

Sobre la necessitat d'estudiar trets genètics que influeixen en l'organització i l'estructura del genoma en projectes de seqüenciació de plantes

Jaume Pellicer,^{1,2} Oriane Hidalgo,^{1,2} Joan Vallès^{3,4} i Teresa Garnatje¹

¹ Institut Botànic de Barcelona (IBB), CSIC - Ajuntament de Barcelona

² Royal Botanic Gardens, Kew, Regne Unit

³ Laboratori de Botànica, Unitat Associada al CSIC, Facultat de Farmàcia i Ciències de l'Alimentació, Institut de la Biodiversitat IRBio, Universitat de Barcelona

⁴ Secció de Ciències Biològiques, Institut d'Estudis Catalans

Correspondència: Jaume Pellicer. Institut Botànic de Barcelona (IBB). Parc de Montjuïc. Passeig del Migdia, s. n. 08038 Barcelona. Adreça electrònica: jaume.pellicer@ibb.csic.es.

DOI: 10.2436/20.1501.02.211

ISSN (ed. impresa): 0212-3037

ISSN (ed. digital): 2013-9802

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

Rebut: 16/01/2022

Acceptat: 12/03/2022

Resum

El coneixement dels mecanismes responsables de mantenir la biodiversitat requereix, entre d'altres, iniciatives que faciliten l'estudi en profunditat del genoma dels organismes eucariotes. D'aquesta manera, podrem assolir una visió crítica i robusta sobre la seua estructura i el seu funcionament. En el cas de les plantes, a més, es fa palesa la necessitat de continuar generant dades fonamentals referents al contingut de DNA nuclear i el nombre de cromosomes. Aquests trets són molt importants en l'evolució de les plantes, atesa la seua gran diversitat. A més, aporten informació bàsica, i alhora essencial, per a plantejar l'assemblatge d'un genoma. Aquests reptes, però, necessiten el suport d'estratègies d'investigació coordinades a escala global com l'Earth BioGenome Project (EBP) i les seues filials regionals, que faciliten l'accés a la tecnologia, a l'intercanvi d'informació i al desenvolupament de plans de formació. Prioritzar l'estudi comparatiu d'espècies amb un mostratge equilibrat i inclusiu, representatiu de la diversitat taxonòmica i geogràfica de les plantes, així com dels seus trets genòmics, morfològics i ecològics serà, per tant, un dels grans reptes en els pròxims anys.

Paraules clau: cromosoma, evolució, grandària del genoma, poliploidia, seqüenciació.

On the need to study genetic traits influencing the organization and structure of the genome in plant sequencing projects

Abstract

Understanding the mechanisms responsible for maintaining biodiversity requires, among other things, initiatives that facilitate in-depth genomic surveillance across eukaryote organisms. These initiatives support critical research programs for gaining robust insights into how genomes are structured and operate. In relation to land plants, there is also a clear need to continue generating fundamental data regarding both nuclear DNA contents and chromosome numbers. As a result of their great diversity, these traits are unquestionably important drivers of plant evolution. In addition, they provide basic yet essential information for the planning of a genome assembly. These challenges, however, need the support of globally coordinated research strategies such as the Earth BioGenome Project (EBP) and its regional subsidiaries, which facilitate access to technology, information exchange and the development of training plans. Prioritizing comparative studies across species with a balanced and inclusive taxonomic sampling that covers the geographical diversity of plants as well as their genomic, morphological and ecological features will consequently be one of our major challenges in the years to come.

Keywords: chromosome, evolution, genome size, polyploidy, sequencing.

Introducció

Vivim un present de grans projectes de seqüenciació de genomes, com l'Earth BioGenome Project (EBP, <https://www.earthbiogenome.org>), i alguns dels seus ramals com, per exemple, el Darwin Tree of Life (DTOL, <https://www.darwintreeoflife.org>) o la iniciativa catalana per a l'EBP (CBP, de l'anglès Catalan Initiative for the Earth BioGenome Project, <https://www.biogenoma.cat>). Aquestes iniciatives, entre moltes altres, juntament amb el desenvolupament constant de les noves tècniques de seqüenciació d'alt rendiment, contribueixen de

