

INFORMÀTICA A LES TEVES CÈL·LULES

*“Abans pensàvem que el nostre
futur es trobava a les estrelles.
Ara sabem que es troba als
gens.*

James Watson

Aclariments

Totes les paraules marcades amb un (*) es troben al glossari. En alguns casos, quan s'ha considerat que algunes paraules necessiten un aclariment, portaran incorporat un número i un petit aclariment en el peu de pàgina.

Les informacions que es considerin importants es trobaran marcades en negreta.

Abreviatures

| | |
|----------------|---|
| A | Adenina |
| ADN/DNA | Àcid desoxiribonucleic |
| ARN/RNA | Àcid ribonucleic |
| ARNm | Àcid ribonucleic missatger |
| ARNt | Àcid ribonucleic de transferència |
| C | Citosina |
| F1 | Primera generació filial |
| F2 | Segona generació filial |
| G | Guanina |
| NCBI | National Center for Biotechnology Information |
| P | Generació parental |
| preARNm | Transcrit primari de l'àcid ribonucleic missatger |
| T | Timina |
| U | Uracil |

ÍNDEX

| | Pàgina |
|--|--------|
| 0. Índex de figures, taules i gràfics | 7 |
| 1. Introducció | 10 |
| 2. Marc teòric | 13 |
| 2.1. La Genètica | 14 |
| 2.1.1. Què és la genètica? | 14 |
| 2.1.2. Història de la Genètica | 15 |
| 2.1.2.1. Les lleis de Mendel | 17 |
| 2.1.2.2. James Watson i Francis Crick | 23 |
| 2.1.2.3. Francis Crick i Gamow George | 24 |
| 2.1.2.4. Arthur Kornberg descobreix com es duplica l'ADN | 25 |
| 2.1.2.5. Marshall Nirenberg i Heinrich Matthaei | 26 |
| 2.1.2.6. Creació del Centre Nacional d'Informació sobre Biotecnologia (NCBI) | 27 |
| 2.1.2.7. Publicació del Projecte Genoma Humà | 27 |
| 2.1.3. Àcids nucleics | 28 |
| 2.1.3.1. Nucleòtids | 28 |
| 2.1.3.2. Tipus d'Àcids Nucleics | 29 |
| 2.1.3.2.1. Àcid desoxiribonucleic (ADN) | 29 |
| 2.1.3.2.2. Àcid ribonucleic (ARN) | 31 |
| 2.1.4. El Dogma Central de la biologia molecular | 34 |
| 2.1.4.1. Replicació | 35 |
| 2.1.4.1.1. La necessitat de la replicació de l'ADN | 35 |
| 2.1.4.1.2. Les primeres hipòtesis de la replicació de l'ADN | 35 |
| 2.1.4.1.3. El mecanisme de la replicació de l'ADN | 36 |
| 2.1.4.1.3.1. La replicació en organismes procariotes | 36 |

| | | |
|--------------|---|----|
| 2.1.4.1.3.2. | La replicació en organismes eucariotes | 37 |
| 2.1.4.2. | La teoria “un gen→un enzim” | 39 |
| 2.1.4.2.1. | La formulació de la teoria | 39 |
| 2.1.4.2.2. | La formulació de la teoria per Beadle i Tatum | 40 |
| 2.1.4.3. | L’expressió del missatge genètic | 41 |
| 2.1.4.4. | Transcripció | 42 |
| 2.1.4.4.1. | La transcripció en procarïotes | 42 |
| 2.1.4.4.2. | La transcripció en eucariotes | 43 |
| 2.1.4.5. | La clau genètica o codi genètic | 45 |
| 2.1.4.6. | Traducció | 46 |
| 2.1.5. | Les mutacions | 50 |
| 2.1.5.1. | Mutacions gèniques | 51 |
| 2.1.5.1.1. | Causes de les mutacions gèniques | 52 |
| 2.1.5.1.2. | Les mutacions gèniques i els seus sistemes de reparació | 53 |
| 2.1.5.2. | Mutacions cromosòmiques | 53 |
| 2.1.5.3. | Mutacions genòmiques | 54 |
| 2.2. | La Bioinformàtica | 56 |
| 2.2.1. | Què és la bioinformàtica? | 56 |
| 2.2.2. | Història de la Bioinformàtica | 58 |
| 2.2.2.1. | Les primeres dècades: anys 60 i 70 del segle XX | 58 |
| 2.2.2.2. | Anys 80 | 59 |
| 2.2.2.3. | Anys 90 | 59 |
| 2.2.2.4. | Primers anys del segle XXI | 60 |
| 2.2.3. | Relació entre Biologia i Informàtica | 60 |
| 2.2.4. | Objectius de la Bioinformàtica | 62 |
| 2.2.5. | Aplicacions i àrees d’investigació | 63 |
| 2.2.6. | Importància de la Bioinformàtica | 64 |
| 2.2.7. | Perfils d’un bioinformàtic | 65 |
| 2.2.8. | Els programes més utilitzats en Bioinformàtica | 65 |
| 3. | Marc pràctic | 66 |

| | |
|--|----|
| 3.1. Exercici d'evolució: <i>Evolució dels óssos</i> | 67 |
| 3.2. PERL | 72 |
| 3.2.1. Què és un llenguatge de programació? | 72 |
| 3.2.2. Per què PERL? | 72 |
| 3.2.3. El meu programa amb PERL | 73 |
| 3.3. Enquestes | 78 |
| 4. Conclusions | 83 |
| 5. Fonts | 85 |
| 6. Annex | 90 |
| 6.1. Glossari | 90 |
| 6.2. Estructura de la cèl·lula eucariota animal | 93 |

0. ÍNDEX DE FIGURES, TAULES I

GRÀFICS

FIGURES:

| | Pàgina |
|---|---------|
| Figura 1: ADN. | 14 |
| Figura 2: <i>Pisum sativum</i> . | 17 |
| Figura 3: Gregor Mendel cultivant pèsols. | 18 |
| Figura 4: Observació i interpretació de Mendel en la seva primera llei. | 20 |
| Figura 5: Observació i interpretació de Mendel en la seva segona llei. | 21 |
| Figura 6: Observació i interpretació de Mendel en la seva tercera llei. | 22 |
| Figura 7: James Watson i Francis Crick. | 23 |
| Figura 8: Francis Crick i Gamow George. | 24 |
| Figura 9: Arthur Kornberg. | 25 |
| Figura 10: Marshal Nirenberg i Heinrich Matthaei. | 26 |
| Figura 11: Símbol de l'NCBI. | 27 |
| Figura 12 | 27 |
| Figura 13: Estructura química de l'ADN. | 30 |
| Figura 14: Estructura química de l'ARN. | 32 |
| Figura 15: Figura comparativa entre l'estructura química de l'ADN i de l'ARN. | 33 |
| Figura 16: Figura comparativa dels tres models de replicació de l'ADN. | 36 |
| Figura 17: Procés de la replicació. | 38 |
| Figura 18: Procés de la transcripció. | 44 |
| Figura 19: Codi genètic. | 46 |
| Figura 20: Procés de la traducció. | 48 |
| Figures 21 i 22: Exemples de mutacions gèniques. | 51 i 52 |

| | |
|---|----|
| Figura 23: Tipus de mutacions cromosòmiques. | 54 |
| Figura 24: Cariotip del Síndrome de Down, exemple de mutació genòmica. | 55 |
| Figura 25 | 57 |
| Figura 26: Bacteriòfag fagoφ-X174. | 58 |
| Figura 27: PCR. | 59 |
| Figura 28 | 59 |
| Figura 29 | 60 |
| Figura 30 | 62 |
| Figura 31 | 65 |
| Figura 32: Primera part dels resultats que ens retorna CrustalW. | 69 |
| Figura 33: Segona part dels resultats que ens torna CrustalW. | 70 |
| Figura 34: Arbre filogenètic que ens elabora CrustalW, amb els codis corresponents de cada animal. Les línies corresponen al temps. | 71 |
| Figura 35: Exemple d'script amb PERL. | 73 |
| Figura 36: Parts de la cèl·lula eucariota animal. | 94 |

TAULES:

| | Pàgina |
|---|--------|
| Taula 1: Formulació de la teoria “un gen → un enzim” establint un paral·lelisme entre gen i substància. | 40 |
| Taula 2: Quadre comparatiu de la transcripció i la traducció en cèl·lules procariotes i eucariotes. | 49 |
| Taula 3: Nom científic i comú dels organismes que s'estudiaran i número d'accés a la seqüència del gen d'interès. | 67 |

GRÀFICS

| | Pàgina |
|---|--------|
| Gràfic 1: Resultats de la pregunta <i>Saps què és la Genètica?</i> | 80 |
| Gràfic 2: Resultats de la pregunta <i>En cas afirmatiu, quina creus que és la millor definició?</i> | 80 |
| Gràfic 3: Resultats de la pregunta <i>Saps què és la Bioinformàtica?</i> | 81 |

Gràfic 4: Resultats de la pregunta *En cas afirmatiu, saps quines aplicacions té en el camp de la investigació?*

81

1.INTRODUCCIÓ

Per què Bioinformàtica i Genètica?

El primer cop que vaig sentir parlar del “treball de recerca” em vaig començar a embolicar i vaig decidir que a partir d’aquell moment havia d’organitzar-me, tenir molta paciència i constància i, sobretot, tenir les coses molt clares. Em van explicar que és un treball que implica molt temps i molta feina. A més, és un treball molt important i requereix molt esforç si vols treure-li un bon profit i que influeixi de forma positiva en la nota del Batxillerat.

Cal dir que per fer un treball de recerca és important que el tema que s’escull sigui interessant i apassionant per a qui el fa, ja que se li ha de dedicar moltes hores. En el meu cas, jo tenia molt clar que el tema del meu treball de recerca havia de tenir alguna relació amb la Genètica, ja que és el que vull estudiar després del Batxillerat i sabia que si era relacionat amb aquesta em seria molt més amè realitzar-lo.

Abans de tot, volia fer el meu treball de recerca sobre les Lleis de Mendel amb les mosques de la fruita (*Drosophila melanogaster*), i quan ja havia començat aquest treball, el programa Argó de la Universitat Autònoma de Barcelona (UAB) em va acceptar la sol·licitud per a l’estada durant tres setmanes a la Universitat. Vaig triar el tema “Desenvolupar un software bioinformàtic per analitzar i seqüenciar mostres d’ADN”, em va semblar molt interessant i innovador i no me’n penedeixo. En aquesta estada vaig estar realitzant molts exercicis d’evolució amb programes bioinformàtics i finalment em van ensenyar a crear el meu propi programa. Durant aquelles tres setmanes vaig aprendre moltíssim i vaig decidir canviar el meu treball de recerca i fer-lo de quelcom que fusionés Genètica i Bioinformàtica.

Mentre feia l’estada, me’n vaig adonar que la Bioinformàtica era poc coneguda respecte a altres disciplines. Això em va permetre plantejar les **hipòtesis** del meu treball de recerca: la Bioinformàtica ens ajuda a estudiar Genètica i no és coneguda com es mereix. Vaig decidir fer una enquesta per a poder comprovar-ho. Havent

analitzat els resultats, vaig pensar que estaria bé fer conèixer aquesta ciència a l'institut. Així doncs, em vaig proposar fer una pàgina web útil per a l'institut, de manera que es podria introduir als alumnes en aquest món.

Els meus objectius són: conèixer més profundament el món de la Genètica, aprendre els aspectes fonamentals de la Bioinformàtica, realitzar un exercici d'evolució mitjançant eines bioinformàtiques ja conegudes, crear un programa Bioinformàtic i aconseguir que la Bioinformàtica sigui més coneguda.

La metodologia i el procediment que he seguit per a poder resoldre les meves hipòtesis i assolir els meus objectius ha estat la següent: en primer lloc, he realitzat una part teòrica explicant els conceptes necessaris per a entendre aquest treball. En segon lloc, he fet una part pràctica que consta de diferents parts: primerament, he dut a terme un exercici d'evolució de l'espècie dels óssos, seguidament, he realitzat un software bioinformàtic per analitzar i seqüenciar mostres d'ADN, una enquesta per comprovar si la Bioinformàtica és coneguda i una pàgina web per divulgar-la.

Aquests en són els resultats.

ABSTRACT

The first time I heard about "Treball de Recerca" I was concerned about being able to manage my time so I decided I had to organize my time in order to meet the hand in dates. Students from other years had told me it was a project that required a lot of effort and time.

I was strongly convinced I wanted to do my project on Genetics because this is what I want to study after I have finished Baccaureate (Batxillerat).

At first I wanted to do my "Treball de Recerca" about Mendel's laws on flies. When I had already started this, the "Programa Argó" of the UAB accepted my application to spend three weeks of the summer there. I tried the subject "Developing bioinformatic software to analyze and sequence samples of DNA", it surprised me when I discovered this project and I was curious and enthusiastic to do it. Throughout, the following weeks I did some evolution exercises with bioinformatic softwares and I learnt how to create my own software. At that moment, I decided I wanted to do my "Treball de Recerca" about a combination of Genetics and Bioinformatics.

When I was at UAB I realized there is little known about Bioinformatics in comparison to other subjects. This made me set out the hypothesis of my "Treball de Recerca": Bioinformatics helps us to study Genetics and it is less recognized than it deserves to be. I decided to do a survey to prove this. Having analyzed my results, I thought it could also be useful to create a webpage to make students a little more aware of Bioinformatics.

These are my results.

2. MARC TEÒRIC

2.1. LA GENÈTICA

2.1.1. QUÈ ÉS LA GENÈTICA?

La Genètica estudia la forma com les característiques dels éssers vius, siguin morfològiques, fisiològiques o bioquímiques, es transmeten, es generen i s'expressen, d'una generació a una altra segons les condicions que els envolten.

La Genètica, doncs, intenta explicar com s'hereten i modifiquen les característiques dels éssers vius, que poden ser de forma (l'alçada d'una planta, el color dels pètals, la forma de la flor...), fisiològiques (per exemple, la constitució d'una determinada proteïna que fa una funció concreta dins d'un ésser viu) i inclús el comportament (la forma d'aparellar-se entre animals, per exemple). D'aquesta forma, la Genètica tracta d'estudiar com aquestes característiques passen de pares a fills, a néts... i per què, alhora, varien de generació en generació.

Aquesta disciplina és la que unifica totes les ciències biològiques (o biociències), ja que tots els seus principis generals s'apliquen a tots els éssers vius. En tots els camps de la Biologia es recorre als conceptes de l'herència, quan es tracta d'explicar la variabilitat existent a la naturalesa. La millora de plantes i animals, la comprensió de la patologia humana i la producció de medicaments per mitjà de la biotecnologia, en són alguns exemples.



Figura 1: ADN.

2.1.2. HISTÒRIA DE LA GENÈTICA

Abans d'endinsar-nos ens cal saber una mica d'història d'aquesta gran ciència, la Genètica. El que veureu a continuació és una llista amb les fites més importants de la història de la Genètica i, més endavant, una explicació de les més importants d'aquestes.

- 1859: Chales Darwin va publicar *L'origen de les espècies*.
- 1865: Gregor Mendel va establir les lleis de l'herència genètica.
- 1902: Walter Sutton va crear el terme gen per descriure els factors situats en els cromosomes, va observar el moviment que feien els cromosomes durant la meiosi* i la mitosi* i va desenvolupar la teoria cromosòmica de l'herència.
- 1944: Oswald Avery, Maclyn McCarty i Colin McLeod van descobrir l'àcid desoxiribonucleic (ADN) com a molècula relacionada amb el material hereditari.
- 1953: James Watson i Francis Crick van proposar el model de doble cadena, helicoïdal, complementària i antiparal.lela del DNA.
- 1955: Frederick Sanger va anunciar la primera seqüència completa d'una proteïna: la insulina bovina.
- 1956:
 - Francis Crick i Gamov George van desenvolupar el Dogma Central per explicar la síntesi de proteïnes a partir de DNA.
 - KornBerg descobreix com es replica el DNA.
- 1958: Meselson i Stahl van demostrar que la replicació de l'ADN és semiconservativa.
- 1961:
 - Sidney Brenner, François Jacob i Matthew Meselson van identificar l'RNA missatger.
 - Marshall Nirenberg i Heinrich Mathaei van desxifrar el codi genètic
- 1975: Frederick Sanger va desenvolupar mètodes per seqüenciar el DNA.
- 1977: Es va seqüenciar el bacteriòfag (o virus) Φ X-174, el primer genoma*

- 1988: Es va crear el Centre Nacional d'Informació sobre Biotecnologia (NCBI)¹.
- 1990:
 - Es va crear BLAST, un programa que fa una cerca ràpida de similitud de seqüències, busca les seqüències més semblants dins la base de dades GenBank, que és pública i conté tota la informació obtinguda pels investigadors fins ara.
 - Es va iniciar el Projecte del Genoma Humà.
- 1995: Es seqüencien els primers genomes bacterians completament (*Haemophilus influenzae*).
- 1996:
 - Ian Wilmut produeix el primer mamífer clonat, l'ovella Dolly, a partir d'una cèl.lula adulta.
 - Es va seqüenciar el primer genoma de llevat, la primera seqüència de genoma eucariòtic.
- 1998: Es va seqüenciar el genoma d'un cuc completament (*Caenorhabditis elegans*).
- 1999: Es va seqüenciar completament el genoma de la mosca (*Drosophila melanogaster*).
- 2000:
 - Es va publicar el genoma de *Pseudomonas aeruginosa*.
 - Es va seqüenciar el genoma d'*Arabidopsis thaliana*
- 2001: Es va publicar el genoma humà.
- 2003:
 - Va fer cinquanta anys del descobriment de l'estructura del DNA.
 - S'anuncia la finalització de la seqüenciació del genoma humà.
- 2007: James Thompson i Shinya Yamanaka transformen cèl.lules de la pell humana en cèl.lules mare embrionàries.

¹ <http://www.ncbi.nlm.nih.gov>

2.1.2.1. Les lleis de Mendel

L'herència

L'herència és el fenomen de transmissió d'un conjunt de caràcters que té un individu en una determinada espècie. És el pas dels trets que caracteritzen un ésser viu als seus descendents. És el procés pel qual una cèl·lula filla adquireix les característiques de la seva cèl·lula progenitora.

Els experiments de Mendel

En la segona meitat del segle XIX, el monjo Gregor Mendel va fer un seguit d'experiments que el van dur a descobrir els principis bàsics que regeixen la transmissió dels caràcters hereditaris. Per fer els experiments, Mendel va escollir el pèsol d'olor (*Pisum sativum*). L'elecció va ser molt encertada per diverses raons:

- És una planta fàcil de cultivar i de creixement ràpid, de manera que en poc temps se n'obtenen diverses generacions.
- Presenta varietats amb característiques observables fàcilment, om el color de la llavor, la forma de les tavelles, la posició de les flors, etc.



Figura 2: *Pisum sativum*

- Els pèsols es poden autopol·linitzar, però també són fàcils de fecundar de manera artificial per fecundació encreuada.

A més, Mendel va utilitzar un mètode innovador a l'època:

- Va seleccionar, per estudiar-los, set caràcters, i en cada encreuament només es fixava en un o dos. Eren caràcters que variaven completament, per exemple: llavors verdes o grogues; llavors llises o rugoses. D'aquesta forma, els resultats van ser fàcils d'identificar.
- Va utilitzar línies pures. Les plantes que són línies pures, en autopol·linitzar-se, produeixen una descendència igual, de la mateixa varietat, i que es manté de generació en generació.

- Va estudiar la descendència al llarg de diverses generacions, no es va fixar només en la primera. D'aquesta manera, va poder observar la transmissió dels caràcters que havia escollit al llarg del temps.
- Va analitzar les dades resultants dels encreuaments de manera quantitativa. Així va obtenir proporcions numèriques fàcils d'interpretar.



Figura 3: Gregor Mendel cultivant pèsols.

PRIMER GRUP D'EXPERIMENTS

Mendel va iniciar els experiments estudiant la transmissió d'un únic caràcter entre la generació parental (P) i els seus descendents.

Per fer-ho va fecundar, artificialment, dues línies pures que diferien únicament en el color de la llavor, que podia ser groga o verda.

Va obtenir una descendència de plantes híbrides en què totes tenien les llavors grogues. Va anomenar primera generació filial (F1) a aquesta primera descendència. L'altre caràcter, el color verd, havia desaparegut en aquesta generació.

Va repetir l'experiment per als set caràcters que havia seleccionat i va trobar-se, en tots els casos, que dels dos caràcters alternatius paternals només apareixia en tots els individus de la F1.

Va anomenar dominant* aquest caràcter. Per contra, el caràcter que no es manifestava en la descendència el va anomenar recessiu.

SEGON GRUP D'EXPERIMENTS

Per continuar els experiments, Mendel va deixar que es produís l'autofecundació dels híbrids* de la F1 obtinguts en els encreuaments anteriors entre línies pures de pèsols amb llavors grogues i llavors verdes.

En estudiar la descendència d'aquesta segona generació filial (F2), va veure que de cada 4 llavors produïdes en cada planta, 3 eren grogues i 1 era verda. El caràcter

recessiu* que no s'havia manifestat en la F1 reapareixia en la segona generació filial.

Va repetir l'experiment amb la resta de caràcters escollits, i va trobar resultats similars en tots els casos, sempre en una proporció propera a 3:1.

Per explicar els resultats, Mendel va proposar que cada caràcter estava determinat per dos factors hereditaris, cada un provinent d'un progenitor. Per tant, el que s'hereta no són els caràcters, sinó els factors que els determinen i que es poden manifestar o no en la descendència.

