

Protists, la principal font de diversitat genòmica en eucariotes

Ramon Massana,¹ Ramiro Logares,¹ David López-Escardó¹ i Javier del Campo²

¹ Institut de Ciències del Mar (CSIC)

² Institut de Biologia Evolutiva (CSIC - Universitat Pompeu Fabra)

Correspondència: Ramon Massana. Institut de Ciències del Mar (CSIC). Passeig Marítim de la Barceloneta, 37-49. 08003 Barcelona. Tel.: +34 932 309 500. Adreça electrònica: ramonm@icm.csic.es.

DOI: 10.2436/20.1501.02.216

ISSN (ed. impresa): 0212-3037

ISSN (ed. digital): 2013-9802

<http://revistes.iec.cat/index.php/TSCB>

Rebut: 24/03/2022

Acceptat: 17/04/2022

Resum

La genòmica, la determinació de la seqüència de DNA d'una espècie, avui dia està en expansió gràcies a noves eines de seqüenciament. Les iniciatives per a obtenir els genomes de totes les espècies vives sovint oblidem els protists, eucariotes unicel·lulars cabdals per a entendre l'evolució de la vida i el funcionament dels ecosistemes. En aquest treball repassarem el que se sap de genòmica de protists, sovint a partir d'organismes cultivats, i presentarem com es pot obtenir el genoma d'espècies no cultivades que conformen una part enorme de la biodiversitat. Després explicarem exemples de com la genòmica de protists permet abordar qüestions com l'origen del llinatge eucariota, l'adquisició i expansió del cloroplast, la multicel·lularitat, els processos de simbiosi, l'estudi d'activitats biogeoquímiques a partir de gens funcionals o la genètica de poblacions. Aquesta contribució vol donar a conèixer l'extraordinària diversitat de protists que hi ha i promoure la consideració que mereixen en els estudis genòmics.

Paraules clau: biodiversitat, ecologia, evolució, genòmica, protists.

Introducció

Un dels grans misteris desxifrats al segle passat va ser el mecanisme de manteniment i transmissió de la informació que codifica els éssers vius. El dogma central de la biologia estableix que la informació està escrita en la seqüència (ordre) de les quatre bases que formen el DNA, i que les unitats d'aquesta informació (els gens) es tradueixen a proteïnes (fent servir RNA com a missatger), les quals formen les estructures cel·lulars i controlen les reaccions químiques de les rutes metabòliques. En aquest context, esbrinar la seqüència del DNA d'una espècie podrà donar respostes a la seva ecofisiologia i història evolutiva. En contrast amb la genètica, que estudia el paper de gens concrets, la genòmica pretén caracteritzar conjuntament tots els gens de l'espècie, així com les regions no codificants, la seva posició relativa i la seva estructura tridimensional. En procariotes, el DNA acostuma a trobar-se en una sola cadena circular, un sol cromosoma, mentre que en eucariotes hi sol haver un nombre concret de cromosomes lineals embotellats dins del nucli que varia segons l'espècie.

La genòmica tractaria, doncs, d'obtenir la seqüència completa del DNA de tots els cromosomes i la seva interpretació funcional.

Vist el gran potencial científic de la informació genòmica en pràcticament tots els camps de la biologia (evolució, ecologia, biomedicina, biotecnologia), hi ha hagut moltes iniciatives per a obtenir els genomes de les espècies de la biosfera. Aquesta informació serà fonamental per a esbrinar la història natural i els patrons evolutius de cada espècie, permetrà conèixer-ne les potencialitats genètiques i, en alguns casos, identificar mutacions en gens concrets responsables de disfuncions o malalties. Aquest esforç és encara més crític en la situació actual de canvi climàtic, que podria acabar derivant en la sisena extinció massiva degut a les activitats humanes. És en aquest context en què s'emmarca l'Earth BioGenome Project (EBP), començat el 2018 i que té l'ambició de seqüenciar, catalogar i caracteritzar el genoma de totes les espècies eucariotes existents en un termini de deu anys (Lewin *et al.*, 2022). El projecte està estructurat en diverses iniciatives, i els autors d'aquest treball en són

Protists, the primary source of genomic diversity in eukaryotes

Abstract

Genomics, which unravels the DNA sequence of species, is an expanding discipline thanks to new sequencing tools. Initiatives aimed at obtaining the genomes of all living species often overlook protists, which are unicellular eukaryotes that are crucial for understanding the evolution of life and the functioning of ecosystems. In this paper we review what is known about protist genomics, often on the basis of cultured species, and present new ways to obtain the genomes of uncultured species, which represent a large part of eukaryotic diversity. We then present some examples of how protist genomics is crucial to address key issues in evolution and ecology, such as the origin of the eukaryotes, the acquisition and expansion of chloroplasts, the emergence of multicellularity, symbiotic events, the study of biogeochemical activities on the basis of functional genes, and population genetics. This paper seeks to highlight the vast diversity of protists and to underscore the fact that they should be given greater consideration in genomic studies.

Keywords: biodiversity, ecology, evolution, genomics, protists.

part d'una: la iniciativa catalana per a l'EBP (CBP, de l'anglès Catalan Initiative for the Earth BioGenome Project, <https://www.biogenoma.cat>). Tot i que l'EBP proposa caracteritzar exhaustivament la biodiversitat d'eucariotes, és evident que fa un èmfasi especial en animals, plantes i fongs, i que posa poca atenció en els eucariotes unicel·lulars, els protists. Enlloc no es diu que no siguin importants, però fàcilment s'obvien. Gran part de la comunitat implicada en aquests projectes genòmics passa per alt que els protists representen la major part de la diversitat eucariota i que són clau per a entendre algunes de les principals transicions evolutives, com l'origen de la fotosíntesi, que permetrà l'aparició de les plantes superiors, o l'origen de la multicel·lularitat, clau per a l'aparició d'animals i plantes. Així, aquests organismes microscòpics i estructuralment senzills queden desenfocats del lloc que els pertoca en l'evolució de la vida, el funcionament dels ecosistemes i la seva enorme contribució a la biodiversitat. Amb aquest article volem posar una mica de llum a la foscor, i explicar en quina situació es troben els eucariotes unicel·lulars.

Projectes genòmics, de virus a tota la biodiversitat

La capacitat d'obtenir dades genòmiques ha anat lligada al desenvolupament de mètodes per a seqüenciar el DNA. Amb l'aparició de la tècnica de Sanger, el 1977 es va obtenir el primer genoma, el del bacteriòfag ϕ X174 (5,4 kb). Així es va iniciar la cursa per a dur a terme projectes cada cop més complexos i ambiciosos: el genoma del mitocondri humà el 1981 (16,6 kb), el de cloroplasts de plantes el 1986 (155,8 kb), un cromosoma de fong el 1992 (315 kb) i, finalment, els primers genomes d'un organisme, el bacteri *Haemophilus influenzae* (1,8 Mb) el 1995 i el llevat *Saccharomyces cerevisiae* el 1996 (12,1 Mb). Poc després, la comunitat científica es va plantejar seriosament el Projecte Genoma Humà, que va esdevenir una altra cursa que va acabar amb la publicació d'una primera versió incompleta el 2001, una versió acabada el 2003 i la versió final corregida (menys d'un error en 20 kb) el 2007.