manera significativa a accelerar el descobriment i el coneixement de la gran diversitat de genomes presents en el nostre planeta. Entre aquests genomes, les plantes (*Viridiplantae*) tenen un paper fonamental en la dinàmica dels ecosistemes, amb implicacions directes en el benestar dels éssers humans i amb un gran impacte en l'agroecologia global (Schaal, 2019; Hidalgo *et al.*, 2021). A mesura que l'activitat humana impacta de manera descontrolada sobre els ecosistemes, accelerant la pèrdua d'hàbitats, i consegüentment d'espècies, es fa més patent la necessitat de generar coneixement

crític i evidències sòlides que faciliten la implementació d'estratègies de gestió i conservació per tal de mitigar l'emergència climàtica i de diversitat a la qual ens enfrontem actualment.

Ateses la gran diversitat i la rellevància de les plantes, són moltes les investigacions centrades en el seu estudi. Un punt clau és el coneixement del seu genoma, pel que fa a l'estructura i al funcionament, en un context tant aplicat com evolutiu. Si bé la majoria de treballs se centren en organismes model, d'interès econòmic i/o sanitari, l'accés a les grans

plataformes de seqüenciació, que estan revolucionant el camp de la genòmica, ha facilitat en gran manera la possibilitat de plantejar i dur a terme iniciatives cada vegada més ambicioses, centrades en llinatges no estudiats fins al moment amb aquestes noves tècniques d'anàlisi genòmiques. De fet, actualment ja s'han publicat al voltant de 300 genomes de plantes assemblats a escala cromosòmica (Kress *et al.*, 2022), tot i que encara representen una petita fracció (al voltant de 900 espècies) de les aproximadament 400.000 espècies de plantes terrestres que coneixem (Lughadha *et al.*, 2016), però que, alhora, conformen un dipòsit global de coneixement sense precedents. L'esforç actual, tant tecnològic com científic, centralitzat a través de la creació de consorcis de recerca, té com a objectiu principal la coordinació i l'estandardització de metodologies, amb la finalitat de continuar millorant el procés d'assemblatge de genomes, resultat de la millora continuada de les eines de seqüenciació. L'obtenció de genomes de referència, però, actualment encara està esbiaixada, amb uns llinatges molt més explorats que d'altres, com per exemple les *Poaceae*, atesa la nombrosa presència d'espècies d'interès agronòmic com l'arròs (*Oryza sativa*). L'objectiu final ha de ser, per tant, intentar representar el màxim nombre de llinatges vegetals coneguts, la qual cosa ha d'implementar-se paral·lelament a la millora del coneixement sobre les relacions taxonòmiques entre les plantes. Certament, aquesta és una altra parcel·la de recerca que també està vivint una revolució substancial, a través d'iniciatives com el Plant and Fungal Trees of Life (PAFTOL, <https://www.kew.org/science/our-science/projects/plant-and-fungal-trees-of-life>), i la creació de nous kits de seqüenciació dirigida d'alt rendiment, com l'Angiosperms353 (Johnson *et al.*, 2019) o el GoFlag451 (Breinholt *et al.*, 2021), que han suposat un revulsiu en la millora del coneixement de la història evolutiva de les plantes terrestres.

Sense cap mena de dubte, la revolució genòmica actual està sotmesa a les exigències tècniques d'una comunitat científica decidida a conèixer els secrets més íntims que s'amaguen en el genoma. L'arribada de tecnologies que permeten seqüenciar fragments de DNA cada vegada més llargs ha estat el punt de partida del que es coneix com l'era dels genomes de qualitat plati, amb assemblatges contigus de gran precisió a escala cromosòmica, que ens permetran estudiar en profunditat patrons de reestructuració i evolució a gran escala. Ini-

