

INTRODUCCIÓ

El treball de recerca que teniu a les mans forma part del conjunt de treballs presentats a l'Institut Guillem Catà durant el curs 2013-2014 pels alumnes de 2n de batxillerat.

La meua hipòtesi de treball és demostrar la teoria endosimbiòtica, coneguda també amb el nom d'endosimbiosi seriada: a grans trets, l'objectiu és verificar que la cèl·lula eucariota prové d'una simbiosi entre un arqueu (cèl·lula hoste), i un bacteri (l'organisme englobat) que va derivar als posteriors mitocondris de la cèl·lula eucariota actual.

Aquesta teoria té una certa importància: si bé tan sols és l'explicació més coherent per a l'origen de les cèl·lules eucariotes, aquestes han derivat a organismes més complexos. Per tant aporten un pas important a l'evolució, i a la vida tal com la coneixem avui en dia.

Per a estudiar la meua hipòtesi i arribar a les conclusions, seguiré una sèrie de processos bioinformàtics. Mitjançant aquests, compararé en diversos organismes representatius un gen comú a tots ells: l'aminoacil-ARNt-sintetasa.

"Les seqüències de l'endosimbiosi" queda inclòs dins una àrea concreta de les ciències: la biologia. Tanmateix, cal puntualitzar que treballaré amb les bases d'algunes branques més específiques d'aquesta ciència, com ara l'evolució, la biologia molecular, certs components de zoologia i taxonomia, la genètica, i un component prou fort de bioinformàtica. Aquest fet em permet aprendre sobre aspectes molt diferents de les ciències biològiques aplicades.

A banda de la part experimental del treball, hi ha un contingut de bases teòriques necessàries. Per a comprendre el perquè i les justificacions dels processos que es duen a terme, es necessiten uns coneixements mínims, que he intentat anar explicant quan era necessari al llarg del text.

Els principals motius que m'han portat a fer el treball han estat:

- L'accessibilitat als medis amb els que he realitzat el treball: la base de dades del GenBank és de lliure accés. De totes maneres, és un món força peculiar, i s'hi necessita una prèvia familiarització per a poder treballar-hi bé.
- La biologia, ja que és la temàtica principal del treball i una ciència que m'agrada, i la bioinformàtica. Aquesta última, tot i estar estrictament relacionada amb les ciències de la vida, queda lluny dels temes tractats a classe. A més, amb la realització del treball tinc la oportunitat d'endinsar-me en tots aquests camps que em resulten prou motivadors i interessants, i ampliar coneixements.
- L'interès per ampliar els meus coneixements sobre la teoria endosimbiòtica: és una de les més esteses per a entendre l'evolució de la vida tal com l'entenem. A partir d'aquí puc arribar a conèixer més a fons el seu raonament, i derivar-ne la hipòtesis.
- L'autonomia que em permet un treball d'aquest tipus. La informació és completament al meu abast: és a les meves mans la capacitat per a aprendre, comprendre, establir el treball i desenvolupar-ne la hipòtesis amb les eines de les que dispo.

BASE DEL TREBALL: LA TEORIA

ENDOSIMBIÒTICA

1. EUCARIOGÈNESI

Es coneix com a Eucariogènesi al procés que va donar lloc a les cèl·lules eucariotes. Aquestes, a la seva vegada, provenen de les cèl·lules procariotes, uns organismes més senzills i diferents dels eucariotes, amb unes clares diferències morfològiques. A grans trets, per obtenir una idea general, les principals diferències que trobem entre aquests dos tipus de cèl·lules són:

- El tipus d'organismes que formen, ja que la gran majoria de procariotes són bacteris, concretament eubacteris i arqueobacteris; mentre que les eucariotes formen els protistos, fongs, plantes i animals.
Per tant, l'organització de les cèl·lules procariotes tendeix a ser unicel·lular, i en les eucariotes pot ser unicel·lular, o pluricel·lular. A més, la majoria dels individus pluricel·lulars que aquestes formen existeixen diversos tipus de cèl·lules especialitzades.
- La grandària, ja que les cèl·lules eucariotes són més grans que les procariotes (les primeres tenen una mida que va dels 10 als 100 micròmetres; les procariotes, d'1 a 10 micròmetres).
- Pel que fa a l'ADN, en el cas dels procariotes és circular; en els eucariotes el conformen molècules lineals associades a histones¹ (un tipus de proteïna). Pel que respecta als mitocondris i els cloroplasts, com més endavant es veurà, el seu ADN és circular, fet que els fa més similars als procariotes, i per tant als bacteris.
- A més, tenen una divisió més simple, i una complexitat menor (els eucariotes disposen de més orgànuls cel·lulars (tal com es veu a les figures 1.1.1 i 1.1.2 de la pàgina 4), i la complexitat de l'estructura citoplasmàtica és també major: membranes internes, citoesquelet...)

¹ Concretament, les histones no són associades tan sols a l'ADN de les cèl·lules eucariotes, sinó també a una part dels procariotes (els arqueobacteris). Tot i així, els organismes procariotes més representatius són els bacteris.

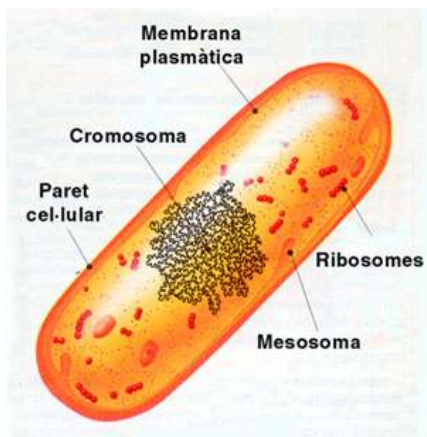


Fig. 1.1.1 Components de la cèl·lula procariota

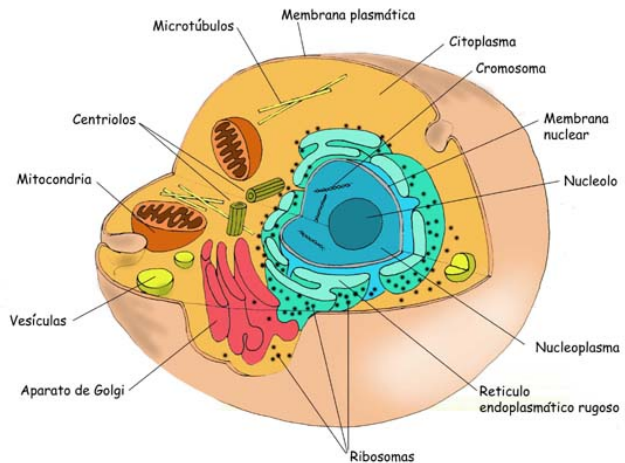


Fig. 1.1.2 Components de la cèl·lula eucariota (animal)

1.1 Tipus de teories

Les teories eucariogèniques es diferencien en diversos tipus: teories simbiogenètiques, autògenes, i autogeno-simbiogenètiques.

1.1.1 Teories simbiogenètiques

Són les que determinen que l'origen eucariota es troba en la simbiosi² entre 2 o més procariotes.

- **Principi de simbioticisme:** Paul Potier, el 1918, defineix per primera vegada l'origen bacterià simbiòtic dels mitocondris, a l'observar prèviament la simbiosi en termites i rizobis.
- **Teoria de l'endosimbiosi seriada de Lynn Margulis:** aquesta teoria, àmpliament acceptada, és tractada a fons a l'apartat "2. La teoria endosimbiótica" (pàgina 7). Postula que l'origen dels eucariotes es troba en successius processos simbiogenètics entre bacteris.

² Entenem com a simbiosi a "la vida en conjunció de dos organismes diferents, normalment en íntima associació, i en general amb efectes benèfics per a almenys un d'ells" tal com va exposar el botànic alemany Anton de Bary el 1873.

- **Teoria del crenarqueot:** dóna suport a la teoria de l'origen eucariota per simbiogènesi, pas que s'hauria donat entre un crenarqueot i el bacteri que va donar lloc als mitocondris.
- **Hipòtesi de l'origen quimèric:** a partir d'un estudi que comparava els tres dominis (arqueobacteris, eubacteris i eucariotes), s'insinua que el genoma nuclear eucariota podria haver tingut lloc a partir d'un origen quimèric per fusió genètica entre un arqueobacteri i un eubacteri
- **Hipòtesi de l'hidrogen:** es va proposar que els primers eucarionts es van originar per relació simbiogenètica entre un arqueu metanògen que utilitzava hidrogen i diòxid de carboni, i un bacteri anaeròbic facultatiu (simbiosi anaeròbica).
- **Hipòtesi sintròfica:** proposa que l'origen dels eucariotes es va donar per simbiosi metabòlica (sintròfia) entre un arqueu metanògen i un delta-proteobacteri (de tipus moxibacteri reductor de sulfat).
- **Hipòtesi de l'eucariogenèsis viral:** és possible que l'origen del nucli cel·lular eucariota estigui relacionat amb un procés de simbiogènesi entre un arqueu i un gran virus d'ADN com el mimivirus. La influència de l'ancestre viral va ser determinant per a la transformació de l'ADN circular típic procariota en un lineal eucariota, així com als canvis en l'ARNm.

1.1.2 Teories autògenes

Expliquen la formació de la cèl·lula eucariota com a un desenvolupament evolutiu on no hi ha simbiogènesi, o si hi és, no té una importància fonamental per a la seva constitució.

- **Hipòtesi de Meyer-Abich:** postula que els orgànuls i el nucli cel·lular haurien evolucionat a partir d'invaginacions de la membrana plasmàtica.
- **Hipòtesi dels tres dominis:** basant-se en l'anàlisi de seqüències moleculars de l'ARN ribosomal, postulen que els eucariotes, al igual que els eubacteris i els arqueobacteris, descendeixen de progenotes als albors de l'origen de la vida, tenint per tant una antiguitat de més de 3.000 milions d'anys. Tot i no descartar la evolució simbiogenètica, se li resta importància ja que era un fenomen, fins aquell moment, no demostrat.

1.1.3 Teories autogeno-simbiogenètiques

Postulen una simbiogènesi tardana amb un desenvolupament autogen precursor, ja sigui per formacions prèvies d'invaginacions o de la fagotrofia.

- **Teoria de la endosimbiosi seriada de Taylor:** com a resposta a les teories de Margulis, F. J. R. Taylor postula el 1974 que la versió de Margulis implica que són dos i no tres els processos simbiòtics que es van produir, originant-se així els mitocondris i cloroplasts. Aquests orgànuls no tenen una relació independent amb la cèl·lula eucariota, sinó una connexió indivisible i en conseqüència una fusió biològica, donant origen a un nou organisme.
- **Teoria de les invaginacions:** sorgeix com a pas previ a l'origen simbiogenètic, on s'argumenta que el desenvolupament eucariota es va produir mitjançant invaginacions. Les membranes internes de les cèl·lules eucariotes, sobretot la membrana nuclear i el reticle endoplasmàtic, s'expliquen millor com a invaginacions.
- **Teoria del fagocit primitiu:** segons aquesta teoria, hi hauria hagut un model primitiu de fagocitosi pre-eucariogènesi. Aquesta fagocitosi inicial es relaciona amb un fagosoma primogeni i els lisosomes.
- **Hipòtesis del cronocit:** un cronocit hauria estat un organisme pre-eucariota amb citoesquelet, sistema endomembranós i sistema de senyalització intern, que va desenvolupar el nucli cel·lular per fagocitosi d'arqueus i eubacteris. Aquesta teoria dóna suport, d'aquesta manera, al sistema de tres dominis.
- **Teories de Cavalier-Smith:** Thomas Cavalier-Smith descriu la evolució eucariota amb un extensa exposició de cadascun dels processos que van tenir lloc, postulant que algunes fases autògenes pre-eucariotes van tenir lloc abans de la simbiogènesi.

2. LA TEORIA ENDOSIMBIÒTICA

2.1 Antecedents

El 1883, el biòleg alemany Andreas Schimper va proposar que la capacitat fotosintètica de les cèl·lules vegetals podia procedir de cianobacteris encara presents a la natura i amb iguals capacitats.

Les teories de Margulis van ser prefigurades ja al segle XIX pel naturalista rus Konstantin Merezhkovsky (1855-1921). Aquest científic oblidat, que va néixer abans de la publicació de l'Origen de les espècies (Darwin) i que tenia ja 27 anys quan va morir Darwin, va ser el primer autor que va proposar l'extravagant idea de la simbiogènesi. Segons aquesta, alguns òrgans, i fins i tot alguns organismes, no sorgien en l'evolució pel gradual mecanisme de la selecció natural, sinó mitjançant associacions simbiòtiques entre una espècie animal o vegetal i algun tipus de microbi. Merezhkovsky va arribar a postular que el nucli de la cèl·lula eucariota provenia d'un antic microorganisme, un bacteri de vida lliure, fet que possiblement és erroni, al menys dit així sense més matisos. En qualsevol cas, les seves idees no van tenir la menor repercussió, i els seus treballs van passar inadvertits.

Anys després, Ivan Wallin (anatomista estadunidenc), va arribar a la mateixa conclusió, publicant l'any 1927 el llibre *Simbiosi i l'origen de les espècies*. En aquell moment, les seves conclusions van ser enteses com a absurdes, costant-li el seu prestigi professional.

A França, el biòleg Paul Portier, el 1918, també va arribar a conclusions semblants sobre l'origen simbiòtic dels eucariotes. En aquest cas, Portier va patir els atacs del llavors influent microbiòleg August Lumière.

Aquests treballs, menyspreats en el seu temps, va romandre oblidats fins que Lynn Margulis, donant-se suport en ells i posant èmfasis en les capacitats de les bactèries i la potencialitat de la simbiosi, formulés el 1967 la Teoria Endosimbiòtica.

2.2 La teoria endosimbiòtica de Lynn Margulis

La teoria endosimbiòtica, o endosimbiosi seriada, és una teoria proposada per Lynn Margulis, biòloga nord-americana, l'any 1967, àmpliament acceptada.

Segons la teoria, alguns orgànuls de la cèl·lula eucariota es van originar a partir d'organismes procariotes, que van ser englobats per altres microorganismes i posteriorment haurien establert una relació de simbiosi amb aquests. Els procariotes englobats (endosimbionts, o simbiotes interns), després de diverses etapes van esdevenir orgànuls cel·lulars d'un nou tipus de cèl·lula més complexa: la cèl·lula eucariota.

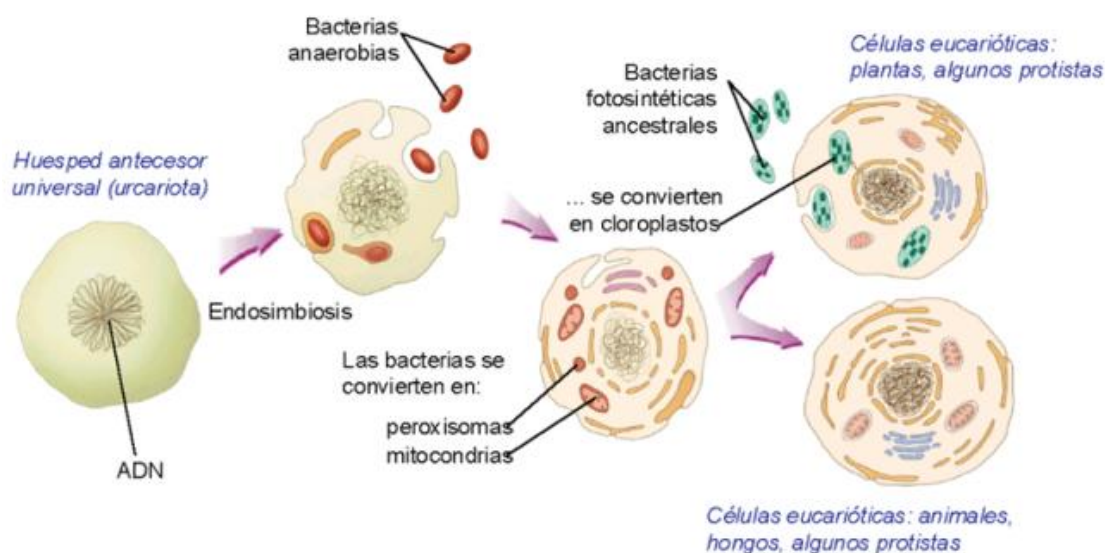


Fig. 1.2.1 Procés d'endosimbiosi de la teoria endosimbiòtica

El principal argument d'aquesta teoria és el fet de que els mitocondris i cloroplasts³ es divideixen independentment de la divisió cel·lular i tinguin ADN i ribosomes de tipus bacterià, implicats en la síntesi de proteïnes pròpies d'aquests orgànuls.

Actualment s'accepta que els mitocondris i cloroplasts dels eucariotes procedeixen de l'endosimbiosi, però l'idea (també proposada per Margulis) de que una espiroqueta⁴ endosimbiòtica es convertís en els flagels i cilis dels eucariotes, no ha rebut gaire acceptació: aquests orgànuls no mostren semblances ultraestruturals amb les espiroquetes i manquen d'ADN.

³ Explicats amb més detall als apartats "2.2.1 Els mitocondris" i "2.2.2 Els cloroplasts".

⁴ Fil de bacteries Gram-negatives que tenen cèl·lules allargades i enrotllades helicoidalment (en forma d'hèlix).

2.2.1 Els mitocondris

Els mitocondris són orgànuls presents en totes les cèl·lules eucariotes, amb una funció principal fonamental: són els encarregats de produir energia a partir de reaccions químiques diverses.

En altres paraules, realitzen la respiració cel·lular: obtenen energia a partir de l'oxidació de molècules orgàniques que la cèl·lula ha ingerit. L'oxidació comporta la degradació o catabolització de les molècules orgàniques (mitjançant processos com la hidròlisi⁵), per obtenir molècules més senzilles i energia.

Tenen un diàmetre mitjà de 0,5-1 μm , i encara que normalment són descrits com a cilindres allargats i rígids, poden ser de mida i forma molt variables. Per tant, s'han de considerar orgànuls molt polimòrfics. Estan formats per una doble membrana: la externa és llisa, mentre que la interna presenta estructures membranoses anomenades crestes, que són replècs en forma de doblecs o dits de guant. L'espai intern, anomenat matriu o estroma mitocondrial, conté dues o més molècules circulars d'ADN i ribosomes.

Aquest ADN mitocondrial no segueix la genètica mendeliana en tres aspectes importants: l'herència maternal, l'heteroplasmia (dins d'una mateixa cèl·lula, les mutacions patològiques són presents en algunes molècules d'ADN però no en

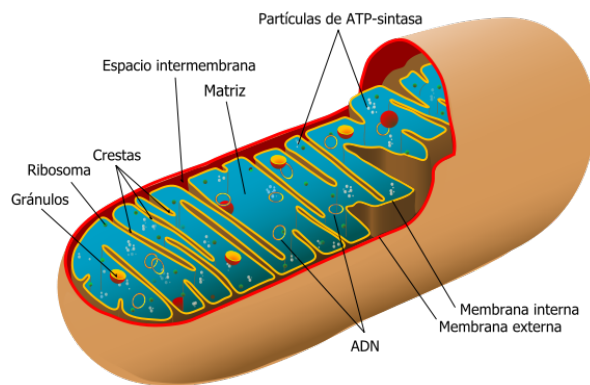


Fig. 1.3.2 Mitocondri

altres) i la segregació mitòtica (la distribució aleatòria dels orgànuls a la divisió cel·lular fa que variï la proporció d'ADN amb les mutacions que rep cada cèl·lula filla).

El nombre de mitocondris a una cèl·lula depèn del tipus cel·lular, així que en el cas d'una cèl·lula que necessita molta energia (com les

cèl·lules musculars), el nombre d'aquests orgànuls al seu interior serà elevat.

Pel que fa a la seva divisió, es realitza per bipartició (el mitocondri “mare” origina per divisió dos mitocondris “fills” mitjançant un allargament, i posteriorment una divisió per estrangulació), i independentment de la divisió de la cèl·lula.

⁵ La hidròlisi és un procés de trencament d'una molècula, amb concurs de l'aigua.

2.2.2 Els cloroplasts

Tot i que al treball ens centrarem en comparar l'ADN mitocondrial amb el del nucli de la cèl·lula en qüestió (ja que els mitocondris, com s'ha dit, els trobem en totes les cèl·lules eucariotes a diferència dels cloroplasts), val a dir que la teoria s'ajusta als mitocondris i cloroplasts com a orgànuls d'origen bacterià.

Els cloroplasts, uns orgànuls cel·lulars exclusius de les cèl·lules vegetals, són plasts (orgànuls que solen contenir pigments i intervenen en processos anabòlics⁶ i d'assimilació del carboni) que realitzen la fotosíntesi gràcies a la capacitat de captar energia lluminosa amb pigments fotosintètics.

El pigment més important que contenen aquests és la clorofil·la (un pigment verd), que com a conseqüència, dóna color als organismes que el contenen. També contenen altres pigments fotosintètics, però en proporció menor, com ara els carotenoides (de color vermell o ataronjat), les xantofil·les (groguenques), i la luteïna.

Cal dir que aquests orgànuls, com els mitocondris, també realitzen al seva divisió independentment de la cèl·lula.

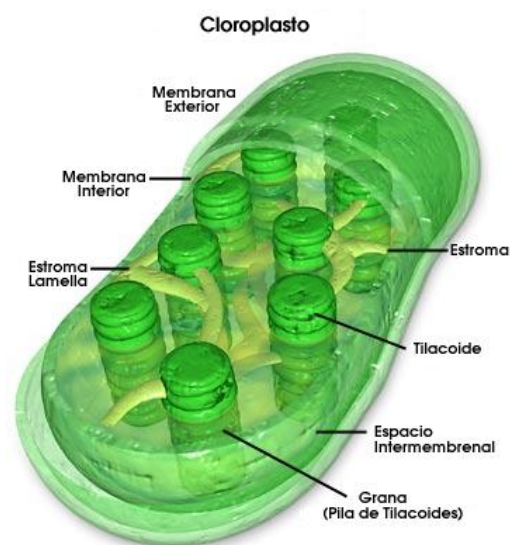


Fig. 1.2.3 Cloroplast

Un cop coneguts els orgànuls implicats en la teoria, passem a analitzar i conèixer-la de manera més fonamentada.

2.3 Les incorporacions simbiogenètiques: les hipòtesis originals de Lynn Margulis

En aquest apartat es mostren de manera més detallada les tres etapes descrites per Lynn Margulis a l'hora de fonamentar i postular la seva teoria.

⁶ Processos metabòlics que tenen com a objectiu sintetitzar biomolècules orgàniques a partir d'unitats més petites, com ara molècules o ions.