Paral·lelament, es van anar triant espècies concretes que servien com a model per a representar grans grups d'organismes. Dins dels mamífers es van completar els genomes del gos (*Canis familiaris*), la rata (*Rattus norvegicus*), el ratolí (*Mus musculus*) i el ximpanzé (*Pan troglodytes*), tots utilitzats en recerca biomèdica. Respecte d'altres animals emprats com a organismes model, el primer genoma obtingut va ser el del nematode *Caenorhabditis elegans*, després el de la mosca del vinagre *Drosophila melanogaster* i, finalment, el genoma del peix zebra, *Danio rerio*. Com a model de planta amb flor, es va triar *Arabidopsis thaliana*, i de fong, el llevat *Saccharomyces cerevisiae*. Amb relació a protists, els genomes complets disponibles estan esbiaixats cap a espècies d'interès biomèdic, com l'agent causant de la malària, l'apicomplex *Plasmodium falciparum*, o el responsable de la leishmaniosi, el cinetoplastid *Leishmania major*. L'objectiu dels projectes genòmics amb espècies model era determinar el nombre de cromosomes, seqüenciar-los de punta a punta, identificar-ne els gens i predir-ne la funció. Donat que cada individu de la mateixa població té genomes lleugerament diferents, la seqüència de referència per a l'espècie és la del component haploide d'un sol individu, i aquesta referència serveix per a cartografiar la variabilitat intrínseca dins la població.

Un salt crític en la història dels projectes genoma ha estat l'aparició de les tècniques de seqüenciació massiva (Reuter *et al.*, 2015). El rendiment en lectures de DNA d'aquestes noves màquines ha anat evolucionant amb el temps, inicialment produint fragments curts (primer la

tecnologia 454 i avui dia Illumina), i més endavant, fragments llargs (tecnologies com PacBio o ONT). Aquest increment espectacular de rendiment ha anat acompanyat d'avenços igualment trepidants en el maquinari computacional, amb ordinadors cada cop més potents que permeten emmagatzemar i processar l'allau d'informació generada, i en les eines bioinformàtiques per a gestionar i analitzar les dades. La seqüenciació massiva s'ha aplicat ràpidament a moltes disciplines, també per a obtenir el genoma d'una gran varietat d'espècies. Tot i que sovint ha anat acompanyada d'una qualitat inferior de les seqüències i de menys ambició a l'hora de finalitzar els projectes genòmics, és innegable que la seqüenciació massiva ha promogut una explosió en el camp de la genòmica. A més, ha permès avançar en el coneixement dels microorganismes dominants als ecosistemes, molts dels quals encara no cultivats. És en aquest nou escenari de capacitat enorme de seqüenciació que s'emmarca l'Earth BioGenome Project.

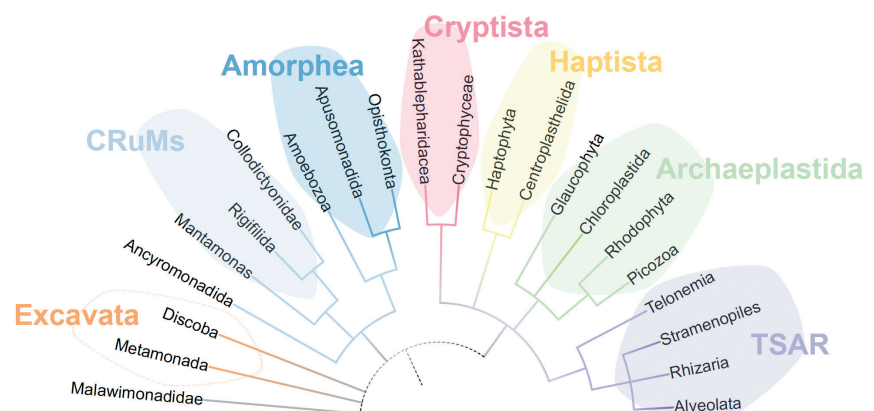
Els protists, els grans oblidats

Diversitat dels protists i la seva importància ecològica

L'obtenció de dades genòmiques d'un ampli espectre d'espècies de macro- i microorganismes, juntament amb la millora d'algorismes de reconstrucció filogenètica, han permès arribar a una visió realista de la biodiversitat d'eucariotes (Burki *et al.*, 2020). L'arbre resultant mostra com els animals, els fongs i les plantes formen una petitíssima part dels llinatges existents (vegeu la figura 1) i evidencia la descompensació en el coneixement que tenim dels protists respecte a la triada multicel·lular, que representa al vol-

tant del 95 % de les espècies descrites (Pawlowski *et al.*, 2012). Aquest biaix té arrels històriques i metodològiques. Des d'abans dels treballs de Darwin, els protists eren un calaix de sastre on es posaven tots els eucariotes unicel·lulars, equiparats com a regne al mateix nivell que animals, fongs i plantes, i obviant que darrere aquesta simplicitat aparent hi havia diferents orígens evolutius i dissenys cel·lulars. La informació genòmica ha permès caracteritzar aquesta gran diversitat i fer-li justícia. A escala evolutiva, és clar que l'últim ancestre comú de tots els eucariotes va ser un organisme unicel·lular i que durant una gran part del temps geològic només hi havia protists, que anaven diferenciant-se i ocupant nous nínxols ecològics. Vista aquesta presència seminal, no és estrany que siguin molt importants en tots els ecosistemes i en els cicles biogeoquímics com a productors primaris, depredadors, paràsits o descomponedors. Conèixer, doncs, l'enorme diversitat de protists és fonamental per a entendre la diversitat eucariota i el funcionament de la biosfera.

L'arbre de la vida eucariota està estructurat en uns quants supergrups, cadascun dels quals inclou força llinatges rellevants i multitud d'espècies. Un dels supergrups més importants i diversos és el dels TSAR, format per telonèmids, estramenòpils, alveolats i rizaris (vegeu la figura 1). Tots ells contenen grups de flagel·lats heterotròfics que tenen impacte en els ecosistemes com a depredadors de bacteris i eucariotes microbians. A més, també s'hi troben llinatges fotosintètics amb una gran rellevància ecològica, com les diatomees dins dels estramenòpils i els dinoflagel·lats dins dels alveolats, o importants paràsits, com els apicomplexos. Un segon supergrup és el dels Amorphea, que in-



† Figura 1. Esquema de l'arbre de la vida eucariota en forma de supergrups i els seus principals components. L'arbre mostra com els animals i fongs (dintre d'Opisthokonta) i les plantes (dintre de Chloroplastida) representen una petita fracció del disseny evolutiu dels eucariotes. Elaboració pròpia a partir de Burki *et al.* (2020).

clou els opistoconts, amb animals, fongs i els seus parents unicel·lulars, i els amebozous, que engloben la majoria d'amebes. Altres supergrups són els arqueplàstids, que contenen les plantes i les algues vermelles i verdes, els Cryptista i Haptista, que, al seu torn, inclouen algues unicel·lulars importants a escala ecològica i geològica, i els excavats, que contenen majoritàriament protists heterotròfics, molts dels quals paràsits. Finalment, hi ha una sèrie de llinatges orfes que tenen una posició filogenètica poc clara a l'arbre, generalment formats per poques espècies molt distants entre si.