TERCER GRUP D'EXPERIMENTS

Després de comprovar els resultats de l'encreuament entre línies pures que diferien només en un caràcter, Mendel es va proposar investigar si les conclusions obtingudes en els experiments anteriors es complien també quan s'estudiaven simultàniament dos caràcters. Per fer-ho va encreuar dues línies pures de pèsols per a dos caràcters: una amb llavors llises de color groc i l'altra amb llavors rugoses de color verd.

El resultat va ser una primera generació filial (F1) uniforme, en la qual tots els descendents tenien llavors grogues i llises, tal com calia esperar partint dels experiments anteriors. D'aquesta manera va comprovar que ara també es complien els resultats obtinguts en el primer grup d'experiments.

A continuació va deixar que s'autofecundessin els híbrids obtinguts en la F1, i va obtenir una segona generació filial F2, en què apareixien plantes que presentaven totes les combinacions possibles i sempre en la mateixa proporció. De cada 16 llavors, 9 eren grogues i llises, 3 grogues i rugoses, 3 verdes i llises i 1 verda i rugosa. La proporció que obtenia era de 9:3:3:1.

Va tornar a fer els experiments amb plantes que es diferenciaven en uns altres dos caràcters (flors de color porpra i tija alta; flors de color blanc i tija nana). En tots els encreuaments va obtenir la mateixa proporció en els resultats, propera a 9:3:3:1. Això va dur Mendel a plantejar que cada factor s'hereta de manera independent de la resta i que es pot combinar amb els altres i generar combinacions de caràcters que no existien en la generació parental.

Un cop explicats els tres grups d'experiments i entenent els conceptes explicats anteriorment, ara parlarem de la interpretació dels experiments de Mendel.

Interpretació dels experiments de Mendel

Mendel va realitzar els tres experiments esmentats sense conèixer la gran majoria de conceptes que he explicat anteriorment, però amb els coneixements de genètica que es tenen actualment, les experiències de Mendel es poden resumir en tres lleis.

PRIMERA LLEI. ENCREUAMENT D'HOMOZIGOTS PER A UN SOL CARÀCTER.

En el primer experiment, Mendel va encreuar dos individus homozigots* per al caràcter del color de la llavor.

Si s'encreua un individu homozigot dominant (AA) de llavors grogues amb un individu recessiu (aa) de llavors verdes, s'origina una F1 formada per individus heterozigòtics* (Aa) que tenen llavors de color groc.

La planta AA produeix gàmetes que porten l'al·lel A. Els gàmetes de la planta aa porten l'al·lel a. El zigot que resulta de la unió de tots dos gàmetes donarà lloc a un individu heterozigot (Aa) per a aquest caràcter.

En conclusió i per entendre'ns, **quan s'encreuen dues línies pures que es diferencien en un caràcter, la primera generació filial és uniforme.**

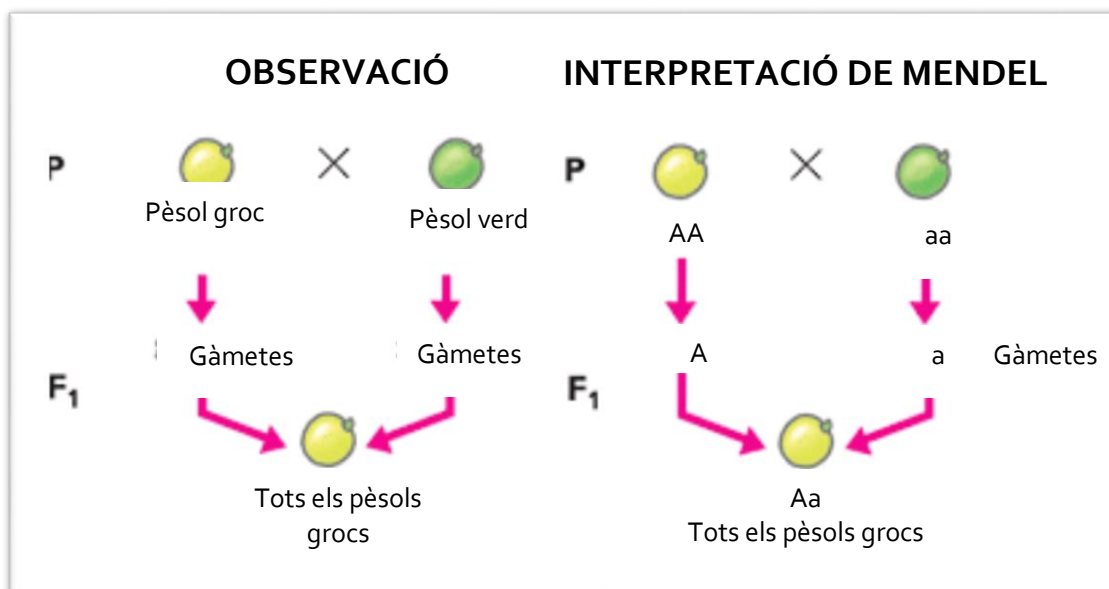


Figura 4: Observació i interpretació de Mendel en la seva primera llei

SEGONA LLEI. ENCREUAMENT D'HÍBRIDS PER A UN SOL CARÀCTER.

En el segon grup d'experiments Mendel va deixar que s'autofecundessin individus obtinguts en la F1, tots heterozigòtics (Aa) de fenotip* groc. Cada individu dóna lloc a dos tipus de gàmetes*, uns porten l'al·lel* A, i els altres, l'al·lel a. Després de la fecundació s'origina una F2 amb combinacions que no es manifestaven en la F1, en una proporció en què apareix un al·lel recessiu per cada tres de dominants (3:1). Per tant, **en encreuar els híbrids de la primera generació, els al·lells se separen i es distribueixen en els gàmetes de manera independent.**

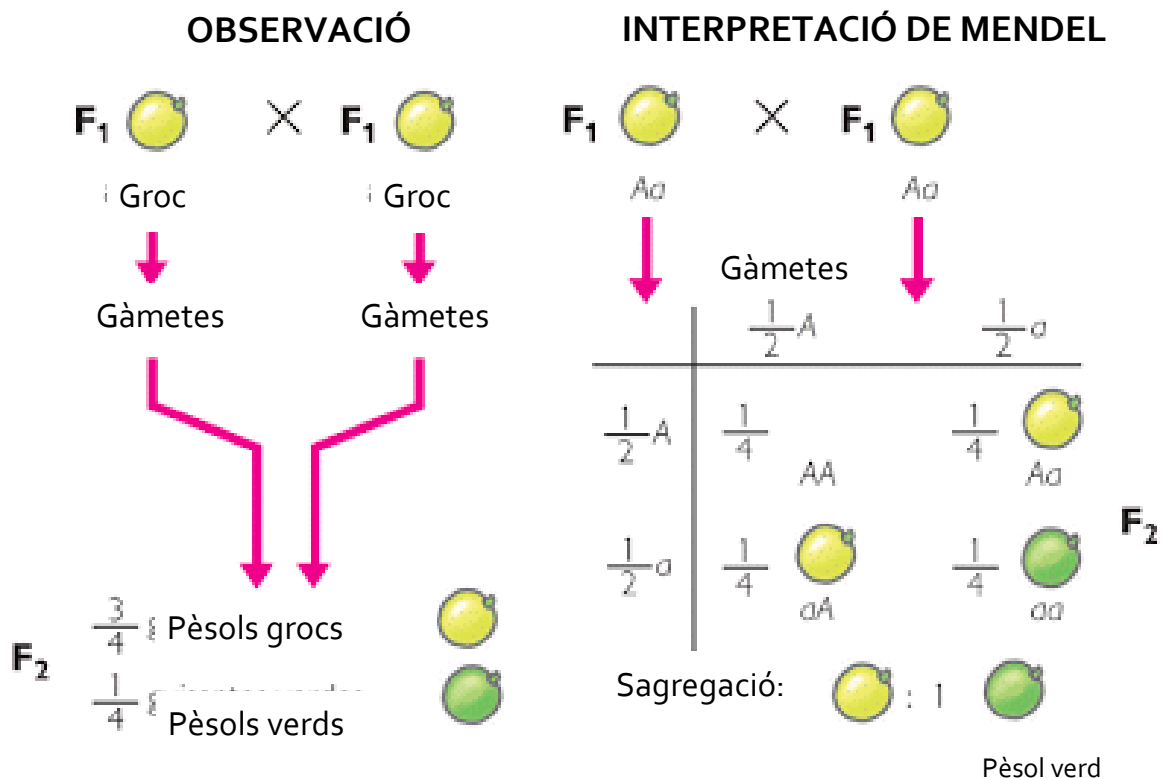


Figura 5: Observació i interpretació de Mendel en la seva segona llei.

TERCERA LLEI. HERÈNCIA DE DOS CARÀCTERS.

En el tercer grup d'experiments Mendel va escollir progenitors homozigòtics que diferien en dos caràcters: el color i l'aspecte de la llavor. La generació parental tenia els fenotips següents: AALL (groc llis) i aall (verd rugós). Tots dos progenitors produeixen un sol tipus de gàmetes que contenen un gen* de cada parell d'al·lells: AL i al.

La unió de gàmetes AL i al genera una F1 uniforme, d'individus dihíbrids de genotip AaLl i fenotip groc llis.

Els individus de la F1, després de la meiosi, originen gàmetes de quatre tipus: AL, aL, Al i al.

Després de l'autofecundació dels individus de la F1, s'obtenen setze combinacions genotípiques (degut a les diferents combinacions d'al·lels) i quatre fenotips diferents, en proporció 9:3:3:1.

La conclusió d'aquest experiment és que els dos caràcters es comporten de forma independent al formar els gàmetes, i en la fecundació es poden combinar de manera aleatòria, donant lloc a setze combinacions genotípiques i quatre fenotípiques.

Per tant, **els al·lels responsables de diferents caràcters s'hereten independentment els uns dels altres i es combinen a l'atzar en la descendència.**

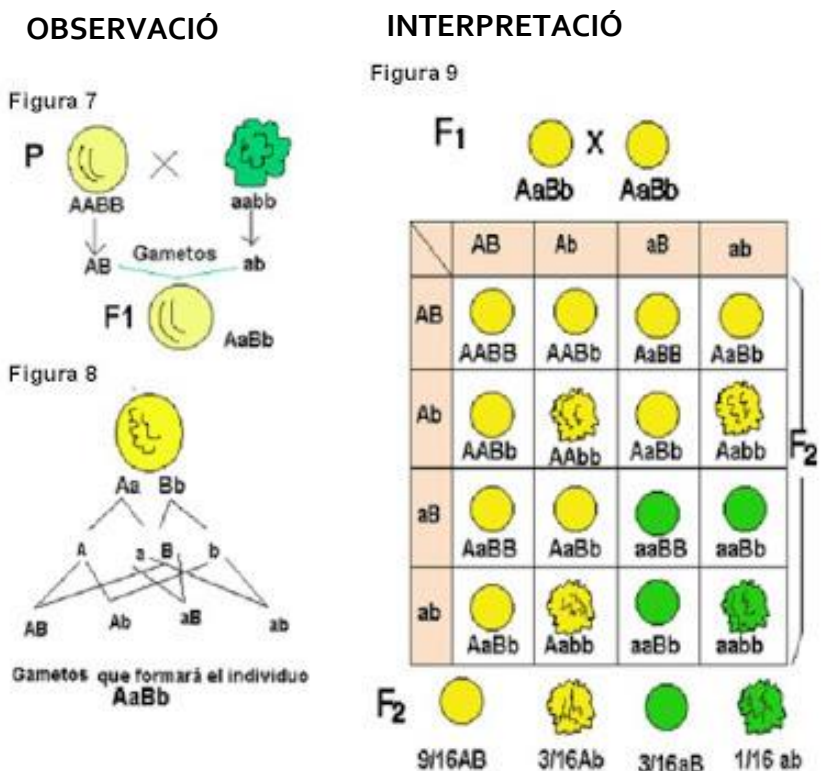


Figura 6: Observació i interpretació de Mendel en la seva tercera llei.

2.1.2.2. James Watson i Francis Crick

El descobriment de l'estructura de l'ADN² va involucrar més noms que no pas només Watson i Crick, com Linus Pauling i Maurice Wilkings, que, juntament amb Rosalind Franklin van captar imatges amb rajos X de la molècula. Van descobrir l'estructura de l'ADN: una doble hèlix amb dues cadenes complementàries i antiparal·leles, protagonistes de l'emmagatzematge i la transmissió de la informació hereditària. A més, aquest gran descobriment permetia començar a explicar el mecanisme amb el qual l'ADN es separa en dues cadenes per reproduir-se en dues molècules idèntiques.

No obstant, en aquell moment ja se sabia que la composició química de la molècula d'ADN està formada per quatre bases: timina (T), guanina (G), citosina (C) i adenina (A)³. Totes aquestes bases estan unides per febles enllaços fosfat-glúcid⁴, però encara no se sabia quina era l'estructura de la molècula ni de quina forma s'unien aquestes bases.

Al veure les imatges amb rajos X, Watson va observar que els patrons que es veien en creu deuriem, tridimensionalment, estar construïts en forma de doble hèlix. Juntament amb Crick van confeccionar un model amb metalls, en la que la doble hèlix d'ADN es constituïda per parells de quatre molècules.

Les implicacions del seu treball van ser essencials pel desenvolupament posterior de la biologia i la genètica.

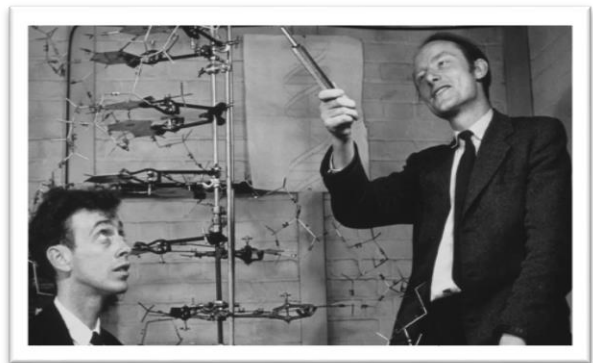


Figura 7: James Watson i Francis Crick

² S'explicarà posteriorment en l'apartat 2.3.

³ S'explicaran posteriorment en l'apartat 2.3.

⁴ Enllaços en què s'uneixen el grup fosfat d'un nucleòtid amb la pentosa del següent nucleòtid.

2.1.2.3. Francis Crick i Gamow George

Després del descobriment de l'estructura de l'ADN, Gamow va fer una contribució amb Crick, va ajudar-lo a resoldre el problema de com quatre tipus de bases (adenina, timina, guanina i citosina) en l'ADN, poden controlar la síntesi de proteïnes* a partir d'aminoàcids*. Va proposar que seqüències curtes formades per aquestes bases formaven un "codi" on cada seqüència especificava un dels vint aminoàcids⁵. La hipòtesi del físic Gamow suggeria que si els àcids nucleics contenen fonamentalment un grup de quatre bases com material únic per desxifrar 20 aminoàcids (que formen les proteïnes), el nombre clau havia de combinar tres de les quatre bases, segons la següent combinació: $4^3 = 64$. Tot i aquesta idea no era del tot correcta, Crick va dir que la seva col·laboració li va servir per resoldre el problema.



Figura 8: Francis Crick i Gamow George

⁵ Com que ell era físic, amb diferents càlculs va deduir que només hi havia 20 aminoàcids.

2.1.2.4. Arthur Kornberg descobreix com es replica la molècula de l'ADN

L'any 1956, Kornberg va descobrir un enzim en el bacteri *Escherichia coli*, la **DNA polimerasa**, amb la qual va sintetitzar per primera vegada àcid desoxiribonucleic (ADN) en un tub d'assaig. Posteriorment, als inicis dels anys setanta, va postular i va descobrir que per a la iniciació de la replicació de l'ADN es necessitava una petita molècula d'àcid ribonucleic, l'ARN iniciador. Més tard va desenvolupar un sistema de replicació 'in vitro' utilitzant la DNA polimerasa, el que va tenir una gran importància pel desenvolupament de l'enginyeria genètica i altres avenços tecnològics, com la reacció en cadena de la polimerasa (PCR)⁶.

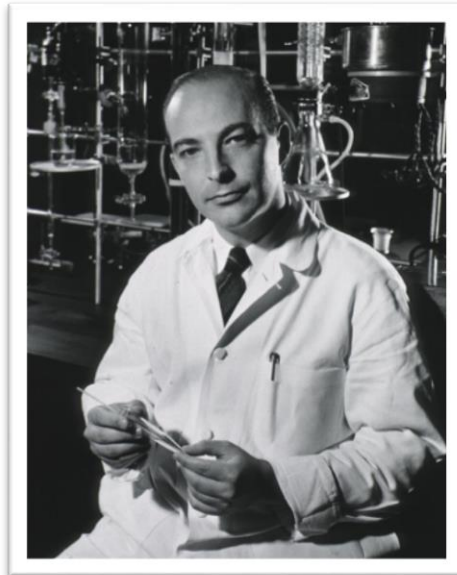


Figura 9: Arthur Kornberg

⁶ És una tècnica de biologia molecular amb la qual s'obtenen un gran nombre de còpies d'un fragment d'ADN específic a partir d'una quantitat mínima.

2.1.2.5. Marshall Nirenberg i Heinrich Matthaei

El desxiframent del codi genètic, és a dir, el trobar la relació existent entre les 64 possibles combinacions dels quatre tipus de nucleòtids d'ADN (adenina, guanina, citosina i timina) en grups de tres nucleòtids (triplets o codons) i les seqüències dels 20 aminoàcids en les proteïnes va ser descobert per Nirenberg i Matthaei l'any 1961. El codi genètic estableix la relació entre les seqüències de nucleòtids d'ADN (o els seus transcrits d'ARN) i les seqüències d'aminoàcids en les proteïnes.

Marshall Nirenberg i Heinrich Matthaei van descobrir que afegir una cadena de nucleòtids amb base d'uracil (U) conduïa a la síntesi del polipèptid (proteïna) fenilalanina. D'aquesta forma, es va fer evident que la cadena de poli-U actuava com ARN missatger. Per tant, la primera paraula desxifrada del codi va ser el triplet UUU, que codifica la fenilalanina. Aquest experiment va obrir camí pel complet desxiframent de la clau genètica.

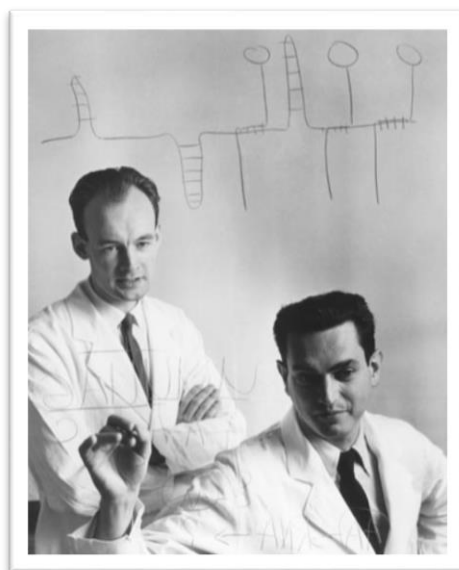


Figura 10: Marshall Nirenberg i Heinrich Matthaei

2.1.2.6. Creació del Centre Nacional d'Informació sobre Biotecnologia (NCBI)

L'NCBI (de les sigles en anglès, *National Center for Biotechnology Information*) va ser creat l'any 1988 amb l'objectiu de ser una font d'informació fonamental en el camp de la biologia molecular. S'encarrega d'emmagatzemar i actualitzar constantment la informació de seqüències genòmiques en el *GenBank*, de publicar un índex d'articles científics referents a biomedicina, biotecnologia, bioquímica, genètica i genòmica a *PubMed*⁷ i de realitzar una recopilació de malalties genètiques humanes en l'*OMIM*⁸. L'NCBI ofereix algunes eines bioinformàtiques per a l'anàlisi de seqüències d'ADN, ARN i proteïnes; una de les més utilitzades és BLAST.



Figura 11:
Símbol de
l'NCBI.

2.1.2.7. Publicació del Projecte Genoma Humà

El Projecte Genoma Humà va ser un esforç internacional de recerca per determinar la seqüència del genoma humà i identificar els gens que conté. En aquest projecte, promogut i coordinat pels Estats Units, hi van participar altres països com el Regne Unit, França, Alemanya, Japó i la Xina. Tècnicament va consistir en determinar les posicions dels nucleòtids (o parells de bases) de l'ADN humà i identificar els gens que hi són presents.



Figura 12

⁷ És una base de dades bibliogràfiques de citacions i resums d'articles de recerca en biomedicina i ciències de la vida.

⁸ *Mendelian Inheritance in Man (OMIM)* és una base de dades que cataloga totes les malalties genètiques conegudes amb enllaços als gens rellevants del genoma humà.

2.1.3. ÀCIDS NUCLEICS

Els àcids nucleics són biomolècules que contenen Carboni, Hidrogen, Oxigen, Nitrogen i Fòsfor. La seva unitat bàsica s'anomena **nucleòtid**.

Són les molècules encarregades d'emmagatzemar, transmetre i expressar la informació genètica, és a dir, la informació que es necessita per a la síntesi de proteïnes.

Es troben en totes les cèl·lules de tots els éssers vius, des dels bacteris fins als éssers humans, i amb la mateixa composició química.