2.3.1 Primera incorporació simbiogenètica

“En primer lugar, un tipo de bacteria amante del dióxido de carbono y del calor, llamada arqueobacteria anaerobia (o termoacidófila), se fusionó con una bacteria nadadora. Juntos, los dos componentes integrados de la fusión se convirtieron en el nucleocitoplasma, la sustancia base de los ancestros de las células animales, vegetales y fúngicas. Este temprano protista nadador era, como sus descendientes actuales, un organismo anaerobio. Envenenado por el oxígeno, vivía en arenas y lodos donde abundaba la materia orgánica, en grietas de las rocas, en charcos y estanques donde este elemento estaba ausente o era escaso.”

Lynn Margulis, *Una revolución en la Evolución*, Cap.: *Individualidad por incorporación*.¹⁵

Segons el seu enunciat, un bacteri consumidor de sofre, que utilitzava el sofre i la calor com a font d'energia (arqueu fermentador o termoacidòfil), s'hauria fusionat amb un bacteri nadador (espiroqueta) havent passat a formar un nou organisme i sumant les seves característiques inicials de forma sinèrgica (en la que el resultat de la incorporació de dues o més unitats adquireix més valor que la suma dels seus components). El resultat seria el primer eucariont (unicel·lular eucariota) i ancestre únic de tots els pluricel·lulars. El nucleoplasma de les cèl·lules animals, plantes i fongs, seria el resultat de la unió d'aquestes dues bactèries.

A les característiques inicials d'ambdues cèl·lules se'ls sumaria una nova morfologia més complexa amb una nova i destacable resistència a l'intercanvi genètic horitzontal⁷. L'ARN quedaria confinat en un nucli intern separat de la resta de la cèl·lula per una membrana.

Sobre aquest primer pas, a dia d'avui, encara existeixen discrepàncies sobre com va succeir (segons la hipòtesis de Margulis), i no existeix consens científic per part de la comunitat científica. A finals dels anys vuitanta i principi dels noranta diversos treballs no admetien les homologies proposades entre els flagels dels eucariotes i de les espiroquetes. Margulis defineix que les associacions entre espiroquetes i protists donen suport a la seva teoria, i “la comparació de gens i genomes arqueobacterians amb seqüències d'eucarionts han demostrat la seva relació filogenètica en ambdós

⁷ Procés en que un organisme transmet informació genètica a una altre cèl·lula sense que sigui descendent.

grups". No obstant, des de la formulació de Margulis, han sorgit innumerables interrogants i noves teories sobre com va succeir aquesta etapa.

Margulis va admetre que aquest és el punt de la seva teoria amb més dificultats per a defensar i Antonio Lacano, el 2002, va preveure que per a comprendre l'origen d'aquest primer pas, s'accepti o no el seu origen simbiogenètic, "és indispensable seqüenciar no tan sols els genomes de la gama representativa de protists sinó també reconèixer la importància de l'estudi de la biologia en aquests organismes".

Ja als anys 20 va sorgir, com a alternativa a l'origen simbiogenètic d'aquest primer pas, la hipòtesis de que aquest s'hagués produït mitjançant invaginacions, proposta que no contradiu el paradigma neodarwinista i que, encara avui, es considera plausible per amplis sectors del món acadèmic.

Recurrentment s'han proposat diferents hipòtesis, també simbiogenètiques, en les que el propi nucli seria resultat de la incorporació d'un altre simbiot, com en el cas dels mitocondris i els cloroplasts. Fins i tot actualment s'ha proposat per a la formació del nucli una endosimbiosi d'una cèl·lula amb un virus (teoria de la eucariogènesi viral, vista a l'apartat "1.1.1 Teories simbiogenètiques" de la pàgina 4).

2.3.2 Segona incorporació simbiogenètica

"Después de que evolucionara la mitosis en los protistas nadadores, otro tipo de microorganismo de vida libre fue incorporado a la fusión: una bacteria que respiraba oxígeno. Surgieron células todavía más grandes, más complejas. El triplemente complejo respirador de oxígeno (amante del calor y del ácido, nadador y respirador de oxígeno) se volvió capaz de engullir alimento en forma de partículas. Estas células con núcleo, seres complejos y asombrosos que nadaban y respiraban oxígeno, aparecieron por primera vez sobre la Tierra quizá tan pronto como hace unos 2.000 millones de años. Esta segunda fusión, en la que el anaerobio nadador adquirió un respirador de oxígeno, condujo a células con tres componentes cada vez más preparadas para soportar los niveles de dióxido de carbono libre que se acumulaban en el aire. Juntos, el delicado nadador, la arqueobacteria tolerante al calor y al ácido y el respirador de oxígeno, formaban ahora un único y prolífico individuo que produjo nubes de prole."

Margulis, *Una revolución en la Evolución*, Cap.: Individualidad por incorporación.

Es postula que existia un organisme amb nucli que encara era anaeròbic, incapaç de metabolitzar l'oxigen, ja que aquest gas suposava un verí per a ell, raó per la qual viuria en medis on aquest oxigen, cada vegada més present, fos escàs. En aquest punt, una nova incorporació dotaria a aquest primogènit eucariont de la capacitat de metabolitzar oxigen. En aquest nou simbiot, originàriament bacteri respirador d'oxigen de vida lliure, es convertiria en els actuals mitocondris i peroxisomes presents en les cèl·lules eucariotes dels pluricel·lulars, possibilitant tot el seu èxit en un medi ric en oxigen com ha arribat a convertir-se el planeta Terra. Els animals i fons som el resultat d'aquesta incorporació.

2.3.3 Tercera incorporació simbiogenètica

“En la adquisición final de la serie generadora de células complejas, los respiradores de oxígeno engulleron, ingirieron, pero no pudieron digerir bacterias fotosintéticas de color verde brillante. La «incorporación» literal tuvo lugar tras una gran lucha en la que las bacterias verdes no digeridas sobrevivieron y la fusión completa prevaleció. Con el tiempo las bacterias verdes se convirtieron en cloroplastos (paso 4, figura 1.1). Como cuarto miembro, estos productivos amantes del sol se integraron con los demás socios anteriormente independientes. Esta fusión final dio lugar a las algas verdes nadadoras. Estas antiguas algas verdes nadadoras no sólo son los ancestros de las células vegetales actuales; todos sus componentes individuales todavía están vivos y en buena forma, nadando, fermentando y respirando oxígeno.”

Margulis, *Una revolución en la Evolución*, Cap.: *Individualidad por incorporación*.

Aquesta tercera incorporació va originar el Regne vegetal: les recentment adquirides cèl·lules respiradores d'oxigen fagocitarien bacteris fotosintètics i alguns d'ells, fent-se resistents, passarien a formar part de l'organisme, originant a la seva vegada un nou organisme capaç de sintetitzar l'energia procedent del Sol. Aquests nous pluricel·lulars, les plantes, amb el seu èxit, van contribuir i contribueixen a l'èxit d'animals i protists.

A l'actualitat romanen en els bacteris descendents d'aquells que haurien de, per incorporació, haver originat les cèl·lules eucariotes; així com aquells protists que no van participar en alguna de les successives incorporacions.

2.4 Arguments a favor i en contra de la teoria endosimbiòtica

A continuació s'analitzen els arguments principals que han fet d'aquesta teoria una de les més acceptades.

Entre els arguments a favor, trobem:

- La mida dels mitocondris és similar a la d'alguns bacteris.
- Els mitocondris i els cloroplasts contenen ADN bicatenari circular tancat covalentment –com els procariotes- mentre que el nucli eucariota compta amb varis cromosomes bicatenaris lineals.
- Estan rodejats per una doble membrana, fet que concorda amb la idea de la fagocitosi: la membrana interna seria la membrana plasmàtica originària del bacteri, mentre que la membrana externa correspondria a aquella porció que l'hauria englobat en una vesícula.
- Els mitocondris i els cloroplasts es divideixen per fissió binària, igual que els procariotes (els eucariotes ho fan per mitosis, un procés més complex i elaborat). En algunes algues, tals com l'Euglena, els plastids poden ser destruïts per certs productes químics o l'absència prolongada de llum sense que la resta de la cèl·lula es vegi afectada. En aquests casos, els plastids no es regeneren.
- Als mitocondris i cloroplasts, els centres d'obtenció d'energia es troben situats a les membranes, igual que en el cas dels bacteris. D'altra banda, els tilacoides que trobem als cloroplasts són similars a uns sistemes elaborats d'endomembranes presents en cianobacteris.
- En general, la síntesi proteica (formació de proteïnes) en mitocondris i cloroplasts és autònoma, (ja que no depenen de cap organisme exterior o orgànel·la diferent per a realitzar-la).
- Determinades proteïnes codificades al nucli es transporten als orgànuls, i els mitocondris i cloroplasts tenen genomes petits en comparació amb el de les bacteris. Això és consistent amb la idea d'una dependència creixent cap a l'amfitrió eucariòtic després de l'endosimbiosi. La majoria dels gens als genomes dels orgànuls s'han perdut o s'han mogut al nucli. És per això que amb el transcurs dels anys, l'amfitrió i l'hoste no podrien viure per separat.

- En mitocondris i cloroplasts trobem ribosomes⁸ 70s, característics dels procariotes, mentre que a la resta de la cèl·lula eucariota, els ribosomes que hi trobem són del tipus 80s⁹.
- L'anàlisi de l'ARNr¹⁰ 16s de la subunitat petita¹¹ del ribosoma de mitocondris i plastids revela escasses diferències evolutives amb alguns procariotes.
- Una possible endosimbiosi secundària (és a dir, implicant plastids eucariotes) ha estat observada per Okamoto e Inouye (2005). El protist heteròtrof *Hatena* es comporta com a un depredador i ingereix algues verdes, que perden els seus flagels i citoesquelet, mentre que el protist, ara un amfitrió, adquireix nutrició fotosintètica, fototàxia¹² i perd el seu aparell d'alimentació.

Segons l'enunciat original de Lynn Margulis:

“Las bacterias, fusionadas en simbiosis, nos dejan pistas de su anterior independencia. Tanto las mitocondrias como los plastos son bacterianos en su tamaño y forma. Todavía más importante es que estos orgánulos se reproducen de manera que hay muchos presentes a la vez en el citoplasma pero nunca dentro del núcleo. Ambos tipos de orgánulos, los plastos y las mitocondrias, no sólo proliferan dentro de las células sino que se reproducen de forma distinta y en momentos distintos a los del resto de la célula en la que residen. Ambos tipos, probablemente 1.000 millones de años después de su fusión inicial, retienen sus propias reservas reducidas de ADN. Los genes del ácido desoxirribonucleico (ADN) de los ribosomas de las mitocondrias todavía recuerdan sorprendentemente a los de las bacterias respiradoras de oxígeno que viven actualmente por su cuenta. Los genes ribosómicos de los plastos son muy parecidos a los de las cianobacterias. A principios de los setenta, cuando se compararon por primera vez las secuencias de nucleótidos del ADN de los plastos de las células algales con las secuencias de las cianobacterias de vida libre, ¿se descubrió que el ADN del cloroplasto era mucho

⁸ Els ribosomes són estructures complexes que es troben al citoplasma. La seva funció principal i és sintetitzar proteïnes

⁹ Un svedberg (símbol S, o Sv), és una unitat no inclosa al SI que s'utilitza en ultracentrifugació. És una unitat per a mesurar el coeficient de sedimentació d'una partícula o macromolècula quan són centrifugats en condicions normals. És una magnitud amb dimensions de temps, de manera que un svedberg equival a 10⁻¹³ segons.

¹⁰ ARN Ribosòmic (Àcid ribonucleic), un tipus d'àcid ribonucleic (àcid nucleic format per una cadena de ribonucleòtids). A l'apartat “2.1 Àcids nucleics” (pàgina 19) s'explica amb més profunditat: és un dels conceptes bàsic a l'hora de comprendre el treball.

¹¹ Cada ribosoma està format per dues subunitats, que canvien de forma i composició durant la síntesi proteica. Cada subunitat té una velocitat de sedimentació característica, tal com s'ha comentat, i cadascuna d'elles conté proteïnes i ARN.

¹² Habilitat que posseeixen moltes cèl·lules per dirigir-se cap a la llum.

más parecido al ADN de la cianobacteria que al ADN del núcleo de la propia célula algal!”

Margulis, *Planeta simbiótico*.

D'altra banda, també trobem altres arguments en contra de la teoria:

- Els mitocondris i els cloroplasts contenen introns¹³, una característica exclusiva de l'ADN eucariòtic. Per tant ha d'haver passat algun tipus de transferència entre l'AND nuclear i l'ADN mitocondrial/cloroplàstic.
- Ni els mitocondris ni els plastids poden sobreviure fora de la cèl·lula. De tota manera, aquest fet es pot justificar pel gran nombre d'anys que han passat: els gens i els sistemes que ja no eren necessaris van ser suprimits; part de l'ADN dels orgànuls va ser transferit al genoma de l'amfitrió, permetent a més que la cèl·lula hostatgera regulés l'activitat mitocondrial.
- La cèl·lula tampoc pot sobreviure sense els seus orgànuls: això és degut a que al llarg de l'evolució gràcies a la major energia i carboni orgànic disponible, les cèl·lules han desenvolupat metabolismes que no podrien sustentar-se solament amb les formes anteriors de síntesi i assimilació.

2.5 Fonaments del treball (aplicació de la teoria endosimbiòtica)

A l'hora de realitzar el treball i comprovar aquesta teoria, l'adaptaré a exemples més concrets per a poder fer comparacions adients. A la segona part del treball, es tracten els mètodes, el perquè dels gens escollits i els diferents organismes seleccionats per a les similituds.

Per a comprendre millor el procés, posem en situació les situacions prèvies i posteriors a l'eucariogènesi.

- A partir de l'estat evolutiu anterior a l'eucariogènesi: partirem de la cèl·lula que va fer d'hoste, i la que s'hi va incorporar. En el cas de l'hoste, escollirem

¹³ Els introns són zones no codificants dels gens, és a dir, que tot i ser-hi, no codifiquen proteïnes. Durant el procés de transcripció de l'ARN té lloc un procés anomenat *splicing*, a través del qual es tallen els introns, i s'uneixen els exons (zones codificants).

com a regne els arqueobacteris, que tal com enuncia Margulis al primer postulat de la teoria endosimbiòtica, era una de les cèl·lules originals que va formar el nucleocitoplasma de la cèl·lula eucariota (en altres paraules, la cèl·lula “englobadora”); i en el cas de la cèl·lula incorporada, prendrem com a mostra alguns tipus de bacteris (o també coneguts com a eubacteris¹⁴).

- Partint de l'estat posterior: una cèl·lula eucariota, de la qual seleccionarem uns gens determinats del seu ADN nuclear, i de l'ADN mitocondrial.

Si la resultats corresponen amb la teoria, esperem obtenir una similitud entre els gens seleccionats de:

- L'ADN d'un arqueobacteri i l'ADN nuclear d'una cèl·lula eucariota concreta, ja que a partir d'aquest primer es va formar el posterior nucleocitoplasma de la cèl·lula resultant.
- L'ADN d'un bacteri i l'ADN mitocondrial (ja que tal com s'ha dit, no només trobem ADN al nucli de la cèl·lula, sinó també als mitocondris i cloroplasts) d'una cèl·lula eucariota concreta, ja que segons la teoria, els bacteris van derivar als posteriors orgànuls cel·lulars esmentats.

D'aquesta manera, veurem les relacions evolutives per a comprovar la nostra hipòtesi.

¹⁴ Per a una major comprensió dels termes bacteri, arqueobacteri i eubacteri, vegeu l'apartat “3.1 Els Dominis”.

BIOINFORMÀTICA

1. EXPERIMENTACIÓ

Partint dels fonaments establerts a l'apartat "2.5 Fonaments del treball (aplicació de la teoria endosimbiòtica)" (pàgina 16), seguiré uns passos per a realitzar les comparacions.

- Escolliré uns organismes representatius de cada grup: "3. Selecció d'organismes a comparar" (pàgina 23).
- Seleccionaré un gen comú i estudiat que em permeti treballar: "4. Selecció del gen a comparar: l'aminoacil-ARNt sintetasa" (pàgina 36).
- Realitzaré la recerca dels gens triats al GenBank, i n'obtindrè les seqüències: "5. El GenBank" (pàgina 44).
- Compararé les seqüències obtingudes, i en veuré els tipus de resultats: "6. Comparació de seqüències: el Clustal" (pàgina 53).
- Analitzaré els resultats a partir dels arbres filogenètics: "7. TreeView" (pàgina 60).
- Arribaré a les conclusions finals del treball: "Conclusions" (pàgina 69).

A cada apartat s'aporta una base científica suficient com per a poder comprendre el perquè i el com dels passos realitzats, a més de proporcionar una base de coneixements suficients com per a justificar el que es fa en cada moment.

2. CONEIXEMENTS PREVIS: CONCEPTES BÀSICS DE GENÈTICA

Abans de començar, però, cal introduir una sèrie de conceptes per a comprendre els mitjans en el que es basa el treball.

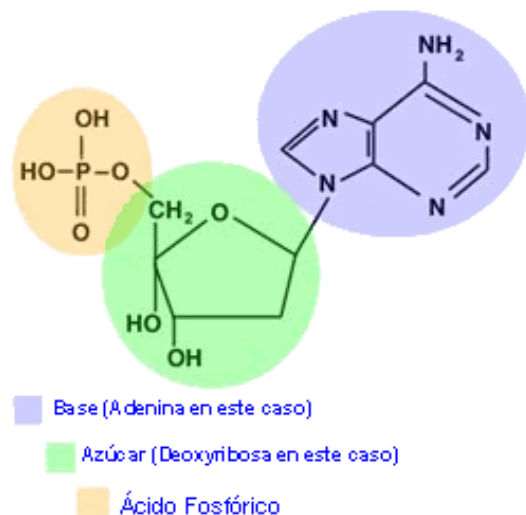
La cerca i selecció de gens ha estat possible gràcies a la prèvia seqüenciació del genoma de determinats organismes: als apartats següents tenen uns fonaments importants en aquests temes, i sense tenir una idea general de que és el material genètic, és força abstracte comprendre la importància d'aquest al treball: el perquè de les seqüències escollides, què són i on es troben, que entenem amb quan ens referim a "seqüenciar el genoma" d'un organisme determinat, que en podem obtenir d'aquest, com ho podem relacionar amb les vies evolutives, etc.

Aquest punt tractarà per sobre alguns d'aquests conceptes bàsics per a comprendre millor alguns d'aquests temes, fonamentals durant el transcurs del treball.

2.1 Àcids nucleics

Els àcids nucleics són biomolècules orgàniques formades per C, H, O, N i P. Són polímers que, per hidròlisi, es poden separar en nucleòtids, formats per:

- Àcid fosfòric
- Una pentosa (sucre amb cinc àtoms de carboni).
- Una base nitrogenada, que caracteritza els diferents nucleòtids que formen l'àcid nucleic.



Hi ha dos tipus d'àcids nucleics, l'ADN i l'ARN.

- L'ADN (àcid desoxiribonucleic) es diferencia de l'ARN pel seu sucre, la desoxiribosa, i les seves bases nitrogenades (Adenina, Timina, Citosina i Guanina).

Fig. 2.2.1 Nucleòtid

Pel que fa a la seva localització i funció, l'ADN es troba als cromosomes, mitocondris i cloroplasts, i "dicta" les ordres perquè la cèl·lula elabori les proteïnes. Les cadenes d'ADN poden arribar a ser molt llargues.

- En el cas de l'ARN (àcid ribonucleic), el sucre que el compona és la ribosa, i les seves bases nitrogenades difereixen de les de l'ADN: l'Uracil substitueix a la Timina pròpia de l'ADN (per tant es forma amb Adenina, Uracil, Citosina i Guanina).

A diferència de l'ADN, l'ARN es troba localitzat tant al nucli com al citoplasma, especialment als ribosomes, i "rep" les ordres, provinents de l'ADN, per a executar-les.

A més, les cadenes moleculars d'ARN són molt més curtes que les de l'ADN, ja que s'obtenen a partir de la informació continguda en alguns gens de l'ADN mitjançant la transcripció: únicament es passa a ARN el fragment d'ADN que es vol utilitzar per a formar unes proteïnes determinades.

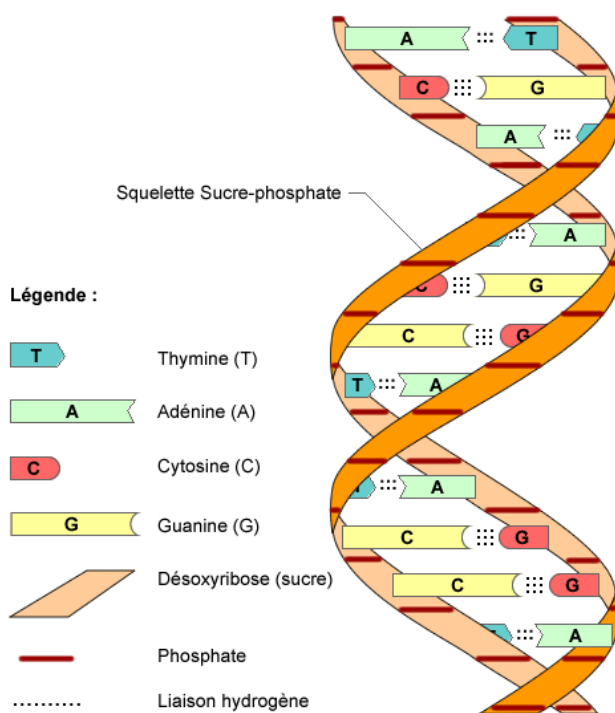


Fig. 2.2.2 Estructura de l'ADN

Aquestes bases nitrogenades caracteritzen els diferents nucleòtids, i són sovint el llenguatge amb el qual "s'escriuen els gens", mitjançant les seves lletres (A, G, C, i T en el cas de l'ADN, i A, G, C i U en el cas de l'ARN).

Sovint s'utilitza la paraula *base* (com a abreviació de base nitrogenada) per a referir-se als elements que formen el genoma. Tot i que aquest costum està molt estès, no és recomanable ja que pot induir a error: pot fer pensar que l'ADN és una cadena de

bases nitrogenades exclusivament, quan en realitat es forma per nucleòtids.

2.2 Els gens

Els gens són fragments d'àcid nucleic (ADN, generalment) que contenen les diferents "instruccions" necessàries per fer possible la vida¹⁵.