ciatives com el 10,000 Plant Genome Sequencing Project, també conegut com a 10KP (Twynford, 2018), representen una fita històrica en el progrés del coneixement del genoma vegetal, i ofereixen una plataforma incomparable per a abordar qüestions sobre els processos biològics al llarg de l'arbre de la vida. En aquest sentit, l'estudi de genomes de llinatges generalment poc investigats és un dels objectius d'aquest projecte. Les plantes, però, plantegen un gran repte, atesa l'extraordinària diversitat pel que fa a una sèrie de trets específics de la seua biologia, com són: *a*) la grandària del genoma, o valor C, que es refereix a la quantitat de DNA total present en un nucli cel·lular haploide no replicat; *b*) el nombre de cromosomes, és a dir, en quants cromosomes es distribueix la seqüència de DNA, i *c*) el nivell de ploïdia, que indica el nombre total de genomes (o còpies) presents en el nucli cel·lular. Aquests trets bàsics, però de gran rellevància en l'evolució de les plantes (Pellicer *et al.*, 2018), s'han de tenir en compte a l'hora de plantejar-se seqüenciar un genoma, ja que poden representar una sèrie de reptes computacionals que creixen de manera exponencial en funció de la mida (Kelly *et al.*, 2012).

Genomes i cromosomes: la diversitat cromosòmica en les plantes

Els cromosomes de les plantes presenten una gran variabilitat, ja siga en nombre com en forma i mida. Encara que pugua semblar una informació menys important, el nombre cromosòmic és un paràmetre fonamental en la botànica sistemàtica i evolutiva (Stuessy, 2009 i 2011). El nombre de cromosomes més petit descobert fins avui és de $2n = 4$, present tant en llinatges de monocotiledònies —per exemple, *Zingiber biebersteiniana* (*Poaceae*) o *Ornithogalum tenuifolium* (*Liliaceae*)— com d'eudicotiledònies —*Haplopappus gracilis* i *Brachyscome dichromosomatica* (*Asteraceae*)— (Castiglione i Cremonini, 2012). A l'altre extrem podem trobar nombres cromosòmics molt grans, com per exemple al gènere *Ophioglossum* (*Ophioglossaceae*, pteridòfits), en què s'han reportat nombres cromosòmics de més de 1.000 cromosomes, com en l'espècie *O. reticulatum*, amb $2n = 1.400$ (Khandelwal, 1990). Aquest valor no només representa el nombre de cromosomes més gran entre les plantes, sinó que segurament és un rècord entre els organismes eucariotes, almenys entre tots els estudiats fins al moment present. Si el comparem amb el nostre propi genoma, amb

una seqüència de DNA distribuïda en 46 cromosomes, aquesta falguera presenta trenta vegades més cromosomes en cada cèl·lula que l'ésser humà. La diversitat de nombres cromosòmics no només és evident entre espècies. De fet, es coneixen exemples en què una sola espècie pot presentar una gran diversitat de nombres cromosòmics, com per exemple *Cardamine pratensis* (*Brassicaceae*), amb aproximadament 65 nombres cromosòmics descrits, que varien des de $2n = 16$ fins a $2n = 96$. Si posem en perspectiva aquests nombres, en comparació amb la diversitat que s'ha trobat entre els animals, observem que les plantes són molt més variables. No obstant això, el nombre de cromosomes més petit que s'ha trobat entre els animals és inferior que en les plantes, $2n = 2$, reportat a *Myrmecia pilosula* (*Formicidae*), una espècie de formiga nativa d'Àustràlia (Crosland i Crozier, 1986). En canvi, el més gran no supera els 500 cromosomes a l'espècie de papallona *Polyommatus atlanticus* (*Lycaenidae*) (Lesse, 1970), amb una tercera part de la dotació cromosòmica de la falguera abans esmentada.

En gran manera, la diversitat de nombres cromosòmics que trobem en les plantes es deu a fenòmens de multiplicació del genoma (o poliploïdia), que són extremament freqüents, i sense cap dubte estan considerats un dels principals motors de la seua diversificació (Wood *et al.*, 2009; Soltis i Soltis, 2016; Clark *et al.*, 2016). Els organismes poliploïdes s'originen en moltes ocasions a través d'un fenomen d'hibridació entre tàxons diferents, i que en molts casos dona lloc a la formació d'un híbrid estèril, com per exemple l'híbrid d'*Urospermum* (*Asteraceae*), descobert recentment a Montjuïc, Barcelona (Pellicer *et al.*, 2022). Si aquest organisme híbrid pateix una duplicació genòmica, a més de restaurar-ne la fertilitat, donarà lloc a la formació d'un individu poliploide (Hegarty i Hiscock, 2008). Els organismes poliploïdes originats arran de l'encreuament entre dues espècies són anomenats *al·lopoliploïdes* i són molt comuns entre els pteridòfits i les angiospermes. De fet, al llarg de l'evolució, i lligat als fenòmens de domesticació, els humans hem tret profit de la poliploïdia (atesos molts dels seus efectes en el fenotip) i actualment molts dels vegetals que utilitzem diàriament són d'origen poliploide (p. ex. blat, plàtan o cotó, entre molts altres). A banda de la gran diversitat de ploïdies descobertes en plantes actuals, el coneixement cada vegada més profund que tenim sobre les relacions evolutives entre les plantes, juntament amb la