Accés als genomes de protists i cobertura dins de l'arbre de la vida

El coneixement de la gran biodiversitat d'eucariotes microbians, present en cultius de laboratori, en situacions de parasitisme o en estudis de biodiversitat ambiental, no ha anat acompanyat d'esforços genòmics equivalents. Això es reflecteix en les llistes de genomes disponibles, on estan clarament subrepresentats, i on hi ha un biaix claríssim cap a espècies d'interès econòmic o mèdic. Un bon lloc de referència amb genomes de bona qualitat és la plataforma ENSEMBL (www.ensembl.org), inicialment plantejada per a genomes de vertebrats, però que s'ha ampliat per a incloure tots els éssers vius. En aquest repositori, hi trobem 311 genomes d'espècies de vertebrats, 123 d'altres animals, 1.506 de fongs, 119 de plantes i 237 de protists. Dins d'aquests darrers, el 75 % són de patògens humans (100 genomes d'apicomplexos i 25 de cinetoplàstids) o de plantes (53 d'oomicets). Tot i així, ENSEMBL té també representants d'un ampli ventall de grups de protists, i recull aspectes cabdals de les mides de genoma, nombre de cromosomes i nombre de gens (vegeu la figura 2). Amb tot, és evident que la representació de la veritable biodiversitat d'eucariotes, sobretot els microbians, encara queda lluny.

En els darrers anys hi ha hagut iniciatives que han intentat omplir aquest buit de coneixement genòmic. Una de molt rellevant és el projecte MMETSP, The Marine Microbial Eukaryotic Transcriptome Sequencing Project (Keeling *et al.*, 2014), que va fer una crida a la comunitat científica internacional per aconseguir el transcriptoma de tants protists marins cultivats com fos possible. D'aquesta manera, es va afegir informació de més de tres-centes espècies d'un gran ventall de grups marins. Aquesta iniciativa no resol·lia, però, un dels grans reptes que va aparèixer en els estudis ambientals: la gran presència d'espècies no cultivades en les comunitats naturals. En els darrers anys s'han

Espècie	Grup taxonòmic	Mida del genoma (Mb)	Gens codificants	Nombre de cromosomes	Scaffolds	Interès	Referència
<i>Plasmodium falciparum</i>	Apicomplexa (Alveolata)	23,3	5.362	14	16	Paràsit de la malària	Gardner <i>et al.</i> , 2002
<i>Dictyostelium discoideum</i>	Mycetozoa (Amoebozoa)	34,1	13.243	6	39	Ameba social	Eichinger <i>et al.</i> , 2005
<i>Chlamydomonas reinhardtii</i>	Chlorophyta (Archaeplastida)	111,1	17.743	17	53	Model alga verda	Merchant <i>et al.</i> , 2007
<i>Guillardia theta</i>	Cryptophyta (Cryptista)	87,1	24.945	3	669	Presència de nucleomorf	Curtis <i>et al.</i> , 2012
<i>Leishmania major</i>	Discoba (Excavata)	32,9	8.315	36	36	Paràsit de la leishmaniosi	Ivens <i>et al.</i> , 2005
<i>Emiliana huxleyi</i>	Haptophyta (Haptista)	167,7	38.544	–	7795	Alga coccolitòfor	Read <i>et al.</i> , 2013
<i>Monosiga brevicollis</i>	Choanoflagellata (Opisthokonta)	41,6	9.172	–	218	Parent unicel·lular d'animals	King <i>et al.</i> , 2008
<i>Nannochloropsis gaditana</i>	Eustigmatophyceae (Stramenopile)	27,6	10.496	30	684	Bioteconologia, biofuel	Carpinelli <i>et al.</i> , 2014
<i>Phytophthora infestans</i>	Oomycetes (Stramenopile)	228,5	17.785	–	4921	Paràsit de plantes	Haas <i>et al.</i> , 2009