Són seqüències de nucleòtids que formen part d'una molècula d'ADN. Aquests contenen dos tipus de regions:

- Codificants: codifiquen una molècula d'ARN, i en la majoria també polipèptids a partir de l'ARN missatger que s'ha codificat prèviament.
- No codificants: estan formades per fragments d'ADN situats abans o després de les regions codificants, i contenen informació per a controlar les regions que s'activaran, determinant també quan ho faran i quant temps es mantindran activades, i les que no (fins a un 98% del total de l'ADN humà). També és anomenada com a matèria fosca del genoma, o ADN escombraries. Actualment se sap que un 80% d'aquest ADN està implicat en la regulació de l'expressió dels gens.

Els gens codifiquen, a la seva vegada, tres tipus d'ARN principals:

- L'ARN missatger (ARNm): codifica els polipèptids
- L'ARN de transferència
- L'ARN ribosòmic (ARNr): intervé, junt a l'ARNm, a la síntesi de polipèptids, però sense codificar-los.

Per tant, la majoria de gens codifiquen l'ARNm, i a través d'ell, les proteïnes.

2.3 El genoma

El genoma és el conjunt de la informació hereditària (o material genètic) d'un individu o d'una espècie. Aquesta informació és codificada per l'ADN, o en la majoria de virus, per ARN.

Cal remarcar que el genoma no és el conjunt de gens, ja que inclou tot el l'ADN (o ARN), i per tant totes les seves zones: tant codificants com no codificants.

¹⁵ Aquesta definició no serveix per a la majoria de virus: els seus gens estan fets d'ARN i no d'ADN.

En el cas del procariotes, que no tenen nucli, el seu genoma està compost normalment per una sola macromolècula d'ADN associada a proteïnes (tot i que en trobem amb dos), anomenada genòfor. La majoria de genòfors dels procariotes són circulars (ja que el seu ADN és circular, sense extrems), tot i que també existeixen genòfors lineals, i contenen la major part del seu genoma. La part petita que manca es troba continguda als plasmidis, unes molècules d'ADN que no estan lligades a proteïnes, i que es repliquen i transcriuen independentment de l'ADN dels genòfors (o dels cromosomes en els eucariotes).

En els eucariotes trobem un genoma al nucli cel·lular (genoma nuclear), que forma part dels cromosomes, formats cadascun per una cadena linear d'ADN associada a histones i altres proteïnes.

Però també trobem genomes ¹⁶ no nuclears: el dels mitocondris (genoma mitocondrial), el dels cloroplasts i el dels plasmidis.

¹⁶ Sovint el terme "genoma" s'utilitza com a sinònim de genoma nuclear.

3. SELECCIÓ D'ORGANISMES A COMPARAR

3.1 Dominis

Tal com s'ha explicat a l'apartat "2.5 Fonaments del treball (aplicació de la teoria endosimbiòtica)" (pàgina 16), els organismes amb els que treballarem són molt diversos, amb els quals trobem exemples de cadascun dels tres dominis de la vida: arqueobacteris, eubacteris, i eucaris. A continuació es fa una breu explicació de cadascun dels dominis per a veure'n les principals diferències i obtenir-ne una idea general.

Abans però, a la imatge següent es mostra un arbre filogenètic de la vida, amb les tres principals branques de les que parteixen tots els organismes. D'aquesta manera, s'obté una imatge general dels tres dominis que ajuda a comprendre l'estructura de la vida abans de començar a profunditzar en la matèria.

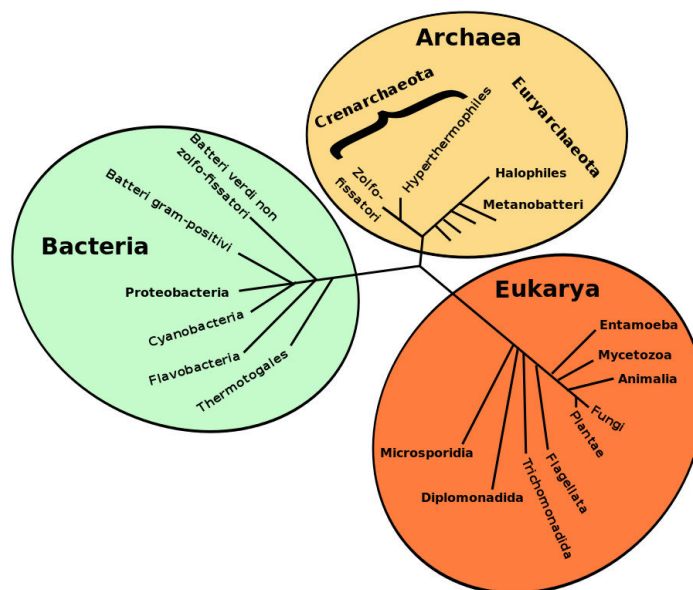


Fig. 2.3.1 Arbre de la vida: els tres dominis en són la divisió inicial

3.1.1 Els arqueobacteris

Els arqueobacteris o arqueus són un grup de microorganismes unicel·lulars que pertanyen al domini Archaea. Dins d'aquest, es troben dividits en cinc fílums, dels quals dos, Crenarchaeota i Euryarchaeota, són els estudiats més intensament.

Els arqueobacteris, com els bacteris, són procariotes que manquen de nucli cel·lular o qualsevol altre orgànel dins les cèl·lules.

Al passat, se'ls va considerar com a un grup inusual de bactèries, però actualment se'ls classifica com a un domini diferent en el sistema de tres dominis. Aquest fet és degut a la seva història evolutiva independent, i a les múltiples diferències en la seva bioquímica respecte a la resta de formes de vida.

A grans trets, els arqueobacteris i els eubacteris són bastant similars en forma i dimensions, tot i que alguns arqueus tenen formes molt inusuals (com les cèl·lules planes i quadrades d'*Haloquadrada walsbyi*). Tot i la semblança visual amb els eubacteris, els arqueobacteris estan dotats de gens i diverses rutes metabòliques que són més properes a les dels eucariotes, en especial els enzims implicats en la transcripció i la traducció. Exploten una varietat de recursos molt més gran que els eucariotes, des de compostos orgànics comuns com els sucres, fins l'ús de l'amoniac, ions de metall o fins i tot l'hidrogen com a nutrients.



Fig. 2.3.2 Arqueobacteri (*Halobacteri*)

coneix cap espècie d'arqueu que formi espores.

Els tolerants a la sal (les halobacteries) utilitzen la llum solar com a font d'energia, i altres espècies d'arqueus fixen carboni. Tot i així, a diferència de les plantes i els cianobacteris, no es coneix cap espècie d'aquest domini que sigui capaç de realitzar ambdues coses. Els arqueus es reproduïxen asexualment i es divideixen per fissió binària, fragmentació o gemmació; a diferència dels bacteris i els eucariotes, no es

Originalment, els arqueus es consideraven extremòfils¹⁷ que vivien en ambients severos, com aigües termals (en el cas dels termòfils extrems, que viuen a temperatures superiors a 80°C) o llacs salats (en el cas dels halòfils, que viuen en aigües amb una concentració de NaCl propera a la saturació).

Posteriorment se'ls ha observat en una gran varietat d'hàbitats, com els sòls, oceans o zones humides. Les Archaea són especialment nombroses als oceans, i les que es troben al plàncton podrien ser un dels grups d'organismes més abundants del

¹⁷ Que viu o necessita viure en condicions extremes que serien excessives per a les condicions òptimes de vida, creixement i reproducció de la majoria d'organismes.

planeta. Actualment es consideren una part important de la vida a la Terra i podrien jugar un paper important tant al cicle del carboni com al del nitrogen. No es coneixen exemples clars d'arqueus patògens o paràsits, però solen ser mutualistes o comensals. Són exemples els arqueus metanògens¹⁸ que viuen a l'intestí dels humans i remugants, on són presents en grans quantitats i contribueixen a digerir l'aliment.

Aquests organismes tenen també la seva importància en la tecnologia, ja que hi ha metanògens que són utilitzats per a produir biogàs, i als processos de depuració d'aigües. A més, els enzims d'arqueus extremòfils són capaços de resistir elevades temperatures i solvents orgànics, raó per la qual s'utilitzen en la biotecnologia.

El coneixement actual sobre la diversitat dels arqueus és fragmentari, i no es poden estimar amb cap tipus de precisió el nombre total d'espècies existents. Fins i tot les estimacions del nombre total de fílums d'arqueobacterians varien entre 18 i 23, dels quals tan sols 8 tenen representats que s'han cultivat i estudiat directament. Molts d'aquests grups hipotètics són coneguts únicament a partir d'una sola seqüència d'ARNr, fet que indica que la diversitat d'aquests organismes roman completament desconeguda.

El problema a l'hora d'estudiar i classificar microbis no cultivats també es dona en bacteris. Recentment, tot i que el projecte es troba amb les dificultats comentades anteriorment, el consorci públic GEBA (*Genomic Encyclopedia of Bacteria and Archaea*) està intentant completar i anotar la major quantitat de genomes d'aquests dos dominis. Amb aquesta informació, la finalitat del projecte és aconseguir dur a terme una classificació basada en el genoma.

Els arqueobacteris, generalment, tenen un únic cromosoma circular, igual que els bacteris.

3.1.2 Els bacteris (Eubacteris)

Tot i que terme "bacteri" es referia tradicionalment a tots els procariotes, la classificació científica canvià després del descobriment a la dècada del 1990 que la vida procariota es compon de dos grups d'organismes molt diferents que evolucionaren independentment a partir d'un avantpassat comú ancestral. Aquests

¹⁸ Tipus d'arqueobacteris que formen metà.

dominis evolutius, com s'ha comentat anteriorment, són els eubacteris i els arqueobacteris.

Els eubacteris són un gran grup de microorganismes unicel·lulars que clàssicament pertany al regne *moneres*. És el regne més primitiu de tots. El seu nom prové de les paraules gregues *pro* (abans) i *karyos* (nucli), és a dir, “abans del nucli”, referent a la falta de membrana nuclear, principal diferència amb els organismes eucariotes. Solen mesurar uns quants micròmetres de llargada i presenten una gran varietat de formes: segons aquesta els podem trobar classificats en cocs (esfèrics), bacils (bastonets rectes), espirils (espiral rígida), espiroquetes (espirals més flexibles, llargues i primes), o vibrions (en forma de coma).

Els eubacteris es troben presents pràcticament a tots els hàbitats de la Terra, on viuen al sòl, en aigües termals àcides, residus radioactius, l'aigua i les profunditats de l'escorça terrestre, a més de la matèria orgànica i els cossos vivents de plantes i animals. Conformen aproximadament cinc nonilions (5×10^{30}) d'eubacteris a la Terra, representant gran part de la biomassa del planeta.

Els eubacteris tenen un paper vital en el reciclatge de nutrients, i molts passos importants dels seus cicles depenen d'ells, com ara la fixació de nitrogen de l'atmosfera i la putrefacció. També són els responsables de moltes de les malalties infeccioses provocades per microorganismes.

Cal puntualitzar que la majoria d'eubacteris no han estat descrits, i aproximadament, només la meitat dels filums d'eubacteris inclouen espècies que es poden cultivar al laboratori.

3.1.3 Eucaris (Eucariotes)

Els eucariotes són el domini d'organismes cel·lulars amb nucli diferenciat, anomenats eucarionts. Aquests organismes consten d'una o més cèl·lules eucariotes, i trobem des d'organismes unicel·lulars fins a complexos pluricel·lulars, als quals les diferents cèl·lules s'han especialitzat per a diferents tasques i que, en general, no poden sobreviure de forma aïllada.

Dins del domini trobem els regnes dels animals, plantes, fongs i protozous. Tots ells presenten semblances a nivell molecular (estructura dels lípids, proteïnes i genoma), i comparteixen un origen comú¹⁹.

Aquests són comunament anomenats organismes superiors, a més d'altres organismes que no constitueixen teixits.

Les anàlisis seqüenciadores dels ribosomes 80S dels eucariotes, avalen la teoria que aquestes seqüències són evolutivament més properes als arqueobacteris que no pas als eubacteris, i que van compartir un avantpassat comú amb els primers durant més temps que amb els segons. Aquest fet és una de les bases, com s'ha vist, del treball: el nucli de la cèl·lula eucariota prové, segons la teoria endosimbiòtica, d'un arqueobacteri simbiònt.

Tanmateix altres grups de gens dels eucariotes estan més relacionats amb els eubacteris, i altres gens semblen ser una diversificació anterior als dos altres dominis. Això podria ser a causa de l'elevada transferència original que es va donar en el brou primitiu²⁰.

Degut a la gran varietat evolutiva dins d'aquest camp, s'ha procurat agafar diferents exemples: mamífers, llevats i plantes, tot i que no en tots els casos s'ha trobat el corresponent gen a l'ADN mitocondrial (veure apartat "3.2 Els organismes").

3.2 Els organismes

Per a obtenir uns resultats més significatius, el nombre d'organismes escollits per a treballar són múltiples: no em limitaré només a escollir un de cada tipus, ja que és possible que no compti amb la seqüència que conté el gen que vull trobar a tots ells.

Aquestes carències venen donades per la manca de recursos a la base de dades i investigació científica (no tots els mapes genòmics són complets, i si hi són no s'han identificat absolutament tots els gens que contenen), raons completament alienes a mi. D'aquesta manera, és possible que no trobi una determinada seqüència a l'ADN nuclear i mitocondrial corresponent per al gen de la mateixa cèl·lula.

¹⁹ La divisió en regnes es pot veure a la figura 2.3.1 de la pàgina 23.

²⁰ Medi aquós on abundava la matèria orgànica, format espontàniament sense la intervenció d'éssers vius (ja que la composició de l'atmosfera era molt diferent de l'actual), on es van formar les primeres cèl·lules fa probablement uns 3900 milions d'anys.

Per tant, treballaré amb més d'un gen²¹, i diversos organismes per a poder obtenir uns resultats múltiples i generals, suficients per a comprovar el que volem i obtenir un resultat adient. Els principals organismes amb els que es treballarà han estat escollits pel seu grau d'importància i informació que tenim a l'abast, procurant que siguin prou coneguts i estudiats per a optimitzar les cerques i resultats a obtenir.

Molts d'ells tenen importància en camps com la genètica i la investigació: alguns han estat els primers organismes del seu regne a tenir el genoma completament seqüenciat, degut al seu interès en investigació; d'altres són molt característics i representatius, sent ells un dels models principals amb els que s'estudien els seus parents més propers, donant una imatge general del regne.

Els organismes escollits han estat:

3.2.1 Arqueobacteris

3.2.1.1 *Desulfurococcus fermentans*

És el primer arqueu cel·lular conegut. Aquest crenarqueot hipertermofílic²² i estrictament anaeròbic produeix hidrogen de la fermentació de diversos hidrats de carboni i pèptids sense inhibició acumulant hidrogen. El seu genoma ha estat seqüenciat completament.

3.2.1.2 *Methanothermus fervidus*

És un tipus de soca²³ del gènere *Methanothermus*. Aquest gènere hipertermofílic és considerat endèmic²⁴ de les primaveres caloroses Islandeses. L'*M. fervidus* no va ser només el primer organisme caracteritzat per tenir una temperatura de creixement màxima (97°C) propera al punt de bullició de l'aigua, sinó també el primer arqueu del qual es va informar sobre la seva anàlisi funcional detallada de la seva proteïna histònica, i el primer al qual la funció del 2,3-ciclodifosfoglicerat a la termoadaptació va ser descrita.

²¹ De tipus semblant, com s'explica a l'apartat Aminoacil-ARNt-sintetases específiques amb les que es treballa.

²² Apliquem el terme hipertermofílic en el cas dels organismes per als quals la temperatura òptima de creixement es troba estrictament per sobre dels 80°C. Dins dels procariotes, trobem hipertermofílics tant al domini dels Bacteris com dels Arqueus, sent en aquest últim grup al que es troben en major nombre.

²³ Població genèticament uniforme d'organismes. En microbiologia i virologia es considera com a soca a un subtipus o variant genètica d'un virus o bacteri.

²⁴ En biologia s'utilitza el terme "endèmic" per a denominar a una espècie exclusiva d'una zona.

És la primera seqüència genòmica completa de la família de les Metanotermàcees (*Methanothermaceae*).

3.2.2 Eubacteris

3.2.2.1 *Escherichia coli*

També conegut com a colibacil o per la seva abreviació *E.Coli* (el nom és degut al seu descobridor, Theodore Escherich), és una enterobacteria que es troba generalment a l'intestí dels animals, on és un dels anaeròbics facultatius (encara que tan sols és un petit constituent de la totalitat de la microflora intestinal). De tota manera se'l pot trobar a molts altres llocs (és molt versàtil).

E. Coli és capaç de causar diverses malalties sobre el seu hoste, especialment quan adquireix característiques virulentes. Algunes soques d'*E. Coli* poden provocar infeccions al tracte urinari, meningitis neonatal, i diverses malalties intestinals, normalment a l'atacar la cèl·lula hoste i introduint toxines que poden pertorbar els processos cel·lulars.

A més, *Escherichia Coli* és el model de procariota més àmpliament estudiat, i una espècie important en els camps de la biotecnologia i la microbiologia, on ha servit com a organisme hoste per a la majoria de treball amb ADN recombinant. També és un dels organismes models principals utilitzats en l'estudi de la genètica bacteriana, la fisiologia i la bioquímica, ja que és emprada com a organisme model per als bacteris en general.

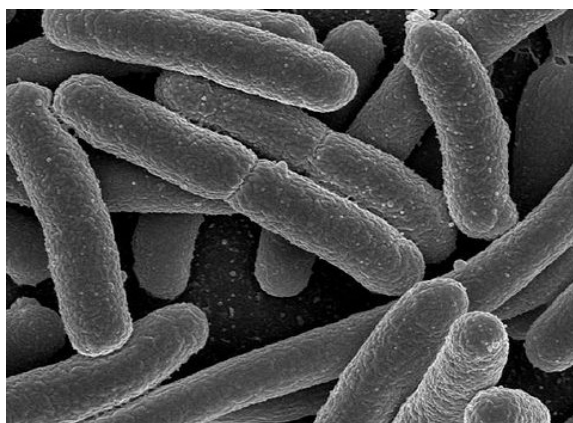


Fig. 2.3.3 Escherichia Coli al microscopi electrònic

3.2.2.2 *Methylobacterium extorquens* CM4

Methylobacterium extorquens és un bacteri capaç de créixer i alimentar-se a partir de compostos d'un sol carboni com el metanol. Mitjançant una oxidació, el transforma, i passa de metanol a formaldehid (també conegut amb el nom de

metanal), que després utilitza metabòlicament per a generar energia o biomassa. Aquests bacteris es troben normalment al medi ambient, i especialment associats a plantes que produeixen metanol a l'hora de metabolitzar pectina durant la síntesi de la paret cel·lular. Es requereixen com a mínim 25 gens per a realitzar el complex procés de transformar el metanol a formaldehid, i aquesta via metabòlica especialitzada és de gran interès.

Pel que fa al CM4 del nom, correspon a una soca concreta de *Methyobacterium Exorquens* utilitzada en anàlisis comparatius, ja que ha estat més estudiada, i per tant seqüenciada.

3.2.3 Eucariotes

Dins del domini dels eucariotes, hem seleccionat individus de cadascun dels seus regnes:

- Animals (*Animalia* o *Metazoa*): *Homo sapiens* i *Rattus norvegicus*.
- Plantes (*Plantae*): *Arabidopsis thaliana*
- Fongs (*Fungi*): *Candida albicans*, *Candida tropicalis* i *Saccharomyces cerevisiae*.

3.2.3.1 *Homo sapiens*

Els humans, coneguts com a *Homo sapiens* o *Homo sapiens sapiens* són els primats de la família *Homínids*, i la única espècie restant del gènere *Homo*. Els humans es distingeixen d'altres primats per la seva locomoció bípeda, i especialment pel seu cervell relativament més gran amb els seus particularment ben desenvolupats neocòrtex, prefrontal còrtex i glòbuls temporals, que permeten alts nivells de raonament abstracte, llenguatge, resolució de problemes, i cultura a través de l'aprenentatge social.

Com tots els mamífers, els humans són una espècie diploide eucariota. Cada cèl·lula somàtica té 23 parells de cromosomes, provenint cada conjunt d'un progenitor: els gàmetes tan sols tenen un exemplar de cada cromosoma, que és una barreja dels 23 parells parentals. D'aquests 23 parells, 22 són autosomes, i 1 està compost per heterocromosomes, o cromosomes sexuals (XX en el cas de les dones, i XY en el dels homes).

El primer mapa genòmic humà va ser seqüenciat completament l'any 2003, i actualment s'estan realitzant nombrosos esforços per aconseguir mostres de la biodiversitat genètica de les espècies (com al Projecte Internacional HapMap). Segons les estimacions actuals, els humans tenen aproximadament 22.000 gens. La variació de l'ADN humà és diminuta comparada amb altres espècies, fet que probablement suggereix un "cul d'ampolla" Durant el Pleistocè Superior (fa uns 100.000 anys), on la població humana va ser reduïda a un petit nombre de parells en cria.

Mitjançant una comparació de les parts del genoma que no es troben sota la selecció natural, i que per tant les seves mutacions són acumulades de manera constant, és possible reconstruir l'arbre genètic incorporant totes les espècies humanes des de l'últim ancestre comú.

Al comparar l'ADN mitocondrial, que és heretat únicament de la mare, els genetistes han arribat a la conclusió de que la última ancestre comuna (anomenada amb el nom d'Eva mitocondrial), el marcador genètic de la qual es troba a tots els humans moderns, va viure fa uns 200.000 anys.

Les forces de la selecció natural han continuat funcionant a les poblacions humanes, ja que determinades zones del genoma mostren una selecció direccional des de fa 15.000 anys.

3.2.3.2 *Rattus norvegicus*

La rata comuna (coneguda amb noms com a rata marró, gris, de claveguera, etc.) és una espècie de rosegador del subordre miomorf de la família *Muridae*. És una de les rates més conegudes i comunes; està lligada a les activitats humanes, i gràcies a això, ha colonitzat tot el món, esdevenint una veritable plaga. La característica de plaga no la té només pel fet de que acostuma a devorar els aliments de les cases i les bodegues, sinó especialment perquè transmet malalties greus, com Infeccions per hantavirus, leptospirosis, criptosporidiosis, la febre hemorràgica viral i la febre Q.

Malgrat tot, aquesta rata, que ha donat origen a les rates albes i altres varietats utilitzades com a animal de laboratori, és un organisme model important per a ajudar a comprendre la psicologia i malalties humanes com ara l'artritis, la hipertensió, la diabetis, i les malalties cardiovasculars.

3.2.3.3 *Arabidopsis thaliana*

Arabidopsis thaliana (coneguda també com a *Thale cress*) és una planta petita amb flors, de la família de les *Brassicàcia* (Brassicaceae), anteriorment anomenada *Crucíferes*.