millora d'eines de seqüenciació i transcriptòmica, han permès descobrir evidències que indiquen una alta incidència d'episodis de multiplicació del genoma en el passat. Aquests fenòmens es coneixen com a *episodis de paleopoliploidia*, i se n'han evidenciat més d'un centenar al llarg de la història evolutiva de les plantes terrestres (Landis *et al.*, 2018).

La poliploidia, però, no requereix sempre un fenomen d'hibridació, ja que aquesta duplicació del genoma pot tenir lloc en un mateix individu, per autofecundació, o per entrecreuament d'individus diferents de la mateixa espècie. Aquest fenomen rep el nom d'*autopoliploidia*, com és el cas, per exemple, de *Senecio carniolicus* (*Asteraceae*, $2n = 2x = 90$) (Sonnleitner *et al.*, 2016). Molts autors suggereixen que el paper evolutiu de l'autopoliploidia no es pot menysprear, atès el seu impacte en el genoma i la capacitat d'induir processos de diversificació (Parisod *et al.*, 2010). De fet, al voltant d'un 13% de les angiospermes presenta citotips poliploides, i les conseqüències en el genoma poden incloure canvis en l'expressió de gens, el metabolisme i la fisiologia (Panchy *et al.*, 2016). Per exemple, en el cas d'*Artemisia annua* (*Asteraceae*), una herba de gran importància en el tractament de la malària, ja que se n'extrau l'artemisinina, s'ha comprovat que individus tetraploides produeixen 1,5 vegades més artemisinina que els individus diploides (Banyai *et al.*, 2010), fet que podria tenir un impacte econòmic rellevant en la producció d'aquest compost químic d'una manera més sostenible amb una reducció de la biomassa necessària per a la seua producció.

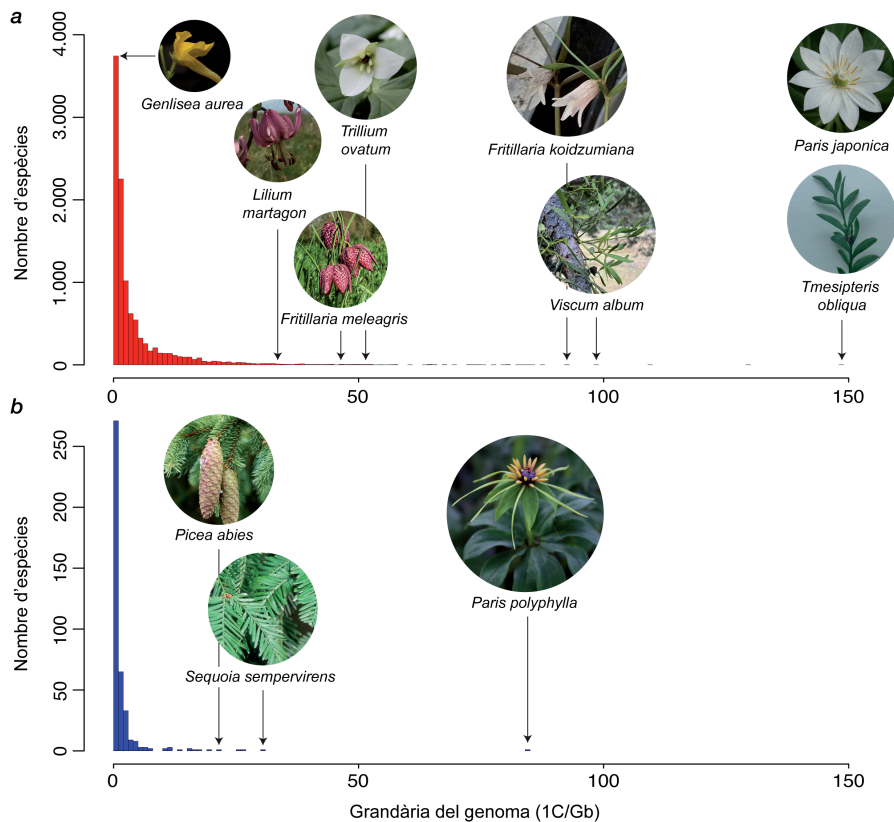
L'impacte de la poliploidia en la biodiversitat global és patent, ja que pot influenciar interaccions ecològiques, processos evolutius i patrons de distribució. De fet, hi ha estudis que demostren que la distribució de la freqüència de poliploides al planeta no és aleatòria, amb una incidència més elevada segons ens allunyem de l'equador i ens aproximem a zones properes als pols terrestres (Rice *et al.*, 2019). Si comparem la incidència de la poliploidia en les plantes amb la que trobem en els animals, és evident que és molt inferior entre aquests últims, tot i que, en els vertebrats, és present en rèptils, amfibis i peixos (Otto i Whitton, 2000). Pel que fa als invertebrats, alguns llinatges de mol·luscs, crustacis, anèl·lids o insectes també s'han reportat com a poliploides (Gregory i Mable, 2005). De fet, en aquest últim grup s'ha posat en evidència l'existència de fins a divuit esdeveniments de paleopoliploidització al llarg de la seua història evolutiva (Li *et al.*, 2018). Tota