2.1.3.1. Nucleòtids

Un nucleòtid és una molècula formada per:

- Una pentosa (que és un glúcid) i pot ser:
 - Una ribosa (ARN)
 - Una desoxiribosa (ADN)
- Un àcid fosfòric
- Una base nitrogenada, que es poden classificar segons si són púriques o pirimidíniques:
 - Púriques:
 - Adenina (A)
 - Guanina (G)
 - Pirimidíniques:
 - Timina (T)
 - Uracil (U), exclusiu en ARN
 - Citosina (C)

2.1.3.2. Tipus d'àcids nucleics

Distingim dos tipus d'àcids nucleics, l'ADN i l'ARN. Les seves principals diferències són:

- **La pentosa:** l'ADN està format per una desoxiribosa mentre que l'ARN està format per una ribosa.
- **Les bases nitrogenades:** l'ADN està format per A,T,C i G, mentre que l'ARN està format per A,U,C i G.
- **L'estructura de les cadenes:** l'ADN és una cadena doble (doble hèlix), mentre que l'ARN acostuma a ser una sola cadena senzilla.

2.1.3.2.1. Àcid desoxiribonucleic (ADN)

L'ADN és aquella molècula on s'emmagatzema de forma codificada la informació genètica de tots els éssers vius, també és l'encarregada de transmetre tota aquesta informació generació a generació.

L'àcid desoxiribonucleic està format per una pentosa (desoxiribosa), un grup fosfat i quatre bases nitrogenades (A,T,C i G).

Aquesta molècula, formada per seqüències de nucleòtids, conté tota la informació que es necessita per a poder controlar el metabolisme, les funcions vitals dels éssers vius i fabricar les seves proteïnes. Cal dir que és una molècula molt complexa!

La molècula de l'ADN pot adoptar diferents estructures: la primària, la secundària i la terciària. L'estructura primària correspon a la seqüència lineal de nucleòtids, la secundària correspon al model explicat anteriorment de Watson i Crick, és a dir: és una doble hèlix formada per dues cadenes de nucleòtids enrotllades antiparal·lelament⁹ i són complementàries. Finalment, l'estructura terciària és l'empaquetament de l'ADN en diferents subnivells sobre unes proteïnes anomenades histones, formant nucleosomes.

⁹ Les dues cadenes d'ADN tenen enllaços 5' → 3' orientades en sentit diferent. El nom de 5' i 3' és el número del carboni del glúcid (pentosa) que es troba lliure.

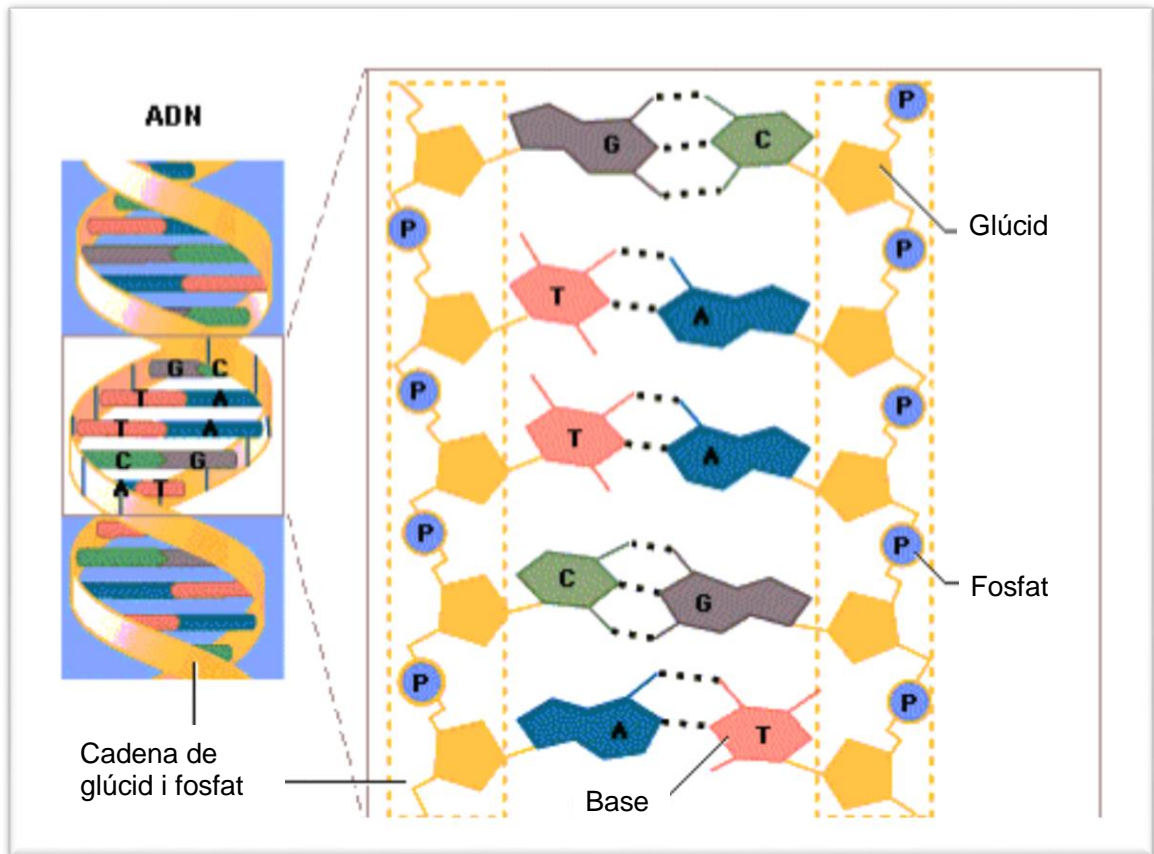


Figura 13: Estructura química de l'ADN.

2.1.3.2.2. Àcid ribonucleic (ARN)

L'ARN és una molècula d'una sola cadena, a diferència de l'ADN, en estructura primària. Aquesta molècula està formada per: una pentosa (ribosa), un grup fosfat i quatre bases nitrogenades (A, U, C i G). A diferència de l'ADN, la molècula d'ARN és molt menor que la d'ADN.

En la cèl·lula podem trobar diferents tipus d'ARN, amb diferents funcions:

- **ARN missatger:** és un ARN lineal que conté la informació, copiada de l'ADN, per sintetitzar una proteïna. Es construeix al nucli cel·lular, a partir d'una seqüència d'ADN. Seguidament, surt del nucli i s'associa a ribosomes*, on es formaran les proteïnes. Cada tres nucleòtids (codó*) es forma un aminoàcid diferent, el que vol dir que la seqüència d'aminoàcids està formada a partir de la seqüència de nucleòtids de l'ARNm, configurat a partir de l'ADN.
- **ARN ribosòmic:** forma els ribosomes, que són les estructures cel·lulars on s'uneixen els aminoàcids per formar proteïnes a partir de la informació proporcionada pel ARNm.
- **ARN de transferència:** no és lineal (té forma de trèvol). En l'ARNt apareix una seqüència de tres nucleòtids (anticodó), aquesta seqüència és complementària a la seqüència proporcionada per l'ARNm (el codó) i posteriorment s'hi unirà un aminoàcid específic codificat pel codó. En resum, la funció d'aquest tipus d'ARN és unir-se al ribosoma, fer que l'ARNm s'hi uneixi també i fer que es formi la proteïna mitjançant la unió dels aminoàcids.

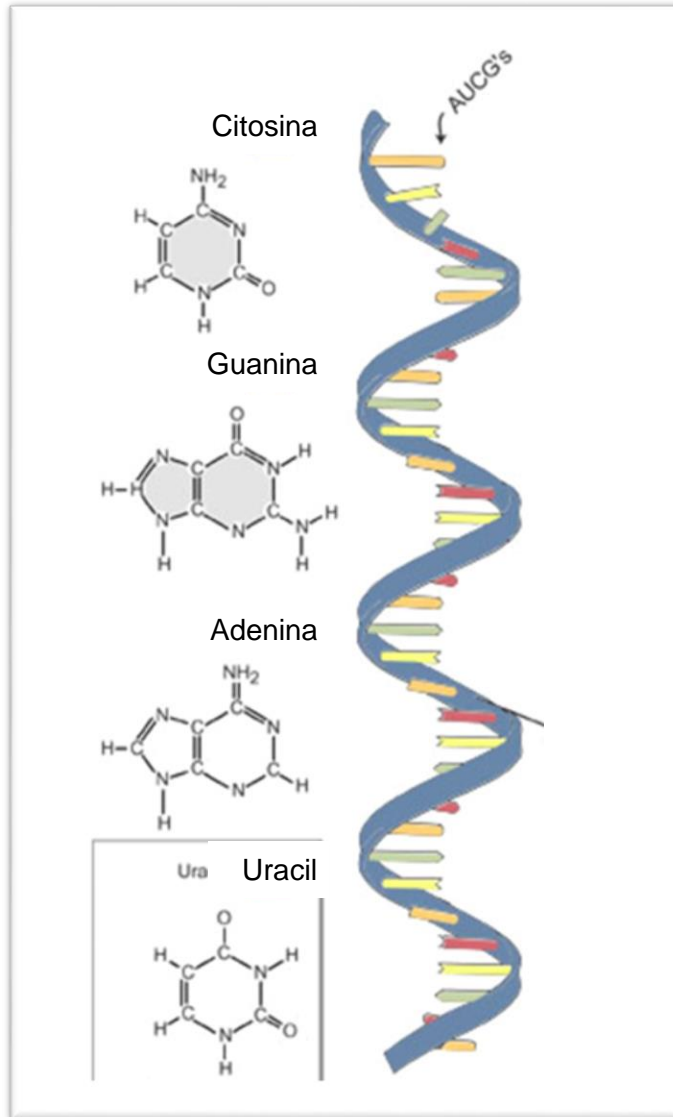


Figura 14: Estructura química de l'ARN.

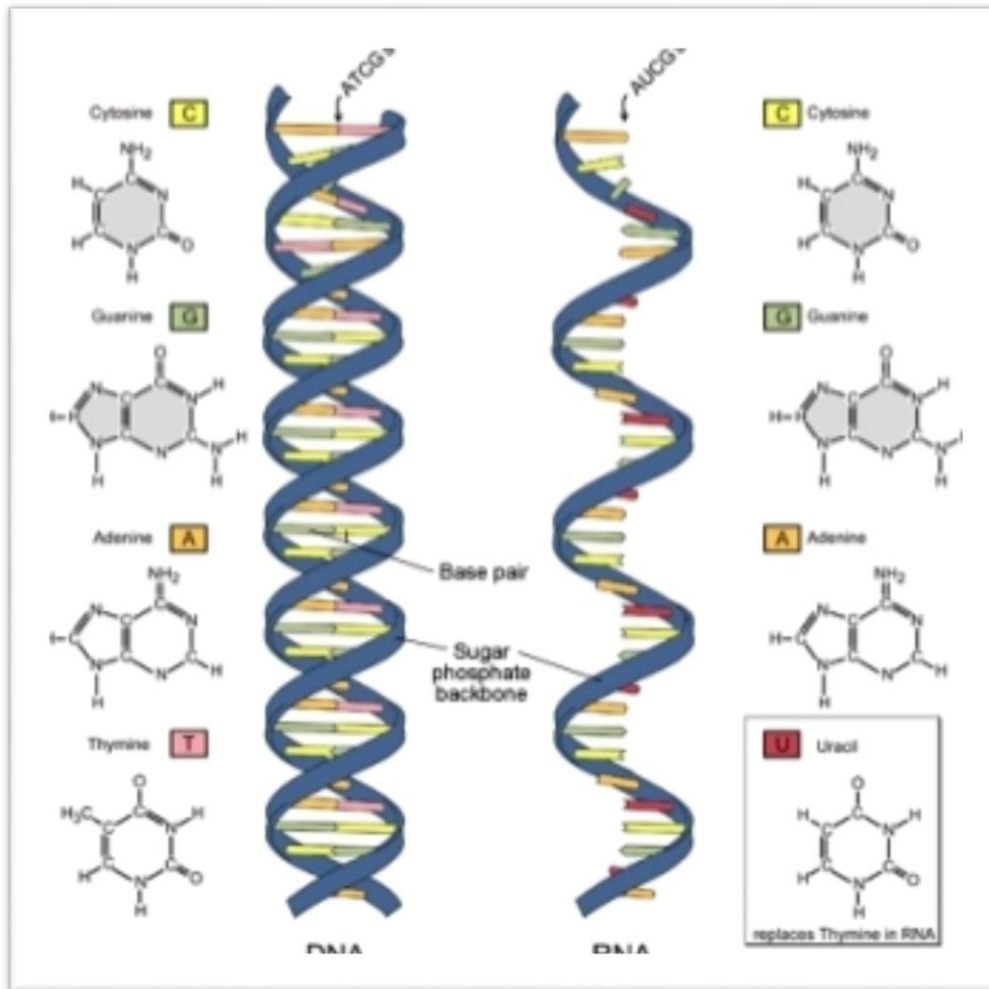


Figura 15: Figura comparativa entre l'estructura química de l'ADN i de l'ARN.

2.1.4. EL DOGMA CENTRAL DE LA BIOLOGIA MOLECULAR

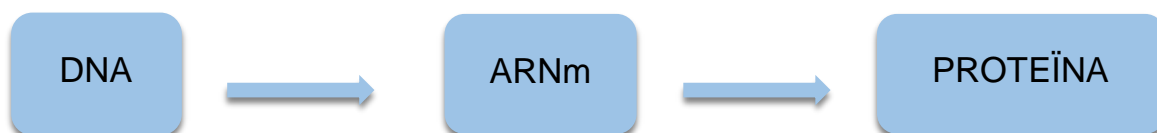
El dogma central de la biologia molecular és el fet de com totes les biomolècules són emmagatzemades i transmeses generació a generació mitjançant la informació genètica.

En cadascun del nucli de les nostres cèl·lules trobem la doble hèlix d'ADN, emmagatzemada de diferents formes com hem vist anteriorment. L'anticomplementarietat de les cadenes permet la **REPLICACIÓ** perfecta de l'ADN, necessària per a l'herència. No obstant això, a vegades també és necessària una replicació imperfecta per a poder garantir evolució a les espècies; aquestes imperfeccions s'anomenen mutacions. Com hem dit abans, les cadenes de la doble hèlix son antiparal·leles, el que vol dir que una d'elles va en direcció 5'→3', mentre que l'altre ho fa en direcció 3'→5'.

Quasi totes les funcions cel·lulars les duen a terme les proteïnes, de forma que la majoria dels gens¹⁰ codifiquen la informació necessària per a produir una proteïna. Aquest procés necessita una biomolècula intermèdia, l'RNA, que és necessària per traduir la informació genètica de l'ADN a proteïna. L'RNA encarregat de fer-ho, com hem dit abans, és l'ARN missatger i el procés pel qual copiem un gen a RNA s'anomena **TRANSCRIPCIÓ**.

L'ARN és semblant a l'ADN, però està format per una sola cadena i conté la base nitrogenada U¹¹ en comptes de T¹². Com hem dit a l'apartat anterior, l'RNAm s'uneix als ribosomes, que llegeixen la informació i formen una proteïna. Aquest procés s'anomena **TRADUCCIÓ**.

Aquest procés és irreversible, per tant:



¹⁰ Regió determinada de l'ADN.

¹¹ Uracil.

¹² Timina.

2.1.4.1. La replicació

2.1.4.1.1. La necessitat de la replicació de l'ADN

La vida de tots els organismes és molt variable: pot ser d'uns quants minuts, com la dels bacteris; d'uns quants anys, com la dels humans; d'uns quants segles, com la de les tortugues marines, o fins i tot de més d'un mil·lenni, com la de les oliveres. Però sempre hi ha un moment en què els organismes moren, és a dir, tots els éssers vius són temporals. Per això, perquè l'espècie no s'extingeixi, hi ha d'haver un moment en què els individus es reproduïxin i donin lloc a nous individus que continuïn vivint.

2.1.4.1.2. Les primeres hipòtesis sobre la replicació de l'ADN

L'estructura de l'ADN en doble hèlix permet comprendre que aquesta molècula és idònia per donar lloc a còpies. D'una banda, l'estructura presenta dues cadenes complementàries entrelaçades, la qual cosa li dona una gran estabilitat; d'altra, n'hi hauria prou que un enzim* específic les separés perquè cadascuna pogués servir com a motlle per sintetitzar el filament complementari a partir de nucleòtids solts i sota l'acció d'un altre enzim.

Per explicar aquest procés es van proposar tres hipòtesis: la hipòtesi semiconservativa, la conservadora i la dispersora.

- La hipòtesi **semiconservativa** sobre la duplicació de l'ADN és deguda als científics Watson i Crick. Proposa que les dues cadenes d'ADN inicial es separen i cadascuna serveix de motlle per a la síntesi, segons la complementarietat de les bases nitrogenades, d'una cadena nova. Per tant, en les dues molècules de l'ADN produïdes, un dels filaments seria l'antic, i l'altre, el nou.
- En la hipòtesi **conservativa** es proposa que després de la duplicació queden, d'una banda, els dos filaments antics junts i, de l'altre, els dos filaments nous.
- En la hipòtesi **dispersiva** es proposa que els filaments, al final, estan construïts per fragments diferents d'ADN antic i d'ADN nou.

L'experiment definitiu per resoldre quina d'aquestes tres hipòtesis era la correcta va ser el de Meselson i Stahl. La hipòtesi correcta va ser la semiconservativa.

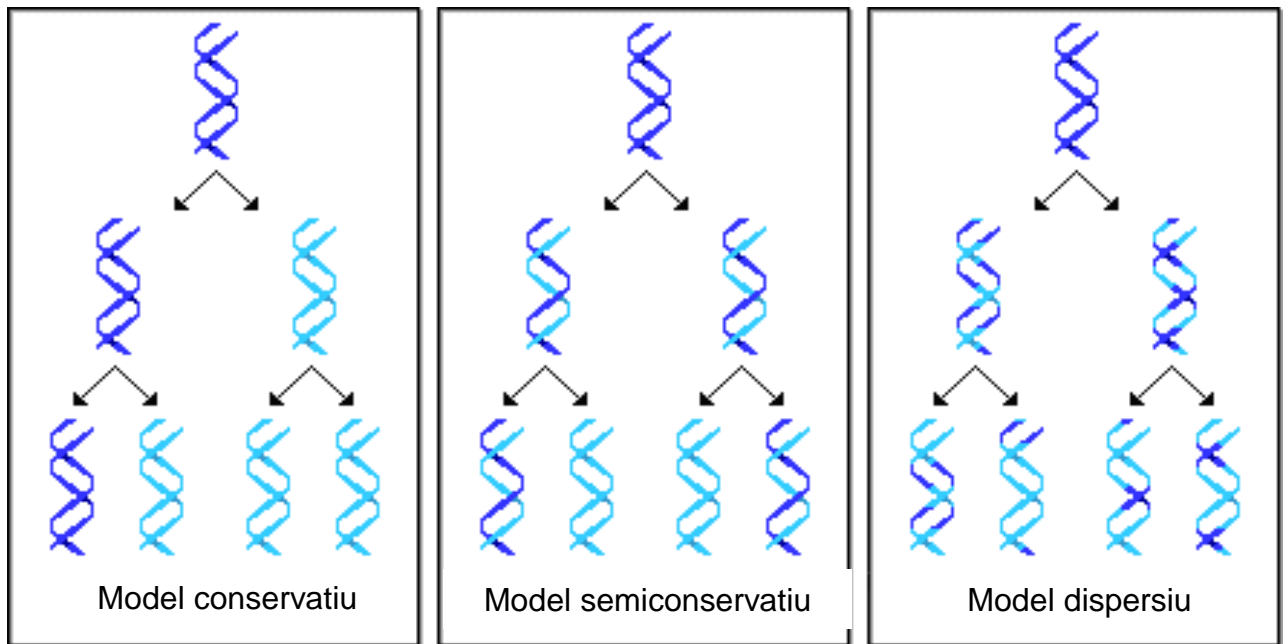


Figura 16: Figura comparativa dels tres models de replicació de l'ADN.

Però, com és capaç l'ADN de copiar-se formant dues cadenes noves idèntiques a la cadena original?

2.1.4.1.3. El mecanisme de la replicació de l'ADN

La duplicació de l'ADN presenta algunes diferències en organismes procariotes* i en eucariotes*.

2.1.4.1.3.1. La replicació en organismes procariotes

- En primer lloc, hi ha una seqüència de nucleòtids a l'ADN, anomenada "origen de la replicació", que actua com a senyal d'iniciació del procés.
- En segon lloc, el procés s'inicia amb un enzim anomenat helicasa, que trenca els ponts d'hidrogen¹³ entre els dos filaments complementaris i els separa perquè serveixin de patrons o motlles. Com que el desenrotllament de la doble hèlix fa que apareguin superenrotllaments en la resta de la molècula, que podrien aturar el procés, calen topoisomerases¹⁴ per a reduir les tensions de la fibra.

¹³ Tipus d'enllaç, és una interacció atractiva entre molècules. És un enllaç feble.

¹⁴ Enzims capaços d'actuar sobre l'ADN; poden embolicar-lo per permetre que s'emmagatzemi de forma més compacte o bé desembolicar-lo per a controlar la síntesi de proteïnes i facilitar la replicació.