S'ha utilitzat durant més de cinquanta anys per a l'estudi de les mutacions en plantes i per a l'anàlisi genètic, ja que va ser la primera planta a tenir tot el seu genoma seqüenciat, l'any 2000. Això va ser gràcies a que el seu genoma és relativament petit (uns 120 Mb), té una estructura simple (és diploide, i compta amb 10 cromosomes, un nombre inusualment petit²⁵) i conté poques seqüències repetides.

La seqüenciació va ser feta per un col·lectiu internacional anomenat "Arabidopsis genome Initiative"²⁶ (AGI). Tot i que no tenia importància econòmica, és un recurs invaluable per a cultius agricultors importants, particularment per a membres de la mateixa família, com la cànula, una font important d'oli vegetal.

Actualment s'utilitza com a organisme model per a estudiar diferents aspectes en la biologia de les plantes, incloent la genètica, el desenvolupament i el fototropisme. La gran riquesa dels diferents tipus de dades comparatives que s'han reunit sobre aquesta planta tant discreta la fan immensament significant en cada aspecte de la biologia de les plantes; a més, el seu curt temps de generació (la llavor germina en 6 setmanes) i la seva gran producció de llavors en faciliten l'estudi.

3.2.3.4 *Candida albicans*

Candida és un gènere de fongs unicel·lulars també anomenats llevats. L'espècie *Candida* més significativa per la seva importància clínica és *Canida albicans*. Les infeccions causades per fongs s'anomenen micosis. *Candida albicans* és un comensal de les mucoses humanes, sobretot de la mucosa oral, digestiva i genital. Les micosis causades per *Candida albicans* o per altres espècies de *Candida* s'anomenen Candidiasis, i les trobem en humans i en altres animals, especialment en pacients amb immunosupressió.

Candida albicans és un fong diploide sexual del gènere *Candida*, causant d'infeccions humanes, trobant entre les més comunes les orals o vaginals.

²⁵ Tenint en compte que el seu nombre haploide de cromosomes, per tant, és n=5.

²⁶ "Iniciativa del genoma de l'*Arabidopsis*".

Sota circumstàncies normals, *C. albicans* viu en el 80% de la població humana sense fer danys, malgrat que un sobrecreixement resulta en la candidiasi. La candidiasi sovint s'observa en individus immunocompromesos: pacients positius en VIH, tractats amb quimioteràpia per a un càncer, o que han passat per

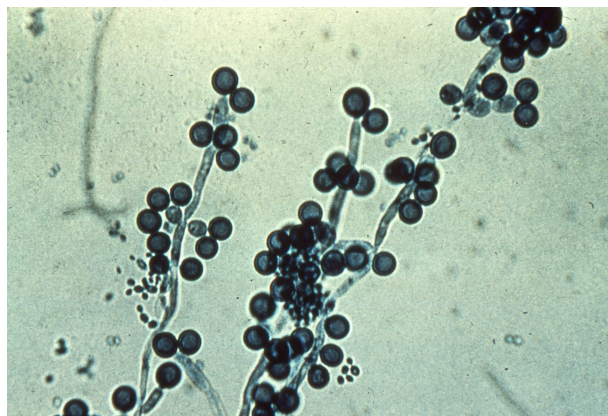


Fig. 2.3.4 *Candida Albicans*

transplantaments d'òrgans o medul·la òssia. En persones sense cap immunodeficiència, és una condició comuna que es guareix fàcilment: A més, les infeccions originades en hospitals en pacients que no havien estat considerats en risc (com els pacients a una unitat de cures intensives), han passat a ser una causa de preocupació sanitària major.

El genoma de *Candida albicans* té un cariotip²⁷ variable. La soca SC5314²⁸ conté 8 parells de cromosomes, 7 dels quals tenen una mida constant i un altre polimòrfic, amb una grandària que va dels 3.2 als 4.0 Mb. La presència de gens per al ARN ribosòmic empaquetats per parells causa aquesta variació en les seves dimensions. La mida del genoma d'aquesta soca és aproximadament d'uns 16 Mb.

Una de les característiques més interessants sobre el genoma de *Candida albicans* és el fet de que el nombre i l'estructura a la redistribució dels cromosomes provoca la diversitat genètica (polimorfisme de la llargada del cromosoma per contracció i expansió, les translocacions recíproques²⁹, les delecions cromosòmiques³⁰ i la trisomia de cromosomes individuals.

Aquestes alteracions del cariotip (que generen diversitat genètica), condueixen a canvis del fenotip, esdevenint una estratègia d'adaptació d'aquest fong. Una vegada els científics hagin realitzat l'anàlisi i estudi complet del genoma de *C. albicans*, es podran comprendre molt millor aquests canvis i mecanismes.

²⁷ El cariotip és el nombre i morfologia dels cromosomes condensats al nucli d'una cèl·lula eucariota. Aquest terme també és utilitzat per al conjunt de cromosomes d'una espècie, o d'un organisme individual.

²⁸ Actualment s'ha seqüenciat el genoma de *C. albicans* de la soca SC5314, i WO1.

²⁹ Mutació cromosòmica que es dona quan un fragment d'un cromosoma passa a formar part d'un altre cromosoma no homòleg.

³⁰ Pèrdua d'un fragment d'ADN d'un cromosoma.

3.2.3.5 *Candida tropicalis*

Candida tropicalis és una espècie de llevat del gènere *Candida*. És fàcilment reconegut per ser un llevat comú patògen, existent com a part de la flora humana normal. És un llevat fermentador de glucosa i maltosa, però no fermenta lactosa ni hidrolitza urea, i té resultats variables quan és analitzat per la fermentació de la sacarosa, galactosa, o la trehalosa.

El llevat *Candida tropicalis* és l'espècie *Candida* més patògena després de *Candida albicans*; s'associa més freqüentment a infeccions fúngiques profundes que a les de la mucosa normal. És un organisme asexual diploide.

També és un organisme important per a l'estudi de la biogènesi dels peroxisomes³¹ i de l'expressió proteica peroxisomal. La mida exacte del genoma i el nombre de cromosomes de la *Candida tropicalis* són desconeguts, però s'ha estimat que pot contenir un genoma haploide d'uns 15 Mb, organitzat en 5 o 6 cromosomes.

3.2.3.6 *Saccharomyces cerevisiae*

Saccharomyces cerevisiae és un dels models d'organisme eucariòtic més intensivament estudiats en biologia molecular i cel·lular, per a entendre els processos d'aquests components en eucariotes, fins i tot més que *Escherichia Coli* (en aquest cas com a model d'organisme procariota, tal com es veurà més endavant).

Aquest organisme ha estat molt útil a l'hora d'estudiar el cicle cel·lular ja que es tracta d'un llevat fàcil de cultivar, però, a causa del fet que pertany als eucariotes, comparteix el complex de l'estructura interna cel·lular de plantes i animals.

Prova de la utilitat d'aquest organisme en el món científic és que *S. cerevisiae* va ser el primer organisme eucariota el genoma del qual va ser completament seqüenciat, amb una grandària aproximada de 12 Mb, organitzat en 16 cromosomes. A la base de dades de genomes de llevats es troba altament representat i encara actualment és una eina molt important pel desenvolupament del coneixement bàsic sobre la funció i l'organització de la genètica i fisiologia de les cèl·lules eucariotes. Aquest organisme es troba darrera de la majoria de fermentacions comunes. Les cèl·lules

³¹ Els peroxisomes són uns orgànuls citoplasmàtics de les cèl·lules eucariotes que compleixen funcions de desintoxicació cel·lular, ja que ajuden a que la cèl·lula a eliminar els peròxids (una espècie química molt lesiva) mitjançant els enzims que contenen (catalases i oxidases). A més, participen en el metabolisme dels àcids grassos i d'altres metabòlits resultants d'una respiració aeròbia.

de *Saccharomyces cerevisiae* poden prendre formes tant circulars com ovals, d'uns 5-10 micròmetres de diàmetre. La reproducció d'aquest microorganisme és principalment asexual i té lloc mitjançant un procés de divisió anomenat gemmació.

És conegut també amb el nom de llevat de cervesa, o de forner: probablement es tracta del tipus de llevat més important de la història, ja que s'ha utilitzat des de fa molts anys per a fermentacions en begudes i aliments com la cervesa i el vi, o el pa. A més, també es molt important per a la producció d'enzims i fàrmacs.

4. SELECCIÓ DEL GEN A COMPARAR: L'AMINOACIL-ARNT SINTETASA

El gen seleccionat ha de ser un gen comú a tots els tipus d'organismes que volem comparar: entre ells trobem tant bacteris, procariotes, com organismes més complexos amb cèl·lules eucariotes. Si bé són diferents, hi ha certs processos que són comuns en ambdós tipus.

Un d'aquests processos és el mecanisme d'expressió gènica de les cèl·lules, que consisteix en transformar la informació genètica en proteïnes. Consta de tres parts: la transcripció, la traducció i la regulació gènica.

Tot i que l'ADN es troba en diferents formes segons l'organisme (molècules circulars en el cas de la majoria de procariotes, i lineal en els eucariotes, tal com s'ha explicat a l'apartat "1. Eucariogènesi" (pàgina 3), ambdós tipus realitzen aquest procés, tot i que trobem algunes diferències. Tot i així, aquestes diferències no ens suposen cap problema, ja que la part amb la que treballarem, és comuna a tots dos tipus de cèl·lula.

El gen escollit amb el que treballaré, conté la informació genètica que codifica la síntesi d'uns enzims que realitzen un paper important dins del procés de traducció³²: les aminoacil-ARNt-sintetases (*aminoacyl-tRNA synthetase* en anglès, o també coneguts com a AARSs).

Aquests enzims del citoplasma catalitzen la càrrega de l'ARNt, és a dir, el procés mitjançant el qual un aminoàcid s'uneix al seu corresponent ARNt (aminoacilació). Aquest procés es mostra amb més detall més endavant, a la figura 2.4.1 (pàgina 38).

Les aminoacil-ARNt-sintetases han fascinat durant molt temps als biòlegs. Són l'eix de la traducció, l'enllaç entre els mons de les proteïnes i els àcids nucleics. Les seves estructures i funcions, que tenen tant significat pràctic i bàsic, mereixen i han rebut molta atenció. De totes maneres, no és tan sols l'aspecte de la funció estructural d'aquests enzims el que ha capturat la imaginació dels biòlegs; també ho

³² S'entén per traducció el conjunt de processos que permeten la síntesi d'una proteïna a partir de la informació continguda en un ARNm (ARN missatger).

són les possibilitats de que podrien arribar a mostrar-nos els secrets del codi genètic³³.

Hi ha almenys 20 aminoacilsintetases diferents³⁴, de manera que cadascuna d'elles reconeix un dels 20 aminoàcids i l'ARNt que hi ha de ser carregat. Els ARNt poden separar-se en vint grups diferents i tots els d'un mateix grup es carreguen amb el mateix aminoàcid (són isoacceptadors, d'iso, que vol dir 'igual'). Encara que hi ha 61 triplets diferents de bases, no hi ha tants ARNt, ja que un mateix ARNt pot ser carregat amb diferents aminoàcids, d'acord amb l'efecte de balanceig³⁵, en la tercera base de l'anticodó.

Al principi de l'apartat s'ha comentat que aquest procés és comú tant a les cèl·lules procariotes com a les eucariotes. Però cal puntualitzar que l'altre tipus d'ADN amb el que treballem, l'ADN mitocondrial (ADNmt), té diferències a l'hora de traduir-se respecte l'ADN nuclear.

La síntesi de les proteïnes mitocondrials té lloc a ribosomes específics del mitocondri, els components dels quals estan codificats a l'ADNmt (ARNr 12s i 16S), i al genoma nuclear (84 proteïnes ribosòmiques). En aquest sistema de traducció es sintetitzen les tretze proteïnes codificades a l'ADNmt utilitzant un codi genètic que difereix lleugerament del codi genètic universal. Així, UGA codifica l'aminoàcid triptòfan (Trp) en comptes de ser un codó de terminació, i els codons AUA i AUU s'utilitzen també com a codons d'iniciació.

La biogènesi dels mitocondris depèn de l'expressió coordinada dels genomes mitocondrial i nuclear, però fins ara se'n coneix ben poc pel que fa als mecanismes que regulen la interacció d'ambdós sistemes genètics. L'expressió de l'ADNmt sembla estar regulada pel factor d'iniciació de la transcripció mtTFA, codificat pel genoma nuclear. Aquest factor podria ser el responsable tant dels nivells d'ARN com del nombre de còpies d'ADNmt, ja que la replicació depèn de la síntesi d'un iniciador d'ARN a partir del promotor³⁶ de la cadena lleugera (la que acaba de ser sintetitzada). La regulació de la relació entre els ARNr i els ARNm mitocondrials es realitza fonamentalment mitjançant la selecció del lloc d'iniciació de la transcripció de la cadena pesada (la cadena parental), que a la seva vegada està relacionada amb

³³ Vegeu apartat "4.3 Importància evolutiva" (pàgina 42).

³⁴ Els tipus d'aminoacil-ARNt es tracten més àmpliament a l'apartat "4.2 Tipus" (pàgina 40).

³⁵ El mateix ARNt es pot aparellar amb codons que només difereixen en l'últim nucleòtid, ja que l'aparellament amb l'última base és menys rígid. Aquest efecte pot arribar a derivar en la degeneració del codi genètic (existeixen triplets sinònims).

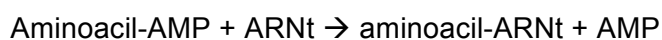
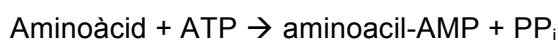
³⁶ Els promotors són seqüències d'ADN que dirigeixen la transcripció d'un gen adjacent a aquest a ARN.

el factor mtTERF (que causa terminació de la transcripció després de la síntesi dels ARNr) i amb el processament dels ARN primaris. Així mateix, l'activitat transcripcional pot ser regulada per estímuls hormonals, especialment per hormones tiroidees que actuen tant d'una manera indirecta (per activació de gens nuclears) com directament sobre el propi ADNmt.

4.1 Funció

La principal funció de l'aminoacil-ARNt-sintetasa és catalitzar la càrrega de l'ARNt, o en altres paraules, unir els aminoàcids corresponents als ARNt.

Per a fer-ho, el grup carboxil de l'aminoàcid forma un enllaç èster amb el grup hidroxil del carboni 2' o 3' de la ribosa del nucleòtid terminal (en l'extrem 3' dels ARNt, en què sempre hi ha la seqüència de bases CCA). L'energia necessària per a l'enllaç la proporciona l'ATP, i la reacció és la següent:



Reacció total:

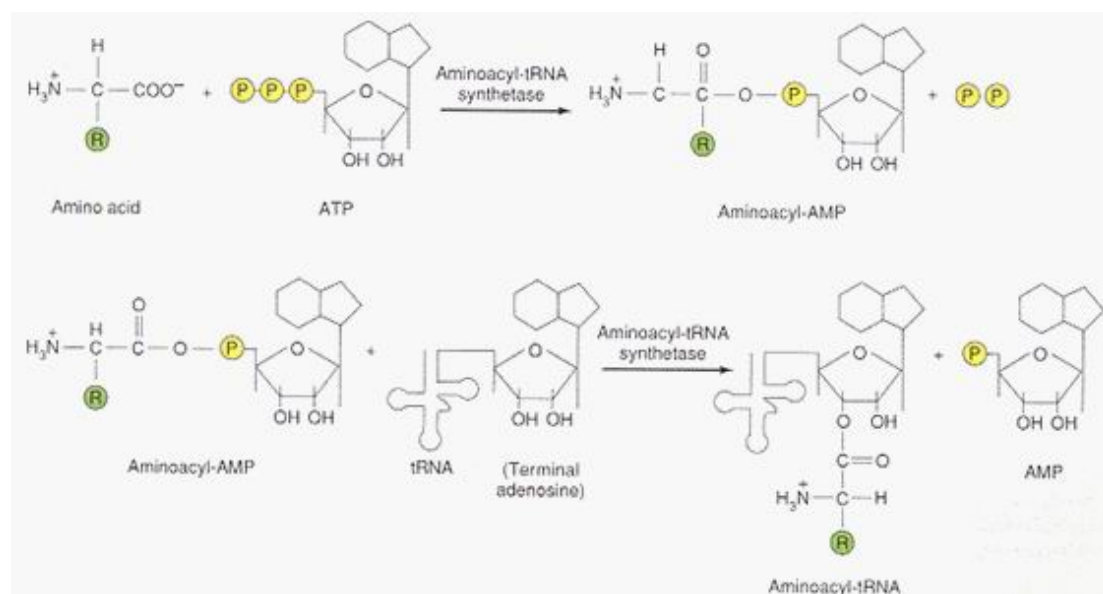
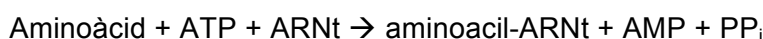


Fig. 2.4.1 Passos de l'aminoacilació

La sintetasa primer s'uneix a l'ATP i al corresponent aminoàcid o al seu precursor per a formar un aminoacil-adenilat, alliberant una molècula de pirofosfat orgànic (PP_i). El complex adenilat-aaRS s'uneix després a la molècula apropiada de l'ARNt, i l'aminoàcid es transfereix des del complex aminoacil-AMP ja sigui a l'extrem –OH 2' o 3' de l'últim nucleòtid (A/&) a l'extrem 3' de l'ARNt. Algunes sintetases també poden realitzar una reacció de correcció de lectura per a garantir una alta fidelitat a la càrrega de l'ARNt, si la sintetasa troba que l'ARNt no es troba ben carregat, hidrolitza la unió aminoacil-ARNt.

La càrrega descrita té lloc al procés de traducció, al qual participen:

- L'ARN missatger (ARNm): transmet la informació genètica emmagatzemada a l'ADN. A la transcripció, el procés previ, les seqüències específiques d'ADN són copiades en forma d'ARNm que transporta el missatge contingut a l'ADN als llocs de la síntesi proteica (els ribosomes).
- L'ARN de transferència (ARNt) i aminoàcids: els aminoàcids (components de les proteïnes) són units als ARN de transferència (ARNt) que els portaran fins al lloc de síntesi proteica, on seran encadenats un rere l'altre. L'enzim aminoacil-ARNt-sintetasa, el protagonista del punt, és l'encarregat d'aquesta unió, tal com s'ha dit, en un procés que consumeix ATP.
- Els ribosomes: són uns orgànuls citoplasmàtics encarregats de la biosíntesi proteica; són els encarregats de la unió dels aminoàcids que transporten els ARNt seguint la seqüència de codons de l'ARNm segons les equivalències del codi genètic.

Per a realitzar aquesta càrrega, l'ARNm s'uneix a la subunitat menor dels ribosomes. A aquests se'ls associa l'aminoacil-ARNt, gràcies a que l'ARNt té en una de les seves ases un triplet de nucleòtids anomenat anticodí, que s'associa al primer codó de l'ARNm segons la complementarietat de les bases (aquest aparellament de cofons es fa de manera antiparal·lela). A aquest grup de molècules se'ls uneix la subunitat ribosòmica major, format-se el complex ribosomal o complex actiu. Tots aquest processos estan catalitzats pels anomenats factors d'iniciació (FI). El primer codó que es tradueix és generalment l'AUG, que correspon a l'aminoàcid metionina en eucariotes. En el cas dels procariotes, l'aminoàcid és la formilmetionina.

El complex ribosòmic compta amb dos llocs d'unió o centres. El centre peptidil o centre P, on se situa el primer aminoacil-ARNt i el centre acceptor de nous

aminoacil-ARNt o centre A. El carboxil terminal (-COOH) de l'aminoàcid iniciat s'uneix amb l'amino terminal (-NH₂) de l'aminoàcid següent mitjançant un enllaç peptídic. Aquesta unió és catalitzada per l'enzim peptidil transferassa. Així, el centre P queda ocupat per un ARNt sense aminoàcid. L'ARNt sense aminoàcid surt del ribosoma. Es produeix la translocació ribosòmica. El dipeptil-ARNt queda ara al centre P. Tot això és catalitzat pels factors d'elongació (FE) i precisa GTP. Segons la terminació del tercer codó, apareix el tercer aminoacil-ARNt i ocupa el centre A. Després, es forma el tripèptid en A i posteriorment el ribosoma realitza la seva segona translocació. Aquests passos es poden repetir múltiples vegades, fins a centenars, segons el nombre d'aminoàcids que contingui el polipèptid. La translocació del ribosoma implica el desplaçament del ribosoma al llarg de l'ARNm en sentit 5'→3'.

Els codons UAA, UAG i UGA són senyals de stop que no especifiquen cap aminoàcid i es coneixen com a codons de terminació: determinen el final de la síntesi proteica. No existeix cap ARNt tal que el seu anticodó sigui complementari a aquests codons, i per tant, la biosíntesi del polipèptid s'interromp. Indiquen que la cadena polipeptídica ja ha acabat. Aquest procés ve regulat pels factors d'alliberació, de naturalesa proteica, que es situen al lloc A i fan que la peptidil transferasa separi, per hidròlisi, la cadena polipeptídica de l'ARNt. Un ARN, si és suficientment llarg, pot ser llegit o traduït, per diversos ribosomes a la vegada, un darrere l'altre. Al microscopi electrònic, s'observa com un rosari de ribosomes, anomenat poliribosoma o polisoma.

Un cop acabada la síntesi d'una proteïna, l'ARN missatger queda lliure i pot ser llegit de nou. De fet, és molt freqüent que abans de que finalitza un a proteïna ja estigui començant una altre, i per tant, una mateixa molècula d'ARN missatger, està sent utilitzada per diversos ribosomes simultàniament.

4.2 Tipus

Troben dues classes d'aminoacil ARNt sintetasa, que comparteixen en comú la seva funció general i un domini estructural anomenat WHEP implicat en la unió d'ATP. Tanmateix, compten amb certes diferències:

- **Classe I:** compta amb dos motius de seqüències molt conservadores. La seva aminoacilització té lloc en un extrem 2'-OH d'un nucleòtid d'adenosina i és normalment una proteïna monomèrica o dimèrica (posseeix una o dues subunitats respectivament). Contenen un plegament de Rossman catalític característic, i són majoritàriament monomèrics.

- **Classe II:** Aquesta té tres motius de seqüències altament conservadores. La seva aminoacilització té lloc a l'extrem 3'-OH de la mateixa adenosina, i normalment són dimèrics o tetramèrics (dos o quatre unitats respectivament). Són diferents estructuralment respecte els enzims de la Classe I, ja que en comptes del plegament de Rossman, tenen una làmina-beta³⁷ antiparalela.

