fai click su di esso.

L'interfaccia zoomerà sullo ione calcio, evidenziando le altre molecole contenute nella struttura con cui esso interagisce. Le interazioni dirette dello ione sono rappresentate da bastoncini tratteggiati color viola (che partono dallo ione), mentre altre interazioni che avvengono nelle vicinanze sono rappresentate da bastoncini tratteggiati color verde acqua.

Ruota o sposta la telecamera per visualizzare tutti gli amminoacidi della proteina con cui interagisce lo ione calcio.

1. Con quali amminoacidi interagisce lo ione calcio? Indica per ciascun amminoacido il codice a tre lettere e la posizione nella sequenza, ad esempio “GLY 100”. (4 punti)

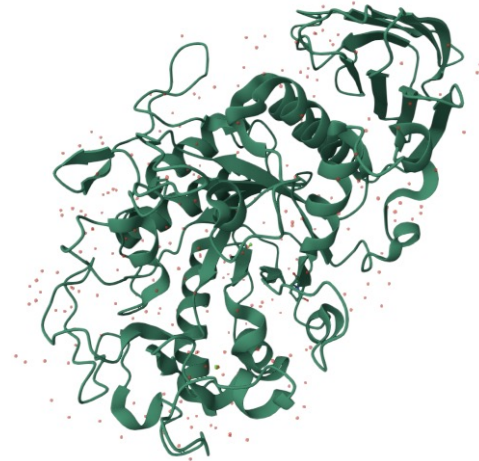
2. La pagina seguente mostra le strutture dei 20 amminoacidi standard. Per ciascuno degli amminoacidi con cui interagisce lo ione calcio, cerchia tutti gli atomi dell'amminoacido con cui interagisce. Se lo ione calcio interagisce con più amminoacidi dello stesso tipo, cerchia comunque una sola volta ciascun atomo coinvolto. (5 punti)

1C8Q

STRUCTURE SOLUTION AND REFINEMENT OF THE RECOMBINANT HUMAN SALIVARY AMYLASE

Sequence of 1C8Q | STRUC... Chain 1: ALPHA-AM... A

PCA YSSNTQQGRISIVHLFEWRWVDIALECEVYLAPKGFQGVQVSPNENVAIHNFFRPMWERYQFVSVKICTRSGNEDEFNNAVTRCNVGVRIYVDVINHMOGNVAVSAG
TSSTCGSYFNGSRDFFPAVTSQWDFNDGKCKTSGSDIENVNDATQVRDCLSGLLDLALGQDVRSKIAEYNNHLIDIGVAGFRIDASKHWPQDIKAILDKLNLNSNWF
PEGSKPFTIQEVIDLGGSEFIKSSDYFQNGRVTEFKYGARLGTIVIRSNWSEMSYLNQWGEQWGFMPSDRALVFVDNHDNQSGHGAGGASLITFWDARLYRQAVGFMLAHFYG



Bioinformatics practical 2023: α -amylase

- ♦ **Challenge 1:** students + web browser = cheating
 - ♦ **Solution 1:** small number of students (10), and someone supervising them at all times.
 - ♦ **Solution 2:** data mining test.
 - ♦ **Solution 3:** tight timings.
- ♦ **Challenge 2:** in a school computer lab, we cannot use custom software
 - ♦ **Solution:** we used a browser-based test.
- ♦ **Challenge 3:** bioinformatics pre-requirements
 - ♦ **Solution:** we provided very detailed instructions. Someone was available to help students if necessary.

Bioinformatics practical 2023: α -amylase

- **Unforeseen challenge:** VERY old PCs in the computer lab!
- Computers had a Core i3-2100 processor (discontinued in 2013).
- They had been updated to Windows 10, for which there are no compatible graphics drivers.
- Deceivingly, the computers appeared to work *mostly* fine...
- When we did a preliminary test, we coincidentally used one of the few computers that had a separate GPU (ironically, because the integrated GPU had previously failed).
- **When students arrived at the structure visualisation phase, most of the computers did not work!**

Bioinformatics practical 2023: α -amylase

- We **frantically tested all PCs** in the computer lab, and we managed to find **three** that worked (we had groups of 10 students).
- At the same time, we tried to gather as many laptops as possible...
 - Side problem: laptop touchpads are not great for moving structures around, so we also had to move mice from the desktop PCs...
- We managed to get 7 working PCs. Students started the test on a random computer and were then moved to a working PC when they arrived at the final question that required it.