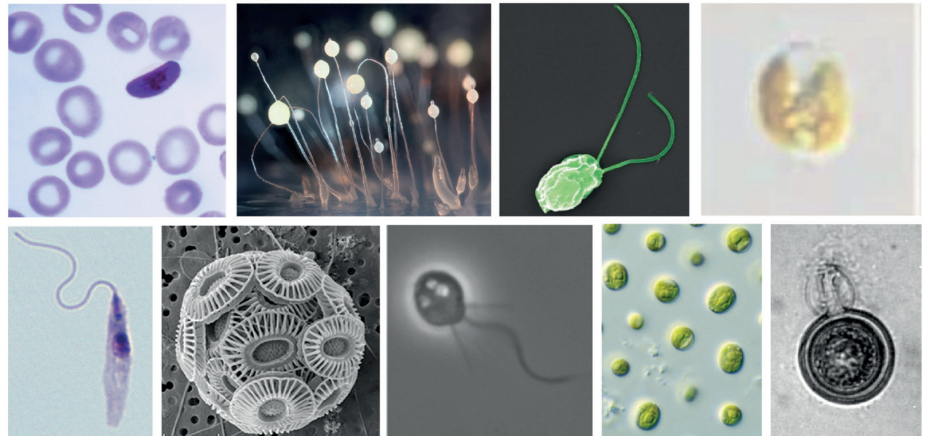


Figura 2. Selecció de protists amb genomes seqüenciats. A dalt, taula amb un resum de les dades genòmiques disponibles al repositori ENSEMBL (elaboració pròpia). Cal destacar que per a alguns gèneres hi ha múltiples soques seqüenciades: *Plasmodium* spp., amb seixanta soques; *Phytophthora* spp., amb vint-i-tres, i *Leishmania* spp., amb quatre. A baix, microfotografies de les espècies anteriors, seguint l'ordre en què apareixen en la taula. Imatges descarregades sota llicència de domini públic o de CC de Wikimedia Commons: *Dictyostelium discoideum* (imatge extreta d'https://commons.wikimedia.org/wiki/File:Dictyostelium_discoideum_fb_2.jpg [autor: Tyler Larsen] i publicada sota una llicència Attribution-ShareAlike 4.0 International (CC BY-SA 4.0), <https://creativecommons.org/licenses/by-sa/4.0/>), *Chlamydomonas reinhardtii* (fragment d'una imatge extreta d'https://commons.wikimedia.org/wiki/File:Biohybrid_Chlamydomonas_reinhardtii_microswimmers_2.jpg [autors: Mukrime Birgul Akolpoglu, Nihal Olcay Dogan, Ugur Bozuyuk, Hakan Ceylan, Seda Kizilel i Metin Sitti; doi:10.1002/adv.202001256] i publicada sota una llicència Attribution-ShareAlike 4.0 International (CC BY-SA 4.0), <https://creativecommons.org/licenses/by-sa/4.0/>), *Guillardia theta* (fragment d'una imatge extreta d'https://commons.wikimedia.org/wiki/File:41598_2017_2668_Fig2d_HTML.jpg [autors: Ryo Onuma, Neha Mishra, Shin-ya Miyagishima; doi:10.1038/s41598-017-02668-2] i publicada sota una llicència Attribution-Share Alike 4.0 International (CC BY-SA 4.0), <https://creativecommons.org/licenses/by-sa/4.0/>), *Leishmania major* (fragment d'una imatge extreta d'https://commons.wikimedia.org/wiki/File:Leishmania_spp_-_promastigote.jpg [autor: Stefan Walkowski] i publicada sota una llicència Attribution-ShareAlike 4.0 International (CC BY-SA 4.0), <https://creativecommons.org/licenses/by-sa/4.0/>), *Emiliana huxleyi* (fragment d'una imatge extreta d'https://commons.wikimedia.org/wiki/File:Emiliana_huxleyi_coccosphere_and_coccolith.jpg [autors: Griet Neukermans i Georges Fournier (Dr. Jeremy Young, University College London, London, amb permís); doi: 10.3389/fmars.2018.00146] i publicada sota una llicència Attribution-ShareAlike 4.0 International (CC BY-SA 4.0), <https://creativecommons.org/licenses/by-sa/4.0/>), *Monosiga Brevicollis* (imatge extreta d'https://commons.wikimedia.org/wiki/File:Monosiga_Brevicollis_Phase.jpg [autor: Stephen Fairclough] i publicada sota una llicència Attribution-ShareAlike 2.5 Generic (CC BY-SA 2.5), <https://creativecommons.org/licenses/by-sa/2.5/>), *Nannochloropsis gaditana* (fragment d'una imatge extreta d'https://commons.wikimedia.org/wiki/File:CSIRO_Sciencemage_10697_Microalgae.jpg [autor: CSIRO, <http://www.sciencemage.csiro.au/pages/about/>] i publicada sota una llicència Creative Commons Attribution 3.0 Unported (CC BY 3.0), <https://creativecommons.org/licenses/by/3.0/>), *Phytophthora infestans* (imatge extreta d'https://commons.wikimedia.org/wiki/File:Oospore_of_Phytophthora_infestans.jpg i publicada sota una llicència Attribution-ShareAlike 2.5 Generic (CC BY-SA 2.5), <https://creativecommons.org/licenses/by-sa/2.5/>).

desenvolupat aproximacions per a aconseguir els genomes d'espècies no cultivades; potser la més rellevant és la basada en la separació de cèl·lules de la comunitat per citometria de flux, se-

guida per l'amplificació i seqüenciació del DNA de cada cèl·lula (Labarre *et al.*, 2021), que genera SAG (*single amplified genomes*). La genòmica de cèl·lules individuals permet, doncs, accedir a

genòmica comparada i filogenòmica, s'ha identificat *Gloeomargarita lithophora*, aïllat fa pocs anys i relacionat amb *Synechococcus*, com el cianobacteri actual més proper al que va donar lloc al primer cloroplast (Ponce-Toledo *et al.*, 2017). Aquesta endosimbiosi s'anomena primària i va originar els arqueplastids, que inclouen glaucòfits, rodòfits i cloròfits, i dintre d'aquests darrers, les plantes. Posteriorment, els cloroplasts d'altres llinatges, com estrapèls, alveolats, rizaris, euglènids, haptistes i criptistes, es van originar per endosimbiosis secundàries. Aquí estem parlant d'una relació entre dos eucariotes, on un organisme fagòtrof n'incorpora un de fotosintètic amb cloroplast primari. Les endosimbiosis primàries i secundàries es poden diferenciar pel nombre de membranes que embolcallen el cloroplast, dues o quatre, respectivament, o per genòmica comparada, ja que hi ha una transferència constant de gens entre els diferents genomes implicats en la simbiosi (Archibald, 2015). En l'endosimbiosi primària veiem senyals provinents del nucli de l'hoste i del cianobacteri inicial, mentre que en la secundària la imatge és més complicada, ja que inclou el senyal del nucli de l'hoste, i del nucli i el cloroplast del fotòtrof eucariota incorporat. Normalment en les endosimbiosis secundàries el nucli de la cèl·lula integrada desapareix amb el temps (només en uns pocs llinatges roman com a nucleomorf), però hi ha parts del seu genoma que s'integren dins del nucli de la cèl·lula hoste, cosa que permet esbrinar quins organismes hi han estat implicats. En la majoria de casos, l'alga implicada en l'endosimbiosi secundària és un rodòfit (alga vermella), mentre que en pocs casos, com els clorarcniòfits (dins els rizaris) i els euglènòfits (dins d'excavats), és un cloròfit (alga verda). L'endosimbiosi secundària no és, però, el darrer estadi, i en grups com els dinoflagel·lats hi ha força casos d'endosimbiosis terciàries, on el cloroplast deriva d'una alga amb endosimbiosi secundària com diatomees, haptòfits o criptomonads. La major part de la fotosíntesi al planeta, i per tant bona part de la captura de CO₂ atmosfèric, és duta a terme per eucariotes, tant plantes com protists fotosintètics. Sense aquest procés d'endosimbiosi que va incorporar la fototrófia dins el llinatge eucariota no podríem entendre el planeta ni els seus cicles biogeoquímics tal com avui els coneixem.

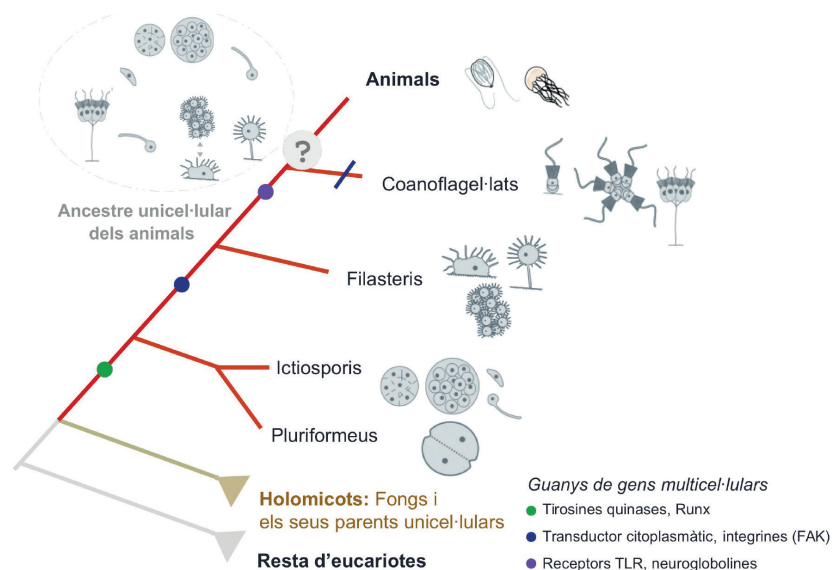
La multicel·lularitat

Després de la fotosíntesi, una de les transicions evolutives més significatives és l'origen de la multicel·lularitat. De multicel·lularitats n'hi ha

de molts tipus, i de fet no és exclusiva dels eucariotes; se'n troben els primers rastres en els estromatòlits, apareguts fa uns 3.500 milions d'anys, per agregació de cianobacteris filamentosos. En eucariotes, la multicel·lularitat ha aparegut independentment en almenys vint-i-cinc llinatges (Grosberg i Strathmann, 2007), en sis dels quals amb multicel·lularitat complexa, que dona lloc als organismes pluricel·lulars: animals, fongs, plantes i macroalgues verdes, vermelles i marrons. La resta pertanyen a esdeveniments de clonalitat reversible simple o multicel·lularitat agregativa. Un dels exemples paradigmàtics el trobem en l'ameba *Dictyostelium discoideum*, en què diferents cèl·lules s'ajunten per a formar un cos fructífer. Així doncs, mentre que la multicel·lularitat agregativa ha succeït força vegades al llarg de l'evolució dels eucariotes, la pluricel·lularitat és un fet més singular i relativament recent per als grups més emblemàtics (fa 600 milions d'anys en animals, 470 en plantes i 300 en fongs; per a algues vermelles, les primeres, fa 1.000 milions d'anys). La genòmica permet comparar les seqüències del DNA dels organismes pluricel·lulars i dels seus parents unicel·lulars propers, la qual cosa dona com a resultat una comprensió millor dels mecanismes de transició cap a la multicel·lularitat.