Els aminoàcids es troben units al grup hidroxil (-OH) de l'adenosina mitjançant un grup carboxil.

Independentment d'on l'aminoacil estigui inicialment connectat al nucleòtid, el 2'-O-aminoacil-ARNt finalment migra a la posició 3' a través d'un procés de transesterificació.

La classificació de les aminoacil-ARNt-sintetases és la següent:

Classe I	Classe II
Leucina	Serina
Isoleucina	Treonina
Valina	Alanina
Arginina	Glicina
Cisteïna	Prolina
Metionina	Histidina

³⁷ És la segona forma d'estructura secundària de les proteïnes, només una mica menys comú que l'hèlix alfa. Les làmines beta consisteixen en cadenes beta connectades lateralment com a mínim amb dos o tres enllaços d'hidrogen.

Glutàmic	Aspartat
Glutamina	Asparagina
Lisina-I	Lisina-II
Tirosina	Fenilalanina
Tryptòfan	

4.3 Importància evolutiva

Les aminoacil-ARNt sintetases són transcendents a la vida: a banda de la seva essencial funció, tenen un paper central en l'establiment del codi genètic. Es creu que aquests enzims es troben entre les primeres proteïnes que van aparèixer en l'evolució. Partint d'aquesta dada, i mitjançant anàlisis filogenètiques, s'ha vist que les aminoacil-ARNt sintetases s'agrupen segons la seva especificitat d'aminoàcid i no segons la seva disposició en l'arbre filogenètic universal. Aquest fet és fonamental, ja que indica que les aminoacil-ARNt sintetases van aparèixer i evolucionar abans que l'arbre de la vida es dividís en els tres dominis reconeguts actualment (*Archaea*, *Bacteria* i *Eukarya*, tal com s'ha vist a l'apartat "3.1 Dominis" a la pàgina 23).

4.4 Aminoacil-ARNt-sintetases específiques amb les que es treballa

Per a escollir els tipus concrets d'ARNt sintetases he fet una tria a l'atzar, guiant-me per les que em semblaven més conegudes. Aquesta tria l'he hagut d'anar modificant quan no he trobat una seqüenciació prou abundant dels gens que codificaven aquell tipus d'enzim específic. Molts d'ells no han estat prou analitzats, i no s'han determinat en un nombre prou significatiu d'organismes.

Aquest procés l'he realitzat també amb els organismes utilitzats: no he partit directament d'uns organismes i tipus de aminoacil-ARNt-sintetases concrets, sinó que he realitzat una recerca prèvia per a comprovar que els proposats des d'un principi estaven seqüenciats a la base de dades. Si no ha estat així, aquests han estat substituïts, i n'he cercat de nous afins a la investigació que volia fer fins a obtenir un grup suficientment representatiu.

Cal dir que el treball ha estat realitzat amb el que he trobat: no es pot partir de l'idea de que trobarem les seqüències que volem d'uns individus determinats, si no que he hagut d'adaptar el treball al nivell de seqüenciació existent, ja que moltes d'aquestes seqüènciacions no han estat realitzades encara. Tot i que en alguns organismes les seqüènciacions del seu genoma siguin completes, no s'han acabat d'identificar, potser, els gens i les funcions de cada fragment d'ADN.

De les 20 aminoacilsintetases, en treballarem amb dos:

- Metionina-ARNt-sintetasa (en anglès Methionyl-tRNA synthetase o Methionine--tRNA ligase)
- Tirosina-ARNt-sintetasa (Tyrosyl-tRNA synthetase o Tyrosine--tRNA ligase)

5. EL GENBANK

Un cop realitzats aquests processos de selecció, ja comptava amb uns organismes determinats sobre els quals cercar les seqüències dels gens dels aminoacil-ARNt-sintetases escollits.

Aquestes seqüències d'ADN proporcionen un punt de partida fonamental per a descriure i entendre l'estructura, la funció i el desenvolupament d'organismes diversos genèticament.

L'eina amb la que he realitzat la cerca és el GenBank.

És una base de dades de seqüències genètiques d'accés lliure, que compta amb una col·lecció anotada de totes les seqüències de nucleòtids disponibles públicament i les seves traduccions a proteïnes. S'actualitza cada dos mesos.

Aquesta base de dades forma part de l'*International Nucleotide Sequence Database Collaboration* (INSDC), integrada per les bases de dades d'ADN de:

- La base de dades d'ADN de Japó (DNA DataBank of Japan (DDBJ))
- El Laboratori Europeu de Biologia Molecular (*European Molecular Biology Laboratory (EMBL)*)
- El GenBank, produït i mantingut pel Centre Nacional per a la Informació Biotecnològica (*National Center for Biotechnology Information (NCBI)*). L'NCBI és una part dels Instituts Nacionals de la Salut (*National Institutes of Health (NIH)*) dels Estats Units.

Aquestes organitzacions intercanvien dades diàriament, ja que el GenBank i els seus col·laboradors reben seqüències genètiques produïdes als laboratoris de tot el món, procedents de més de 100.000 organismes diferents. Aquestes seqüències arriben a la base de dades per la submissió directa des de laboratoris individuals i en gran part per centres de seqüenciació a gran escala.

El GenBank continua creixent a ritme exponencial, doblant la quantitat d'informació continguda cada 18 mesos (figura 5.1).

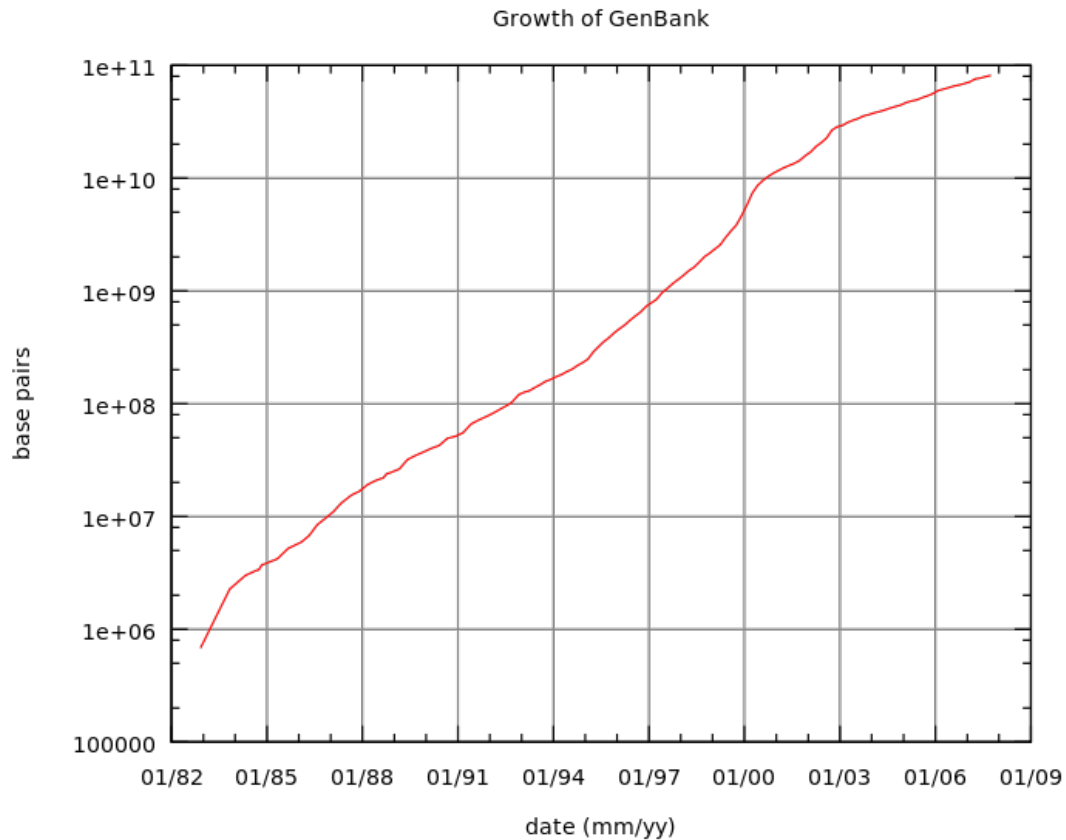


Fig. 2.5.1 La gràfica mostra el creixement de la base de dades del GenBank de l'NCBI, en una escala on es veu l'increment exponencial.

L'edició 194, produïda al febrer del 2013, contenia més de 150 bilions de bases de nucleòtids en més de 162 milions de seqüències.

Des de la seva fundació fa més de 30 anys, el GenBank s'ha tornat la base de dades de recerca més important i amb més influència a casi tots els camps de la biologia, ja que les seves dades accedeixen i són citades per milions d'investigadors de tot el món.

Malgrat tot, la pàgina web del GenBank mostra des de fa poc un missatge força desesperançador:

"The information on this web site remains accessible; but, due to the lapse in government funding, the information may not be up to date, and the agency may not be able to respond to inquiries until appropriations are enacted. For updates regarding government operating status see USA.gov."

(“La informació a aquesta pàgina web roman accessible; però, degut a un lapsus del finançament del govern, la informació podria no ser actualitzada a temps, i l'agència podria no arribar a respondre les preguntes fins que les assignacions siguin promulgades. Per a veure les actualitzacions en relació amb l'estat operatiu del govern, consulteu USA.gov.”)

La crisi econòmica mundial es fa vigent a tot arreu, i la recerca científica n'és una de les principals afectades, com a mínim pel que fa a la recerca pública (deixant de banda la recerca de farmacèutiques privades que prioritzen els productes amb gran benefici).

A més dels nucleòtids que fem servir en aquest apartat, la relació directa del GenBank amb l'NCBI (el *National Center for Biotechnology Information*, Centre Nacional per a la Informació Biotecnològica) ha permès que molta de la informació concreta sobre diversos organismes que hem necessitat s'hagi trobat amb aquesta eina. Com s'ha comentat, les bases de dades són múltiples, i la web compta amb un registre prou alt d'organismes registrats i altres dades com per a que hagi esdevingut una referència important a l'hora de realitzar el treball.

5.1 La recerca

El GenBank permet la cerca d'informació o referències a diverses bases de dades: des de publicacions mèdiques, fins a proteïnes, gens, nucleòtids, o genomes complets de diversos organismes, dels quals també podem consultar les seves dades taxonòmiques, i una gran varietat d'informació addicional.

Per a trobar les seqüències que volem, seleccionem la base de dades dels Gens (*Gene*), i inserim el criteri que volem al camp de cerca. En el nostre cas, el nom de l'organisme, i el del gen concret.

Per exemple:

“Methanothermus fervidus tyrosyl-trna synthetase”

El procés es bastant similar per a tots, i és prou lògic. Tot i així, es requereix una familiarització adient amb tot el món de la bioinformàtica per a comprendre el que s'està fent, hores per a veure les possibilitats que ofereix aquest recurs, i en definitiva, poder extreure'n el màxim d'informació possible.

Cal comentar que la web i tota la informació que conté està en anglès, per tant es requereix cert domini de l'idioma, i una determinada familiarització amb els termes científics en anglès.

Un cop obtinguts els resultats de la cerca, si és molt específica i només hi ha un resultat, la web ens condueix directament a la pàgina del gen.

Si el terme no ha estat prou específic o comptem amb diversos resultats, la web ens dirigeix als titulars de les diferents entrades que contenen el criteri de cerca.

Quan seleccionem l'entrada que volem, ens comencem a adonar de la quantitat d'informació que ens aporta la web: el símbol del gen, el seu locus específic, tipus, genoma, publicacions relacionades, etc.

NCBI Resources How To Sign in to NCBI

Gene [Save search](#) [Limits](#) [Advanced](#) [Help](#)

Display Settings: ☒ Full Report [Send to:](#)

Mfer_0472 tyrosyl-tRNA synthetase [*Methanothermobacter fermentans* DSM 2088]
Gene ID: 9962192, updated on 25-May-2013

Summary

Gene symbol: Mfer_0472
Gene description: tyrosyl-tRNA synthetase
Locus tag: Mfer_0472
Gene type: protein coding
RefSeq status: PROVISIONAL
Organism: [Methanothermobacter fermentans DSM 2088 \(strain: DSM 2088\)](#)
Lineage: Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales; Methanothermobacteraceae; Methanothermobacter

Genomic context

Location: country: Iceland, isolation source: hot sulfataric spring
Sequence: NC_014658.1 (451269..452234)

Genomic map showing the location of Mfer_0472 on the NC_014658.1 sequence. The map includes coordinates 448882 and 453289, and labels for Mfer_0469, Mfer_0470, Mfer_0471, Mfer_0472, Mfer_0473, and Mfer_0474.

Genomic regions, transcripts, and products

Genomic Sequence: [Go to reference sequence details](#)
[Go to nucleotide](#) [Graphics](#) [FASTA](#) [GenBank](#) [PubMed](#) [PubMed\(nucleotide/PMC\)](#)

Related information

- BioProjects
- BioSystems
- Conserved Domains
- Full text in PMC
- Full text in PMC_nucleotide
- Gene neighbors
- Genome
- Nucleotide
- Protein
- Protein Clusters
- PubMed
- PubMed(nucleotide/PMC)

Fig. 2.5.2 Captura d'imatge de l'entrada al GenBank per al gen tirosina-ARNt sintetasa de l'organisme *Methanothermobacter fermentans*

Per a obtenir la seqüència, em dirigeixo a “Genomic regions, transcripts, and products” (“Regions genòmiques, traduccions, i productes”), i a “Go to nucleotide” (“anar al nucleòtid”). Dins d'aquest apartat, tal com es veu a la figura 2.5.2, és possible dirigir-se als enllaços “Graphics”, “FASTA” i “GenBank”. Per a obtenir la seqüència desitjada, em dirigeixo a “FASTA”.

5.2 El format FASTA

El format FASTA és un format de fitxers informàtics basats en text, utilitzat en bioinformàtica, per a representar seqüències tant d'àcids nucleics com pèptids. Aquests es representen utilitzant codis d'una única lletra. El format també permet incloure noms de seqüències i comentaris que precedeixen a les seqüències en si.

La simplicitat del format FASTA fa que manipular i analitzar les seqüències sigui fàcil, utilitzant eines per a processar texts i llenguatges d'*script*³⁸ com el *Python* o el *PERL*.

Una seqüència FASTA comença amb una descripció en una única línia (línia d'encapçalament), seguida per línies de dades de seqüència. La línia de descripció es distingeix de les dades de la seqüència per un símbol ">" (més gran que) a la primera columna. La paraula inserida a continuació d'aquest símbol és l'identificador de la seqüència, i la resta de la línia és la descripció (ambdós són opcionals). No hauria d'existir un espai entre el ">" i la primera lletra de l'identificador. Es recomana que totes les línies de text siguin menors de 80 caràcters. La seqüència acaba si apareix una altra línia iniciada amb el símbol ">": això indica l'inici d'una altra seqüència.

```
>gi|312136230:451123-452378 Methanothermus fervidus DSM 2088 chromosome, complete genome
TCAAATATTAGATCCTAAAACCTTTCAACCTATAGATTTAGACATGAAATCATGGTTCCAAAATTTAAAC
ATTGGAGATGAAGTAAAAGTTATAAAAATTAATAAAAGGATTTACATAGTAGATTGACGCAACATATGGT
GATTTTCATGAACAAACAACTTGAATTAGCTAAAAAGGCACAGTGGAAATAATTACAATTGAAGAAGCTTA
AAGAAGTCCTAAAAAGAAAAATCCTGTGGCATATATAGGTTATGAACCTTCAGGAAAAATACATCTCGG
ACATTTAATTACCATAAAGAAGATGATAGATTTACAAAAAATTGGATTTAAAATTAAAATTCCTCTAGCA
GATTATCATGCATATCTAAATTATAAAGGAGAGTTAGACGAAATTAGAGAAATTGCAGAAATATAATAAGC
GTTGTTTTTTAGCTATGGGATTATCTAAGGATACAGAGTTCATATTAGGTTCTGAGTTTCAAACAGATGA
TGATTATGTTTTAGATCTATATAGGTTAGCACATATAACTACCTTAAATAGAGCTAAAAGGAGTATGTCA
CAAATAATAAGAGAAGCTGGTGAACATAAAGTTTCTGAAGTGCTTTATCCATTAATGCAAGTTATTGACA
TGGTTCATTTGGGAGTAGACGTTTCAGTGGGAGGGATGGAACAAAGAAAAATTCATATGTTAGCTCGTGA
CAACTTACCAAAAATTTGGATACACACCACCACTGTGTATACACACACCTATAATCCATGGAAGTATGGA
TCAGAAAAATGTCATCAAGTAAAGGGAACCTTTATTGCTGTTGATGATCCACCTAAAGTTATAAAGATA
AAATAATGAAAAGTTATTGTCCAGCAGCTAAAAATAGATGGAAATCCTGTAATAGAAATGCAAAATATAT
TATTTTTGATGAAAACAAAAAATTTTAATAGAAAGGCTGAAAAATATGGTGGCAATTTAGAAGTTAAT
TGTGAAGAATTAATCAACTTATATTCTAGTGGAGAGTTACATCCATTAGATCTCAAGAATGCTGTTGCAA
AATATCTTATAAAGATTTTAGAACCTATCAGAGAAAAATTTAACTTCTAAAATGGGTGATTAAAAAATGAA
AAAGTCACGTTTAAAGTTATTTAAAATTACTTCACTGACAATTTCAATTTTAGGACTAATATTAATAATT
GGAAGTGTGATGCTTTTGCATATGTGGGTTTAACTTTTATCAACAAAGTTGGTTCAGAAATT
```

Fig. 2.5.3 Seqüència FASTA del gen tirosina-ARNt sintetasa de l'organisme *Methanothermus fervidus*

³⁸ Un llenguatge script o llenguatge de *scripting* és un llenguatge de programació que controla aplicacions. Els *scripts* són executats directament des del seu codi font, que generalment són fitxers de text que contenen llenguatges de marcatge específics.

Com a exemple de seqüència FASTA, a la figura 2.5.3 es veu la de la seqüència cercada com a l'apartat anterior. Aquest gen concret té una seqüència relativament curta; altres seqüències, com per exemple la d'aquest mateix gen a l'*Homo sapiens*, té una mida molt més gran.

Aquestes seqüències FASTA són directament a l'apartat dels nucleòtids concrets que volem del GenBank, junt a altres formats. Un cop la seleccionem i ens trobem a dins, guardem la seqüència FASTA per a mantenir un ordre i un cert inventari de seqüències, les que cerquem per a comparar-les posteriorment.

Per a fer-ho, obrim un editor de textos (qualsevol editor simple, com ara TextEdit en el cas de Mac OS, o el Bloc de Notes de Windows). Copiem la seqüència, inserim un nom més representatiu a l'identificador (a continuació del ">", sense cap espai) i a l'hora de desar-la, canviem el format de l'arxiu, concretament a ".fasta".

Repetint aquest procés obtenim totes les seqüències que volem: les dues aminoacil-ARNt-sintetases de tots els organismes utilitzats al treball.

De tota manera, com s'ha dit, no sempre trobem tant els dos aminoàcids, com tots els gens a tots els genomes: manquen algunes seqüències de certs ADNs mitocondrials, que no han estat encara seqüenciades, o altres casos. Degut a aquest motiu, el nombre d'organismes amb els que treballem és prou gran com per a compensar aquestes petites mancances: treballem amb les eines que podem, i tot i que la base de dades és actualitzada periòdicament, encara no disposem de totes.

Tots els arxius FASTA amb els que hem realitzat el treball es troben al l'Annex 2, el CD proporcionat junt al treball, organitzats pels aminoàcids concrets, i els organismes (dins dels seus adients dominis). A més, per a facilitar la posterior cerca d'informació, s'ha apuntat el *Gene ID*³⁹ (número d'identificació) dels gens per a trobar-los més fàcilment al GenBank: cadascun té una xifra assignada que en facilita la seva localització a la base de dades.

Ja que els arxius estat desats en format FASTA (".fasta") per a una millor compatibilitat amb el programari bioinformàtic, es poden obrir amb qualsevol editor de textos (en cas de que l'ordinador des del que es veuen les seqüències no detecti cap programa per si mateix per a obrir-les).

³⁹ El *Gene ID* es pot veure clarament a la figura 2.5.2 (pàgina 47), sota el nom del gen cercat (*Gene ID*: 9962192)

5.3 Seqüències obtingudes

Les seqüències que he trobat i adaptat per a comparar i realitzar l'experimentació del treball (proporcionades a l'Annex 2, el CD adjunt al treball), són:

- Desfe_1341 tyrosyl-tRNA synthetase [*Desulfurococcus fermentans* DSM 16532] (Gene ID- 13061741).fasta
- Mfer_0472 tyrosyl-tRNA synthetase [*Methanothermus fervidus* DSM 2088] (Gene ID- 9962192).fasta
- tyrS tyrosyl-tRNA synthetase [*Escherichia coli* str. K-12 substr. MG1655] (Gene ID- 948855)
- Mchl_3213 tyrosyl-tRNA synthetase [*Methylobacterium extorquens* CM4].fasta
- YARS tyrosyl-tRNA synthetase [*Homo sapiens* (human)] (Gene ID 8565).fasta
- CTRG_00338 tyrosyl-tRNA synthetase [*Candida tropicalis* MYA-3404] (Gene ID 8299913).fasta
- TYS1 tyrosine--tRNA ligase TYS1 [*Saccharomyces cerevisiae* S288c] (Gene ID- 853097).fasta
- TYS1 tyrosyl-tRNA synthetase, cytoplasmic [*Candida albicans* SC5314] (Gene ID 3638065).fasta
- Yars tyrosyl-tRNA synthetase [*Rattus norvegicus* (Norway rat)] (Gene ID- 313047).fasta
- AT2G33840 tyrosyl-tRNA synthetase [*Arabidopsis thaliana* (thale cress)] (Gene ID- 817952).fasta
- emb2768 tyrS tyrosyl-tRNA synthetase [*Arabidopsis thaliana* (thale cress)] (Gene ID- 821282).fasta
- YARS2 tyrosyl-tRNA synthetase 2, mitochondrial [*Homo sapiens* (human)] (Gene ID 51067).fasta
- Yars2 tyrosyl-tRNA synthetase 2 (mitochondrial) [*Rattus norvegicus* (Norway rat)] (Gene ID- 287924).fasta
- MSY1 likely tRNA synthetase similar to *S. cerevisiae* MSY1 (YPL097W) mitochondrial tyrosyl-tRNA synthetase [*Candida albicans* SC5314] (Gene ID 3644295).fasta
- Desfe_0577 methionyl-tRNA synthetase [*Desulfurococcus fermentans* DSM 16532] (Gene ID 13062267).fasta

- Mfer_0606 methionyl-tRNA synthetase [*Methanothermus fervidus* DSM 2088].fasta
- metG methionyl-tRNA synthetase [*Escherichia coli* str. K-12 substr. MG1655] (Gene ID- 946643).fasta
- Mchl_4558 methionyl-tRNA synthetase [*Methylobacterium extorquens* CM4] (Gene ID 7117953).fasta
- MARS methionyl-tRNA synthetase [*Homo sapiens* (human)] (Gene ID 4141).fasta
- CTRG_06073 methionyl-tRNA synthetase [*Candida tropicalis* MYA-3404] (Gene ID 8298722).fasta
- MSM1 methionine--tRNA ligase MSM1 [*Saccharomyces cerevisiae* S288c] (Gene ID 853081).fasta
- MES1 likely tRNA synthetase similar to *S. cerevisiae* MES1 (YGR264C) cytoplasmic methionyl tRNA synthetase [*Candida albicans* SC5314] (Gene ID 3636516).fasta
- Mars methionyl-tRNA synthetase [*Rattus norvegicus* (Norway rat)] (Gene ID 299851).fasta
- OVA1 methionyl-tRNA synthetase [*Arabidopsis thaliana* (thale cress)] (Gene ID 824706).fasta
- MARS2 methionyl-tRNA synthetase 2, mitochondrial [*Homo sapiens* (human)] (Gene ID 92935).fasta
- Mars2 methionyl-tRNA synthetase 2, mitochondrial [*Rattus norvegicus* (Norway rat)] (Gene ID 316403).fasta
- MSM1 mitochondrial tRNA methionyl-tRNA synthetase similar to *S. cerevisiae* MSM1 (YGR171C) mitochondrial methionyl-tRNA synthetase [*Candida albicans* SC5314] (Gene ID 3636398).fasta
- CTRG_06069 methionyl-tRNA synthetase, mitochondrial [*Candida tropicalis* MYA-3404] (Gene ID 8298718).fasta

Una vegada obtingudes les seqüències FASTA, arriba el moment de realitzar les comparacions per a obtenir els resultats i percentatges de similitud entre elles. Aquestes comparacions es realitzen mitjançant alineaments de seqüències.