aquesta diversitat impressionant de nombres cromosòmics i d'incidència de fenòmens de duplicació del genoma té una sèrie de conseqüències a l'hora de plantejar-se la seqüenciació d'un genoma, especialment si es vol obtenir un assemblatge del genoma d'alta qualitat. Bàsicament, i de manera rutinària, n'hauríem de saber, amb seguretat, el nombre de cromosomes i el nivell de ploïdia, ja que d'això dependrà la base de l'assemblatge que construïrem de qualsevol organisme (i/o individu) del qual ens plantejarem seqüenciar el genoma. Aquest punt és especialment rellevant en espècies per a les quals s'ha descrit un nombre variable de cromosomes i nivells de ploïdia, ja que poden tenir un impacte directe en la interpretació de les dades de seqüenciació.

La grandària del genoma en les plantes: una extraordinària diversitat que planteja grans reptes bioinformàtics

La quantitat de DNA nuclear en un nucli gametofític no replicat és el que es coneix com a *mida del genoma* o *valor C* (Greilhuber *et al.*, 2005), i és un caràcter de gran diversitat, del qual disposem de dades per a aproximadament unes 20.000 espècies d'organismes eucariotes. Entre aquests organismes, les plantes han estat estudiades de manera extensa des de 1951, quan s'estimà, per primera vegada, la grandària del genoma en l'espècie *Lilium longiflorum* (*Liliaceae*) (Ogur *et al.*, 1951). Des d'aleshores, i fins avui, s'han investigat al voltant de 12.000 espècies. Un esforç que s'ha dut a terme en paral·lel al desenvolupament de bases de dades en què la informació és recopilada i accessible al públic en general i a la comunitat investigadora en particular com, per exemple, la Plant DNA C-values Database (<https://cvalues.science.kew.org>) (Pellicer i Leitch, 2020), que conté dades dels grans llinatges de plantes terrestres (angiospermes, gimnospermes, pteridòfits i briòfits), sense deixar de banda les algues. A més, la base inclou dades cariològiques que permeten interpretar millor la mida del genoma. Com s'ha indicat en apartats anteriors, aquest paràmetre és immensament divers entre els organismes eucariotes en general ($> 64.000 \times$), i particularment entre les plantes (aproximadament, $2.400 \times$, vegeu la figura 1a) (Pellicer *et al.*, 2018). Entre aquestes plantes, el genoma més petit reportat s'ha trobat en l'espècie *Genlisea aurea* (figura 1), una petita planta carnívora de la família de les *Lentibulariaceae*, amb un genoma de només 64 Mb (Fleischmann *et al.*, 2014). En canvi, a