- A continuació, hi intervenen unes proteïnes que s'enllacen sobre l'ADN de filament únic. Són les proteïnes estabilitzadores (SSB), que tenen com a funció mantenir la separació dels dos filaments complementaris. D'aquesta manera s'inicia la formació de l'anomenada "forquilla o bombolla de replicació". El procés és bidireccional, és a dir, hi ha una helicasa que treballa en un sentit i una altra que treballa en sentit oposat. Com que cap **DNA-polimerasa** pot actuar sense encebador, primer hi intervé una **RNA-polimerasa** que sí que ho pot fer. Aquest enzim s'anomena **primasa** i sintetitza un fragment curt de RNA format per uns deu nucleòtids, anomenat **primer**, que actua com a encebador.
- Després hi intervé la **DNA-polimerasa III**, que, a partir del primer, comença a sintetitzar en direcció 5'→3' (el que anomenem cadena contínua). Sobre l'altre filament, el discontinu, l'RNA-polimerasa sintetitza un filament d'uns quaranta nucleòtids d'RNA en un punt separat uns mil nucleòtids del senyal d'iniciació. A partir d'aquests, la DNA-polimerasa III sintetitza uns mil nucleòtids d'ADN, i aleshores es forma un **fragment d'Okazaki**.
- Aquest procés es va repetint a mesura que es van separant els dos filaments motlles. Posteriorment, hi intervé la **DNA-polimerasa I**, que primer, retira els fragments d'ARN, i després, gràcies a la funció polimerasa, omple els buits amb nucleòtids d'ADN. Finalment, hi intervé la **DNA-ligasa**, que empalma els diferents fragments. Així doncs, aquest fragment és de creixement discontinu. Com que necessita que es desespiralitzï un segment d'uns quants milers de nucleòtids perquè s'iniciï la síntesi, tarda més a créixer que l'altre, i per això s'anomena filament retardat.

2.1.4.1.3.2. La replicació en organismes eucariotes

El procés en aquests organismes és semblant al dels organismes procarïotes. Cal destacar algunes diferències:

- La primera gran diferència és que l'ADN dels organismes eucariotes està associat a histones. S'ha observat que, durant la replicació, el filament continu es queda amb les histones "velles" i el filament discontinu amb les noves histones.

- La segona gran diferència és que, tenint en compte que la longitud de l'ADN d'un cromosoma eucariòtic és molt més gran que l'ADN d'un cromosoma procariòtic i que, segurament per la presència d'histones, el procés és bastant més lent. Això és degut a que, en cada ADN d'un cromosoma no hi ha un sol origen de replicació, sinó aproximadament uns cent, el que vol dir que s'hi formen unes cent forquilles de replicació.

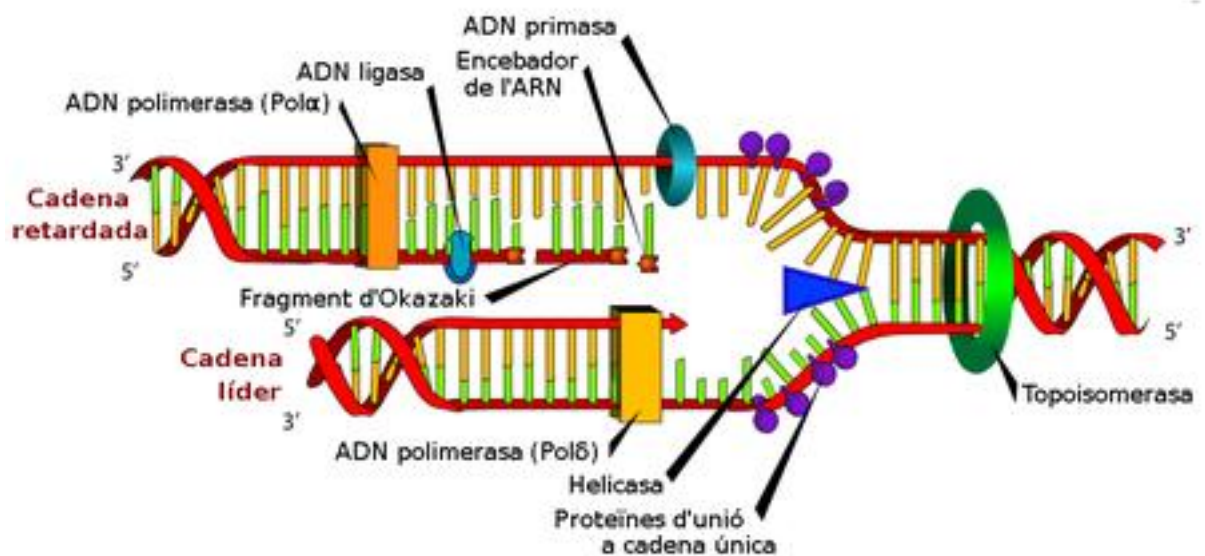


Figura 17: Procés de la replicació.

2.1.4.2. La teoria "un gen → un enzim"

Un cop coneguda l'estructura de l'ADN i com es duplica, faltava descobrir com el seu missatge es comunica en les característiques de l'organisme. El punt de partida va ser la recerca realitzada pel metge anglès A. E. Garrod l'any 1901. Ell havia estudiat diverses malalties humanes hereditàries. Una d'aquestes, l'alcaptonúria, es caracteritza per artritis i ennegriment dels cartílags i de l'orina (quan entrava en contacte amb l'aire). Un any abans, els científics Correns, De Vries i Tschermak havien redescobert les lleis de l'herència proposades per Mendel l'any 1866, i el món científic estava molt interessat a detectar caràcters hereditaris i saber com s'heretaven.

Havent llegit els conceptes del glossari: gen, locus*, diploide*, haploide*, gen dominant* i al·lel*:

En l'alcaptonúria, es tracta d'un locus que té dos al·lells: l'al·lel normal i l'al·lel alcaptonúria, en què l'al·lel normal (N) és dominant i l'al·lel alcaptonúria (n) és recessiu.

Estudiant la genealogia* dels individus malalts, Garrod va observar que l'alcaptonúria era una malaltia hereditària, ja que sempre hi havia algun avantpassat que n'havia patit. La malaltia tan sols apareixia quan les dues informacions rebudes, és a dir, la que rebia per part del pare i la que rebia per part de la mare, indicaven alcaptonúria (homozigot recessiu nn). Quan les dues informacions eren normals, o quan simplement ho era una, l'individu era normal (homozigot dominant NN o heterozigot Nn). Així doncs, es tractava d'una malaltia provocada per una informació genètica anormal, que era recessiva respecte de la informació correcta.

2.1.4.2.1. La formulació de la teoria

A partir d'anàlisis clíniques, es va esbrinar que el motiu de l'ennegriment de l'orina i els cartílags era degut a la presència de l'àcid homogentístic¹⁵. Llavors es va suposar que en els individus sans aquesta substància es transformava en d'altres

¹⁵ Àcid 2,5-dihidroxigenilacètic.

i desapareixia. D'aquesta manera es poder passar d'un paral·lelisme entre gen i caràcter a un paral·lelisme entre gen i substància.

| Gens | Caràcter | Substància a l'organisme |
|------|--------------|--------------------------|
| AA | Sa | No àcid homogenístic |
| Aa | Sa | No àcid homogenístic |
| aa | Alcaptonúria | Sí àcid homogenístic |

Taula 1: formulació de la teoria establint un paral·lelisme entre gen i substància.

La idea va ser revolucionària en l'època de Garrod. Uns anys més tard, Beadle i Tatum va confirmar la hipòtesi de Garrod.

2.1.4.2.2. La formulació de la teoria per Beadle i Tatum

L'any 1948, G. Beadle i E. Tatum van publicar un conjunt d'experiments que havien realitzat amb la floridura del pa (*Neurospora crassa*). Per esbrinar i poder explicar què fan els gens, la idea de la parella de científics era alterar un gen transformant-lo en defectuós i després deduir el que fa, observant el fenotip de l'individu mutant.

Estudi de la síntesi d'arginina en *Neurospora crassa*

Van exposar el fong a radiacions d'alta energia que van produir mutacions en el seu ADN i després van seleccionar mutants que no poguessin sintetitzar compostos específics, entre altres, l'aminoàcid arginina. Segons la hipòtesi d'aquests investigadors: *La incapacitat de sintetitzar arginina es devia a un defecte d'un gen, i la incapacitat de sintetitzar altres substàncies es devia a defectes en altres gens.* Beadle i Tatum van proposar que els mutants de *Neurospora* que no podien produir arginina era perquè no tenien l'enzim necessari per sintetitzar aquest compost i que l'absència d'aquest enzim es devia a un defecte genètic: *Cada gen conté la informació necessària per fer un enzim diferent.*

Les conclusions que la parella d'investigadors van obtenir van ser:

- Els diferents mutants estan mancats per determinats enzims. Sense un enzim no poden sintetitzar una substància.
- Els mutants no tenen un enzim perquè s'ha modificat el gen que porta la informació per a sintetitzar-lo.

En conclusió, el que estableix la teoria “un gen → un enzim” és un paral·lelisme entre els gens i els enzims. Per tant, el que proposa és que si s'altera la seqüència de nucleòtids d'un gen, s'altera l'enzim que aquesta seqüència codificava. En conseqüència, com els enzims controlen les substàncies i les substàncies les característiques dels organismes, apareixen malalties genètiques derivades de la manca de determinades substàncies.

Actualment, aquesta teoria no és considerada correcta degut a que un gen produeix més d'una proteïna diferent, que, en conseqüència, tindran funcions diferents.¹⁶

2.1.4.3. L'expressió del missatge genètic

Quan Beadle i Tatum van establir el paral·lelisme entre gens i enzims i després Watson i Crick van proposar el model de doble hèlix, Crick va proposar l'anomenada “**hipòtesi de la col·linealitat**”, en la que s'estableix una correspondència entre la seqüència de nucleòtids del gen i la seqüència d'aminoàcids de l'enzim que el gen codifica. En el mecanisme pel qual es passa d'una seqüència a l'altra es poden diferenciar dos processos: un es dóna al genoma¹⁷, i l'altre als ribosomes.

Al DNA cel·lular, que en les cèl·lules eucariotes es troba al nucli, es passa d'una seqüència de bases nitrogenades d'un gen a una seqüència de bases nitrogenades complementàries pertanyents a un RNAm. Aquest procés s'anomena **TRANSCRIPCIÓ**.

¹⁶ <https://embryo.asu.edu/pages/george-w-beadles-one-gene-one-enzyme-hypothesis>

¹⁷ O DNA cel·lular.

Als ribosomes es passa d'una seqüència de ribonucleòtids d'aquest RNAm a una seqüència d'aminoàcids, que formaran una proteïna. Aquest procés s'anomena **TRADUCCIÓ**.

Més tard es va descobrir que alguns virus podien produir ADN a partir del seu ARN, aquest procés s'anomena retrotranscripció.

2.1.4.4. La transcripció

La transcripció és el pas d'una seqüència d'ADN a una seqüència d'ARN. Per realitzar-ho, hi intervenen l'ADN, nucleòtids d'ARN (per tant, que contenen bases A,U,C,G), les RNA-polimerases (RNAP) i els cofactors. Podem distingir dos mecanismes diferents en funció dels organismes: procariotes o eucariotes.

2.1.4.4.1. La transcripció en procariotes

Els organismes procariotes només tenen un tipus d'RNA-polimerasa. En la síntesi de l'ARNm, es distingeixen diferents etapes:

Etapa d'iniciació

Abans de cada regió d'ADN que es transcriu, el que s'anomena unitat de transcripció, hi ha una regió d'ADN que no es transcriu, el que s'anomena promotor. Aquest conté unes seqüències de nucleòtids, que en diem seqüències de consens, a les quals s'associa l'RNA-polimerasa i el primer nucleòtid que serà transcrit. El promotor determina quina de les dues cadenes es transcriurà.

Un cop l'RNA-polimerasa s'ha fixat sobre el promotor, desenrotlla més o menys una volta d'hèlix, i inicia la polimerització d'ARN seguint els dos filaments de l'ADN, l'anomenada cadena o filament patró.

Etapa d'elongació o allargament

El procés continua a una velocitat d'uns quaranta nucleòtids per segon. A mesura que l'enzim RNA-polimerasa avança pel filament d'ADN patró cap a l'extrem 5', se sintetitza una cadena d'ARN en direcció 5' → 3'.

Vegem un exemple:

- Seqüència d'ADN: 3' ... TACGCT ... 5'
- Seqüència d'ARNm: 5' ... AUGCGA ... 3'

Etapa de finalització

L'etapa de finalització es dona quan l'RNA-polimerasa arriba a una seqüència que s'anomena terminador. Aleshores es separa i l'ADN torna a formar la doble hèlix.

Etapa de maduració

Si el que estem sintetitzant és un ARNm, no hi ha maduració en el cas dels procariotes. En canvi, si estem sintetitzant un ARNt o un ARNr, hi ha un transcrit primari que després pateix un procés fins a madurar-se.

2.1.4.4.2. La transcripció en eucariotes

Primer, cal ressaltar que existeixen tres tipus d'enzims RNA-polimerasa segons el tipus d'ARN que s'ha de sintetitzar. Seguidament, cal destacar que els gens estan fragmentats, de manera que sempre cal un procés de maduració en el qual s'eliminin les seqüències sense sentit (introns*) i s'empalmin entre elles les seqüències amb sentit (exons*). A vegades ens podem trobar amb gens, com els de les histones, que no presenten introns. Finalment, cal recordar que en els organismes eucariotes l'ADN està associat a histones* i formen nucleosomes*.

En els organismes eucariotes, la síntesi d'ARNm presenta les següents etapes:

Etapa d'iniciació

Existeix una regió d'ADN anomenada promotor, on es fixa l'RNA-polimerasa II, que consta de dos senyals anomenades seqüències de consens: la CAAT i la TATA, a diferents distàncies del punt d'inici. Perquè es pugui fixar l'RNA-polimerasa, abans s'han de fixar en aquestes seqüències unes proteïnes anomenades factors de transcripció. Tot el conjunt s'anomena complex d'iniciació de la transcripció.

Etapa d'elongació o allargament

El procés de la síntesi continua en sentit 5' → 3'. Al cap d'uns trenta nucleòtids transcrits s'afegeix una mena de caputxa a l'extrem 5' per evitar la degradació.

Etapa de finalització

La síntesi de l'ARNm finalitza quan s'arriba a la seqüència TTATTT d'ADN. Seguidament, hi intervé l'enzim poli-A-polimerasa, que afegeix a l'extrem 3' uns 20 nucleòtids d'adenina, el que s'anomena cua de poli-A, al transcrit primari o preARNm.

Etapa de maduració

La maduració de l'ARNm en els organismes eucariotes es duu a terme al nucli. Un conjunt de proteïnes associades a ARN formen una estructura semblant a la d'un ribosoma formant l'anomenat espliceosoma*. Aquesta estructura elimina els introns i, a continuació, unes RNA-ligases empalmen els exons.

En els organismes eucariotes, l'ARNt i l'ARNr també pateixen un procés de maduració.

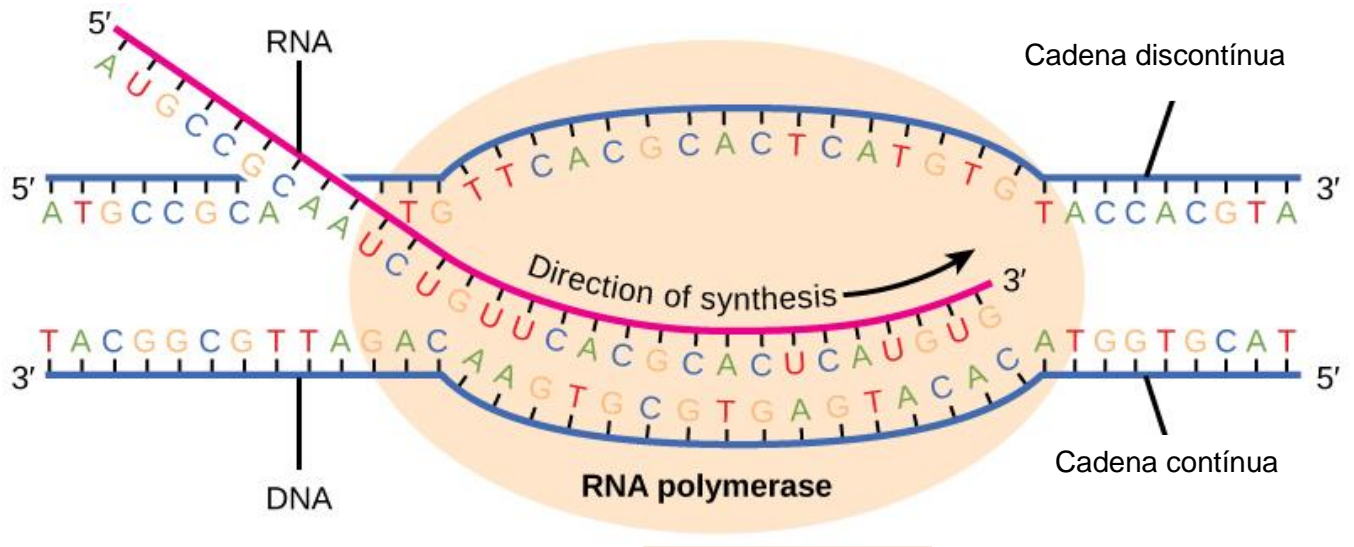


Figura 18: Procés de la transcripció.

2.1.4.5. La clau genètica o codi genètic

La interpretació de la clau genètica, és a dir, la relació que hi ha entre la seqüència de nucleòtids i la seqüència d'aminoàcids, es va aconseguir a partir dels descobriments següents:

- L'any 1955, Severo Ochoa i Grunberg-Manago van aïllar l'enzim polinucleòtid fosforilasa capaç de sintetitzar ARNm sense necessitat de model i a partir de qualsevol tipus de nucleòtids que hi hagués al medi.
- L'any 1961, Nirenberg va disposar una sèrie de vint tubs amb els vint aminoàcids proteics en cadascun. A cada tub hi havia un aminoàcid diferent marcat amb C-14¹⁸.

A tots els tubs va afegir l'enzim poli-U i tots els elements necessaris per a la síntesi proteica.

Va observar que tan sols apareixia un polipèptid radioactiu al tub on s'havia marcat la fenilalanina. A partir d'un poli-C en va identificar la col·linealitat amb la prolina.

A partir d'un poli-G no va obtenir resultats satisfactoris i a partir d'un poli-A va obtenir un polímer de lisines.

Com que tan sols hi ha quatre tipus de nucleòtids i com que hi intervenen vint tipus d'aminoàcids, la col·linealitat no es podia establir d'un en un, ni entre doblets de nucleòtids ($4^2 = 16$), sinó, com a mínim, entre triplets de nucleòtids ($4^3 = 64$) i aminoàcids.

- Posteriorment, i amb mesclades proporcionades de diferents ribonucleòtids difosfat, es va acabar de deduir la clau genètica i es va confirmar que la col·linealitat es devia establir entre els triplets de nucleòtids i els aminoàcids.

En el codi genètic podem observar que per alguns aminoàcids hi ha diversos triplets. Normalment difereixen en un sol nucleòtid (per exemple, UAU i UAC codifiquen l'aminoàcid Tyr). És per això que diem que el codi genètic és degenerat, el que és un avantatge, ja que, encara que es produís un error en la còpia d'un nucleòtid, continuaria la col·linealitat entre el triplet i l'aminoàcid corresponent.

¹⁸ El Carboni-14 és un isòtop radiactiu del carboni que s'utilitza per determinar característiques d'elements orgànics.

D'altra banda, si només hi hagués vint triplets que fossin traduïbles, hi hauria 44 triplets ($64 - 20 = 44$) sense sentit, i un simple error en un nucleòtid d'un triplet probablement el convertiria en un triplet sense sentit, i així s'interrompria la biosíntesi.

Una altra característica important del codi genètic és que és universal, és a dir, tots els éssers vius utilitzem el mateix codi genètic.

| | | Segona lletra | | | | |
|----------------------------|---|--|--------------------------------------|--|---|----------------------------|
| | | U | C | A | G | |
| Primera lletra (extrem 5') | U | UUU } Phe UUC } UUA } Leu UUG } | UCU } UCC } Ser UCA } UCG } | UAU } Tyr UAC } UAA } stop UAG } stop | UGU } Cys UGC } UGA } stop UGG } Trp | Tercera lletra (extrem 3') |
| | C | CUU } CUC } Leu CUA } CUG } | CCU } CCC } Pro CCA } CCG } | CAU } His CAC } CAA } Gln CAG } | CGU } CGC } Arg CGA } CGG } | |
| | A | AUU } AUC } Ile AUA } AUG } Met | ACU } ACC } Thr ACA } ACG } | AAU } Asn AAC } AAA } Lys AAG } | AGU } Ser AGC } AGA } Arg AGG } | |
| | G | GUU } GUC } Val GUA } GUG } | GCU } GCC } Ala GCA } GCG } | GAU } Asp GAC } GAA } Glu GAG } | GGU } GGC } Gly GGA } GGG } | |

Figura 19: Codi genètic.

2.1.4.6. La traducció

La traducció és la tercera etapa de la biosíntesi de les proteïnes i en aquest procés, l'ARNm produït en el procés de la transcripció es descodifica pels ribosomes per a produir una cadena d'aminoàcids o un polipèptid, que més tard serà plegada dins d'una proteïna activa.

Activació dels aminoàcids

Els aminoàcids, quan l'enzim aminoacil-RNAt-sintetasa i l'ATP són presents, tenen la capacitat d'associar-s'hi i donar lloc a un aminoacil-RNAt, de manera que

s'allibera ATP i queda lliure l'enzim, que després torna a actuar. La unió de l'aminoàcid al seu ARNt específic es duu a terme entre el seu grup carboxil (-COOH) i el radical -OH de l'extrem 3' de l'ARNt.