Bioinformatics practical 2023: α -amylase

- In the end it was mostly fine – we gave **an extra 5 minutes** to everyone to account for the disruption.
- There was **only one case** of a student that actually had to wait before accessing a working PC (who was given extra time).
- **Key takeaway:** if you are doing anything with PCs, test ALL of them beforehand.

Bioinformatics practical 2024: genome assembly

- This year, Bioinformatics was a **separate task**, lasting 45 minutes.
- Having been traumatised by the previous year, we decided to go for a paper-based test.
- The task involved assembling a (small) bacterial genome by hand.

Assemblare un genoma

Nel corso degli ultimi vent'anni, notevoli innovazioni dal punto di vista delle tecniche sperimentali e del software bioinformatico hanno reso il sequenziamento di un genoma batterico una procedura *di routine*. Più di due milioni di genomi batterici sono stati sequenziati, utilizzando un approccio "**shotgun**", che prevede generalmente i seguenti passaggi:

- **Estrazione del DNA:** Il DNA genomico viene estratto da un campione.
- **Preparazione di una "library":** il DNA viene diviso in una serie di frammenti di piccola dimensione (se richiesto dalla tecnica di sequenziamento, questo passaggio coinvolge anche l'aggiunta di "adattatori" alle estremità dei frammenti).
- **Sequenziamento high-throughput:** i frammenti vengono sequenziati in parallelo, utilizzando varie tecniche. Ciò produce una serie di piccole sequenze, dette "reads".
- **Assemblaggio:** software bioinformatici sfruttano la sovrapposizione tra le reads per ricostruire la sequenza del genoma originale.

In questa prova, il tuo compito sarà assemblare un genoma batterico a partire da una lista di reads. Nota in particolare che le reads che utilizzerai sono "**paired-end**"; ovvero, per ciascun frammento di DNA nella library, sono state sequenziate entrambe le estremità (**Figura 1**).



Figura 1. Reads paired-end.
Un inserto (frammento di DNA) viene sequenziato a entrambe le estremità. La forward read è quella blu (5'-AATCTAA-3'); la reverse read è quella arancione (5'-TACATT-3').

La tabella alla pagina seguente mostra la lista delle reads che sono state ottenute da un processo di sequenziamento. Per semplicità, assumi che tutte le forward reads provengano dallo stesso filamento del DNA genomico, mentre le reverse reads sono state tutte sequenziate sul filamento di DNA a esso complementare. Assumi anche l'assenza di errori di sequenziamento.

Bioinformatics practical 2024: genome assembly

1. Per ciascuna reverse read, indica la sequenza di DNA corrispondente che si trova sul filamento di DNA su cui sono state sequenziate le forward reads (in direzione 5'-3').

Forward read (sul filamento forward)	Reverse read (sul filamento reverse)	Reverse read (sul filamento forward)
5' - CACA - 3'	5' - CACC - 3'	[0.25pt]
5' - GTGA - 3'	5' - GTGT - 3'	[0.25pt]
5' - ACGG - 3'	5' - TGTC - 3'	[0.25pt]

Per assemblare il genoma, utilizzeremo uno strumento matematico chiamato **grafo di de Bruijn**. Il primo passaggio per la costruzione di un grafo di de Bruijn è la costruzione dei *k*-meri, ovvero di tutte le sequenze di nucleotidi lunghe *k* che si possono osservare nelle reads. Il valore di *k* influisce sull'accuratezza del processo di assemblaggio e sulla sensibilità a vari tipi di artefatti; per questa prova, utilizzeremo un valore di *k* = 3.

2. Elenca tutti i 3-meri osservati nelle reads (usa tutte le sei reads sul filamento forward; se un 3-mero è osservato più di una volta, elencalo comunque una sola volta). [1 punto in totale, nota che ci sono più spazi del necessario]

Il passaggio successivo è l'individuazione di tutti i prefissi e i suffissi dei *k*-meri. Il **prefisso** di un *k*-mero è la sequenza di nucleotidi lunga *k* - 1 che si ottiene rimuovendo l'ultimo nucleotide dal *k*-mero. Analogamente, il **suffisso** è la sequenza che si ottiene rimuovendo il primo nucleotide del *k*-mero.

3. Elenca tutti i prefissi e i suffissi che puoi ottenere a partire dai *k*-meri precedentemente individuati. Non distinguere tra prefissi e suffissi, e includi ciascuna sequenza una sola volta. [1 punto in totale, nota che ci sono più spazi del necessario]

A questo punto, possiamo effettivamente costruire il grafo di de Bruijn. Un grafo di de Bruijn è formato da **nodi**, costituiti dai prefissi e dai suffissi, e da **archi**, costituiti dai *k*-meri. Ciascun arco (*k*-mero) collega i due nodi corrispondenti al prefisso e al suffisso del *k*-mero, ed è **orientato** dal prefisso verso il suffisso (ovvero, ha una punta di freccia che indica verso il suffisso). La **Figura 2** alla pagina successiva rappresenta un grafo di de Bruijn analogo a quello che devi creare.