Hi ha hagut unes quantes seqüències que no han estat analitzades encara, i que per tant no podran participar als alineaments.

Aquestes són les que correspondrien als enzims tant tirosina i metionina-ARNt sintetasa dels mitocondris de l'*Arabidopsis thaliana*, i al gen tirosina-ARNt sintetasa del llevat *Saccharomyces cerevisiae*. Això és degut al fet de que tot i comptar amb la seqüència completa del seu genoma (ambdós organismes han estat completament seqüenciats), encara no s'han localitzat i identificat aquests gens al seu ADN.

5.4 La bioinformàtica i la seqüenciació de l'ADN

Per a realitzar el treball, selecciono uns gens concrets d'organismes específics, extraient-ho tot de la base de dades del GenBank. Però com es poden arribar a tenir aquests genomes "escrits"?

La bioinformàtica, que implica la manipulació, la cerca i extracció d'informació de les dades de la seqüència de l'ADN, n'és l'encarregada. El desenvolupament de les tècniques per a emmagatzemar i buscar seqüències d'ADN ha generat molts avanços en el desenvolupament de software dels ordinadors, per a moltes aplicacions, especialment algoritmes de cerca de frases, aprenentatge automàtic i teories de bases de dades. La recerca de frases o algoritmes de coincidències, que busquen l'ocurrència d'una seqüència de lletres dins d'una seqüència més gran, es va desenvolupar per a buscar seqüències específiques de nucleòtids. En altres aplicacions com editors de text, fins i tot algoritmes simples poden funcionar, però les seqüències d'ADN poden generar que aquests algoritmes presentin un comportament pèssim, degut al baix nombre de caràcters. El problema relacionat de l'alineament de seqüències persegueix identificar seqüències homòlogues i localitzar mutacions específiques que les diferencien. Aquestes tècniques, fonamentalment l'alineament múltiple de seqüències, s'utilitzen per a estudiar les relacions filogenètiques i la funció de les proteïnes. Les col·leccions de dades que representen seqüències d'ADN com la d'un genoma, tals com les produïdes pel Projecte Genoma Humà, són difícils d'utilitzar sense anotacions, que marquen la localització dels gens i els elements reguladors a cada cromosoma. Les regions d'ADN que tenen patrons associats com gens que codifiquen proteïnes, o ARN, poden identificar-se per algoritmes de localització de gens, fet que permet als investigadors predir la presència de productes gènics específics en un organisme fins i tot abans de que hagi estat aïllat experimentalment.

6. COMPARACIÓ DE SEQÜÈNCIES: EL CLUSTAL

En aquest apartat s'analitzaran els diferents mètodes de comparacions de seqüències, els tipus d'alineaments, i les eines que em permetran analitzar-los.

6.1 Alineaments

Un alineament de seqüències, en bioinformàtica, és una forma de representar i comparar dues o més seqüències o cadenes d'ADN, ARN, o estructures primàries proteiques per a ressaltar les seves zones de similitud, que podrien indicar relacions funcionals o evolutives entre els gens o proteïnes consultats. Les seqüències alineades s'escriuen amb les lletres (representant aminoàcids o nucleòtids) en files d'una matriu en les que, si es necessari, s'insereixen espais per a que les zones amb idèntica o similar estructura s'alineïn.

Troblem tres tipus d'alineaments amb els que treballar:

- **Alineaments Locals:** s'utilitzen quan només es disposa d'un fragment de la seqüència o quan les dues seqüències a comparar són distants i només s'assemblen en una part. Aquests tipus genera alineaments a les zones més similars de la seqüència, i sol ser la millor opció a no ser que no s'estigui segur que les dues seqüències s'hagin d'assemblar en tota la seva extensió. En molts casos les seqüències homòlogues s'assemblen tan sols en les regions més conservades.
- **Alineaments Globals:** són utilitzats quan es coneix que dues seqüències s'assemblen en parts de tota la seva extensió. S'utilitzen sobretot dins dels alineaments múltiples.

Els tres mètodes principals amb els que es generen alineaments de parells de seqüències són els de matriu de punts, els de programació dinàmica i els de cerca de paraula. La majoria de mètodes d'alineació múltiple poden funcionar també amb tan sols dues seqüències.

- **Alineaments múltiples:** És una extensió de l'alineament de parells que incorpora més de dues seqüències al mateix temps. Els mètodes d'alineament múltiple intenten alinear totes les seqüències d'un conjunt donat. Són utilitzats sovint per a la identificació de regions conservades en un grup de seqüències que hipotèticament estan relacionades evolutivament.

Aquests motius conservats poden servir com a recurs en conjunt amb l'estructura i la informació mecànica per a localitzar llocs actius catalítics dels enzims. Els alineaments són també utilitzats per a ajudar a l'establiment de relacions evolutives mitjançant la construcció d'arbres filogenètics. Els alineaments múltiples de seqüències són informàticament difícils de produir i la majoria de les formulacions del problema condueixen a problemes d'optimització combinatòria NP-complets. Tot i així, la utilitat d'aquests alineaments en la bioinformàtica ha donat pas al desenvolupament d'una varietat de mètodes adequats per a la alineació de tres o més seqüències.

Per a comparar les seqüències obtingudes, en el meu cas es farà servir un programa d'alineaments múltiples: volem incorporar més de dues seqüències alhora per a identificar, com s'ha dit, les regions conservades que hipotèticament estan relacionades evolutivament.

6.2 Eines per a l'alineament de seqüències

Una de les eines més populars en bioinformàtica és el BLAST: aquesta, tot i no realitzar alineaments entre seqüències inserides, realitza una cerca a tota la base de dades a partir d'una seqüència concreta. D'aquesta manera, el programa mostra totes les seqüències similars que existeixen, sent un recurs força utilitzat.

Però per a l'alineament de seqüències, trobem una gran varietat d'aplicacions que ens permeten comparar-les. Entre elles trobem algunes com el CodonCode Aligner, MUSCLE, DIALIGN-TX y DIALIGN-T, ADN Baser, T-Coffee, MEGA, etc.

En el meu cas, els alineaments es fan mitjançant el Clustal Omega, una eina per a realitzar alineaments múltiples de seqüències.

El Clustal (originalment ClustalW2 i ClustalX) és una de les eines més populars per a realitzar aquest tipus de comparacions, i ha estat actualitzat recentment (Clustal Omega). Aquesta aplicació ofereix moltes possibilitats a l'hora de treballar i obtenir resultats.

6.3 Clustal Omega

El Clustal Omega és un programa d'alineament múltiple de seqüències per a proteïnes, ADN i ARN. Produeix aquest tipus d'alineaments, biològicament significatius, de seqüències divergents. Ofereix la possibilitat de veure les relacions evolutives mitjançant cladogrames o arbres filogenètics⁴⁰, i una especificació a l'hora de crear-los si els traslladem al *ClustalW2 Phylogeny*.

Altres opcions que ens proporciona són per exemple el visionat dels alineaments (*Alignments*) o els percentatges de proximitat de les seqüències (*Percent Identity Matrix*). A més, el Clustal permet una varietat d'opcions per a inserir les seqüències, i per a la sortida d'aquestes: es poden establir els valors a la pàgina inicial del propi programa.

Per a facilitar el procés d'alineació de les meves seqüències, he inserit totes les seqüències d'un mateix aminoàcid en un arxiu, obtenint-te dos totals: un per al metionina-ARNt sintetasa, i un altre per al tirosina-ARNt sintetasa. Els resultats els obtindré per a cada aminoàcid-ARNt-sintetasa, realitzant cada alineament per separat.

D'aquesta manera obtindré els resultats i les relacions genètiques per al metionina-ARNt-sintetasa en tots els organismes, i d'altra banda els del tirosina-ARNt-sintetasa.

The screenshot shows the Clustal Omega web interface. At the top, there's a navigation bar with 'EMBL-EBI' logo and links for 'Services', 'Research', 'Training', 'Industry', and 'About us'. Below this is the 'Clustal Omega' title. A secondary navigation bar includes 'Input form', 'Web services', and 'Help & Documentation', along with 'Share' and 'Feedback' links. The main content area is titled 'Multiple Sequence Alignment' and describes the tool. It features two steps: 'STEP 1 - Enter your input sequences' and 'STEP 2 - Set your parameters'. In Step 1, there's a text area for pasting sequences in various formats, with a sample sequence for Desulfurococcus fermentans. Below the text area is an option to upload a file. In Step 2, there's a dropdown for 'OUTPUT FORMAT' set to 'Clustal w/ numbers', and a link for 'More options...'.

Fig. 2.6.1 Captura d'imatge del Clustal Omega amb el principi (zona que es veu del quadre) de les seqüències del gen tirosina-ARNt sintetasa dels organismes amb els que realitzo l'experimentació

⁴⁰ Arbre que mostra les relacions genealògiques entre organismes. Difereix del cladograma en que les arrels es dibuixen proporcionalment a la quantitat de canvi evolutiu.

6.4 Resultats del Clustal Omega per a les seqüències

Una vegada inserides les seqüències, el programa les processa. Depenent de la quantitat i similitud entre elles, varia el temps total per a obtenir els resultats. Un cop finalitzat el procés, el Clustal mostra les seqüències proporcionades ja alineades, oferint múltiples opcions. Els resultats només es desen a la base de dades durant 7 dies, raó per la qual la majoria de dades necessàries i significants s'han desat, i inclòs a l'Annex 2, el CD adjunt al treball.

A la pàgina inicial de l'alineament, el Clustal mostra l'alineament de les seqüències, on es pot veure clarament les zones on no hi ha similitud (figura 2.6.2), i les zones on la semblança és vigent (figura 2.6.3): hi ha més seqüències i fragments alineats.

```
Hom_Sapiens          GGTAGACAGTTGGATATTAGGCATATGGAGTTCAGGAGAGAGTCTCCTGAGGAATTATAAT 14435
Rat_Norvegicus       AATAG-----ATGTGGATTTCAGATGAGCTAT-----GAGACACTAT 9312
Ara_Thaliana         -----26
Met_Fervidus         -----0
Can_Tropicalis       -----0
Can_Albigans_Sac_Cerevisiae -----0
Esc_Coli             -----0
Met_Extorquens       -----0
Hom_Sapiens_Mith     -----0
Rat_Norvegicus_Mith  -----5
Des_Fermentans       -----0
Sac_Cerevisiae       -----0
Can_Albigans_Mith_Sac_Cerevisiae_Mith -----0
Can_Tropicalis_Mith  -----0
```

Fig. 2.6.2 Fragment de l'alineament total: zones de les seqüències sense gran similitud en els organismes amb els que treballem per al gen metionina-ARNt sintetasa

```
Hom_Sapiens          -----18424
Rat_Norvegicus       CCGAAAAAATAAAGATTGTTGTTATGTG-----TATGGGTGTTGCGTG 11041
Ara_Thaliana         ---GCGAGATTAAACATTGA--AGTTAATATTGCC-----GTTTCTCATTGCGAGT 1458
Met_Fervidus         GCAATAGTGCATTACACTTTGGTCAATTACGATCAAC-----CTATATCCAGCAGA 86
Can_Tropicalis       GTTAATATGTTCCACATTAGGTAAATATTGTTGGATC-----AGTTTTATCCGCGGA 644
Can_Albigans_Sac_Cerevisiae GTTAACAATGTTCTCATTGGGTAAACATTATGGGTC-----GTTTTGTCTGCTGA 644
Esc_Coli             ---GAAG---GCAGCAACTCCGCAAGCCGTCACAC-----TGTTGTTCATGTC 1281
Met_Extorquens       GTGAGGGC-----GCGTGCCGGAGC 1150
Hom_Sapiens_Mith     CTGAGACTACCCAGCCTTCGCACTACCTGCTTCCCTAGTGAGCCAGGTTGGTGGGC 1414
Rat_Norvegicus_Mith  CTGGAGCTACCCACTTCTGTCGCTGCTTTCCCAAGTGAGCCGGGTTAAGGGGC 1401
Des_Fermentans       ---ATGCACCAACACATATA--GGGCAAGCGTATACAA-----CAGTGTCCGCTGA 86
Sac_Cerevisiae       ---ATGCCAAACCGCACTC--GGCCTCTTACTCCA-----GCTTATTAAGTGA 119
Can_Albigans_Mith_Sac_Cerevisiae_Mith ---ATGCAGCACCAATATA--GGCCTCTTATTCGA-----TGTTGATGACAGA 134
Can_Tropicalis_Mith  ---ATGCAGCTCCGATATT--GTCCTTGTATTCGA-----TGTTGATGCGGA 134

Hom_Sapiens          -----18408
Rat_Norvegicus       -----AATTGTTTTTATTTTTAAAGAT-----TTTATGGAGGTGGGGAT 10927
Ara_Thaliana         TGGTTACTGCTTA---AGTCTTT-----GTTAT--GTCTCCACACATTGATGAT 1362
Met_Fervidus         -----0
Can_Tropicalis       CAAACACTGGTAACAAATGTTCAAACTG---GTCATTGATTAACAAACAGCTGAT 534
Can_Albigans_Sac_Cerevisiae CCAACACTGGGAAGCAAAATGTTGAACTG---GTCATTAGTGAACAAACAGGCTGAC 534
Esc_Coli             CGTAAGCTTCGATGAATAGTGAAGAAAGACTCCACCGCCGAGCTGACCACTTCACTGG 1190
Met_Extorquens       CAGCGACGGCAATTACGACCAAGAGGCGAT---CATCAACCGCATCAACCGGACCTCGC 1076
Hom_Sapiens_Mith     CTGGGACTGTGACTACTATGATGAGAAAGGT---GGTTAAATGCTGGACTCTGAGCTGGC 1301
Rat_Norvegicus_Mith  CTGGGACTGCGACTACTATGATGAGAAAGGT---GGTTAAATGCTGGACTCTGAGCTGGC 1288
Des_Fermentans       -----0
Sac_Cerevisiae       -----ATGCAT-----GT-----CGATCAATTGTGATCCT 27
Can_Albigans_Mith_Sac_Cerevisiae_Mith -----ATGAGATTCAAAATCAGGGGACCT-----TTAATCAGCTTAGATAT 42
Can_Tropicalis_Mith  -----ATGAAGTTACGAAGTCTTAAACCC-----GTTGTACAAATACGGTTC 42

Hom_Sapiens          GGAAGCTATAGTGAGC-----18424
Rat_Norvegicus       TTAGCTCAGTGGTAGAGCGCTTGCCTAGCAAGTGAAGGCCCTGGGTTCGGTCCCGAGCT 10987
Ara_Thaliana         ATGTCACTAATTATGCTGTTTCATATAAAGAAAGGTGTTCTCT---TAGTT-----1412
Met_Fervidus         -----TGAAAAATTTTGTAACTGTGCACTGCCATAT 33
Can_Tropicalis       AAAATTCTCCAAAGT---CTGATGAAGAAACATTTAATCACTCTGCTTTACTTAT 591
Can_Albigans_Sac_Cerevisiae AAAATTCTCTTAAGT---CTGATGAAGAAACATTTAATCACTCTGCTTTACTTAT 591
Esc_Coli             TAAAGATATTGTTTAC---TTCGAGGCTGTCTGGCCTGCTAT---GCTG-----1236
Met_Extorquens       CAACGACTCGGCAC---CTGCGCCAGCGCTGCTCTGATGAT---CAGGAAGAACT 1129
Hom_Sapiens_Mith     AGATGCTCTGGGAGGT---CTCTGAACCGGTGCACTGCCAAAG---AATAAATCCTT 1354
Rat_Norvegicus_Mith  AGATGCTCTAGGAGGT---CTTTAAACCGGTGCACTGCCAGAC---AATAAACCTT 1341
Des_Fermentans       -----ATGAGTGTG---TTTATGTGACTACACCATATATTAC---CCTA-----40
Sac_Cerevisiae       CTGACTCTAAGGTCT---CCCAGTAAACGACCAATTTTCTAT---CCCA-----73
Can_Albigans_Mith_Sac_Cerevisiae_Mith AAATCTACAAGCGGT---TCTATATAACTACCATATATTGTTAT---GTCA-----88
Can_Tropicalis_Mith  AAGTCACTAAACAT---TTTATATAACACTCTTATTTCTAT---GTTA-----88
```

Fig. 2.6.3 Fragment de l'alineament total: zones amb similitud a les seqüències dels organismes per al gen metionina-ARNt sintetasa

Altres resums dels resultats que ofereix el Clustal Omega són els percentatges de similitud entre seqüències, que es troben desats per a les dues alineacions al CD d'arxius.

```
#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
```

1: Hom_Sapiens	100.00	59.53	38.67	34.86	38.81	38.15	35.55	31.59	36.44	35.60	32.90	34.68	34.91	37.18
2: Rat_Norvegicus	59.53	100.00	37.38	33.75	34.28	34.37	33.97	32.05	34.87	34.78	32.20	34.17	35.02	35.04
3: Ara_Thaliana	38.67	37.38	100.00	35.66	37.50	36.29	32.72	35.37	36.33	37.35	33.49	36.13	37.81	39.40
4: Met_Fervidus	34.86	33.75	35.66	100.00	51.87	51.07	31.50	25.63	34.09	29.44	41.76	41.09	41.67	45.35
5: Can_Tropicalis	38.81	34.28	37.50	51.87	100.00	82.38	34.12	29.47	35.78	36.01	38.43	37.95	39.96	42.75
6: Can_Albigans_Sac_Cerevisiae	38.15	34.37	36.29	51.07	82.38	100.00	34.40	29.47	34.52	35.57	39.00	37.24	39.57	41.31
7: Esc_Coli	35.55	33.97	32.72	31.50	34.12	34.40	100.00	35.80	37.35	37.53	36.87	37.27	36.22	36.50
8: Met_Extorquens	31.59	32.05	35.37	25.63	29.47	29.47	35.80	100.00	51.34	49.68	30.50	33.26	29.37	29.37
9: Hom_Sapiens_Mith	36.44	34.87	36.33	34.09	35.78	34.52	37.35	51.34	100.00	79.58	35.13	38.21	36.77	38.31
10: Rat_Norvegicus_Mith	35.60	34.78	37.35	29.44	36.01	35.57	37.53	49.68	79.58	100.00	32.62	34.86	35.20	35.49
11: Des_Fermentans	32.90	32.20	33.49	41.76	38.43	39.00	36.87	30.50	35.13	32.62	100.00	43.79	46.61	47.18
12: Sac_Cerevisiae	34.68	34.17	36.13	41.09	37.95	37.24	37.27	33.26	38.21	34.86	43.79	100.00	52.98	54.11
13: Can_Albigans_Mith_Sac_Cerevisiae_Mith	34.91	35.02	37.81	41.67	39.96	39.57	36.22	29.37	36.77	35.20	46.61	52.98	100.00	70.42
14: Can_Tropicalis_Mith	37.18	35.04	38.40	45.35	42.75	41.31	36.50	29.37	38.31	35.49	47.18	54.11	70.42	100.00

Fig. 2.6.4 Percentatges de similitud entre les seqüències (Percent Identity Matrix) del gen metionina-ARNt sintetasa

```
#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
```

1: Hom_Sapiens_Mith	100.00	52.26	37.15	33.53	40.17	38.10	36.21	39.63	52.63	33.56	30.73	32.44	34.45	34.68
2: Rat_Norvegicus_Mith	52.26	100.00	39.13	nan	40.17	nan	29.35	39.28	41.13	33.09	28.21	29.18	32.03	30.25
3: Esc_Coli	37.15	39.13	100.00	50.39	46.22	45.14	38.93	31.11	35.38	29.01	29.69	28.24	33.59	32.06
4: Met_Extorquens	33.53	nan	50.39	100.00	42.25	37.75	31.17	32.94	32.82	27.01	22.39	18.25	24.82	23.36
5: Ara_Thaliana2	40.17	40.17	46.22	42.25	100.00	50.11	34.35	35.32	38.20	26.67	29.17	28.00	29.33	29.33
6: Can_Albigans_Mith_Sac_Cerevisiae_Mith	38.10	nan	45.14	37.75	50.11	100.00	36.00	32.15	36.68	35.06	25.97	36.36	32.47	35.06
7: Ara_Thaliana	36.21	29.35	38.93	31.17	34.35	36.00	100.00	44.41	47.76	nan	nan	nan	nan	nan
8: Rat_Norvegicus	39.63	39.28	31.11	32.94	35.32	32.15	44.41	100.00	56.98	42.65	43.16	51.63	50.66	50.65
9: Hom_Sapiens	52.63	41.13	35.38	32.82	38.20	36.68	47.76	56.98	100.00	42.91	45.96	51.34	50.61	51.02
10: Des_Fermentans	33.56	33.09	29.01	27.01	26.67	35.06	nan	42.65	42.91	100.00	50.27	43.90	44.48	44.38
11: Met_Fervidus	30.73	28.21	29.69	22.39	29.17	25.97	nan	43.16	45.96	50.27	100.00	48.08	51.14	51.66
12: Sac_Cerevisiae	32.44	29.18	28.24	18.25	28.00	36.36	nan	51.63	51.34	43.90	48.08	100.00	70.38	67.97
13: Can_Tropicalis	34.45	32.03	33.59	24.82	29.33	32.47	nan	50.66	50.61	44.48	51.14	70.38	100.00	83.12
14: Can_Albigans	34.68	30.25	32.06	23.36	29.33	35.06	nan	50.65	51.02	44.38	51.66	67.97	83.12	100.00

Fig. 2.6.5 Percentatges de similitud entre les seqüències (Percent Identity Matrix) del gen tirosina-ARNt sintetasa

El Clustal Omega també ens mostra l'arbre filogenètic sense arrelar (figura 2.6.6), creat a partir de les seqüències, al que podem donar la opció de mostrar les distàncies. El propi programa ofereix també la possibilitat d'enviar-lo al *Clustal Phylogeny* ("Send to ClustalW2_Phylogeny" des de l'apartat inicial d'Alineaments), on es pot modificar per a obtenir-lo arrelat, en forma de Cladograma, amb distàncies, etc.