Etapa d'iniciació

L'ARNm produït en el procés anterior, la transcripció, s'uneix a la subunitat ribosòmica¹⁹ petita. Tot seguit, aquesta subunitat petita es mou respecte a l'ARNm fins a trobar el codó d'iniciació, que és: 5'... AUG ...3'. A aquests s'hi associa un aminoacil-RNAt iniciador específic, que presenta l'anticodó 3'... UAC ... 5' i que porta l'aminoàcid metionina (Met) en les cèl·lules eucariotes i l'aminoàcid formilmetionina en les cèl·lules procariotes. Seguidament, s'estableixen enllaços d'hidrogen entre el codó 5' ... AUG ...3' i el seu anticodó 3' ... UAC ... 5'. A aquest grup de molècules s'hi uneix la subunitat ribosòmica gran, i així es forma el complex ribosomal o complex actiu. Aquest procés necessita energia, que és aportada per un GTP²⁰ i unes proteïnes anomenades factors d'iniciació. En el complex ribosomal podem diferenciar tres centres:

- El centre **peptidil** o **P**: és el centre on se situa el primer aminoacil-RNAt.
- El centre **acceptor** o **A**: on es situen els aminoacils-RNAt següents.
- El centre de **sortida** o **E**: es situa l'RNAt sense el seu aminoàcid.

La iniciació de la síntesi és diferent entre cèl·lules procariotes i eucariotes. Com que en les cèl·lules procariotes l'ARNm no presenta el procés de maduració, immediatament després de la transcripció comença la traducció. En canvi, en les cèl·lules eucariotes, l'ARNm es sintetitza al nucli i abans de sortir d'aquest experimenta el procés de maduració. A l'extrem 5' duu una caputxa que permet que els ribosomes la identifiquin.

L'ARNm, si és suficientment llarg, pot ser traduït per uns quants ribosomes alhora, un rere l'altre, el que s'anomena poliribosoma.

Etapa d'elongació o allargament de la cadena polipeptídica

El primer triplet que es tradueix és l'AUG, que codifica l'aminoàcid metionina (Met). Al centre A arriba el segon aminoacil-RNAt. El radical carboxil de l'aminoàcid

¹⁹ Les parts de la cèl·lula eucariota animal es troben explicades a l'Annex (6.2)

²⁰ És una molècula que intervé en el metabolisme cel·lular, com l'ATP.

iniciador (metionina) s'uneix amb el radical amino de l'aminoàcid següent per mitjà d'un enllaç peptídic. D'aquesta forma, el centre P queda ocupat per un ARNt sense aminoàcid.

Aleshores es produeix la translocació ribosomal i aquest ARNt passa a ocupar el centre E i surt del ribosoma. El dipeptidil-RNAt ara queda al centre P i el centre A queda lliure a l'espera d'un nou aminoacil-RNAt.

Etapa de finalització de la síntesi

El final de la síntesi està determinat pels anomenats "triplets sense sentit", que són tres: UAA, UAG i UGA. No hi ha cap ARNt l'anticodó del qual en sigui complementari. En canvi, aquests triplets sense sentit o stop són reconeguts pels factors proteics d'alliberació (FR). Aquests s'instal·len sobre el centre A i fan que la cadena polipeptídica quedi alliberada. A continuació, l'ARNm i les dues subunitats ribosomals es separen.

Associació de diverses cadenes polipeptídiques per constituir les proteïnes

A mesura que la cadena polipeptídica es va sintetitzant, aquesta va adoptant una determinada estructura secundària i terciària²¹ mitjançant els enllaços per ponts d'hidrogen i ponts disulfur.

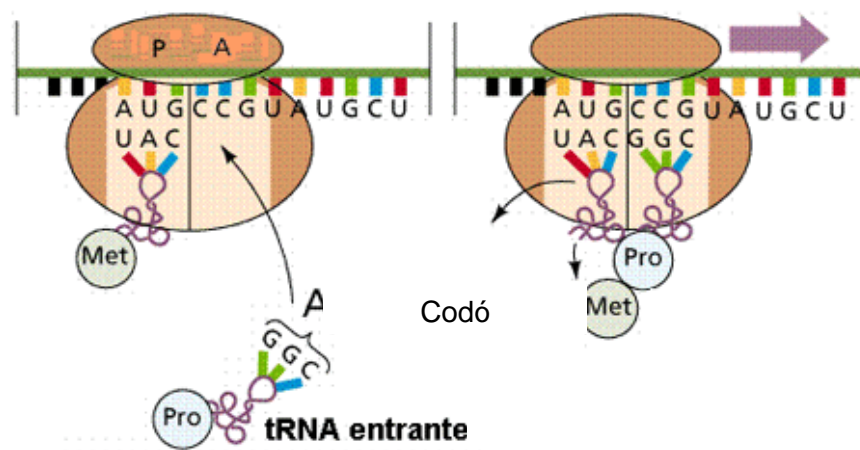


Figura 20: Procés de la traducció.

²¹ Com els àcid nucleics, les proteïnes també adopten diferents tipus d'estructures.

Taula 2: Quadre comparatiu de la transcripció i la traducció en cèl·lules procariotes i eucariotes:

| Cèl·lules procariotes | Cèl·lules eucariotes |
|--|---|
| L'ADN no està enrotllat formant nucleosomes i, per tant, no necessita desenrotllar-se per ser llegit. | L'ADN està enrotllat i forma nucleosomes, per la qual cosa generalment s'ha de desenrotllar. |
| Tan sols hi ha una bombolla de replicació i els fragments d'Okazaki tenen de 1.000 a 2.000 nucleòtids. | Hi ha moltes bombolles de replicació i els fragments d'Okazaki tenen de 100 a 200 nucleòtids. |
| Els gens són continus. | Normalment els gens són discontinus, és a dir, presenten introns intercalats amb exons. |
| La transcripció es fa al citosol. | La transcripció es fa a l'interior del nucli. |
| Hi ha un sol enzim RNA-polimerasa sigui quin sigui el tipus d'ARN que s'ha de sintetitzar. | Cada tipus d'ARN és sintetitzat per un dels tres tipus d'enzims RNA-polimerases. |
| L'ARNt i l'ARNr procedeixen d'un transcrit primari madurat. L'ARNm, a diferència de l'eucariòtic, se sintetitza directament, sense maduració prèvia. | Tots els transcrits primaris experimenten una maduració, però tan sols al preRNAm hi ha addició d'una caputxa a l'extrem 5', eliminació dels introns i addició de la cua de poli-A. |
| La traducció de l'ARNm s'inicia abans que acabi de ser sintetitzat. | L'ARNm ha d'anar des del nucli fins al citosol, on es duu a terme la traducció. |
| L'ARNm pot ser policistrònic, és a dir, pot donar lloc a més d'una proteïna. | L'ARNm sempre és monocistrònic, és a dir, dona lloc a només una proteïna. |
| El primer aminoàcid de la cadena és la formilmetionina. | El primer aminoàcid de la cadena és la metionina. |

2.1.5. LES MUTACIONS

Les mutacions són alteracions a l'atzar del material genètic, que pot ser ADN a les cèl·lules o ARN als virus). Normalment comporten diferències que poden arribar a ser letals, tot i que en general acostumen a ser recessives i queden amagades. Malgrat que normalment són negatives per a l'individu, comporten un aspecte positiu per a l'espècie, ja que aporten variabilitat a la població. Això permet que, si es produeix un canvi en l'ambient i les noves condicions són molt adverses per als individus normals, l'existència dels individus mutants faci possible que alguns suportin aquestes condicions i, gràcies a aquests, que l'espècie no s'extingeixi. Aquest procés s'anomena **selecció natural**. Per tant, les mutacions permeten l'evolució de les espècies i la continuïtat de la vida d'aquestes al llarg de milions d'anys.

Podem classificar les mutacions de diferents formes:

- Segons el tipus de cèl·lules que afecta:
 - Mutacions **somàtiques** : afecten a cèl·lules somàtiques; no tenen gaire importància, ja que si les cèl·lules no són viables, es poden substituir per altres cèl·lules, i si són viables, com que es divideixen per mitosi, donen lloc a una colònia de cèl·lules iguals a la primera.
 - Mutacions **germinals**: afecten a cèl·lules germinals²²; són molt transcendents ja que totes les cèl·lules del nou organisme tindran la mateixa informació que la cèl·lula zigot*, és a dir, afectaran a la descendència.
- Segons el seu origen:
 - Mutacions **naturals**: es donen de forma espontània.
 - Mutacions **induides**: són provocades artificialment per mitjà de radiacions i determinades substàncies químiques, que s'anomenen **agents mutàgens**.
- Poden ser:
 - Mutacions **positives**
 - Mutacions **negatives**
 - Mutacions **neutres**

²² O reproductores.

En els humans, la taxa de mutació espontània és molt alta: es calcula que de mitjana hi ha un gen mutat cada dos gàmetes. Com que cada zigot es forma a partir de dos gàmetes, a cada generació s'incorpora un gen mutat per individu; per tant, milions si es considera tota la població mundial. És l'anomenada **càrrega genètica d'efectes negatius**, que va augmentant a cada generació, sobretot en les poblacions més desenvolupades, ja que la utilització de fàrmacs permet la supervivència dels afectats.

Segons l'extensió del material genètic afectat, distingim tres tipus de mutacions:

- Mutacions **gèniques**: són alteracions en la seqüència de nucleòtids d'un gen.
- Mutacions **cromosòmiques**: són alteracions en la seqüència de gens d'un cromosoma.
- Mutacions **genòmiques**: són alteracions en el nombre de cromosomes.

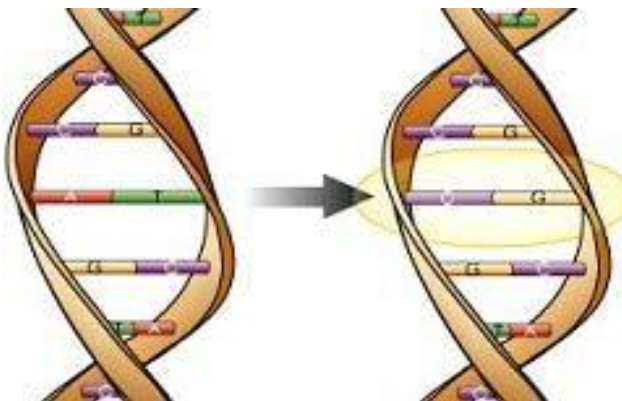


Figura 21: Exemple de mutació gènica.

2.1.5.1. Les mutacions gèniques

Les mutacions gèniques són alteracions en la seqüència de nucleòtids d'un gen. Segons el tipus d'alteració, es classifiquen en mutacions per **substitucions de bases** i en mutacions per **pèrdua o inserció de nucleòtids**.

- Mutacions per **substitució de bases**: són canvis d'una base per una altra. Provoquen l'alteració d'un triplet i, per tant, d'un aminoàcid.
- Mutacions per **pèrdua o inserció de nucleòtids**: s'eliminen o s'afegeixen una o diverses bases. Provoquen corriments de lectura i, per tant, alteren tota la

cadena (tots els triplets). Com que això provoca canvi de tota la proteïna, acostumen a comportar conseqüències greus.

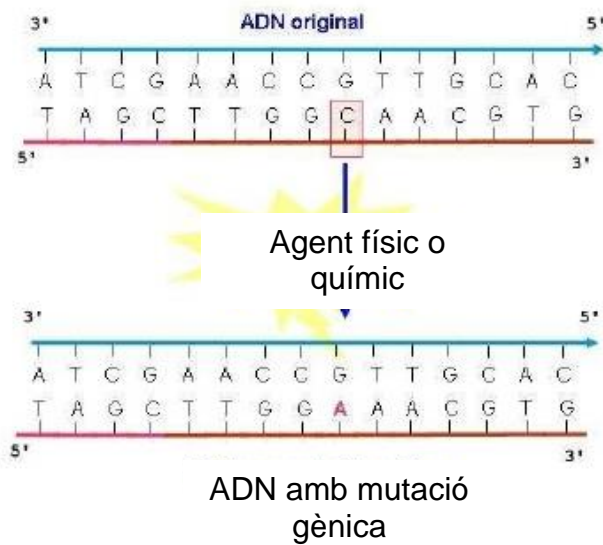


Figura 22: Exemple de mutació gènica

2.1.5.1.1. Causes de les mutacions gèniques

Les mutacions gèniques es poden produir per tres causes: per **errors de lectura** durant la replicació de l'ADN, per **lesions fortuïtes** o per **transposicions** d'uns segments del gen.

- **Errors de lectura:** són tots aquells errors que es produeixen durant la replicació de l'ADN.
- **Lesions fortuïtes:** són alteracions en l'estructura d'un o diversos nucleòtids que apareixen de forma natural. Les més freqüents són:
 - **Despurinitzacions:** pèrdua de bases purines (A o G) degut al trencament de l'enllaç entre elles i el glúcid.
 - **Desaminacions:** pèrdua de grups amino de les bases.
- **Dímers de timina:** enllaç entre dues timines (T) contigües degut a les radiacions solars.
- **Transposicions:** són canvis de lloc espontanis de determinats segments de l'ADN, els anomenats **elements genètics transposables**. Aquests elements poden produir una mutació gènica si l'element genètic transposat se situa a dins d'un gen, o mutacions cromosòmiques si passa a un lloc on no hi ha un gen, tant dins del mateix cromosoma om fins i tot en un altre cromosoma.

2.1.5.1.2. Les mutacions gèniques i els seus sistemes de reparació

El fet que l'enzim DNA-polimerasa presenti una activitat **exonucleasa** permet que aquest comprovi i corregeixi els nucleòtids, ja que la taxa d'addició d'un nucleòtid equivocat és d'un per cada deu milions de nucleòtids, el que és una taxa molt elevada per un ADN eucariota. Degut a això, cal un **sistema de reparació** que revisi constantment l'ADN per disminuir aquesta taxa. Distingim tres sistemes de reparació:

- **Reparació amb escissió d'ADN:** consisteix en un enzim que reconeix els errors i produeix dos talls als dos costats de l'error, després, un altre enzim elimina tots els nucleòtids del segment tallat i, finalment, una DNA-ligasa uneix l'extrem final.
- **Reparació sense escissió d'ADN:** consisteix en enzims fotoreactius capaços de trencar els dímers de timina.
- **Sistemes SOS:** són uns enzims que permeten la duplicació de l'ADN amb errors per evitar que aquesta quedi bloquejada, això comporta que es generin moltes cèl·lules amb moltes mutacions.

2.1.5.2. Les mutacions cromosòmiques

Les mutacions cromosòmiques són aquelles que provoquen canvis en l'estructura interna dels cromosomes. Distingim tres tipus:

- **Deleció:** És la pèrdua d'un fragment del cromosoma. Si el fragment conté una gran quantitat de gens, la deleció pot tenir conseqüències greus.
- **Duplicació:** És la repetició d'un segment del cromosoma. La rèplica pot trobar-se en el mateix cromosoma, haver-se unit a un cromosoma no homòleg, o fins i tot haver-se independitzat amb un centròmer propi. Les duplicacions permeten augmentar el material genètic i, gràcies a les mutacions posteriors, poden determinar l'aparició de nous gens durant el procés evolutiu.
- **Inversió:** És el canvi de sentit d'un fragment en el cromosoma. Les inversions no acostumen a comportar perjudicis a l'individu però sí als descendents si durant la meiosi es produeix un encreuament dins de la inversió.
- **Translocació:** És el canvi de posició d'un segment del cromosoma. Quan es produeix per intercanvi de segments entre dos cromosomes homòlegs, parlem

d'una translocació recíproca. En canvi, quan tan sols hi ha una translocació d'un segment a un altre lloc del mateix cromosoma o d'altres cromosomes, parlem de transposició o translocació no recíproca. Aquestes no solen perjudicar a l'individu que les pateix però si a la descendència, ja que aquesta pot heretar un cromosoma incomplet o amb duplicacions.

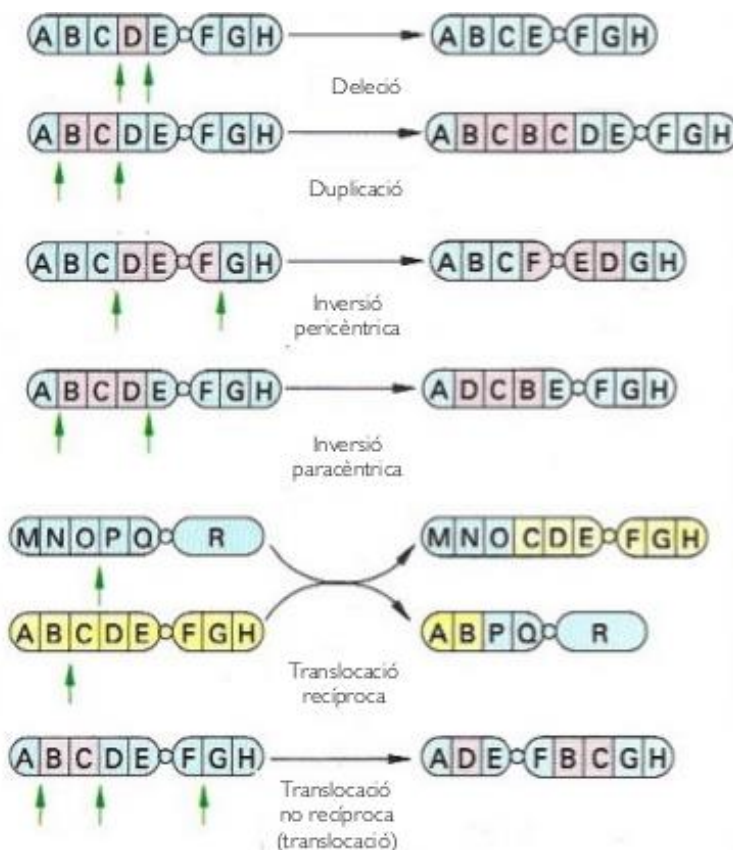


Figura 23: Tipus de mutacions cromosòmiques.

2.1.5.3. Les mutacions genòmiques

Les mutacions genòmiques són les alteracions en el nombre de cromosomes propi d'una espècie. Poden ser:

- **Aneuploïdia:** És l'alteració en el nombre normal d'exemplars d'un o més tipus de cromosomes. Poden ser nul·liploïdies, monosomies, trisomies (com en el cas de la síndrome de Down, en què l'individu té una trisomia en el cromosoma 21), tetrasomies...
- **Euploïdia:** És l'alteració en el nombre normal de dotacions haploides (jocs de cromosomes) d'un individu. Poden ser:

- **Monoplòidia:** una sola dotació cromosòmica.
- **Poliplòidia:** existència de més de dos jocs complets de cromosomes.

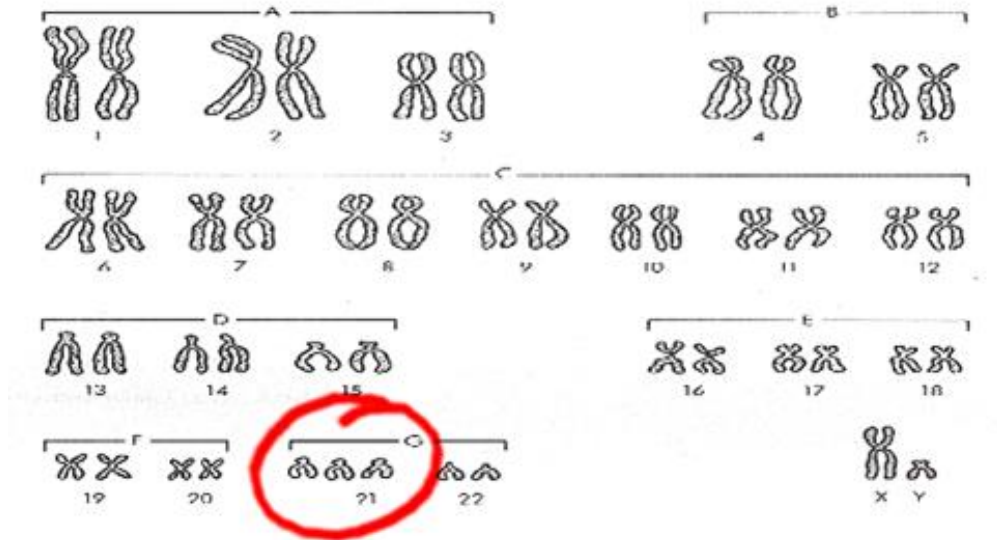


Figura 24: Cariotip del Síndrome de Down, exemple de mutació genòmica.

2.2. LA BIOINFORMÀTICA

2.2.1. QUÈ ÉS LA BIOINFORMÀTICA?

La **Bioinformàtica** és una àrea d'investigació on s'apliquen les ciències de la computació i les tecnologies de la informació per al tractament de dades biològiques. La Bioinformàtica és una àrea d'investigació multidisciplinària, ja que pot ser àmpliament definida com la interfase entre dues ciències: la Biologia i la Computació; a més, està impulsada per la incògnita del genoma humà i l'objectiu d'estudiar aquest per ajudar a millorar la condició i qualitat de vida humana.

Millores en la detecció i el tractament de malalties i la producció d'aliments genèticament modificats són altres exemples dels beneficis més coneguts. La Bioinformàtica involucra la solució de problemes complexos utilitzant eines de sistemes i computació. També inclou la col·lecció, organització, emmagatzemament i recuperació de la informació biològica que es troba a la base de dades.