Segurament el cas més significatiu és l'estudi de l'origen de la multicel·lularitat en animals. Des que es va publicar el genoma del coanoflagel·lat *Monosiga brevicollis* (King *et al.*, 2008),

s'han publicat molts altres genomes o transcriptomes d'aquest grup unicel·lular, el més proper als animals, i també d'altres grups relacionats, com els filasteris, ictiosporis i pluriformes (Ros-Rocher *et al.*, 2021). A més de situar aquests nous llinatges a l'arbre de la vida, la informació genòmica ha revelat que alguns gens involucrats en funcions multicel·lulars i que es pensava que eren exclusius d'animals ja estaven presents al seu ancestre unicel·lular. Aquest ancestre, doncs, tindria una composició genètica relativament complexa i inclouria gens implicats en l'adhesió cel·lular (integrines), en la comunicació cel·lular (receptors tirosina-quinasa, receptors dels sistemes nerviós i immunitari), i factors transcripcionals per a regular la divisió cel·lular, el desenvolupament embrionari o la resposta immunitària (vegeu la figura 4). A més, en els quatre llinatges unicel·lulars parents dels animals hi ha espècies que presenten multicel·lularitat simple: el coanoflagel·lat *Salpingoeca rosetta* pot fer colònies, el filasteri *Capsaspora owczarzaki* presenta un estadi agregatiu en el cycle vital, i els ictiosporis i els pluriformes poden formar cèl·lules multinucleades prèvies a la seva divisió. El pas següent ha estat, doncs, estudiar en detall els gens responsables d'aquests estadis multicel·lulars. Per exemple, les integrines, que són proteïnes d'unió a la matriu extracel·lular en els teixits animals, estan expressades en els agregats de *C. owczarzaki* i implicades en la seva adhesió al substrat (Parra-Acero



† Figura 4. Representació filogenètica dels animals i els seus parents unicel·lulars que mostra les morfologies i els estadis de multicel·lularitat simple dels parents unicel·lulars, juntament amb el possible cycle de vida de l'ancestre unicel·lular dels animals. També s'hi inclouen exemples de gens relacionats amb funcions multicel·lulars adquirits en diferents moments evolutius previs als animals. Els guanys estan indicats amb punts de colors i les pèrdues secundàries, amb una ratlla (FAK = focal adhesion kinase [quinasa d'adhesió focal]; TLR = toll-like receptor [receptor de tipus Toll]). Elaboració pròpia.

et al., 2020). Conèixer els genomes d'aquestes espècies ha donat peu a una sèrie d'estudis funcionals que han portat a una visió única de l'ancestre unicel·lular dels animals: una espècie amb un cicle de vida complex, amb regulació temporal de l'expressió gènica i que hauria evolucionat cap a integrar els estadis temporals dintre d'un estadi multicel·lular amb regulació espacial dins el propi organisme (Ros-Rocher *et al.*, 2021). Les transformacions genòmiques clau que van acompanyar aquesta transició foren l'adquisició de nous gens formats a còpia de reordenar dominis proteics existents (López-Escardó *et al.*, 2019), la duplicació de factors de transcripció i senyalització cel·lular, i el guany d'introns (Grau-Bové *et al.*, 2017).

Genomes de protists en qüestions ecològiques

Les dades genòmiques també permeten aprofundir en el coneixement de l'ecofisiologia de les diferents espècies i expliquen fenòmens de diversificació i de la seva participació en els cicles biogeoquímics. A continuació en presentem alguns casos particularment interessants.

Les simbiosis

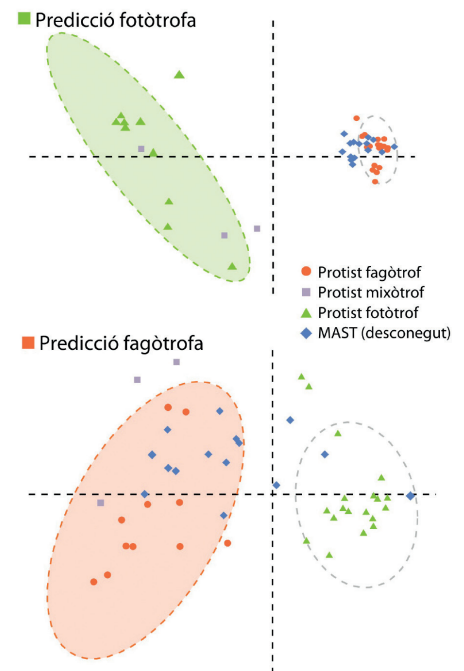
La simbiosi, individus de diferents espècies que estableixen una relació que els duu a viure junts, és molt freqüent a la natura. L'espectre de relacions simbiòtiques va des del parasitisme, on un dels dos membres de la relació es veu perjudicat, fins al mutualisme, on els dos membres en treuen un benefici. Els protists han establert relacions simbiòtiques entre ells, amb bacteris, i amb plantes i animals, i han participat en tot l'espectre de simbiosis. La relació que ha cridat més l'atenció ha estat el parasitisme, que no deixa de ser un cas extrem de simbiosi, i gran part de la informació que tenim de protists prové de l'interès històric en espècies paràsites d'humans i de bestiar. L'especificitat del parasitisme feia pensar que hi havia d'haver una empremta en el genoma del paràsit, i trobem, doncs, que la majoria de genomes de protists pertanyen a paràsits, tant d'humans (com *Plasmodium*, *Toxoplasma*, *Cryptosporidium* o *Leishmania*) com de plantes (com *Phytophthora*). Tanmateix, tot i que si ens basem en la literatura disponible podria semblar que el parasitisme és la forma dominant de simbiosi, l'exploració incipient de l'eucarioma d'animals, la fracció eucariota del microbioma, suggereix que la immensa majoria de relacions no són més que comensals (J. del Campo, J. *et al.*, 2019); és a dir, que els hostes on viuen ni es veuen beneficiats ni perjudicats pels simbiotes. Hi ha, doncs, un gran ventall de rela-

cions simbiòtiques que involucren protists que esperen ser estudiades.