Da questo punto in poi, ignora tutti i dati che hai utilizzato nelle domande 1-4, e utilizza i dati forniti di seguito.

	Forward reads	Reverse reads (sul filamento forward)	K-meri	Prefissi/suffissi
1	5' - TGTC - 3'	5' - TCGA - 3'	TGT	CA
2	5' - TCAT - 3'	5' - GAAT - 3'	GTC	GT
3	5' - ATCG - 3'	5' - ATGT - 3'	TCG	TC
			CGA	TG
			GAA	AA
			AAT	AT
			TCA	GA
			CAT	CG
			ATC	
			ATG	

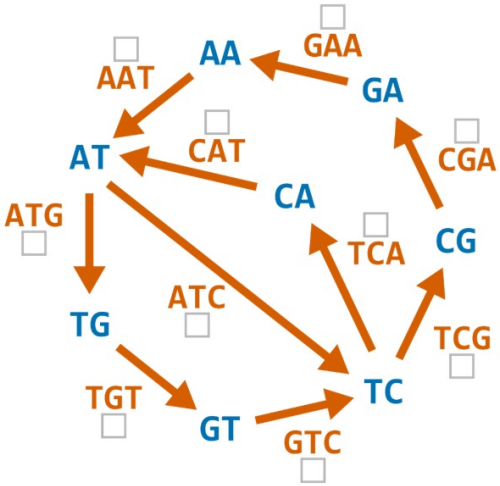


Figura 2. Grafo di de Bruijn.

Bioinformatics practical 2024: genome assembly

Procedimento

- Inizializzazione:
 - Prendi un foglio a quadretti. Scrivi "Inizio " nella prima riga disponibile, che sarà la "riga corrente".
 - Nel grafo, parti da un nodo qualsiasi, **eccetto che AT e TC**, e posiziona il "cursore" su di esso (che sarà il "nodo corrente").
- Nel nodo corrente, considera le frecce che non sono state marcate ("frecce attive"). Quante sono?
 - 0: vai al passaggio 7.
 - 1: "scegli" l'unica freccia attiva e vai al passaggio 3.
 - >1:
 - Disegna un pallino (•) sulla riga corrente sul foglio a quadretti, per indicare che in corrispondenza di questa riga c'è stata una biforcazione.
 - Scegli una freccia attiva qualsiasi e vai al passaggio 3.
- Scrivi sulla riga corrente la sequenza del k -mero corrispondente alla freccia che hai scelto, posizionando i nucleotidi in comune con la sequenza sulla riga precedente in modo che siano allineati verticalmente con essa, come nell'esempio qui a destra (eccetto che nel primo passaggio, ovviamente).

TAG
AGC
GCT
- Sul grafo, marca con una crocetta nel quadratino la freccia che hai scelto.
- Sposta il "cursore" sul nodo verso cui puntava la freccia che hai scelto e, sul foglio a quadretti, vai alla riga successiva.
- Vai al passaggio 2.
- Sei tornato al nodo di partenza?
 - Sì: vai al passaggio 8.
 - No: hai sbagliato qualcosa. Ricontrolla tutti i passaggi, oppure cancella tutto e ricomincia da capo facendo più attenzione.
- Ci sono ancora frecce attive nel grafo?
 - Sì: Il cammino che hai percorso non è un cammino euleriano (perché non passa da tutte le frecce). Vai al passaggio 9 per risolvere.
 - No: Hai trovato un cammino euleriano! Vai al passaggio 10.
- Il problema è che hai scelto la strada sbagliata nell'ultima biforcazione del grafo. Per tornare indietro fino all'ultima biforcazione, devi essenzialmente "annullare" tutti i passaggi che hai effettuato a partire dall'ultimo pallino che hai indicato sul foglio a quadretti:
 - "Riattiva" tutte le frecce che avevi scelto cancellando la loro crocetta dal grafo (utilizza i k -meri che hai scritto sul foglio a quadretti per ricordarti quali sono), fino alla biforcazione.
 - Cancella dal foglio a quadretti i k -meri corrispondenti alle frecce che hai riattivato. Cancella anche il pallino dalla riga con la biforcazione.
 - Posiziona il cursore sulla biforcazione.
 - Scegli la freccia che NON avevi seguito precedentemente e vai al passaggio 3 per continuare.
- Conclusioni:
 - Scrivi sulla riga corrente la sequenza del nodo di partenza, allineando i nucleotidi in comune con la sequenza sulla riga precedente. Scrivi " Fine" accanto alla sequenza.
 - Leggi la sequenza del genoma a partire dal primo nucleotide della riga iniziale e indicala nello spazio apposito della domanda 5. La sequenza che devi indicare **NON** include i nucleotidi del nodo di partenza che hai allineato nella riga finale. Scrivi "Sequenza A" o "Sequenza B" accanto alla parola "Fine" sul foglio a quadretti.
 - A seconda del percorso scelto a ciascuna biforcazione, è possibile ottenere sequenze genomiche differenti. Ripeti il processo una seconda volta, utilizzando lo stesso nodo di partenza, ma seguendo un percorso diverso a partire dalla **prima** biforcazione. Per fare ciò:
 - Cancella tutte le crocette dal grafo, in modo da riattivare tutte le frecce.
 - Parti dallo stesso nodo e ripeti tutti i passaggi che hai già effettuato, fino alla prima riga con il pallino.
 - In corrispondenza di questa biforcazione, scegli la freccia che NON avevi seguito precedentemente, poi vai al passaggio 3 per continuare.