Segons la definició del Centre Nacional per a la Informació Biotecnològica, "National Center for Biotechnology Information" (NCBI per les sigles en anglès, 2001): "La Bioinformàtica és un camp de la ciència en el qual conflueixen diverses disciplines com: biologia, computació i tecnologia de la informació. El principal objectiu d'aquest camp és facilitar el descobriment de noves idees biològiques així com crear perspectives globals a partir de les quals es puguin discernir principis unificadors en Biologia. Al començament de l'anomenada "revolució genòmica", el concepte de Bioinformàtica es referia tan sols a la creació i el manteniment de bases de dades on s'emmagatzemava informació biològica, com seqüències de nucleòtids i aminoàcids. El desenvolupament d'aquest tipus de bases de dades no només significava el disseny de les mateixes sinó també el desenvolupament de programes complexos on els investigadors poguessin accedir a les dades existents i subministrar o revisar-les. Després, tota aquella informació s'havia de combinar per formar una idea lògica de les activitats cel·lulars normals, de tal manera que els investigadors poguessin estudiar com aquestes activitats es veien alterades en una malaltia. D'aquí ve el sorgiment del camp de la Bioinformàtica i ara el camp més

important en la anàlisi i la interpretació de diferents tipus de dades, incloent seqüències de nucleòtids i aminoàcids, dominis de proteïnes i estructura d'aquestes. El procés d'analitzar i interpretar les dades és conegut com Biocomputació. Dins de la bioinformàtica i la Biocomputació existeixen altres subdisciplines importants:

- El desenvolupament d'eines que permetin l'accés, ús i maneig de diferents tipus d'informació.
- El desenvolupament de nous algorismes²³ i estadístiques en els quals es pugui relacionar parts d'un conjunt de dades, com per exemple mètodes per localitzar un gen dins d'una seqüència, predir l'estructura o funció de proteïnes i poder agrupar seqüències de proteïnes en famílies relacionades.”

La Medicina molecular i la Biotecnologia constitueixen dues àrees prioritàries científico-tecnològiques com el desenvolupament i la innovació tecnològica. El desenvolupament de totes dues està estretament relacionat. En les dues àrees es pretén potenciar la investigació genòmica, que és al que es dedica la Bioinformàtica, una eina imprescindible per al desenvolupament d'aquestes.



Figura 25

²³ Fórmules matemàtiques.

2.2.2. HISTÒRIA DE LA BIOINFORMÀTICA

No es pot parlar de la història de la Bioinformàtica sense descriure inicialment la història de la Biologia. En realitat són els biòlegs i els bioquímics qui fan el seu primer apropament a la tecnologia computacional com element fonamental per al seu treball diari.

La biocomputació ha estat la base per ajudar en les grans investigacions sobre la vida, ja que la tecnologia proporciona un element teòric i eines bàsiques, per a que els científics puguin explorar les proteïnes i l'ADN.

2.2.2.1. Les primeres dècades: anys 60 i 70 del segle XX

Als anys 60, L. Pauling va elaborar la seva teoria sobre l'evolució molecular i Margaret Dayhoff, una de les pioneres de la bioinformàtica, va publicar el primer Atles de seqüències de proteïnes, que es va convertir en una obra bàsica en el desenvolupament estadístic i és el precursor de les actuals bases de dades de proteïnes.

Als anys 70 es publica l'algoritme per a alinear seqüències, es crea la primera molècula d'ADN recombinant, es comença la seqüenciació de l'ADN i el desenvolupament de software per a analitzar-lo i es publica la primera seqüència de gens completa d'un organisme, el fago Φ -X174 (5.386 parells de bases que codifiquen 9 proteïnes).

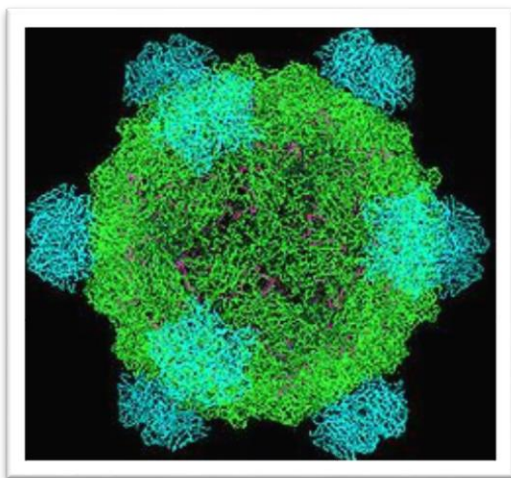


Figura 26 : Bacteriòfag fago Φ -X174.

2.2.2.2. Anys 80

Als anys 80, es publica el mètode d'utilització de la RMN (Ressonància magnètica nuclear) per a determinar estructures de proteïnes i es descobreix la PCR (Reacció en cadena de la polimerasa, és un mètode per a multiplicar mostres d'ADN per a analitzar-les). També apareixen nous algorismes per a cerca en bases de dades de seqüències, com FASTA, que és un algorisme que compara seqüències. Sorgeixen també bases de dades biològiques, com el GenBank.

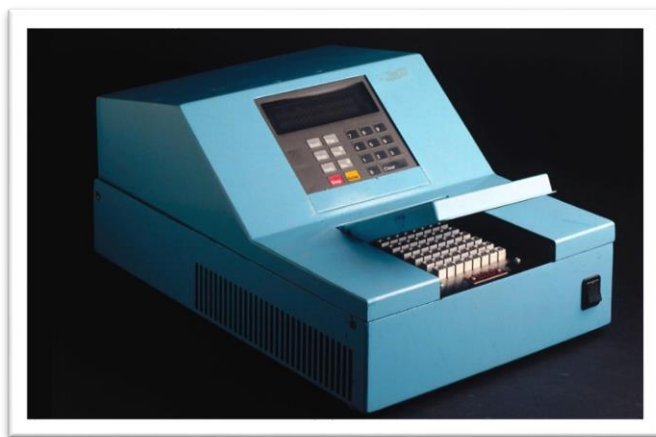


Figura 27: PCR.

2.2.2.3. Anys 90

Pel que fa a esdeveniments científics, es publica el mapa de lligament genètic del genoma humà, s'aconsegueix seqüenciar completament els primer genomes de bacteris i el primer genoma eucariota, el de la llevadura, també el de *Escherichia coli* com el primer genoma d'un organisme pluricel·lular i, finalment, la seqüenciació del primer cromosoma humà, el 22.

Si parlem de Bioinformàtica, es crea BLAST, un programa que fa cerca ràpida de similitud entre seqüències, també es crea CrystalW, orientat a l'alineament de seqüències.



Figura 28

2.2.2.4. Primers anys del segle XXI

En aquests primers anys del segle XXI, cal destacar que s'han seqüenciat molts genomes de diferents organismes, com: *Arabidopsis thaliana*, *Drosophila melanogaster* (mosca de la fruita), *Rattus norvegicus* (rata), el gat domèstic, es completa el Human Genome Project (és a dir, la seqüenciació del genoma humà). Finalment, també es seqüencia el genoma d'una dona.



Figura 29

2.2.3. RELACIÓ ENTRE BIOLOGIA I INFORMÀTICA

Hem de distingir tres interfases en les que s'uneixen la Biologia i la Informàtica, amb objectius i metodologies ben diferenciades:

- **Bioinformàtica o Biologia Molecular Computacional:** investigació i desenvolupament d'infraestructures i sistemes d'informació i comunicacions que requereixen la biologia molecular i la genètica. → Informàtica aplicada a la biologia molecular i la genètica.
- **Biologia Computacional:** computació que s'aplica a la comprensió de qüestions biològiques bàsiques, no necessàriament a nivell molecular, sinó mitjançant la modelació i la simulació d'ecosistemes, models fisiològics... → Informàtica i Matemàtiques aplicades a la Biologia.
- **Biocomputació:** desenvolupament i ús de sistemes computacionals basats en models i materials biològics, com bioxips, biosensors, xarxes de neurones, computació basada en l'ADN... → Biologia aplicada a la computació.

Bàsicament, els sistemes informàtics usats en aquest camp són:

- Bases de dades
- Software per a visualització
- Generació i empalmament de seqüències
- **Programes per a anàlisis de seqüències**
- Programes per a la predicció de l'estructura de les proteïnes
- Creació de mapes genètics
- Software per a classificar i comparar seqüències
- Tècniques d'intel·ligència artificial
- Gestió de dades
- Literatura mèdica i científica de les seqüències
- Distribució de dades
- Xarxes de comunicacions
- Aplicacions
- Gestió de dades del laboratori
- Automatització d'experiments
- Predicció de funcions de determinades seqüències
- **Alineació de seqüències**
- Cerca en les bases de dades d'estructures
- Predicció de gens
- Evolució molecular: arbres filogenètics
- Documents de difusió i recolzament a la Bioinformàtica

2.2.4. OBJECTIUS DE LA BIOINFORMÀTICA

Entre molts d'altres, els objectius principals de la Bioinformàtica són:

- **Organitzar les dades en bases de dades:** la Bioinformàtica organitza les dades permetent als investigadors accedir a la informació ja coneguda i publicar noves dades a mesura que se'n produeixen. La informació que emmagatzemen aquestes bases de dades no serveix fins que s'analitza, el que vol dir que la curació i el manteniment de les dades és una feina essencial realitzada per la Bioinformàtica.
- **Desenvolupament d'eines i recursos que ajuden en l'anàlisi de dades:** per exemple, un dels interessos principals de la Biologia és comparar una seqüència nova d'una proteïna amb seqüències ja caracteritzades. Es necessiten programes com BLAST, que troba les regions de similitud entre les seqüències, el que ens permetria predir funcions.
- **Utilitzar aquestes eines per analitzar les dades i interpretar els resultats:** la Bioinformàtica ha revolucionat els estudis biològics ja que, tradicionalment, aquests estudis es concentraven en els detalls dels sistemes, en canvi, actualment, es poden realitzar anàlisis globals amb totes les dades disponibles, el que permet descobrir principis comuns i noves característiques dels sistemes biològics.



Figura 30

2.2.5. APLICACIONES I ÀREES D'INVESTIGACIÓ

La Bioinformàtica s'està utilitzant en molts camps: des de la medicina molecular als estudis evolutius, de la teràpia gènica al desenvolupament de fàrmacs i fins i tot s'aplica a estudis de canvi climàtic, entre d'altres.

Les aplicacions pràctiques més rellevants són:

- **Cerca d'homòlegs:** és la recerca de similituds entre diferents molècules. Per exemple, això permet trobar una funció hipotètica d'una proteïna mal caracteritzada buscant homòlegs. També s'utilitza en genòmica, per confirmar regions codificants en noves dades genòmiques seqüenciades.
- **Disseny de fàrmacs:** és l'aplicació mèdica més important de la Bioinformàtica. Consisteix en la recerca de nous medicaments basats en el ja coneixement d'alguna substància biològica, com per exemple una proteïna. Podem utilitzar la Bioinformàtica per dissenyar un fàrmac que activi o inhibeixi la proteïna a la que s'uneix, el que acaba sent un benefici terapèutic per al pacient.

En resum, tots els estudis bioinformàtics es basen en dues aproximacions:

- Comparar dades d'acord a similituds biològicament importants
- Deduir i comprendre les observacions de les dades mitjançant l'anàlisi d'altres tipus de dades

Pel que fa a les àrees d'investigació, les més importants són:

- Anàlisi de seqüències
- Anotació de genomes
- Biologia evolutiva computacional
- Estudi de la biodiversitat
- Anàlisi de l'expressió gènica
- Anàlisi de la regulació gènica
- Anàlisi de l'expressió de proteïnes
- Anàlisi de mutacions en el càncer
- Predicció de l'estructura de les proteïnes
- Model de sistemes biològics

2.2.6. IMPORTÀNCIA DE LA BIOINFORMÀTICA

Integració és la paraula clau per entendre la importància de la Bioinformàtica, ja que mitjançant eines i utilitzant la informació ja emmagatzemada en les bases de dades, estem començant a descobrir determinades relacions biològiques amagades en el codi de la vida.

La Bioinformàtica ha començat a ocupar un paper central com la disciplina que uneix a diverses àrees de la ciència, com la enzimologia, la genètica, la biologia estructural, la medicina, la morfologia...entre molts d'altres. La pregunta important és: com es poden aconseguir relacions importants entre tanta informació? Aquesta pregunta i molts altres problemes biològics estan sent resposos a través de la Bioinformàtica , unint o relacionant tota la informació que està emmagatzemada en bases de dades a través de les seves associacions en gens.

A més, a mesura que nous projectes genoma generen grans quantitats de dades, l'aplicació d'ordinadors s'està convertint en una eina essencial en les ciències biològiques. La Bioinformàtica neix de la urgent necessitat de transformar una gran quantitat de dades d'ADN en coneixement.

En conclusió, el fet que la Bioinformàtica ens permeti descobrir moltes relacions biològiques i que existeixin bases de dades amb tota la informació emmagatzemada (com l'NCBI o el GenBank) ens permet avançar i estudiar en molts camps abans desconeguts. A més, les eines de la Bioinformàtica han canviat molts àmbits de l'estudi de les ciències. Actualment, tots els equips d'investigació en Biologia Molecular, Genètica i Medicina utilitzen softwares més especialitzats i, ara, el paper de científic en aquests tres àmbits, més que fer experiments, és interpretar correctament les dades.

2.2.7. PERFILS D'UN BIOINFORMÀTIC

Distingim dos tipus de bioinformàtics que són molt diferents:

- Els que utilitzen eines bioinformàtiques ja existents com BLAST, CrustalW, Transeq...
- Els que desenvolupen noves eines bioinformàtiques

El pas intermedi entre aquests dos tipus de bioinformàtics és aquell que escriu els seus propis scripts, que és el que aprendrem en aquest treball.

L'habilitat de poder crear nous petits programes utilitzant un codi existent és molt útil: ens permet estalviar molt temps, fer millor ciència i obtenir nous resultats.

2.2.8. ELS PROGRAMES MÉS UTILITZATS EN BIOINFORMÀTICA

Els programes més utilitzats en Bioinformàtica i els que jo he usat durant la meva estada a la Universitat Autònoma de Barcelona són: BLAST, CrustalW i Perl.

- **BLAST:** Basic Local Alignment Search Tool; és un programa bioinformàtic que alinea seqüències ja sigui d'ADN o de proteïnes. El programa compara una seqüència "problema" amb seqüències que es troben en una base de dades.
- **CrustalW:** és un programa bioinformàtic que realitza alineaments múltiples de seqüències.
- **Perl:** és un llenguatge de programació d'alt nivell, estable, interpretat i multiplataforma dissenyat per Larry Wall. Els programes escrits amb Perl s'anomenen scripts²⁴.



Figura 31

²⁴ En informàtica, un script és un guió o conjunt d'instruccions que manem a l'ordinador.

3. MARC PRÀCTIC

3.1. Exercici d'evolució: *Evolució dels óssos*

Us heu preguntat mai quina relació tenen l'ós bru americà amb els óssos polars? I amb quin d'aquests dos organismes està més relacionat el panda gegant? Com és

| Organisme | Nom científic | Número d'accés |
|----------------------|------------------------|----------------|
| Ós negre americà | Ursus americanus | Y08520.1 |
| Ós marró americà | Ursus arctos | L21889.1 |
| Ós d'antifaç | Tremarctos ornatus | L21883.1 |
| Ós asiàtic | Selenarctos thibetanus | L21890.1 |
| Ós polar | Thalarctos maritimus | L22164.1 |
| Panda gegant | A.Melanoleuca | Y08521.1 |
| Panda vermell | Ailurus fulgens | S80939.1 |
| Ós rentador del Nord | Procyon lotor | U78345.1 |
| Foca lleopard | Hydrurga leptonyx | AY377288.1 |

la història evolutiva dels óssos? En aquest exercici utilitzarem seqüències de gens reals per esbrinar quins animals estan més estretament relacionats i crearem un arbre genealògic d'aquests.

Taula 3: Nom científic i comú dels organismes que s'estudiaran i número d'accés a la seqüència del gen d'interès.

1. En primer lloc, accedirem a la pàgina de l'NCBI²⁵ i clicarem la fletxa del menú desplegable "All databases", seleccionarem **Nucleotide**. Tot seguit, introduïrem el **número d'accés**²⁶ de l'ós negre americà en l'espai de la cerca.

²⁵ <http://www.ncbi.nlm.nih.gov>

²⁶ Aquest número d'accés ens portarà directament a la seqüència del gen que ens interessa, però si cerquéssim (en anglès) el nom del gent i el corresponent organisme trobaríem el mateix.

2. Veurem un munt d'informació. En els primers resultats, veurem una definició com *U. Americanus* mitocondrial 12S rRNA gene. *U. Americanus* és el nom científic de l'animal i aquest gen forma part de l'ADN mitocondrial i del ribosoma.
3. Cliquem a **FASTA** (sota el menú GenBank, a dalt a l'esquerra) al principi de la pàgina. Veurem la seqüència sencera de l'ADN del gen rRNA 12S.
4. A la dreta, veurem un menú desplegable que posa "Send". Cliquem i escollim **Clipboard**. Tot seguit, cliquem **Add to Clipboard**.
5. Continuem la cerca amb cada animal, seguint el mateix procés (FASTA → clipboard)
6. Quan haguem realitzat la cerca de les 9 seqüències i les haguem afegit en format fasta al "clipboard", cliquem en els "**9 items**" que veurem al costat de l'opció *Clipboard*.
7. Després d'haver afegit les 9 seqüències (en format FASTA) al portapapers, cliquem en els "9 ítems" que veurem al costat de la opció *Clipboard*.
8. Seleccionem cada animal clicant la caixeta de l'esquerra.
9. Després d'haver seleccionat els 9 organismes, clica a "Send to" a la part superior esquerra de la pàgina.
10. Escollim "File" i sota el Format, seleccionem FASTA.
11. Escollim "Organism Name" sota "Sort by".
12. Cliquem a "Create File".
13. El nostre fitxer serà creat amb el nom de **sequence.fasta**, el descarreguem. Tot seguit utilitzarem les seqüències del gen que hem triat, l'rRNA 12S i les alinearem utilitzant un programa anomenat ClustalW. Aquesta web utilitza algorismes matemàtics per trobar similituds entre organismes.
14. Entrem a la web de ClustalW²⁷.
15. Escollim "CRUSTAL" com a format del fitxer de sortida.
16. Cliquem a "SLOW/ACCURATE" per fer l'alineament.
17. Cliquem a "DNA" per introduir les nostres seqüències.
18. Ara pujarem el fitxer escollint "Choose File". Naveguem per l'ordinador per escollir el fitxer que hem descarregat, **sequence.fasta**.
19. Cliquem a "Exercute Multiple Aligment".

²⁷ <http://www.genome.jp/tools/clustalw/>

20. El programa alinearà les seqüències i ens mostrarà els resultats de la següent forma:

```
CLUSTAL 2.1 Multiple Sequence Alignments

Sequence type explicitly set to DNA
Sequence format is Pearson
Sequence 1: gi|1871572|emb|Y08520.1|          966 bp
Sequence 2: gi|418058|gb|L21889.1|URSMTRG12S  347 bp
Sequence 3: gi|418052|gb|L21883.1|TMCMTTRG12S  346 bp
Sequence 4: gi|418045|gb|L21890.1|SNAMTRG12S  347 bp
Sequence 5: gi|418048|gb|L22164.1|TLTMTRG1SA  346 bp
Sequence 6: gi|1871553|emb|Y08521.1|          966 bp
Sequence 7: gi|1911545|gb|S80939.1|           349 bp
Sequence 8: gi|2231359|gb|U78345.1|PLU78345   1073 bp
Sequence 9: gi|37575891|gb|AY377288.1|        964 bp
Start of Pairwise alignments
Aligning...
```

Figura 32: Primera part dels resultats que ens retorna CrustalW.

21.A

l'inici de la pàgina web veurem les 9 seqüències amb els números d'accés i el nombre de parell de bases (bp) per cada seqüència. El programa ens ha numerat els organismes de l'1 al 9:

- 1) Y08520.1 → Ós negre americà
- 2) L21889.1 → Ós marró americà
- 3) L21883.1 → Ós d'antifaç
- 4) L21890.1 → Ós asiàtic
- 5) L22164.1 → Ós polar
- 6) Y08521.1 → Panda gegant
- 7) S80939.1 → Panda vermell
- 8) U78345.1 → Ós rentador del Nord
- 9) AY377288.1 → Foca lleopard

22. Ara anem cap avall i veurem les puntuacions dels alineaments:

Exemple: Sequences (3:2) Aligned. Score: 56.

23. Aquest resultat representa la puntuació entre els organismes 3 i 2: com més alta és aquesta puntuació, més similar és l'ADN d'aquests organismes.

24. Si anem encara més abaix i mirem els alineaments veurem les seqüències de nucleòtids. El símbol "-" indica que no hi ha informació per aquella localització en la seqüència: la base nucleotídica (A,T,C,G) no és coneguda. Això és degut a que algunes seqüències són més llargues que altres, hi ha seccions on la informació no és coneguda. El símbol "*" indica que el nucleòtid és el mateix per aquella posició: la base està conservada, és a dir, que el nucleòtid haurà d'estar en l'ancestre comú.

Resultats:

```
Aligning...
Sequences (1:2) Aligned. Score: 95
Sequences (1:3) Aligned. Score: 92
Sequences (1:4) Aligned. Score: 95
Sequences (1:5) Aligned. Score: 92
Sequences (1:6) Aligned. Score: 87
Sequences (1:7) Aligned. Score: 89
Sequences (1:8) Aligned. Score: 44
Sequences (1:9) Aligned. Score: 86
Sequences (2:3) Aligned. Score: 91
Sequences (2:4) Aligned. Score: 96
Sequences (2:5) Aligned. Score: 92
Sequences (2:6) Aligned. Score: 91
Sequences (2:7) Aligned. Score: 86
Sequences (2:8) Aligned. Score: 89
Sequences (2:9) Aligned. Score: 91
Sequences (3:4) Aligned. Score: 94
Sequences (3:5) Aligned. Score: 98
Sequences (3:6) Aligned. Score: 93
Sequences (3:7) Aligned. Score: 86
Sequences (3:8) Aligned. Score: 88
Sequences (3:9) Aligned. Score: 91
Sequences (4:5) Aligned. Score: 94
Sequences (4:6) Aligned. Score: 92
Sequences (4:7) Aligned. Score: 85
Sequences (4:8) Aligned. Score: 87
Sequences (4:9) Aligned. Score: 91
Sequences (5:6) Aligned. Score: 93
Sequences (5:7) Aligned. Score: 86
Sequences (5:8) Aligned. Score: 88
Sequences (5:9) Aligned. Score: 91
Sequences (6:7) Aligned. Score: 89
Sequences (6:8) Aligned. Score: 44
Sequences (6:9) Aligned. Score: 86
Sequences (7:8) Aligned. Score: 90
Sequences (7:9) Aligned. Score: 91
Sequences (8:9) Aligned. Score: 45
Guide tree file created: [clustalw.dnd]
```

Figura 33: Segona part dels resultats que ens torna ClustalW.