Un aspecte rellevant és que la simbiosi no és un estadi fix. Per dir-ho d'una manera entenedora, un protist no és paràsit o mutualista, més aviat *està* paràsit o mutualista en el moment en què fem l'observació. A més, la posició dins de l'espectre de la simbiosi depèn de l'estat fisiològic de l'hoste i del simbiote i de les condicions ambientals, i un organisme pot tenir més d'un paper al llarg del seu cicle vital. Un cas paradigmàtic és el de l'alga verda *Ostreobium queckettii*, que viu a l'esquelet dels coralls formadors d'esculls (J. del Campo *et al.*, 2017). Durant la major part de la seva vida, *O. queckettii* és comensal, però, com a resultat de canvis fisiològics en el corall, pot passar a actuar com a mutualista i en qüestió de setmanes esdevenir un paràsit. Aquesta gradació d'estats queda també reflectida a escala genòmica. S'havia assumit que el parasitisme comportava una sèrie d'innovacions a escala genòmica que permetien la transició de la vida lliure al parasitisme. En comparar, però, els genomes de protists paràsits amb els d'espècies més properes de vida lliure s'ha vist que la major part de mecanismes associats al parasitisme ja estan presents en les formes de vida lliure (Janouskovec i Keeling, 2016). Pel que fa al mutualisme, en sabem molt menys, degut sobretot a la manca d'informació genòmica, però el que s'intueix és que la situació és similar a la dels paràsits.

Diversitat funcional i adaptacions ecològiques

És fascinant pensar com la informació continguda en el DNA és cabdal per a entendre com és i com es comporta un organisme. Així, una de les tasques òbvies ha estat buscar gens de rellevància ecològica que expliquin la capacitat d'adaptació i funcionament de l'espècie. Aquesta cerca és molt profitosa en procariotes, en els quals es poden traçar papers biogeoquímics dins dels cicles del nitrogen, fòsfor o sofre a partir de gens funcionals (Ferrera *et al.*, 2015). Tanmateix, fer el mateix a partir de genomes eucariotes no és trivial, ja que moltes funcions eucariotes depenen de la mida, de l'estructura i del comportament cel·lular, propietats que involucren molts gens sovint també implicats en altres funcions. Potser l'excepció en seria la fotosíntesi, basada en complexos proteics especialitzats en aquesta funció. Una primera aproximació en eucariotes ha estat fer genòmica comparada entre espècies amb modes tròfics coneguts com fototrofia i fagotrofia i des-



↑ Figura 5. Predicció del mode tròfic d'espècies no cultivades MAST utilitzant genòmica comparada. Cada espècie és representada per un punt, que se situa dins d'un espai bidimensional definit per models de predicció fotòtrofa (a dalt) i fagòtrofa (a baix). La zona verda del panell superior posiciona les espècies fotòtrofes (els MAST no hi són) i la zona carbassa del panell inferior inclou les fagòtrofes (on se situen la majoria de MAST). Figura modificada de Labarre *et al.* (2021) i publicada sota una llicència Attribution-ShareAlike 4.0 International (CC BY-SA 4.0), <https://creativecommons.org/licenses/by-sa/4.0/>.

envolupar un model estadístic per a aplicar a organismes desconeguts (Burns *et al.*, 2018). Certament, avui dia podem tenir el genoma d'una espècie de la qual ni tan sols coneixem el mode tròfic a partir d'una única cèl·lula (SAG) o de la comunitat (MAG). Aquest model prediu que diverses espècies marines no cultivades dels grups MAST (*marine stramenopiles*) (Labarre *et al.*, 2021) són fagòtrofes i no fotòtrofes (vegeu la figura 5), fet ja observat empíricament en algunes espècies, però no en totes.

També s'han fet estudis per a esbrinar la relació entre famílies gèniques i un mode tròfic concret, com seria el cas de les peptidases i la fagotrofia sobre bacteris. Les peptidases són enzims importants durant la digestió de les preses ingerides, ja que més de la meitat del pes sec dels bacteris és proteïna, però també es troben en espècies no fagòtrofes que les fan servir en el reciclatge de les seves proteïnes. Es podria pensar que les espècies fagòtrofes tindrien molta més varietat i un nombre més ele-

vat de gens de peptidases, però l'anàlisi de genomes d'estrampenòpils no va detectar una diferència clara entre espècies fotòtrofes, fagòtrofes i osmòtrofes (Labarre *et al.*, 2021). Per tant, la simple presència d'aquests gens no és suficient per a explicar la fagotrofia. El que sí que ha resultat prometedora ha estat l'estudi de la seva expressió en condicions fisiològiques conegudes. Així, s'ha vist que els gens de peptidases estan més expressats en la fase exponencial d'un bacterívor quan creix digerint bacteris, que en la fase estacionària (Massana *et al.*, 2021). També s'han fet estudis amb organismes mixòtrofs per a avaluar canvis en la seva expressió en funció de la llum i de la disponibilitat de nutrients (Kang *et al.*, 2021). A més, disposar de genomes d'espècies dominants a la natura permet estudiar-ne l'activitat a partir de metatranscriptomes, l'expressió gènica de la comunitat. Per exemple, s'ha vist que un altre grup d'enzims digestius, les glucosidases, presenten diferències rellevants en la seva expressió a l'oceà global entre espècies properes de protists no cultivats (Latorre *et al.*, 2021).

La genòmica també pot servir per a establir noves hipòtesis a partir de la troballa de gens no esperats. Un exemple interessant n'és el gen de la rodopsina, que codifica una proteïna descrita

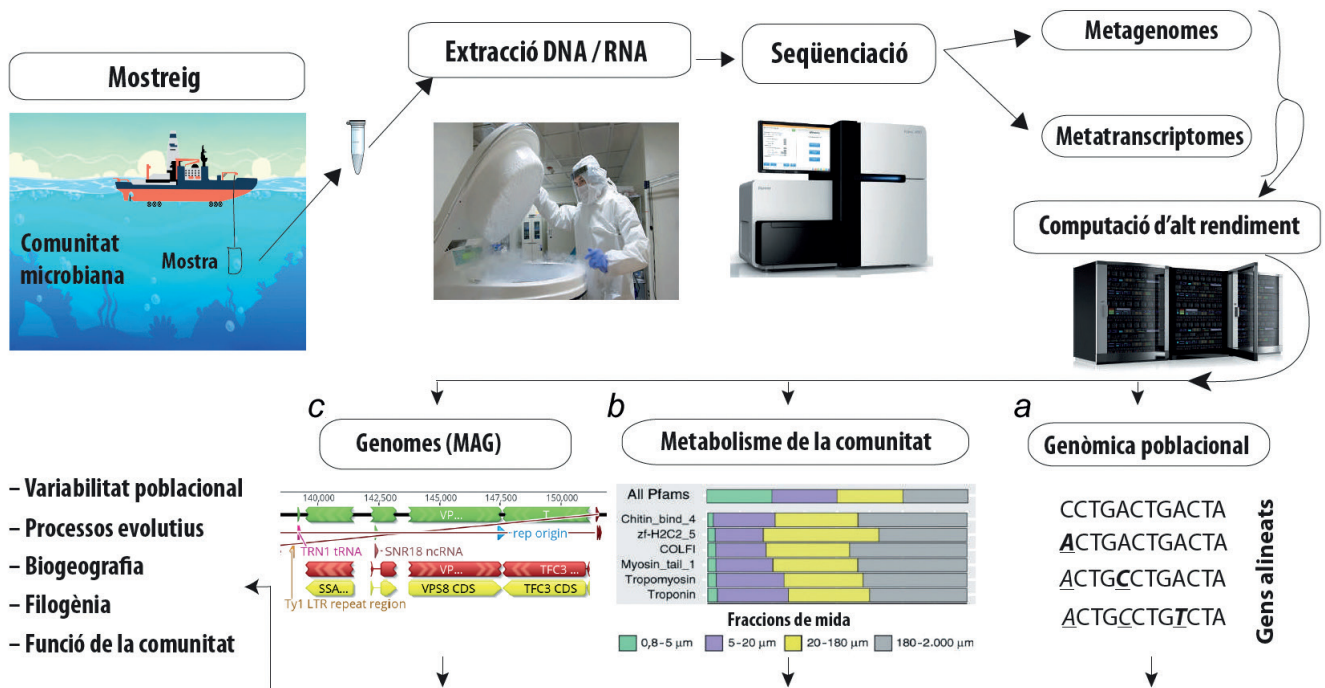
inicialment a la retina dels mamífers, i que va agafar protagonisme a l'ecologia microbiana en trobar-se dins d'un bacteri marí abundant i no cultivat. Es va veure que la rodopsina formava part d'un sistema pigmentari que funcionava com a bomba de protons fent servir la llum solar, cosa que generava ATP independentment de la respiració. Des del seu descobriment, la presència de rodopsina no ha parat de créixer entre llinatges de bacteris, d'arqueus i també de protists (Pinhassi *et al.*, 2016). Recentment s'ha trobat també en els genomes de bacterívors marins no cultivats (Labarre *et al.*, 2021). De fet, es pensava que aquestes espècies vivien indiferents a la llum, però aquesta troballa podria qüestionar-ho i explicar la seva tendència a viure a les capes il·luminades dels oceans. Tenim, doncs, un cas interessant per a investigar i desenvolupar, aparegut a partir de l'observació i anàlisi dels nous genomes.