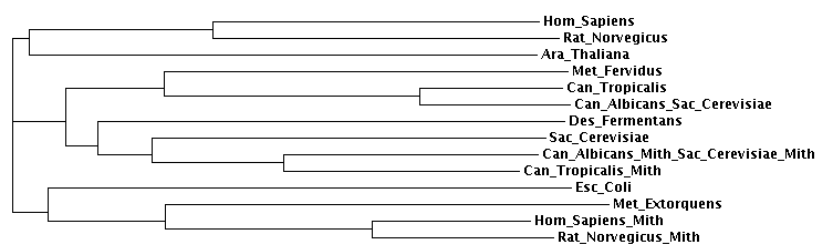


Fig. 2.6.6 Arbre filogenètic (del Clustal Omega) per als resultats dels alineaments del gen metionina-ARNt sintetasa

A partir d'aquest primer arbre, ja es poden començar a intuir els resultats. No obstant, és possible obtenir arbres més visuals i elaborats. Per a fer-ho, utilitzaré el TreeView, un programa específic per a aquest tipus d'arbres filogenètics, en comptes del *ClustalW2_Phylogeny*. A partir d'aquests, en podré veure les relacions més clarament i extreure'n les conclusions definitives.

Per a traslladar la informació d'aquest primer arbre al TreeView, n'obtindrè les dades inicials a partir de les quals s'ha realitzat.

A la part superior de la pàgina es pot accedir a l'arxiu amb la informació de l'arbre (figura 2.6.7). Des de la pàgina inicial del Clustal, s'accedeix a l'apartat "*Phylogenetic tree*", i d'aquí a "*Download Phylogenetic Tree File*", on ens mostra les dades. Després de desar-lo, en qualsevol format (".txt" per exemple), ja se'n poden disposar per a treballar-hi més a fons.

```
(
(
(
Hom_Sapiens_Mith:0.25259,
(
(
(
(
Ara_Thaliana:-0.12221,
Des_Fermentans:0.12221)
:0.08595,
Met_Fervidus:0.16272)
:0.08160,
(
Sac_Cerevisiae:0.16054,
(
Can_Tropicalis:0.08311,
Can_Albigans:0.08571)
:0.06329)
:0.03085)
:0.09378,
Rat_Norvegicus:0.23833)
:0.01185,
Hom_Sapiens:0.19755)
:0.06571)
:0.07956,
(
(
Rat_Norvegicus_Mith:-0.04892,
Met_Extorquens:0.04892)
:0.17226,
Can_Albigans_Mith_Sac_Cerevisiae_Mith:0.13898)
:0.11292)
:0.00955,
Esc_Coli:0.27689,
Ara_Thaliana2:0.26087);
```

Fig. 2.6.7 Captura de les dades mostrades a l'arxiu amb la informació i les distàncies per a formar el posterior arbre filogenètic

6.5 Interpretació dels resultats

Arribar a unes conclusions fonamentades amb els resultats numerals obtinguts (els percentatges de similitud), és complicat. És possible intuir i interpretar les similituds i relacions, però falta una base estadística per a poder fer-ho de forma detallada i fonamentada.

Sempre que alineem un algoritme d'alineació amb un parell de seqüències obtindrem un alineament, fins i tot si les seqüències estan composades per lletres al atzar i no s'assemblen en res. Per tant, a més de fer l'alineament s'ha d'estimar la significació estadística d'aquest mateix.

Si treballem amb seqüències d'ADN és més fàcil fer-se una idea de la qualitat de l'alineament que no pas amb proteïnes. Això és degut al fet de que al valorar positivament només la identitat dels resultats obtinguts, les seqüències per atzar solen ser molt curtes.

En canvi, amb les seqüències de proteïnes és més complex: també es valoren positivament els semblants entre aminoàcids que no són idèntics, però que són químicament similars. Aquest fet contribueix a que els alineaments de les seqüències al atzar es puguin confondre amb els alineaments realment significatius.

Tot i així, casi tots els programes d'alineament solen calcular també els índex que indiquen la significació. S'han d'estudiar aquests índex per a no confondre alineaments espuris amb reals.

Per tant, partirem de les altres possibilitats que ofereix el Clustal Omega per arribar a les conclusions, més visibles i entenedores. Els arbres filogenètics obtinguts ofereixen unes relacions més visuals i entenedores: l'estructura i l'arrelament de l'arbre es crea a partir de les distàncies evolutives estudiades a partir dels alineaments. Com s'ha vist a l'apartat anterior, agafaré les dades fonamentals per a fer els arbres filogenètics, i els traslladaré al programa TreeView per a estudiar-los millor.

7. TREEVIEW

Aprofitaré una de les eines més populars per a fer arbres filogenètics a partir de les dades obtingudes.

El TreeView és un programa per a visualitzar i manipular arbres filogenètics des de l'ordinador. Pot dibuixar arbres arrelats i sense arrelar, mostrar els valors de semblança, i editar arbres mitjançant el trasllat de branques, plegant-les, o tornant-les a arrelar. Per a realitzar-ho, el programa és capaç de llegir i mostrar alguns dels formats per a arbres més utilitzats, com ara "NEXUS", "PHYLIP", "Henning86", "NONA", "MEGA", i "ClustalW/X" (incloent els arxius produïts per programes com el "fasDNAMl" i el "ClustalW". A més, el fet de poder guardar els arbres en aquests tipus d'arxius, permet la conversió entre formats entre ells.

7.1 Arbres filogenètics obtinguts

Amb el Treeview, les opcions per a representar les relacions amb les que treballa entre uns organismes i els altres és més diversa.

Per a fer-ho, agafaré l'arxiu guardat prèviament amb les dades de l'arbre original.

Un cop obert amb el programa, tenim tres modes de visualització del mateix arbre: "Unrooted" (sense arrelar), "Slanted cladogram" (cladograma inclinat), "Rectangular cladogram" (cladograma rectangular) i "Phylogram" (Filograma). Aquest últim no apareixia disponible per a visualitzar-lo amb els meus resultats.

Els arbres resultants són:

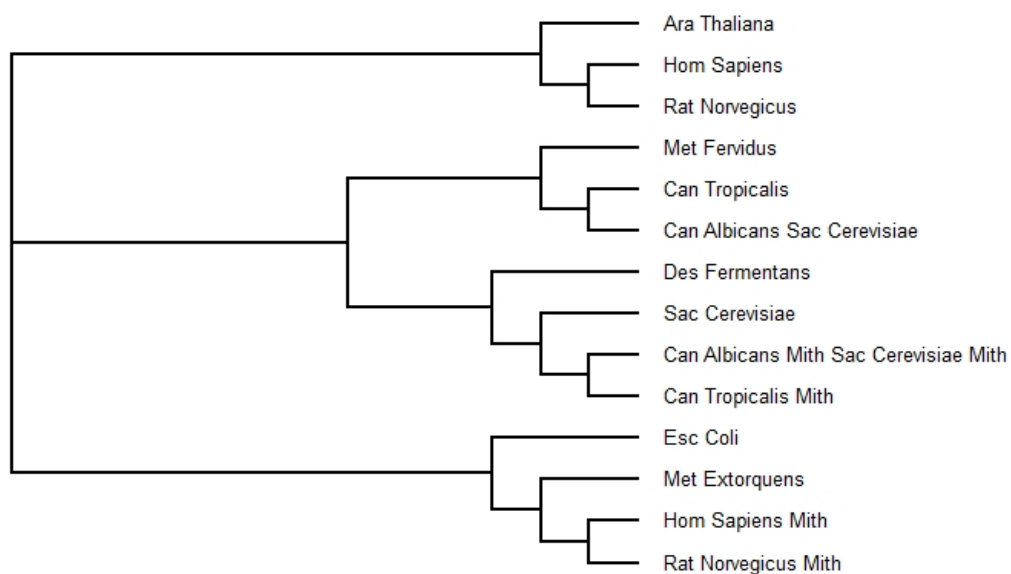


Fig. 2.7.1 Cladograma rectangular del gen metionina-ARNt sintetasa

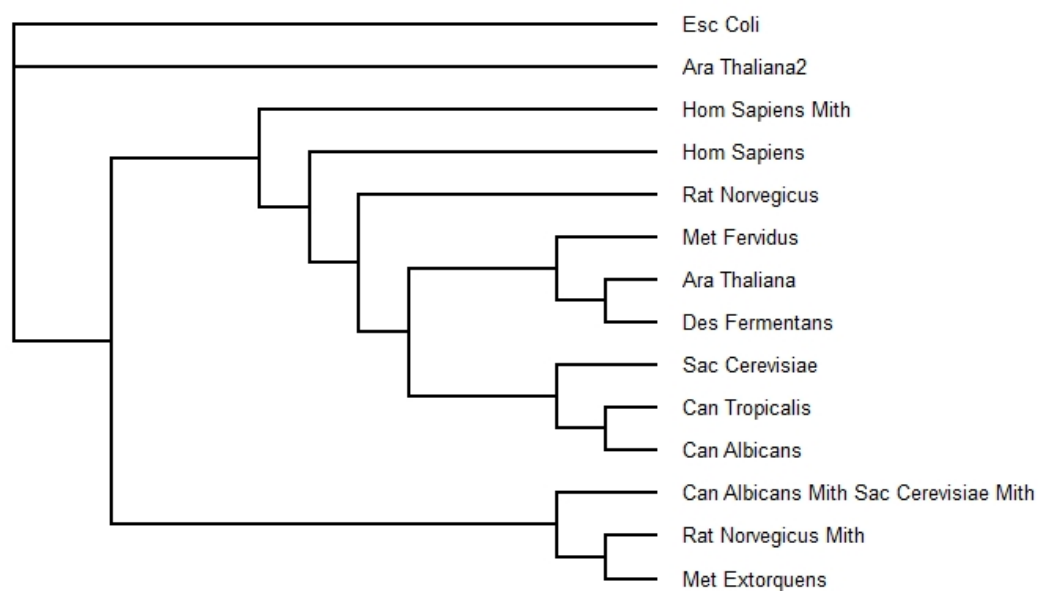


Fig. 2.7.2 Cladograma rectangular del gen tirosina-ARNt sintetasa

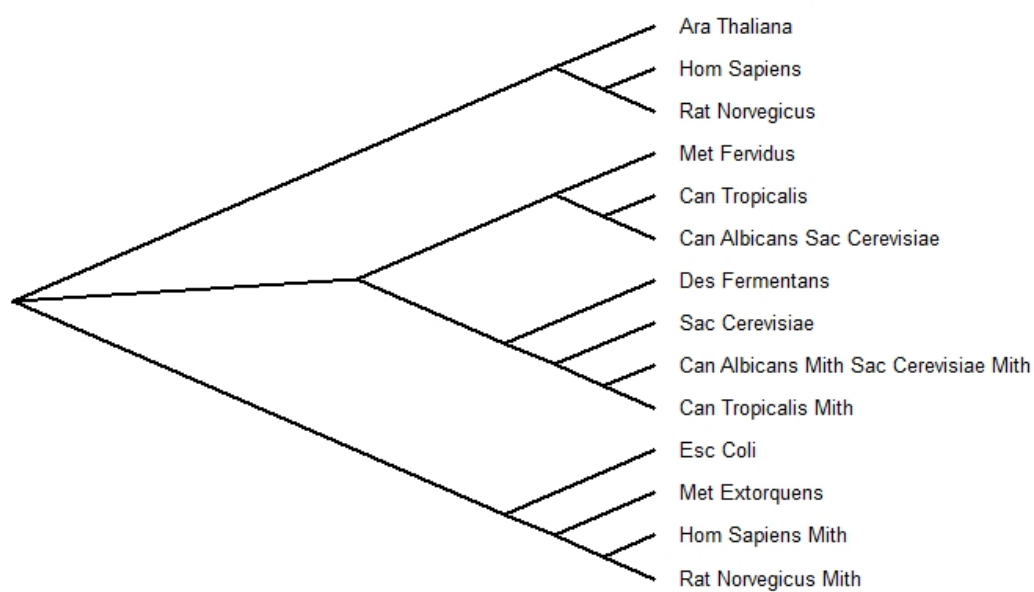


Fig. 2.7.3 Cladograma inclinat del gen metionina-ARNt sintetasa

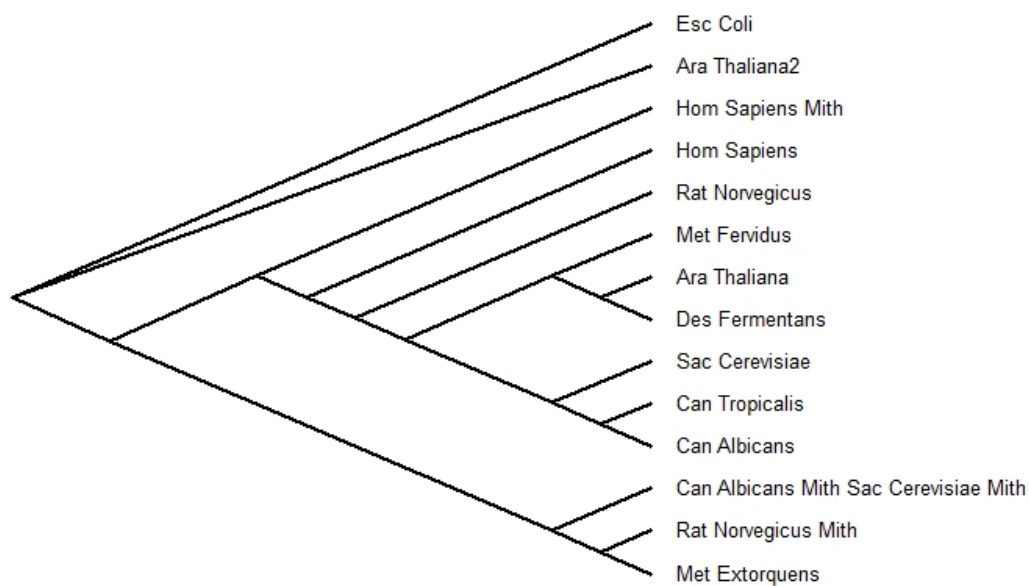


Fig. 2.7.4 Cladograma inclinat del gen tirosina-ARNt sintetasa

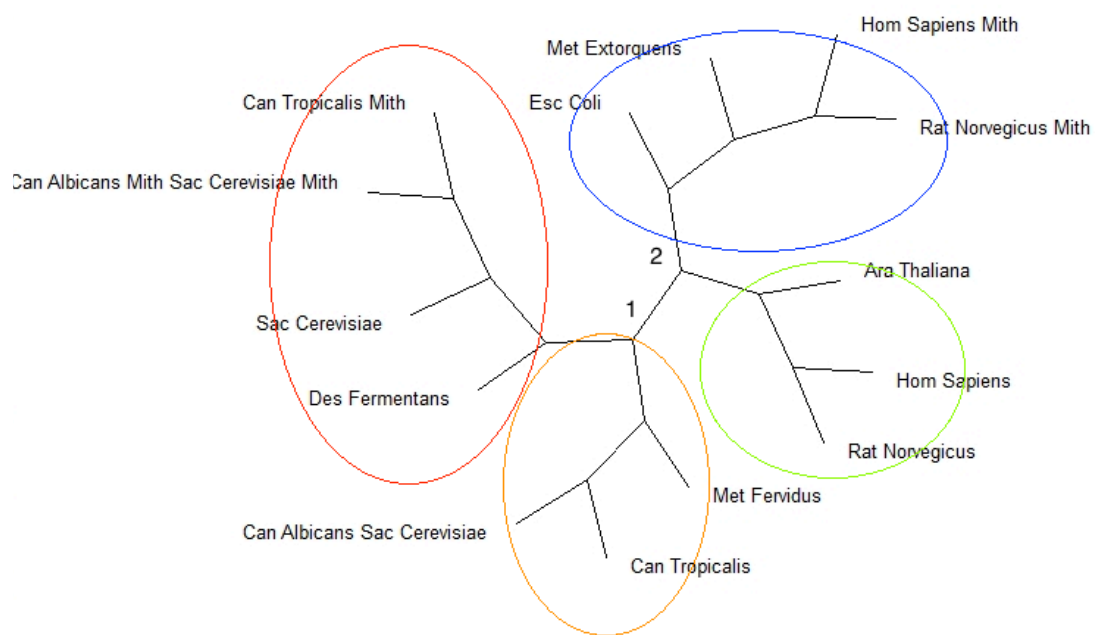


Fig. 2.7.5 Arbre sense arrelar del gen metionina-ARNt sintetasa (amb indicacions descriptives en color)

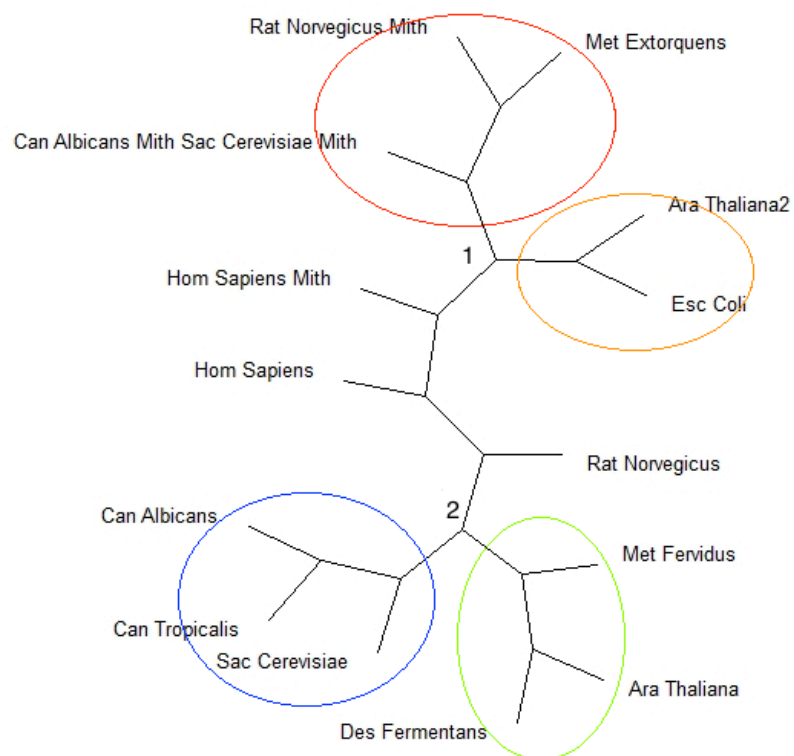


Fig. 2.7.6 Arbre sense arrelar del gen tirosina-ARNt sintetasa (amb indicacions descriptives en color)

7.2 Conclusions: anàlisi dels resultats

Una vegada realitzats els arbres, obtenim les relacions entre els organismes per a cada gen corresponent. L'anàlisi el fonamentaré a partir dels últims arbres, els que no estan del tot arrelats, ja que són més visuals a l'hora de reunir els organismes per la seva similitud i distància entre ells: serà més fàcil entendre les explicacions.

D'altra banda, per a facilitar més l'enteniment de les deduccions, he modificat els arbres originals lleugerament incorporant indicadors: els números 1 i 2 per a les dues divisions principals de l'arbre, i els cercles per a les subdivisions a comentar.

Segons la hipòtesi inicial, hauríem d'obtenir similituds entre:

- ADN Arqueobacteris – ADN Nuclear Eucariotes
- ADN Bacteris – ADN Mitocondrial dels Eucariotes

7.2.1 Resultats per al gen metionina-ARNt sintetasa

Partint de l'últim arbre (figura 2.7.5), el més visual a l'hora de reunir els organismes per similitud, comento els resultats.

Partint del centre de l'arbre, tenim dues zones generals en les que es divideix: l'esquerra (1) i la dreta (2).

A les branques de la primera zona, ens trobem amb dues divisions, la indicada en vermell, i la indicada en taronja.

A la zona vermella, trobem agrupats els dos ADN mitocondrials corresponents als llevats *Candida tropicalis* i *Candida albicans* / *Saccharomyces cerevisiae*⁴¹ (ADN mitocondrial d'Eucariotes) en un extrem. A continuació l'ADN nuclear de *Saccharomyces cerevisiae* (llevat eucariota), prop del seu mitocondrial ja descrit (d'aquí la possible proximitat). De ben aprop, *Desulfurococcus fermentans*, un arqueobacteri.

⁴¹ Durant la cerca al GenBank de l'ADN nuclear i mitocondrial dels dos organismes, la base de dades donava pas a unes seqüències ambigües: les seqüències amb les que treballem en aquest gen per a aquests organismes, són l'ADN nuclear i mitocondrial de *Candida albicans*, tot i que segons els estudis, és similar als corresponent ADN nuclear i mitocondrial en *Saccharomyces cerevisiae*. En el cas del gen Tirosina-ARNt sintetasa, són similars per a l'ADN mitocondrial, però en el cas del nuclear, el GenBank els proporciona diferenciats.

Per tant, en aquesta primera zona vermella de la branca 1, podem dir que hi ha una confirmació:

- L'ADN de l'arqueobacteri (*Desulfurococcus fermentans*) és pròxim a l'ADN Nuclear de l'eucariota (*Saccharomyces cerevisiae*).