El programa compara tots els animals entre ells, de forma que ens permet saber quins s'assemblen més entre ells. Com podem observar, els organismes 3 i 5, l'ós d'antifaç i l'ós polar, són els que més puntuació tenen entre ells, el que vol dir que són els que més similitud tenen. Com en aquest cas, si veiem que dos grups tenen una puntuació molt alta, probablement un deriva de l'altre. Pel contrari, els que tenen la puntuació més baixa són els animals l'1 amb el 8 i el 6 amb el 8. Aquestes semblances i/o diferències es podrien explicar amb l'aïllament de les espècies, és a dir, els que més s'assemblen segurament han viscut menys aïllats en el temps que els que menys s'assemblen.

Ara utilitzarem una pàgina web per construir un arbre filogenètic. És un arbre que mostra les relacions evolutives entre diverses espècies que es creu que van tenir una descendència comuna, és a dir, s'utilitzen per a conèixer com es troben emparentats els organismes i per saber si tenen un ancestre comú. La pàgina utilitzarà les dades que hem utilitzat i dibuixarà un arbre filogenètic.

25. A dalt de la mateixa pàgina de CrustalW, per sobre de les dades de l'alineament, seleccionem "Rooted phylogenetic tree with branch length (UPGMA)" del menú desplegable.

26. Cliquem a "Exec."

27. Observem l'arbre.

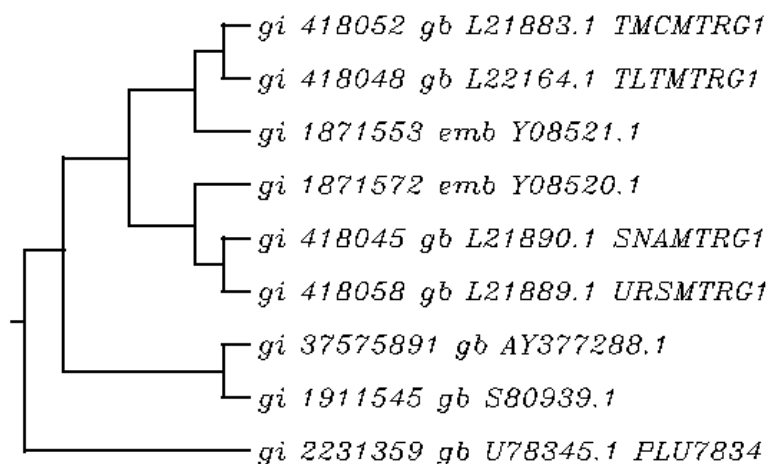


Figura 34: Arbre filogenètic que ens elabora CrustalW, amb els codis corresponents de cada animal. Les línies corresponen al temps.

3.2. PERL

3.2.1. QUÈ ÉS UN LLENGUATGE DE PROGRAMACIÓ?

Un **llenguatge de programació** és un llenguatge informàtic que s'utilitza per controlar el comportament d'una màquina, normalment un ordinador. Amb aquest programa l'ordinador ens entendreà i podrem manar-li que faci tot el que nosaltres li demanem.

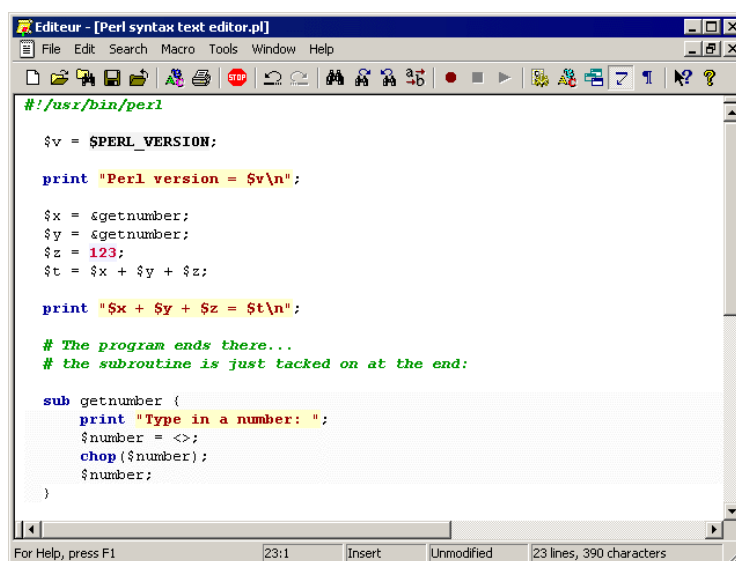
Existeixen molts llenguatges de programació, com FORTRAN, ALGOL, BASIC, C, C++, Pascal, Java, PERL...

3.2.2. PER QUÈ PERL?

PERL és un llenguatge de programació d'alt nivel, estable i interpretat que va ser dissenyat per Larry Wall l'any 1987.

Tot i que és difícil aprendre a programar en poc temps (ja que hi ha moltes funcions per aprendre, molta sintaxi...), el fet que PERL sigui molt fàcil, **podem fer molt sabent poc**. A més d'això, és un llenguatge ràpid d'escriure, multi plataforma (Windows, Mac i Unix) i és un codi lliure.

El meu objectiu en aquesta part del treball és que entengueu les bases de PERL, sapigueu quin tipus de coses pot fer i aprengueu a escriure scripts.

A screenshot of a text editor window titled "Editeur - [Perl syntax text editor.pl]". The window contains a Perl script with the following code:

```
#!/usr/bin/perl

$v = $PERL_VERSION;

print "Perl version = $v\n";

$x = &getnumber;
$y = &getnumber;
$z = 123;
$t = $x + $y + $z;

print "$x + $y + $z = $t\n";

# The program ends there...
# the subroutine is just tacked on at the end:

sub getnumber {
    print "Type in a number: ";
    $number = <>;
    chop ($number);
    $number;
}
```

The status bar at the bottom of the window shows "For Help, press F1", "23:1", "Insert", "Unmodified", and "23 lines, 390 characters".

Figura 35: Exemple d'script amb PERL.

3.2.4. EL MEU PROPI PROGRAMA AMB PERL

A continuació trobareu captures de pantalla del programa que jo he realitzat amb el llenguatge de programació PERL.

El meu programa, que s'anomena *infoProtein* ens escriu la seqüència d'ADN d'interès, calcula la seva longitud ([pb]: parells de bases), ens informa de la seva cadena complementària i de la complementària reversa, ens proporciona l'ARN corresponent, calcula el nombre de bases A,T,C i G i el percentatge de bases CG, el nombre de codons i ens pregunta si volem trobar alguna seqüència de restricció. Finalment, el programa desa la proteïna en format .txt.

En aquesta primera captura de pantalla, podem veure els scripts que obren el programa, escriuen la seqüència, calculen la longitud en [pb], la cadena complementària, la complementària reversa i la corresponent seqüència d'ARN a partir de la seqüència d'ADN que li proporcionem. També trobem els scripts que calculen el nombre de bases A,T,C i G i calculen el percentatge de bases CG:

```
infoProteinFINAL.pl x
1
2  #!/usr/bin/perl
3
4  print "Welcome to infoProtein, introduce the name of your file:\n";
5  #Obrir el fitxer d'interes
6  $fitxer = <>;
7  chomp($fitxer);
8
9  open (FILE2, "$fitxer") or die "Error, the file $fitxer doesn't exist\n";
10 @array = <FILE2>;
11 print "Select a name for your output file\n";
12 $Z= <>;
13 chomp($Z);
14 print "Select a sequence line width for your output FASTA file:\n";
15 $X= <>;
16 chomp($X);
17
18 open (FILE, ">$Z") or die "Error, the file $Z can't be created\n";
19
20 ($DNA, $ID) = obre_fasta(@array);
21 print "ID:\n$ID\n";
22 print "SEQUENCE:\n$DNA\n\n";
23
24 #Longitud del DNA
25 $longitudDNA = length$DNA;
26 print "The length of the DNA is: $longitudDNA pb\n\n";
27
28 #Cadena complementaria
29 $DNAcompl = $DNA;
30 ($DNAcompl) = cadena_compl($DNA);
31 print "The complementary sequence is: \n$DNAcompl\n\n";
32
33 #Cadena complementaria reversa
34 @DNAcompl = split("", $DNAcompl); #Passem de variable escalar a matriu
35 @DNAcompl_reversa = reverse(@DNAcompl);
36 $DNAcompl_reversa = join("", @DNAcompl_reversa);
37 ($DNAcompl_reversa) = cadena_compl_rev($DNAcompl);
38 print "The complementary reverse sequence is:\n$DNAcompl_reversa\n\n";
39
40 #RNA
41 $RNA = $DNA;
42 ($RNA) = DNAaRNA($DNA);
43 print "The RNA is: \n$RNA\n\n";
44
45
46 #Nombre de bases A,T,C i G i percentatge CG
47 @DNA = split('', $DNA);
48 ($cont_A, $cont_T, $cont_C, $cont_G, $cont_errors, $percentatge) = nombre_bases(@DNA);
49 print "Number of bases A: $cont_A\n";
50 print "Number of bases T: $cont_T\n";
51 print "Number of bases C: $cont_C\n";
52 print "Number of bases G: $cont_G\n";
53 print "Number of errors: $cont_errors\n";
54 printf "Percentage of bases GC: %0.2f % \n\n", $percentatge;
55
56
Line 1, Column 1
```

En la següent imatge, trobem els scripts que calculen el nombre de codons i el que ens pregunta si volem trobar alguna seqüència de restricció:

```
infoProteinFINAL.pl
57 #Nombre de codons
58 ($num_codons) = codons($DNA);
59 print "Number of codons: $num_codons\n\n";
60
61
62
63
64 $semaforo = 0;
65 while ( ! $semaforo ) {
66
67 print "Do you want to find any restriccion sequences?";
68 $resposta = <>;
69 chomp($resposta);
70
71 if ($resposta ne "yes" || $resposta ne "no")
72 {
73     print "I don't understand you. Write yes/no\n\n";
74 }
75 if ($resposta eq "yes")
76 {
77     $semaforo = 1;
78
79     print "Which one?\n\n";
80     $fragment = <>;
81     chomp($fragment);
82     if ($DNA =~ /$fragment/)
83     {
84         print "The restriccion sequence has been found.\n\n";
85     }
86     else
87     {
88         print "The restriccion sequence hasn't been found.\n\n";
89     }
90     $posicio = index ($DNA, $fragment);
91     $posicio = $posicio +1; #pq index comença des de 0
92     $longitud = length$DNA;
93     print "It is in the position $posicio of a total of $longitud\n\n"; #sempre sha de fer punt i coma!
94
95
96
97 }
98 if ($resposta eq "no")
99 {
100     $semaforo = 1;
101     print "Continuing...\n\n";
102 }
103 }
104
105
106
107 #Obre fasta i proteina final
108 ($proteina) = traduccio($DNA);
109 print "Final protein:\n$proteina\n\n";
110
111 print "Saving your protein\n\n";
112
```


Si vols aprendre a programar de la forma més senzilla i, com no, a partir de dades nucleotídiques, no ho dubtis i visita la meva pàgina web. Allà trobaràs informació sobre Bioinformàtica i un petit curs relacionat amb la programació amb PERL.

<http://nlupiongarcia.wixsite.com/treballderecerca/gib>

3.3. ENQUESTES

Quan vaig començar el treball tenia la sensació de que tant la Genètica com la Bioinformàtica no són disciplines conegudes o, si més no, tan conegudes com es mereixen. Com que no n'estava segura, vaig decidir realitzar una enquesta per comprovar-ho.

L'enquesta que he fet és la següent:

1) Saps què és la genètica?

- a. Sí
- b. No

2) En cas afirmatiu, quina creus que és la millor definició?

- a. És una biociència.
- b. És la ciència que estudia l'herència biològica dels éssers vius.
- c. És la biologia que estudia l'herència dels arbres genealògics.

3) Saps amb quins mètodes es pot estudiar genètica?

Resposta oberta

4) Saps què és la Bioinformàtica?

- a. Sí
- b. No
- c. Molt poc

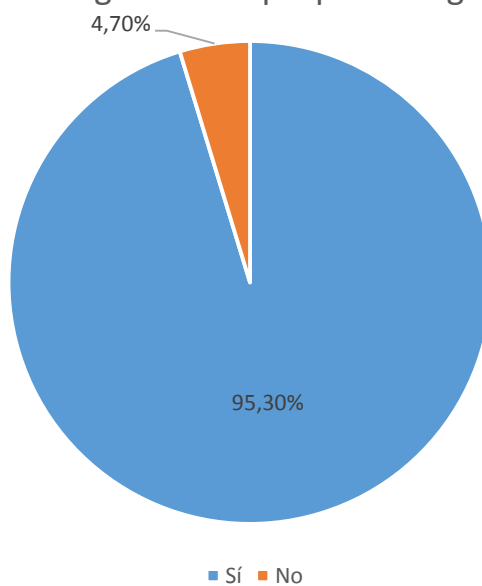
5) En cas afirmatiu, saps quines aplicacions té en el camp de la investigació?

- a. Sí
- b. No

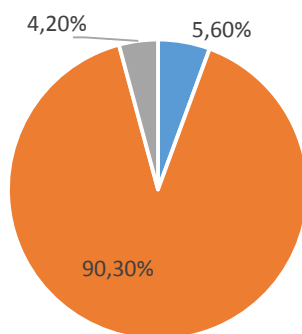
6) Si és que sí, cita algun exemple

Resposta oberta

Gràfic 1- Pregunta 1: Saps què és la genètica?



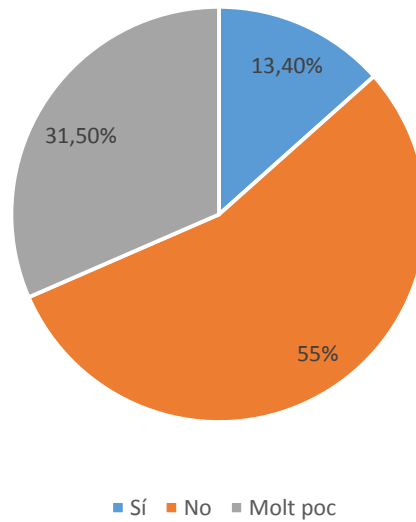
Gràfic 2- Pregunta 2: En cas afirmatiu, quina creus que és la millor definició?



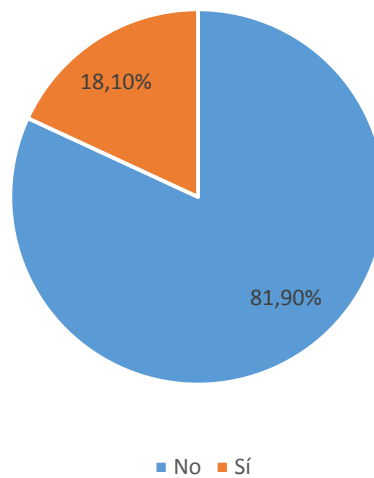
- És una biociència.
- És la ciència que estudia l'herència biològica dels éssers vius.
- És la biologia que estudia l'herència dels arbres genealògics.

Pregunta 3: Saps amb quins mètodes es pot estudiar genètica?

Gràfic 3- Pregunta 4: Saps què és la bioinformàtica?



Gràfic 4- Pregunta 5: En cas afirmatiu, saps quines aplicacions té en el camp de la investigació?



Pregunta 6: Si és que sí, cita algun exemple

Resultats:

1) Saps què és la genètica?

En aquesta pregunta, un 95,3% dels enquestats saben el que és la Genètica, mentre que un 4,7% no ho saben. Cal tenir en compte que de les 149 persones que han realitzat l'enquesta també hi havia persones grans.

2) En cas afirmatiu, quina creus que és la millor definició?

En aquesta pregunta hi havia tres possibles respostes: la primera, semicorrecta; la segona, correcta i la tercera, totalment incorrecta. Com podem observar, un 5,6% ha contestat "És una biociència", un 90,3% ha respost "És la ciència que estudia l'herència biològica dels éssers vius" i un 4,2% ha contestat "És la biologia que estudia l'herència dels arbres genealògics".

Amb aquests resultats interpreto que del 95,3% que sap què és la Genètica, un petit percentatge no ha encertat la definició, el que vol dir que no ho sabien gaire bé.

3) Saps amb quins mètodes es pot estudiar Genètica?

Aquesta era una resposta oberta; el meu objectiu era que algú, i si podia ser molta gent, contestés "Amb bioinformàtica", però de les 149 persones que van contestar l'enquesta ningú ho va pensar. Això em fa pensar que la bioinformàtica no és coneguda, o, si més no, no és coneguda com a eina per a estudiar Genètica.

4) Saps què és la bioinformàtica?

Pel que fa a la quarta pregunta, un 13,4% si ho saben, un 55% no la coneixen i un 31,5% només la coneixen una mica. Amb aquests resultats interpreto que hi ha un petit percentatge que la coneix realment, encara que no sigui en l'àmbit de la Genètica, un percentatge que n'ha sentit parlar, i la resta que no ho saben.

5) En cas afirmatiu, saps quines aplicacions té en el camp de la investigació?

En aquesta pregunta, com podem observar, un 81,9% va respondre que no coneixia cap aplicació en el camp de la investigació, mentre que un 18,1% sí.

6) Si és que sí, cita algun exemple

Aquesta també era una resposta oberta i l'objectiu era que els enquestats pensessin en la recerca genètica, la creació de fàrmacs, anàlisi de genomes... I així va ser, de les persones que coneixien algunes aplicacions en el camp de la investigació, la majoria van pensar en el que jo buscava.

En conclusió i tenint en compte els resultats de l'enquesta, puc dir que la Bioinformàtica no és gaire coneguda o, si més no, no ho és com s'ho mereix, ja que aquesta ens ajuda a resoldre molts problemes complexos utilitzant eines de sistemes i computació.

4. CONCLUSIONS

Després d'haver realitzat el meu treball de recerca, he pogut saber si els meus objectius s'han complert i si les meves hipòtesis són certes.

En primer lloc, m'agradaria parlar de la meva experiència al Programa Argó de la UAB.

Durant l'estada que em va oferir el Programa Argó de la Universitat Autònoma de Barcelona vaig aprendre què és la Bioinformàtica realitzant diferents exercicis en relació a aquesta i aprenent a desenvolupar un software Bioinformàtic per a analitzar i seqüenciar mostres d'ADN.

L'estada es va desenvolupar a l'IBB (l'Institut de Biomedicina i Biotecnologia). Les pràctiques van consistir en fer un programa bioinformàtic. També vaig realitzar diferents exercicis d'evolució que permeten comprovar la relació entre espècies.

Al final de l'estada, vam haver d'exposar un vídeo²⁸ resum a tots els companys del Programa Argó 2016 de la nostra feina durant aquelles setmanes.

Aquest estiu he après moltíssim, m'he familiaritzat amb el món universitari i he pogut acabar de decidir el que vull estudiar després del Batxillerat, Genètica.

Pel que fa als objectius que em vaig proposar a l'inici del treball, puc afirmar que els he complert tots: he après molt més del que sabia dels mons de la Genètica i de la Bioinformàtica i he aconseguit realitzar l'exercici d'evolució i el meu propi programa.

Respecte a les hipòtesis que vaig plantejar al principi del meu treball, puc constatar, gràcies a l'anàlisi que he realitzat (les aplicacions bioinformàtiques més utilitzades en el camp de la Biologia i la Genètica, el meu exercici d'evolució...), que avui en dia la Bioinformàtica és el motor de la Genètica i de l'estudi evolutiu de les espècies actuals, ja que amb una petita mostra de material genètic de diferents animals, es pot elaborar un arbre filogenètic per esbrinar les relacions de parentiu i temporals

²⁸ https://www.youtube.com/watch?v=Vqy-L_11cGo

entre ells. Per tant, avui dia, la Bioinformàtica resulta una eina imprescindible per als biòlegs i genetistes. De fet, si consultem els plans d'estudis dels diferents graus relacionats amb la Biologia, ens trobarem que la Bioinformàtica forma part de les assignatures obligatòries en tots ells.

En quant a la segona hipòtesi, les enquestes em permeten afirmar que la Bioinformàtica no és una disciplina coneguda. La major part dels enquestats no n'han sentit parlar a no ser que formin part de la comunitat científica.

És per aquest motiu que m'ha semblat interessant crear una pàgina web divulgativa sobre el tema que podria fer que la Bioinformàtica arribés a més gent. La pàgina web l'he pensat bàsicament pels estudiants de secundària.

Per finalitzar, m'agradaria dir que he gaudit molt fent el meu treball, tot i que m'hagi portat feina. N'he après un munt i que espero que el meu objectiu principal, fer que la Bioinformàtica es difongui, es compleixi.