Els genomes de la comunitat microbiana

Seqüenciar els genomes del conjunt de microorganismes de la comunitat, el metagenoma, va ser un repte per als pioners en ecologia microbiana i cada cop és més factible gràcies als avenços en seqüenciació. La metagenòmica

de protists ha avançat més lentament que la de procariotes, perquè tenen uns genomes més grans i complexos, però avui dia està esdevenint un camp molt prometedora. Així, pot ajudar a entendre la distribució geogràfica i la variabilitat poblacional, els processos evolutius a diverses escales temporals i el potencial metabòlic de les comunitats (vegeu la figura 6). També, com hem dit anteriorment, els metagenomes permeten reconstruir genomes d'espècies no cultivades, els MAG. L'aplicació simultània de la metagenòmica, la metatranscriptòmica i la genòmica de cèl·lules individuals ofereix un ventall enorme de possibilitats per a respondre preguntes ecològiques, de manera que esdevé una caixa d'eines molecular de gran utilitat.

La genòmica poblacional investiga la variabilitat genètica de les poblacions microbianes, les adaptacions de les poblacions a nínxols ecològics i la diversificació intraespecífica que pot derivar en especiació (Logares, 2011). A més, permet relacionar la variabilitat genètica amb el potencial d'adaptació de cada espècie als canvis ambientals. Cal destacar que, fins fa pocs anys, identificar poblacions de protists naturals era una tasca laboriosa i costosa, i que ara s'ha tornat accessible gràcies a la seqüenciació massiva. El mapatge de metagenomes a



↑ Figura 6. Flux de treball en genòmica ambiental de protists marins. La comunitat de protists es mostra de l'ambient, es fa una extracció del DNA o RNA ambiental i se seqüencia per a produir metagenomes i metatranscriptomes. Aquests són analitzats bioinformàticament en ordinadors d'alt rendiment amb l'objectiu d'estudiar la genòmica poblacional, el metabolisme de la comunitat o d'obtenir nous genomes. En el panell a cada seqüència correspon a un gen alineat de diferents genomes de la mateixa espècie, que indica els SNP (elaboració pròpia); el panell b mostra l'expressió relativa de diversos dominis Pfam en fraccions de mida del plàncton eucariota marí (fragment d'imatge de Carradec *et al.*, 2018, publicada sota una llicència Attribution-ShareAlike 4.0 Internacional (CC BY-SA 4.0), <https://creativecommons.org/licenses/by-sa/4.0/>); el panell c mostra una regió anotada d'un genoma eucariota (elaboració pròpia).

genomes de referència permet detectar el polimorfisme de nucleòtids simples (SNP, *single nucleotide polymorphism*) que representen les variants genòmiques de la mostra analitzada. La detecció i quantificació dels SNP permet determinar poblacions i el grau de diferenciació genètica (Sjöqvist *et al.*, 2021). Les anàlisis d'SNP possibiliten, a més, detectar processos selectius. La proporció de substitucions no sinònimes (dN) enfront de les sinònimes (dS) per a gens de la mateixa espècie pot servir per a detectar selecció positiva (dN/dS > 1) o selecció estabilitzadora (dN/dS < 1). Així, els gens que han experimentat una alta selecció positiva serien els que explicarien les millores adaptatives. En aplicar aquesta aproximació metagenòmica a sèries temporals, hi pot haver canvis en la freqüència dels SNP, cosa que suggereix processos evolutius en directe.

La metagenòmica i la metatranscriptòmica poden ajudar a comprendre el metabolisme de les comunitats de protists. Per exemple, un estudi de l'oceà global fet pel consorci Tara-Oceans va produir prop de 400 metatranscriptomes de les zones epi- i mesopelàgiques, cosa

que va generar un catàleg de ~116 milions de gens no redundants (Carradec *et al.*, 2018). Aquest catàleg va permetre tenir una visió global dels metabolismes eucariotes, sobretot unicel·lulars, actius a l'oceà, a més de constituir un recurs de gran valor per a estudiar l'expressió d'espècies determinades. La reconstrucció de genomes de protists a partir de metagenomes (MAG) continua sent un desafiament, tot i que hi ha dades recents que utilitzen mostres enriquides en protists i seqüenciacions massives que estan donant bons resultats (Delmont *et al.*, 2022). Probablement aquests MAG milloraran quan s'analitzin metagenomes obtinguts a partir de lectures llargues (PacBio, ONT). Aquesta seqüenciació també pot servir per a analitzar els canvis epigenètics en els protists associats a l'acimatació, per exemple, com la metilació en algunes bases pot afectar l'expressió gènica (Colot i Rossignol, 1999). En aquest sentit, diferenciar els canvis epigenètics dels canvis evolutius és important a l'hora d'entendre els mecanismes emprats pels protists per a adaptar-se als canvis ambientals.