5. Assembla il genoma utilizzando il processo indicato e il grafo di de Bruijn nella **Figura 2**. Come indicato nel passaggio 10c, segui il processo due volte e indica due sequenze distinte. Consegna anche i fogli con i passaggi intermedi.

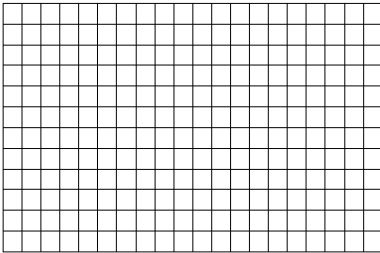
A: _____ [1.5 punti]

B: _____ [1.5 punti]

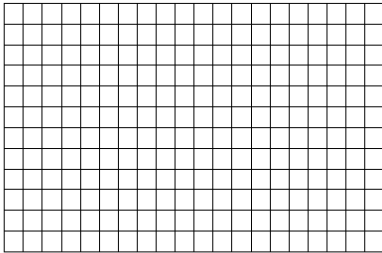
6. Nel passaggio 8, il cursore si trova nel nodo di partenza. Indica due caratteristiche del genoma di un batterio che fanno sì che ciò sia possibile. [0.75 punti]

7. Allinea ciascuna forward e reverse read con le sequenze A e B. Se non è possibile allineare una read con le sequenze che hai ottenuto, indicalo esplicitamente. Per rappresentare l'allineamento, scrivi la sequenza A o B su di una riga, e le reads sotto di essa. Se possibile, rappresenta le reads di ciascuna coppia sulla stessa riga. [1.5 punti, nota che c'è più spazio del necessario]