5. Fonts

Bibliografia

Els llibres que he utilitzat per a realitzar la meva recerca han estat:

- BERG JM, TYMOCZKO JL, STRYER L. (2013) *Bioquímica*, 7ª Edició. Editorial Reverté, Barcelona.
- WATSON JD, BAKER TA, BELL SP, GANN A, LEVINE M, LOSICK R (2016) *Biología molecular del Gen*, 7a edició. Editorial Panamericana.
- JIMENO A., BALLESTEROS M., RODRÍGUEZ S. (2016) *Biología 2 Batxillerat*, 3a edició. Editorial Santillana Promotor.
- PIERCE, B.A. (2015) *Genética : un enfoque conceptual*, 2a edició. Editorial Médica Panamericana, Madrid.
- BROWN, T.A. (2008) *Genomas*, 3a edició. Editorial Médica Panamericana, Buenos Aires.
- LODISH H. (2016) *Biología celular y molecular*, 7a edició (anglesa traduïda). Editorial Panamericana.
- LEWIN B. (2008) *Genes IX*, 9a edició. Editorial McGraw-Hill Interamericana.
- GRIFFITHS, A.J.F. (2012) *Introduction to genetic analysis*, 10a edició. Editorial Freeman, Nova York.

Webgrafia

INFORMACIÓ:

Les pàgines web que he utilitzat per a realitzar la meva recerca han estat:

- <http://adntr.blogspot.com.es/p/descobriments-de-watson-i-crick.html>
- <http://bioinf3.uab.cat/expobi2/>
- <http://bioinformatica.crear-foro.com/t14-importancia-de-la-bioinformatica>
- <http://bioinformatica.uab.es/base/base3.asp?sitio=msbioinformaticasca>
- http://bioinformatica.uab.es/genetica_tfg/bioinformaticaabast/index.html
- http://bioinformatica.uab.es/genetica_tfg/bioinformaticaabast/Perl_2_cat.html

- <http://bioinformaticaumsa.blogspot.com.es>
- <http://biomedbiotec.encb.ipn.mx/bioinformatica/Programas.html>
- <http://blast.ncbi.nlm.nih.gov/Blast.cgi>
- http://bvs.isciii.es/bib-gen/Actividades/curso_virtual/Introduccion/bioinformatica.htm
- <http://definicion.de/bioinformatica/>
- http://genfis40.esalq.usp.br/genfis/index.php?option=com_content&view=article&id=60&Itemid=69
- http://html.rincondelvago.com/james-watson-i-francis-crick_importancia-del-seu-descobrimient.html
- <http://recursos.cnice.mec.es/biosfera/alumno/4ESO/genetica1/contenidos10.htm>
- <http://silvioalejandro.tripod.com>
- <http://teachingbioinformatics.fandm.edu/activities/bear-evolution>
- <http://uvigen.fcien.edu.uy/utem/genmen/Gen%E9ticaMendeliana.pdf>
- <http://www.biologiasur.org/index.php/herencia/genetica-mendeliana>
- <http://www.bionova.org.es/biocast/tema18.htm>
- http://www.cecalc.ula.ve/BIOINFO/red_analisis/que_es_bio.htm
- <http://www.cuevadeloso.com/web-cueva-oso-tella/origen>
- <http://www.ecured.cu/Bioinformática#Historia>
- <http://www.jbc.org/content/280/49/e46>
- <http://www.mhhe.com/biosci/genbio/raven6b/graphics/raven06b/howscientiststhink/09-lab.pdf>
- <http://www.monografias.com/trabajos/genetica/genetica.shtml>
- <http://www.nature.com/ng/index.html>
- http://www.ncbi.nlm.nih.gov/nuccore/470231499?from=874&to=1832&sat=4&start_key=93305367&report=fasta
- http://www.nobelprize.org/nobel_prizes/medicine/laureates/1959/kornberg-bio.html
- <http://www.solociencia.com/biologia/bioinformatica-genomica-funcional.htm>
- https://books.google.es/books?id=e4KyhjDw-goC&pg=PA211&lpg=PA211&dq=crick+i+gamov+george&source=bl&ots=Lee_e_vq1C7&sig=s4_8fos3RvU42xiqy5qZKZukXjs&hl=ca&sa=X&ved=0ahUKEwj

[PgNWKkpiOAhWHCcAKHZm DkMQ6AEIUjAH#v=onepage&q=crick%20i%20gamov%20george&f=false](https://www.nobelprize.org/nobel_prizes/medicine/laureates/1959/kornberg-bio.html)

- https://www.nobelprize.org/nobel_prizes/medicine/laureates/1959/kornberg-bio.html
- <http://learn.genetics.utah.edu>
- http://www.physicsoftheuniverse.com/scientists_gamow.html
- <http://www.omim.org>
- http://nar.oxfordjournals.org/content/33/suppl_1/D514.full
- <https://www.perl.org>
- <http://www.perl.com>
- <http://www.tutorialspoint.com/perl/>
- <http://lagenetica.info/ca/genetica-present-i-futur/projecte-genoma-huma/>
- <http://www.enciclopedia.cat/EC-GEC-0248009.xml>
- <http://www.webconsultas.com/pruebas-medicas/pcr-13299>
- <http://learn.genetics.utah.edu/content/labs/pcr/>
- <http://www.bioinformatics.org>
- <http://www.bioinformaticsbarcelona.eu>
- <http://www.batanga.com/curiosidades/2010/07/20/grandes-cientificos-watson-crick>
- http://www.iesallusser.es/1esobiowp/wp-content/uploads/2013/03/celula_animal.jpg
- <http://apuntesbiologiamol.blogspot.com.es/2014/03/codigo-geneticocomo-se-descifro.html>

FIGURES:

- Figura 1: Emaze, *DNA used in criminal investigations* (<https://www.emaze.com/@ALTLCZCI/DNA-used-in-Criminal-Evidence>)
- Figura 2: <http://www.formulatorsampleshop.com/FSS-Pisum-Sativim-Peptide-p/fss16810.htm>
- Figura 3: Quién, *Biografía de Gregorio Mendel* (<http://www.quien.net/gregorio-mendel.php>)
- Figura 4: Bionova, *Las leyes de Mendel* (<http://www.bionova.org.es/biocast/tema18.htm>)

- Figura 5: Saber es práctico, *Las tres leyes de Mendel* (<http://www.bionova.org.es/biocast/tema18.htm>)
- Figura 6: Saber es práctico, *Las tres leyes de Mendel* (<http://www.bionova.org.es/biocast/tema18.htm>)
- Figura 7: Tresciencias, *James Watson y Francis Crick* (<http://tresciencia.blogspot.com.es/2011/03/biografia-de-j-watson-y-f-crick.html>)
- Figura 8: Columbian College of Arts and Sciences, *George Gamow* (<https://physics.columbian.gwu.edu/george-gamow>)
- Figura 9: Viquipèdia, *Arthur Kornberg* (https://en.wikipedia.org/wiki/Arthur_Kornberg)
- Figura 10: Office of NIH History, *Deciphering the Genetic Code* (https://history.nih.gov/exhibits/nirenberg/HS4_polyU.htm)
- Figura 11: <http://dariosilvestri.me/studi-clinici/>
- Figura 12: <https://cienciaad.wordpress.com/2014/02/13/proyecto-genoma-humano/>
- Figura 13: Cimat, *Introducción a la biología molecular* (http://www.cimat.mx/ciencia_para_jovenes/tcj/2000/biologia/intro.html)
- Figura 14: Science Prof Online, *RNA Structure and function* (<http://www.scienceprofonline.com/genetics/ribonucleic-acid-rna-structure-and-function.html>)
- Figura 15: <http://www.biologyexams4u.com/2012/10/differences-between-dna-and-rna.html>
- Figura 16: <http://www.maph49.galeon.com/adn/classical.html>
- Figura 17: Vikipèdia, *Replicació de l'ADN* (https://ca.wikipedia.org/wiki/Àcid_desoxiribonucleic)
- Figura 18: BC Open Textbooks, *Transcription* (<https://opentextbc.ca/biology/chapter/9-3-transcription/>)
- Figura 19: Aulatres, *El codi genètic* (<http://aulatres.wikispaces.com/El+codi+genètic?responseToken=8a734f04c4a9d416d2e967c518983a8b>)
- Figura 20: Proyecto Biosfera, *La traducción* (<http://recursos.cnice.mec.es/biosfera/alumno/2bachillerato/genetica/contenido12.htm>)

- Figura 21: <http://definicion.de/mutacion/>
- Figura 22: http://www.iespando.com/web/departamentos/biogeo/web/departamento/2BC/H/B4_INFORMACION/T411_MUTACIONES/INDICE.htm
- Figura 23: Slideshare, *Les mutacions cromosòmiques* (<http://es.slideshare.net/daniribo/68-les-mutacions-cromosmiques>)
- Figura 24: <https://biologiemotoc.wikispaces.com/Activitate+a+3?responseToken=0c64c2316900f85921a098a9bf6d178a9>
- Figura 25: <http://dhmri.org/services-support/bioinformatics>
- Figura 26: Viquipèdia, *Phi X-174* (<https://es.wikipedia.org/wiki/Phi-X174>)
- Figura 27: Viquipèdia, *Polymerase chain reaction* (https://en.wikipedia.org/wiki/Polymerase_chain_reaction)
- Figura 28: <http://libguides.westsoundacademy.org/c.php?g=361523&p=2443286>
- Figura 29: Nature Jobs, *The impact of the Human Genome Project* (<http://blogs.nature.com/naturejobs/2015/10/08/big-data-the-impact-of-the-human-genome-project/>)
- Figura 30: <http://www.mincyt.gob.ar/noticias/bioinformatica-aplicada-a-la-atencion-clinica-de-ninos-con-hiv-en-el-hospital-garrahan-10428>
- Figura 31: <http://blog.builtonperl.com/post/perl-is-back-in-the-top-10-programming-languages>
- Figura 32: <https://www.ebi.ac.uk/Tools/msa/clustalo/>
- Figura 33: <https://www.ebi.ac.uk/Tools/msa/clustalo/>
- Figura 35: Duiops, *Morfología de la célula eucariota animal* (http://www.duiops.net/seresvivos/celula_morfo_morf.html)

GRÀFIQUES I TAULES : totes les gràfiques i taules han estat d'elaboració pròpia.

6. ANNEX

6.1. Glossari

- **Àcids nucleics:** són polímers formats per la unió de nucleòtids, que són les unitats més senzilles que es van repetint al llarg de la cadena. Els àcids nucleics poden ser **àcid ribonucleic (ARN)** o **àcid desoxiribonucleic (ADN)**.
- **ADN:** àcid desoxiribonucleic. Format per una doble hèlix de nucleòtids. És la molècula portadora de la informació genètica i és el constituent principal de la cromatina i, per tant, dels cromosomes.
- **Al·lel dominant:** igual que amb el caràcter dominant, l'al·lel dominant és aquell que té més tendència a manifestar-se. Es simbolitza amb lletra majúscula.
- **Al·lel recessiu:** al·lel que no s'expressa amb tanta freqüència ja que un altre (el dominant) té més capacitat de manifestar-se. Es simbolitza amb lletra minúscula.
- **Al·lel:** cadascuna de les diverses varietats de gens que contenen la informació per a un mateix caràcter i que ocupen la mateixa posició en els cromosomes homòlegs.
- **Aminoàcids:** molècules que contenen els grups funcionals amino i carboxil. Formen les proteïnes.
- **Caràcter dominant:** són els més comuns i es produeixen gairebé amb totes les combinacions d'al·lells menys quan s'uneixen dos al·lells mutants que seria quan es produiria el caràcter recessiu.
- **Caràcter recessiu:** caràcter que no s'expressa amb tanta freqüència ja que un altre té una major capacitat de manifestar-se.
- **Cèl·lula eucariota:** tipus de cèl·lula que posseeix membrana nuclear.
- **Cèl·lula procariota:** tipus de cèl·lula que no posseeix membrana nuclear.
- **Codó:** també anomenat triplet, és la seqüència de tres nucleòtids en un àcid nucleic.

- **Cromatina:** material genètic de la cèl·lula eucariota que es troba al nucli.
- **Diploide:** cèl·lula o individu la dotació cromosòmica del qual està constituïda per dues sèries de cromosomes (és a dir, tenen dos exemplars de cada cromosoma). Es simbolitza com “2n”.
- **Enzim:** biomolècules que catalitzen²⁹ les reaccions químiques.
- **Espliceosoma:** complex que elimina els introns del preARNm.
- **Exó:** cadascun dels segments d'ADN codificants d'un gen que contenen informació i no són eliminats en la maduració de l'ARN missatger.
- **Fenotip:** conjunt dels caràcters observables en un organisme. Depèn del genotip i de l'acció ambiental.
- **Fragment d'Okazaki:** sèrie curta d'ADN creada sobre la cadena residual durant el procés de replicació de l'ADN.
- **Gàmetes:** cadascuna de les cèl·lules haploides produïdes per òrgans germinals (mascles i femelles) que s'uneixen en la reproducció sexual. Els gàmetes masculins es produeixen durant l'espermatogènesi, mentre que els gàmetes femenins durant l'oogènesi.
- **Genealogia:** disciplina que estudia l'ascendència i la successió de les persones i determina els seus parentius.
- **Genoma:** ADN contingut als cromosomes, en una cèl·lula o en un organisme. El genoma dels organismes procariotes està format per un sol cromosoma circular. En els organismes eucariotes, el genoma està format pel conjunt dels diversos cromosomes.
- **Genotip:** conjunt de gens presents en un organisme.
- **Gens:** els gens són la unitat bàsica del material hereditari. És un fragment d'àcid nucleic que porta la informació genètica per a un caràcter, una proteïna o una cadena polipeptídica.
- **Haploide:** cèl·lula o individu la dotació cromosòmica del qual està constituïda per una sola sèrie de cromosomes (és a dir, només tenen un exemplar de cada cromosoma). Es simbolitza com “n” (entenen “n” com nombre de cromosomes)

²⁹ És a dir, que n'augmenten la velocitat.

- **Heterozigot:** També anomenat híbrid. Individu que posseeix els dos al·lels diferents per a un caràcter.
- **Híbrid:** organisme viu procedent de l'encreuament sexual entre dues espècies diferents o de varietats genètiques diferents. Per tant, l'híbrid presenta una barreja de les característiques genètiques dels dos progenitors.
- **Histona:** proteïnes bàsiques de baix pes molecular.
- **Homozigot:** També anomenat raça pura. Individu que té dos al·lels iguals per a un caràcter determinat (ja sigui aa o AA).
- **Intró:** regió d'ADN compresa en la part codificant d'un gen però que no s'arriba a expressar, és a dir, la seva seqüència no s'utilitza quan se sintetitza la proteïna corresponent.
- **Locus:** posició fixa d'un gen en un cromosoma.
- **Locus:** posició fixa d'un gen en un cromosoma. Un ADN codificat en un locus que conté un gen és anomenat al·lel.
- **Meiosi:** procés de divisió cel·lular que permet a una cèl·lula diploide generar cèl·lules haploides en cèl·lules eucariotes.
- **Mitosi:** fase del cicle cel·lular en què es produeix la divisió del nucli cel·lular en les cèl·lules eucariotes. Es reparteixen les dues còpies del material genètic en dues parts iguals, per formar els nuclis de les cèl·lules filles.
- **Nucleosoma:** estructura que constitueix la unitat fonamental i essencial de la cromatina, que és la forma d'organització de l'ADN en els eucariotes.
- **Proteïnes:** són compostos orgànics formats per aminoàcids mitjançant enllaços peptídics entre els grups carboxil i amino dels aminoàcids que s'uneixen. La seqüència d'aminoàcids d'una proteïna és definida per la seqüència d'un gen, que està codificada al codi genètic.
- **Ribosoma:** orgànul no membranós que es troba a totes les cèl·lules vives. Reuneix les vint molècules específiques d'aminoàcids per formar determinades proteïnes.
- **Zigot:** primera cèl·lula d'un nou individu que s'obté per la fusió d'un gàmeta femení amb un gàmeta masculí .

6.2. Estructura de la cèl·lula eucariota

animal

Les cèl·lules eucariotes, pròpies dels organismes pluricel·lulars, són les que tenen el nucli cel·lular delimitat per una membrana³⁰, formada per una doble capa lipídica, anomenada **embolcall nuclear**. Jo explicaré les principals parts de la cèl·lula eucariota animal, però també existeix la cèl·lula eucariota vegetal que, tal i com el seu nom indica, és pròpia dels vegetals.

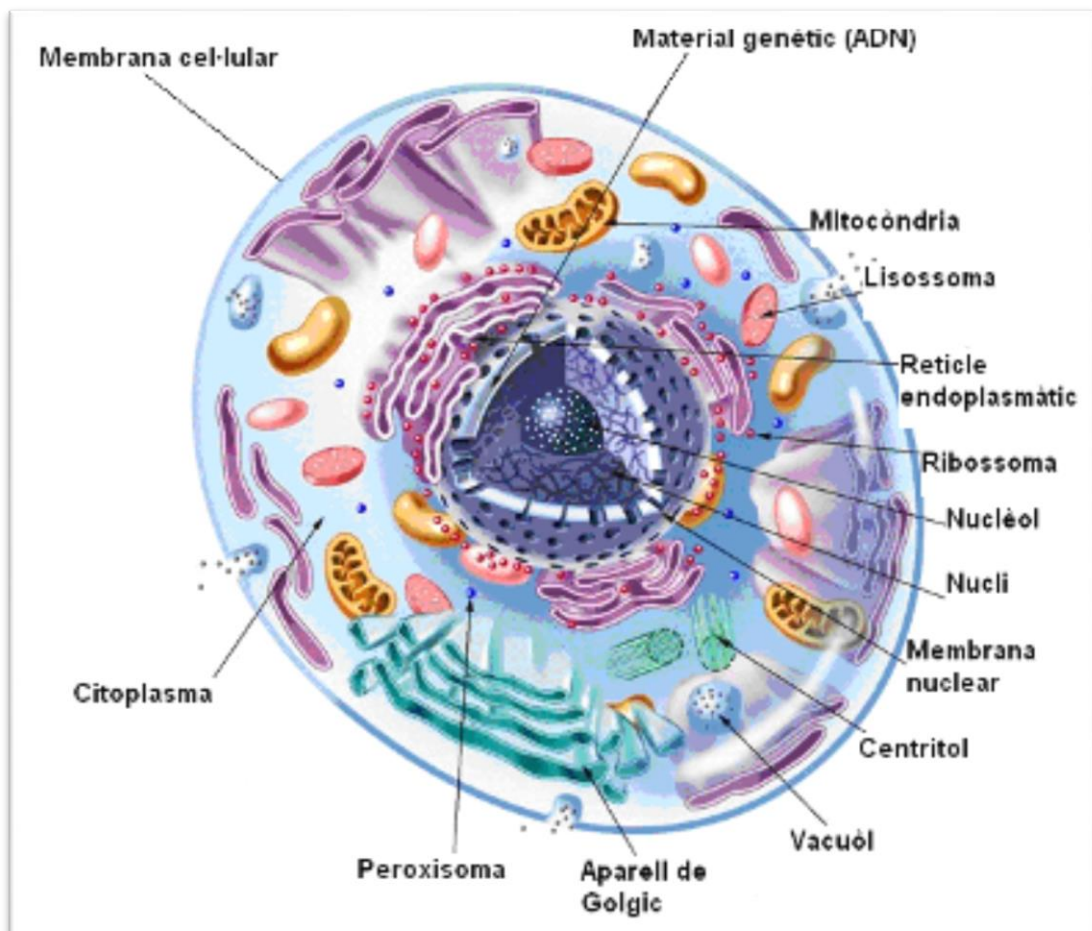


Figura 36: Parts de la cèl·lula eucariota animal

³⁰ A diferència de les cèl·lules procariotes, que no tenen el nucli envoltat per una membrana.

- **Membrana cel·lular:** és una fina membrana que envolta la cèl·lula, la protegeix i permet l'intercanvi de substàncies entre la cèl·lula i el medi.
- **Citoplasma:** està format fonamentalment per aigua i sobre ell hi trobem els orgànuls.
- **Mitocondri:** és l'orgànul que proporciona energia a la cèl·lula per a la seva activitat cel·lular.
- **Reticle Endoplasmàtic Rugós:** format per ribosomes, la seva funció és sintetitzar proteïnes, que seran enviades cap a l'exterior de les cèl·lules.
- **Ribosoma** òrganuls molt petits que es troben a totes les cèl·lules vives. Reuneixen les vint molècules específiques d'aminoàcids per formar proteïnes determinades per seqüències de molècules d'ARN.
- **Nucli:** orgànul embolicat per una membrana en totes les cèl·lules eucariotes. Conté gran part del material genètic de la cèl·lula, organitzat en forma de múltiples llargues molècules lineals d'ADN en combinació amb gran varietat de proteïnes, que contribueixen a formar cromosomes.
- **Membrana nuclear:** doble membrana del nucli que conté el material genètic a les cèl·lules eucariotes. Conté nombrosos porus nuclears per tal de facilitar i regular l'intercanvi de substàncies (com proteïnes i ARN) entre el nucli i el citoplasma.
- **Nuclèol:** orgànul del nucli cel·lular que té com a funció principal la síntesi d'ARN.
- **Centríol:** tipus d'orgànul cel·lular propi de les cèl·lules eucariotes. S'encarreguen de separar els cromosomes durant la mitosi i la meiosi i estan envoltats de proteïnes.
- **Vacúol:** compartiments envoltats de membrana cel·lular presents en les cèl·lules eucariotes. Són considerats entitats diferents al citoplasma perquè tenen funcions diferents: secretora, excretora i d'emmagatzematge.
- **Aparell de Golgi:** orgànul trobat a la major part de les cèl·lules eucariotes, sovint prop del nucli i del centrosoma. La seva funció és modificar i classificar proteïnes

i lípids fabricats a altres parts de la cèl·lula i les empaqueta per dirigir-les a la seva destinació dins o fora de la cèl·lula.