Cloenda

La vida inclou una multitud de formes i llinatges evolutius diferents, i per la seva visibilitat, el seu impacte i la seva proximitat, els animals, les plantes i els fongs han captivat l'atenció de científics i naturalistes. Tanmateix, aquesta diversitat eucariota ha evolucionat a partir d'un substrat previ extremament divers d'eucariotes unicel·lulars, i han seguit coevolucionant plegats. Els protists, doncs, representen baules fonamentals en el coneixement biològic, i mereixen igualment ser inclosos en la revolució genòmica actual. En aquesta contribució hem presentat, a grans trets, què sabem sobre genòmica de protists i el que es pot saber a curt termini gràcies a les noves tècniques disponibles, com la genòmica de cèl·lules individuals i la metagenòmica. També expliquem alguns exemples en què els genomes dels protists han sigut clau per a abordar diferents qüestions biològiques fonamentals i avançar en el seu estudi. Esperem, doncs, que aquest treball contribueixi a posar de nou els protists sota el focus i que tinguin la consideració que es mereixen.

Bibliografia

- ANDERSSON, S. G. E. [et al.] (1998). «The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria». *Nature*, 396: 133-140.
- ARCHIBALD, J. M. (2015). «Genomic perspectives on the birth and spread of plasmids». *Proc. Natl. Acad. Sci. USA*, 112: 10147-10153.
- BURKI, F. [et al.] (2020). «The new tree of eukaryotes». *Trends Ecol. Evol.*, 35: 43-55.
- BURNS, J. A. [et al.] (2018). «Gene-based predictive models of trophic modes suggest *Asgard archaea* are not phagocytic». *Nature Ecol. Evol.*, 2: 697-704.
- CARPINELLI, E. C. [et al.] (2014). «Chromosome scale genome assembly and transcriptome profiling of *Nannochloropsis gaditana* in nitrogen depletion». *Mol. Plant*, 7: 325-335.
- CARRADEC, Q. [et al.] (2018). «A global ocean atlas of eukaryotic genes». *Nat. Commun.*, 9: 373.
- COLOT, V.; ROSSIGNOL, J. L. (1999). «Eukaryotic DNA methylation as an evolutionary device». *BioEssays*, 21: 402-411.
- CURTIS, B. A. [et al.] (2012). «Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs». *Nature*, 492: 59-65.
- DEL CAMPO, J. [et al.] (2017). «The "other" coral symbiont: *Ostreobium* diversity and distribution». *ISME J.*, 11: 296-299.
- (2019). «The eukaryome: Diversity and role of microeukaryotic organisms associated with animal hosts». *Funct. Ecol.*, 34: 2045-2054.
- DELMONT, T. O. [et al.] (2022). «Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean». *Cell Genom.*, 2: 100123.
- EICHINGER, L. [et al.] (2005). «The genome of the social amoeba *Dictyostelium discoideum*». *Nature*, 435: 43-57.
- FERRERA, I. [et al.] (2015). «Prokaryotic functional gene diversity in the sunlit ocean: Stumbling in the dark». *Curr. Opin. Microbiol.*, 25: 33-39.
- GARDNER, M. J. [et al.] (2002). «Genome sequence of the human malaria parasite *Plasmodium falciparum*». *Nature*, 419: 498-511.
- GRAU-BOVÉ, X. [et al.] (2017). «Dynamics of genomic innovation in the unicellular ancestry of animals». *ELife*, 6: e26036.
- GROSBERG, R. K.; STRATHMANN, R. R. (2007). «The evolution of multicellularity: A minor major transition?». *Annu. Rev. Ecol. Evol. Syst.*, 38: 621-654.
- HAAS, B. J. [et al.] (2009). «Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*». *Nature*, 461: 393-398.
- IVENS, A. C. [et al.] (2005). «The genome of the kinetoplast parasite, *Leishmania major*». *Science*, 309: 436-442.
- JANOUSKOVEC, J.; KEELING, P. J. (2016). «Evolution: Causality and the origin of parasitism». *Curr. Biol.*, 26: R174-177.
- KANG, Y. [et al.] (2021). «Transcriptomic responses of four pelagophytes to nutrient (N, P) and light stress». *Front. Mar. Sci.*, 8: 636699.
- KEELING, P. J. [et al.] (2013). «The number, speed, and impact of plastid endosymbioses in eukaryotic evolution». *Annu. Rev. Plant Biol.*, 64: 583-607.
- (2014). «The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing». *PLoS Biol.*, 12: e1001889.
- KING, N. [et al.] (2008). «The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans». *Nature*, 451: 783-788.
- LABARRE, A. [et al.] (2021). «Comparative genomics reveals new functional insights in uncultured MAST species». *ISME J.*, 15: 1767-1781.
- LATORRE, F. [et al.] (2021). «Niche adaptation promoted the evolutionary diversification of tiny ocean predators». *Proc. Natl. Acad. Sci. USA*, 118: e2020955118.
- LEWIN, H. A. [et al.] (2022). «The Earth BioGenome Project 2020: Starting the clock». *Proc. Natl. Acad. Sci. USA*, 119: e2115635118.
- LOGARES, R. (2011). «Population genetics: the next stop for microbial ecologists?». *Open Life Sci.*, 6: 887-892.
- LÓPEZ-ESCARDÓ, D. [et al.] (2019). «Reconstruction of protein domain evolution using single-cell amplified genomes of uncultured choanoflagellates sheds light on the origin of animals». *Phil. Trans. R. Soc. B.*, 374: 20190088.
- MARTIN, W. F. [et al.] (2017). «The physiology of phagocytosis in the context of mitochondrial origin». *Microb. Molec. Biol. Rev.*, 81: e00008-17.
- MASSANA, R. [et al.] (2021). «Gene expression during bacterivorous growth of a widespread marine heterotrophic flagellate». *ISME J.*, 15: 154-167.
- MERCHANT, S. S. [et al.] (2007). «The *Chlamydomonas* genome reveals the evolution of key animal and plant functions». *Science*, 318: 245-251.
- PARRA-ACERO, H. [et al.] (2020). «Integrin-mediated adhesion in the unicellular holozoan *Capsaspora owczarzakii*». *Curr. Biol.*, 30: 4270-4275.
- PAWLOWSKI, J. [et al.] (2012). «CBOL Protist Working Group: Barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms». *PLoS Biol.*, 10: e1001419.
- PINHASSI, J. [et al.] (2016). «Marine bacterial and archaeal ion-pumping rhodopsins: Genetic diversity, physiology, and ecology». *Microbiol. Mol. Biol. Rev.*, 80: 929-954.
- PITTIS, A. A.; GABALDÓN, T. (2016). «Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry». *Nature*, 531: 101-104.
- PONCE-TOLEDO, R. I. [et al.] (2017). «An early-branching freshwater cyanobacterium at the origin of plastids». *Curr. Biol.*, 27: 386-391.
- READ, R. A. [et al.] (2013). «Pan genome of the phytoplankton *Emiliania huxleyi* underpins its global distribution». *Nature*, 499: 209-213.
- REUTER, J. A. [et al.] (2015). «High-throughput sequencing technologies». *Mol. Cell*, 58: 586-597.
- ROS-ROCHER, N. [et al.] (2021). «The origin of animals: An ancestral reconstruction of the unicellular-to-multicellular transition». *Open Biol.*, 11: 200359.
- SJÖQVIST, C. [et al.] (2021). «Ecologically coherent population structure of uncultivated bacterioplankton». *ISME J.*, 15: 3034-3049.
- SPANG, A. [et al.] (2015). «Complex archaea that bridge the gap between prokaryotes and eukaryotes». *Nature*, 521: 173-179.