En aquesta mateixa zona, els ADN mitocondrials dels llevats són pròxims al nuclear del mateix organisme (més que a l'arqueobacteri, relacional amb l'ADN nuclear). Tot i així, com ja podem veure, són pròxims (no per l'arrel, sinó per l'estructuració de l'arbre), als bacteris de la zona blava (*Escherichia coli* i *Methylobacterium extorquens*). Aquesta relació no és totalment directa, per tant la obviarem.

Pel que fa a la zona taronja, trobem més similituds:

L'ADN de *Candida Albicans* / *Saccharomyces cerevisiae* (ADN nuclear), és pròxim al de *Candida tropicalis* (ADN nuclear), i alhora a *Methanothermus fervidus* (arqueobacteri). Per tant:

- L'ADN de l'arqueobacteri (*Methanothermus fervidus*), és similar als ADN nuclears de *Candida albicans* / *Saccharomyces cerevisiae* i *Candida tropicalis*.

Un cop finalitzat l'anàlisi d'aquesta primera branca 1, ja veiem que aquesta meitat relaciona de manera més propera els ADN arqueobacterians i nuclears entre ells (excepte els dos ADN mitocondrials dels llevats, per una possible proximitat al mateix organisme, que amb la resta: corresponen, per deducció i l'estudi que es realitzarà ara, als ADN bacterians i mitocondrials).

A la zona blava, veiem una inicial proximitat entre els bacteris *Escherichia coli* i *Methanothermus fervidus*: els dos són del mateix domini. Just a continuació, tenen una proximitat important amb dos ADNs mitocondrials: els corresponents a *Homo sapiens* i *Rattus norvegicus*.

Per tant:

- L'ADN dels bacteris (*Escherichia coli* i *Methanothermus fervidus*), és pròxim als mitocondrials corresponents als organismes *Homo sapiens* i *rattus Norvegicus*.

De cara a la zona verda, trobem els ADN nuclears de *Arabidopsis thaliana*, *Homo sapiens* i *Rattus norvegicus*. Els tres són ADN nuclears d'organismes eucariotes, i per tant tenen una proximitat genètica. Per la part de sobre (zona blava), són més pròxims als seus ADN mitocondrials que no pas als ADN bacterians; per la banda de sota (zona taronja), són pròxims als organismes *Methanothermobacter thermautotrophicus* (arqueobacteri), i als següents llevats (*Candida albicans* / *Saccharomyces cerevisiae* i *Candida tropicalis*).

- L'ADN nuclear dels eucariotes és similar als seus ADN mitocondrials, i té una proximitat molt més gran amb els arqueobacteris i altres eucariotes, que amb els bacteris.

Una vegada tractades les relacions, la hipòtesi ha quedat prou comprovada: els ADN nuclears són més pròxims als arqueobacterians que no pas als bacterians (tot i haver-hi una proximitat als mitocondrials, ja que pertanyen al mateix organisme); i els mitocondrials als bacterians, i no pas als arqueobacterians.

7.2.2 Resultats per al gen tirosina-ARNt sintetasa

En el cas del gen Tirosina-ARNt sintetasa, la proximitat és lleugerament diferent. Per tant, la idea inicial de treballar amb més d'un gen ha estat acertada: d'aquesta manera puc veure si per més d'una via, puc arribar als mateixos resultats.

Les indicacions parteixen de dues branques: la superior (1), i la inferior (2). A més, es veu una zona intermèdia on queden els dos ADN d'*Homo sapiens*, i el nuclear de *Rattus norvegicus*.

D'una banda, la branca superior (1), ens mostra dues zones diferenciades (vermell i taronja per a aclarir-ho).

Pel que fa a la zona vermella, es mostra una proximitat entre els dos ADN mitocondrials dels organismes *Candida albicans* / *Saccharomyces Cerevisiae* i *Rattus norvegicus* envers el bacteri *Methylobacterium extorquens*.

- L'ADN del bacteri (*Methylobacterium extorquens*) és proper a l'ADN mitocondrial (*Rattus norvegicus*, *Candida albicans* / *Saccharomyces cerevisiae*).

De cara a la zona taronja, trobem una similitud entre *Arabidopsis thaliana* 2 i *Escherichia coli*. Si deixem de banda *Arabidopsis thaliana*, veiem que:

- L'ADN del bacteri (*Escherichia Coli*), és proper a la zona vermella, corresponent a ADN bacterià i mitocondrial.

L'aparició de l'ADN d'*Arabidopsis thaliana* en aquest punt de l'arbre, pot tenir una explicació: durant la cerca d'aquest gen al GenBank, els resultats mostraven algunes seqüències que diferien lleugerament les unes de les altres (moltes d'aquestes seqüències encara són provisionals). En el cas d'aquest organisme, vaig desar dues seqüències, posant el "2" a la seqüència de la que més dubtava. Tot i així, vaig decidir continuar els alineaments i veure si sortiria amb una similitud a la seva seqüència germana, o si pel contrari, destacaria (de forma possiblement errònia) en la suposada hipòtesi esquemàtica que havia de seguir aquest arbre. A hores d'ara, es veu que aquest ADN nuclear no té res a veure envoltat de mitocondrials i bacterians.

A la zona intermèdia trobem els ADN nuclear corresponents a *Homo sapiens* i *Rattus norvegicus*, i l'ADN mitocondrial d'*Homo sapiens*. En aquest cas, el mitocondrial és més proper a la branca 1, corresponent als altres ADN mitocondrials i bacterians.

- L'ADN mitocondrial (*Homo sapiens*), és més proper als altres ADN mitocondrials i bacterians, que no pas a la branca 2 de l'arbre (on possiblement trobem els gens corresponents a organismes eucariotes i arqueobacteris).
- L'ADN nuclear (*Homo sapiens* i *Rattus norvegicus*), és pròxim al seu mitocondrial (en el cas d'*Homo sapiens*), i alhora té més similitud amb la branca 2 que no pas amb la 1.

D'altra banda tenim la branca inferior (2).

A la zona verda, trobem els organismes *Methanothermus fervidus* (arqueobacteri), *Arabidopsis thaliana* (ADN nuclear eucariota) i *Desulfurococcus fermentans* (arqueobacteri).

- L'ADN dels arqueobacteris (*Methanothermus fervidus* i *Desulfurococcus fermentans*), és similar a l'ADN nuclear (*Arabidopsis thaliana*).

Respecte la zona blava, trobem a els ADN nuclears de *Candida albicans*, *Candida tropicalis*, i *Saccharomyces cerevisiae*. Aquests tres organismes són llevats, per tant eucariotes. Són més propers als organismes de la zona verda (es troben localitzats a la mateixa branca 2), que a la resta d'organismes.

- L'ADN nuclear (*Candida albicans*, *Candida tropicalis* i *Saccharomyces cerevisiae*), és més pròxim a l'altre ADN nuclear i a l'ADN arqueobacterià, que a la resta d'organismes (branca 1: mitocondris i bacteris).

Un cop realitzat l'anàlisi d'aquest segon arbre, veiem que exceptuant el cas d'*Arabidopsis thaliana* 2 (sobre la qual ja s'ha comentat la possible raó de la seva localització "errònia" segons la hipòtesi inicial), l'hipòtesi també és certa: l'ADN nuclear és més proper al dels arqueobacteris que al dels bacteris, i l'ADN mitocondrial és, a la seva vegada, més pròxim al bacterià que a l'arqueobacterià.

CONCLUSIONS

Arribat a aquest punt, la meva hipòtesi ha quedat confirmada.

He trobat unes relacions genètiques entre diferents organismes d'acord amb la teoria endosimbiòtica, la qual cosa confirma la meva hipòtesi inicial.

De totes maneres, aquests resultats no són del tot fiables i exactes: els arbres proporcionen una idea general de les proximitats; hi ha altres mètodes més precisos, com ara concretar les zones de les seqüències a alinear, altres mètodes d'alineament, o estadístiques evolutives que em donarien més informació.

Una de les apreciacions que he trobat, ha estat sobre la vigència de l'estudi de les relacions evolutives a partir de l'ADN. En la majoria dels casos, aquestes no parteixen de l'aminoacil-ARNt sintetasa (tot i la seva importància), sinó a partir d'unes altres pautes. Un dels recursos més establerts en aquest tipus d'investigacions són a partir de l'estudi dels ribosomes 16S (un component de la subunitat petita dels ribosomes 30S procariotes).

Tot i haver fet un treball molt exhaustiu, me no n'adono de que la meva feina no acaba aquí: podia haver continuat el meu treball d'investigació indeterminadament. A mesura que el coneixement sobre la mecànica i tema del treball augmenten, les possibilitats del treball es veuen incrementades.

Una de les moltes vies que m'hagués agradat estudiar, si hagués disposat del temps necessari, hagués estat comparar l'altra part de la teoria endosimbiòtica. Partint d'aquesta idea hagués realitzat les comparacions entre cianobacteris i cloroplasts, de la mateixa manera que durant el treball he tractat la similitud entre els eubacteris i els mitocondris. La biologia és una ciència que abraça moltes branques, i tot dóna per més del que pots arribar a imaginar en un principi. El treball realitzat en sí podia haver tingut moltíssims enfocaments diferents, i ampliacions desmesurades. Al cap i a la fi, la biologia estudia el "què" de la nostra existència: la vida.

L'experiència personal de cara al treball ha estat molt satisfactòria: davant la idea inicial del projecte, era molt incrèdula. No veia com partint des de zero, podia arribar a trobar uns resultats coherents i sense errors: era un món força complex i aliè als meus coneixements, sobre el que veia molt relativisme i moltes possibilitats. No era una ciència exacta: una seqüència genòmica seleccionada incorrectament, o qualsevol imprevist, el veia com a un detonant del fracàs del treball. Aquest fet es

veu reflectit a l'ini de la meva experiència personal abans de realitzar el treball.

Des del principi, la cerca de les seqüències va quedar establerta (junt als organismes i gens específics), i la meva por era arribar al final del treball, i topar-me amb uns resultats pèssims després de tanta feina. No entenia que una hipòtesi rebutjada també era un èxit: mentre estigui ben justificat el resultat (tant bo com dolent), la feina és bona.

Per tant, amb aquesta mentalitat, vaig intentar obtenir els resultats abans de realitzar tota la memòria de treball, per a poder canviar la temàtica a temps si el meu sacrifici a la cerca de seqüències resultava un fracàs en els alineaments.

Tot i així, els alineaments que vaig improvisar en aquell moment (amb pocs organismes), els vaig intuir amb els percentatges del Clustal. Em trobava amb unes carències estadístiques per a interpretar els resultats, ja que tot i ser aproximats, eren similars tots entre ells. Va ser aquí quan vaig deixar-me emportar, i esperar al final per a veure els resultats: anava a cegues, i no tenia ni idea de com podia acabar la meva feina.

Durant el transcurs del treball, encara amb una mentalitat escèptica, vaig descobrir la vigència dels ribosomes 16S, abans comentada, en les relacions evolutives d'aquest tipus. Aquest fet em va fer replantejar fins a quin punt el meu treball podia funcionar, fins que el conèixer la importància evolutiva dels aminoacil-ARNt sintetasa em va retornar l'esperança.

I és aquí quan em vaig adonar de la transcendència de la hipòtesi al treball de recerca: no es parteix d'una idea ni de resultats establerts.

Des del moment on la meva implicació de cara al treball va ser total, fins a hores d'ara, vaig comprendre que la gràcia de tota aquesta feina, és realitzar-la en tot moment amb les esperances de confirmar la teva hipòtesi, tot i comptar amb la possibilitat d'un final que la rebutgi. L'empenta que et fa continuar endavant i conduir aquest treball, absolutament teu, hi és fins al final. Una vegada empreses les expectatives i intrigues sobre com acabarà, veus que aquest treball és teu, i comença a prendre forma a les teves mans. Al cap i a la fi, és la teva hipòtesi, la teva recerca, i els teus resultats.

La satisfacció un cop es veu tota la feina realitzada, a partir de la teva idea inicial fins als resultats, tot producte de l'esforç, és indescriptible.

A partir de la realització del treball, també m'emporto altres experiències satisfactòries, com ara l'ampliació dels meus coneixements tant en biologia general, i la introducció en el món de la bioinformàtica. Pel que fa a aquest últim, he passat del complet desconeixement, a tenir unes bases i nocions suficients com per a poder realitzar un treball d'aquestes característiques. Cal comentar que no és ni remotament el mateix dedicar temps a una cosa que no t'interessa, a entregar-t'hi fins al punt en el que no es fa pesat treballar-hi: els conceptes apresos ja no es veuen com a una obligació a nivell educatiu, sinó com a una via per a realitzar un objectiu. D'aquesta manera, les bases vistes no "s'aprenen i es desprenen", sinó que s'assimilen: la temàtica agrada (al cap i a la fi la temàtica del treball és personal, i s'agafa en base a les preferències personals de cadascú), i la retenció dels conceptes és molt millor si no es veu com a una obligació momentània.

"La feina ben feta ni té fronteres, ni té rival".

BIBLIOGRAFIA

TEXTOS

ARIAS, M., BARRACHINA, J., CLOSAS, M. C., FERRER, R. *Biologia 1*. 2a edició. Barcelona: Castellnou Edicions, juny 2011.

ARIAS, M., BARRACHINA, J., CLOSAS, M. C., FERRER, R. *Biologia 2*. 1a edició. Barcelona: Castellnou Edicions, 2009.

JIMENO, A., BALLESTEROS, M. *Biologia 2*. Barcelona: Grup Promotor Santillana, 2009.

MARGULIS, L. *El origen de la cèl·lula* (títol original: *Early Life*). Barcelona: Editorial Reverté, S. A., 1986

VALLÈS, F. *Conceptes bàsics de Biologia* [dossier]

VALLÈS, F. *Arbres filogenètics 2012* [dossier]

VALLÈS, F. *Genètica 2012 (1a part)* [dossier]

WEBGRAFIA

Teories Eucariogèniques

<http://es.wikipedia.org/wiki/Eucariogénesis>

<http://jvilchez2009.blogspot.com.es/2009/04/teoria-de-la-endosimbiosis.html>

<http://www.biologia.edu.ar/botanica/tema8/8-3organulo.htm>

http://ca.wikipedia.org/wiki/Teoria_endosimbiòtica

<http://elblogdeferalfonso.blogspot.com.es/2012/03/teoriaendosimbiotica-lynn-margulis-la.html>

<http://blogcienciasdelmundocontemporaneo.blogspot.com.es/2012/05/la-teoria-endosimbiotica.html>

<http://jvilchez2009.blogspot.com.es/2009/04/teoria-de-la-endosimbiosis.html>

<http://cienciasalmunia.wordpress.com/2010/02/23/teoria-de-la-endosimbiosis/>

http://es.wikipedia.org/wiki/Endosimbiosis_seriada

http://scielo.isciii.es/scielo.php?pid=s0210-56912007000400004&script=sci_arttext

Conceptes

<http://es.wikipedia.org/wiki/Svedberg>

<http://ca.wikipedia.org/wiki/Peroxisoma>

Organismes

http://es.wikipedia.org/wiki/Archaea#Un_nuevo_dominio

[http://phobos.xtec.cat/ieselcastell/intranet/documents/treballs_recerca/tr2/Copia%20\(4\)%20de%20Index.htm](http://phobos.xtec.cat/ieselcastell/intranet/documents/treballs_recerca/tr2/Copia%20(4)%20de%20Index.htm)

<http://www.cienciesnaturals.com/biologia/tema3.html>

<http://es.wikipedia.org/wiki/Eukaryota>

<http://en.wikipedia.org/wiki/Eukaryote>

<http://ca.wikipedia.org/wiki/Eucariota>

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3458677/>

<http://www.jgi.doe.gov/sequencing/why/desulfurococcus.html>

<http://ca.wikipedia.org/wiki/Extremof%C3%ADlia>

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3035299/>

<http://bacteriasyfuentes hidrotermales.jimdo.com/microorganismos/hipertermófilos/>

http://en.wikipedia.org/wiki/Escherichia_coli

<http://www.ncbi.nlm.nih.gov/genome/?term=escherichia%20coli>

<http://www.ncbi.nlm.nih.gov/genome/?term=Methylobacterium%20Extorquens%20CM4>

http://www.ncbi.nlm.nih.gov/genome/693?project_id=58933

http://en.wikipedia.org/wiki/Candida_tropicalis

<http://es.wikipedia.org/wiki/Candida>

<http://www.ncbi.nlm.nih.gov/genome/?term=Candida%20Tropicalis>

<http://www.ncbi.nlm.nih.gov/genome/?term=candida%20albicans>

http://ca.wikipedia.org/wiki/Candida_albicans

<http://en.wikipedia.org/wiki/Karyotype>

<http://www.ncbi.nlm.nih.gov/genome/?term=Saccharomyces%20cerevisiae>

http://ca.wikipedia.org/wiki/Saccharomyces_cerevisiae

<http://www.ncbi.nlm.nih.gov/genome/?term=Arabidopsis%20thaliana>

<http://www.kew.org/plants-fungi/Arabidopsis-thaliana.htm>

<http://en.wikipedia.org/wiki/Human>

<http://www.ncbi.nlm.nih.gov/genome/?term=Rattus%20norvegicus>

http://es.wikipedia.org/wiki/Rattus_norvegicus

http://ca.wikipedia.org/wiki/Rata_comuna

Aminoacil-ARNt sintetasa

<http://www.unav.es/ocw/genetica/tema-1-5.html>

http://es.wikipedia.org/wiki/Bios%C3%ADntesis_proteica

<http://medmol.es/glosario/49/>

http://centrodeartigos.com/articulos-de-todos-los-temas/article_35966.html

<http://bioinformatica.upf.edu/2002/projects/1.6/Catala.htm>

<http://www.biochem.ucl.ac.uk/bsm/xtal/teach/trna/trna.html>

http://en.wikipedia.org/wiki/Aminoacyl_tRNA_synthetases,_class_II#cite_note-PUB00006477-2

<http://www.ncbi.nlm.nih.gov/pubmed/10704480>

<http://mmbr.asm.org/content/64/1/202.full>

<http://www.ncbi.nlm.nih.gov/pubmed/7783229>

Bioinformàtica

http://www.insdc.org/files/feature_table.html#1

http://es.wikipedia.org/wiki/%C3%81cido_desoxirribonucleico#Bioinform.C3.A1tica

<http://en.wikipedia.org/wiki/GenBank>

<http://www.ncbi.nlm.nih.gov/genbank/>

http://es.wikipedia.org/wiki/Formato_FASTA

<http://personales.upv.es/jcanizar/bioinformatica/alineamientos.html>

<http://biologia.uab.es/biocomputacio/tutorial/sessio6/tutorialdnamit.htm#B>

<http://bioinf.ibun.unal.edu.co/documentos/multAlignmentCBIB.pdf>

http://es.wikipedia.org/wiki/Alineamiento_de_secuencias

<http://www.ebi.ac.uk/Tools/msa/clustalo/help/>

<http://perso.wanadoo.es/jjdeharo/sistematica/glosario/glosario.htm>

<http://evolution.genetics.washington.edu/phylip/software.html>

<http://personales.upv.es/jcanizar/bioinformatica/alineamientos.html>

<http://evolution.genetics.washington.edu/phylip/software.etc2.html#TreeView>

<http://akira.ruc.dk/~olesk/sekvens/Treedraw.htm>

http://www.megasoftware.net/WebHelp/part_iv___evolutionary_analysis/constructing__phylogenetic_trees/statistical_tests_of_a_tree_obtained/bootstrap_tests/hc_bootstrap_test_phylogeny.htm

RECURSOS UTILITZATS

Programari i bases de dades

GenBank

<http://www.ncbi.nlm.nih.gov/genbank/>

Clustal Omega

<http://www.ebi.ac.uk/Tools/msa/clustalo/>

http://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=clustalw2_phylogeny-I20131017-202238-0351-55099910-oy&analysis=tree

TreeView

<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>

Diccionaris i consultes lingüístiques

<http://www.diccionaris.cat>

<http://www.wordreference.com>

IMATGES

Figura 1.1.1

http://4.bp.blogspot.com/-X4n1t7SvdnY/TZTQ1weCRsI/AAAAAAAAAY/hJj_oAy4JYs/s1600/procariota.jpg

Figura 1.1.2

http://recursostic.educacion.es/ciencias/biosfera/web/alumno/1bachillerato/organizacion_sv/imagenes/celula%20animal_letreros.jpg

Figura 1.2.1

http://www.geocities.ws/batxillerat_biologia/celori12.jpg

Figura 1.2.2

http://upload.wikimedia.org/wikipedia/commons/thumb/3/3e/Animal_mitochondrion_diagram_es.svg/572px-Animal_mitochondrion_diagram_es.svg.png

Figura 1.2.3

<http://linux.ajusco.upn.mx/fotosintesis/img/cloroplasto.jpg>

Figura 2.2.1

<http://www.virtual.unal.edu.co/cursos/ingenieria/2001832/images/nucleotide.gif>

Figura 2.2.2

http://2.bp.blogspot.com/_bUX0IsD5SF0/TL5mRtW5A8I/AAAAAAAAB5Q/oOCov86zJio/s1600/adn.gif

Figura 2.3.1

http://l2.yimg.com/bt/api/res/1.2/ceNdbkNsL_HGHI.tcfVxww--/YXBwaWQ9eW5ld3M7cT04NQ--/http://mit.zenfs.com/1111/2013/08/Tree_of_life_it.jpg

Figura 2.3.2

<http://upload.wikimedia.org/wikipedia/commons/thumb/a/a1/Halobacteria.jpg/230px-Halobacteria.jpg>

Figura 2.3.3

http://www.euroxpress.es/img/2011/05/ecoli_0104.jpg

Figura 2.3.4

<http://carmelourso.files.wordpress.com/2011/08/candida-albicans.jpg>

Figura 2.4.1

<http://www.monografias.com/trabajos27/aminoacil-sintetasas/amin1.jpg>

Figura 2.5.1

http://en.wikipedia.org/wiki/File:Growth_of_Genbank.svg

Figura 2.5.2

Captura d'imatge de la web:

<http://www.ncbi.nlm.nih.gov/gene/?term=Methanothermus+fervidus+tyrosyl-trna+synthetase>

Figura 2.5.3

Captura d'imatge de la seqüència obtinguda a:

http://www.ncbi.nlm.nih.gov/nuccore/NC_014658.1?report=fasta&from=451123&to=452378

Figura 2.6.1

Captura d'imatge del Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) amb les seqüències que treballem

Figures 2.6.2, 2.6.3, 2.6.4, 2.6.5, 2.6.6 i 2.6.7

Captures pròpies dels resultats dels alineaments realitzats amb el Clustal Omega

Figures 2.7.1, 2.7.2, 2.7.3, 2.7.4, 2.7.5, 2.7.6

Captures pròpies dels resultats dels arbres introduïts al TreeView per als diferents alineaments realitzats