A



B

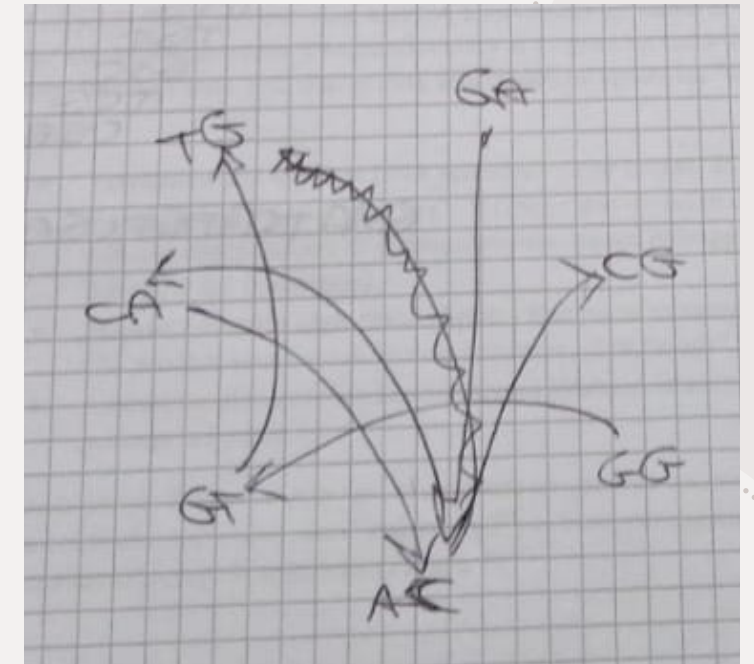
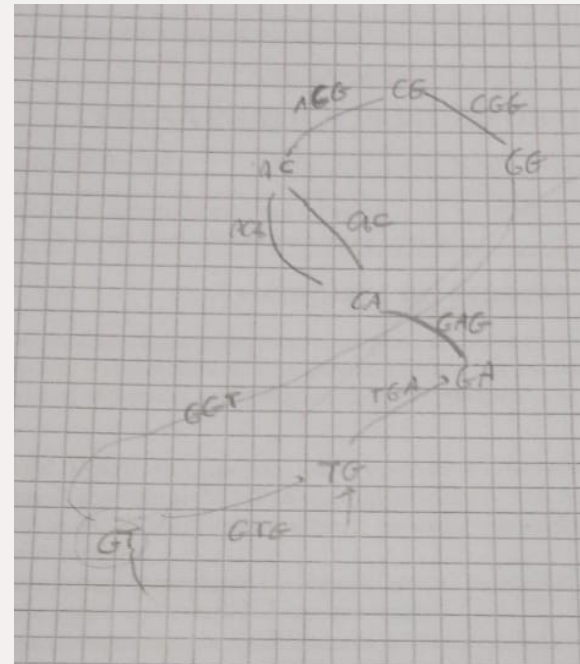
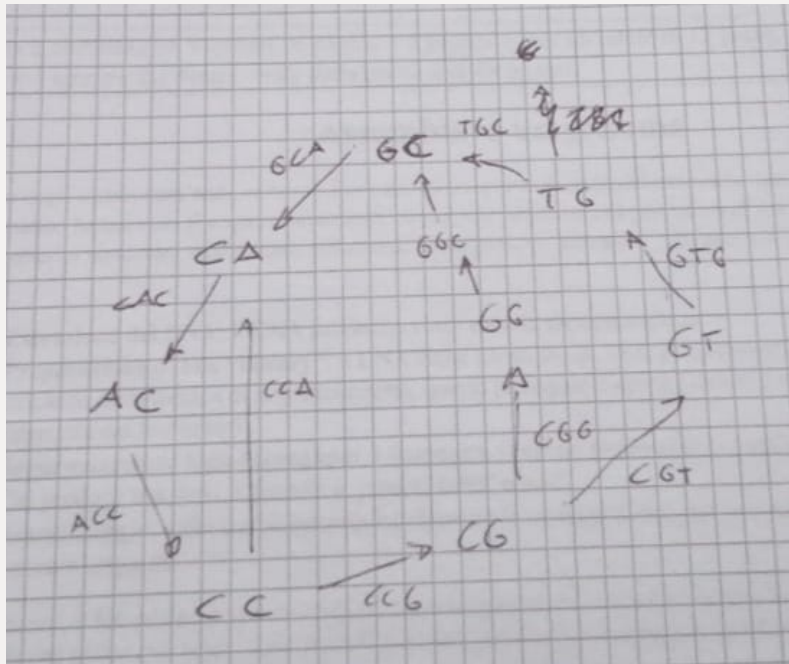


8. Qual è la dimensione degli inserti (v. **Figura 1**)? [1 punti]

9. Quale potrebbe la sequenza di DNA corretta del genoma del batterio? [0.5 punto]

_A _B _né A né B _sia A sia B

Bioinformatics practical 2024: genome assembly



Bioinformatics practical 2024: genome assembly

- **Challenge 1:** developing a paper-based test.
- **Challenge 2:** the test must only last 45 minutes.
 - **Solution 1:** choose a relatively simple case.
 - **Solution 2:** give an example of the expected outcome.
 - **Solution 3:** include some restrictions (e.g., starting from a specific node in the graph) to simplify the problem and remove edge cases.
- **Opportunity:** test skills that are only tangentially related to biology.
 - E.g., many students are surprisingly bad at following instructions.

Conclusion

- ♦ Introducing some Bioinformatics in Biology Olympiads is not particularly difficult.
- ♦ The key **challenges** mainly involve technological preparedness – i.e., ensuring that computers are available and compatible with the task.
- ♦ **Opportunities** include:
 - ♦ Modernising the curriculum
 - ♦ Linking multiple topics (A → Bioinformatics → B)
 - ♦ Testing multiple skills



Bioinformatics at the IBO

- Should Bioinformatics be included **every year**?
- Should the term be **further clarified** (e.g., by referring to a more specific topic or software tool)?