l'altre extrem de l'espectre trobem el genoma gegantesc de l'espècie *Paris japonica* (figura 1), un lliri que pertany a la família de les *Melanthiaceae*, el qual presenta un genoma de 149 Gb (Pellicer *et al.*, 2010). La seqüència del genoma d'aquesta planta és, en aquest cas, aproximadament cinquanta vegades més gran que la del genoma humà i la seua seqüenciació plantejaria un veritable repte tècnic i computacional, malgrat els avanços recents. De nou, les plantes representen un grup extraordinari pel que fa a la diversitat, fet que queda palès si comparem les dades de què disposem per a alguns grups d'animals, com per exemple els ocells, amb una variació de la grandària del genoma que no supera les 2,5 vegades (0,91-2,16 Gb/1C, <https://www.genomesize.com>). Malgrat la gran diversitat de mides del genoma reportades en les plantes, la majoria de les angiospermes, per exemple, es caracteritzen pels genomes petits i/o molt petits (figura 1a), tot i l'elevada incidència d'episodis de poliploidia. Aquest fet contrasta, per exemple, amb la distribució de valors en les gimnospermes, que presenten mides del genoma menys diverses, i relativament més grans que les angiospermes, un fet que s'il·lustra també amb una amplitud de nombres cromosòmics molt més discreta que en les angiospermes (Rastogi i Ohri, 2020). Aquesta tendència contrasta amb el que trobem, per exemple, entre els pteridòfits, en què l'amplitud de valors de quantitat de DNA no és tan elevada com en les angiospermes, tot i que observem que espècies amb genomes gegants de mida similar al d'algunes angiospermes presenten un nombre molt superior de cromosomes (p. ex. *Tmesipteris obliqua* [*Psilotaceae*], $1C = 149 \text{ Gb}$, $2n = 416$, figura 1) (Hidalgo *et al.*, 2017), enfront dels $2n = 40$ de *Paris japonica* (Pellicer *et al.*, 2010). Aquestes diferències naixen del fet que aquests llinatges han mostrat patrons d'evolució diferents. Un genoma poliploide a les angiospermes està generalment sotmès a fenòmens de diploidització, els quals provoquen una pèrdua parcial de DNA repetitiu i còpies de gens, entre d'altres (Wendel, 2015), que en general afavoreixen l'aparellament de cromosomes homòlegs; aquest fet generalitzat és el responsable de la falta de correlació entre la mida del genoma i el nombre de cromosomes (Pellicer *et al.*, 2018; Wang *et al.*, 2021). En canvi, en els pteridòfits s'ha observat que la grandària del genoma està correlacionada amb el nombre de cromosomes (Clark *et al.*, 2016), a pesar de l'elevada incidència de poliploides (Wood *et al.*, 2009). Les diferències, doncs, rau en el



† Figura 1. a) Distribució de la grandària del genoma (valors 1C) en les plantes terrestres (dades extretes de la Plant DNA C-values Database, <https://cvalues.science.kew.org/>); b) distribució de la grandària del genoma de les espècies per a les quals s'ha seqüenciat el genoma (dades extretes de la base de dades PlaBiPD, <https://plabipd.de/index.ep>, i publicacions addicionals). Imatges: Jaume Pellicer (*Genlisea aurea*, *Lilium martagon*, *Fritillaria koidzumiana*, *Fritillaria meleagris*, *Trillium ovatum*, *Paris polyphylla*, *Sequoia sempervirens*, *Picea abies*, *Viscum album*, *Tmesipteris obliqua*) i Laurence Hill (*Paris japonica*).

procés de diploidització, que en els pteridòfits impacta de manera prioritària silenciant gens, sense que hi haja associada una pèrdua significativa aparent de DNA repetitiu, i que, en definitiva, permet que al llarg de l'evolució es puguin mantenir nombres de cromosomes elevats (Haufler, 2014).

Quins són els principals mecanismes responsables de la gran diversitat de mides del genoma en les plantes?

Com s'ha indicat en apartats anteriors, un dels principals motors de l'evolució del canvi en la mida del genoma és la poliploidia, que implica la presència de més de dues còpies del genoma i, per tant, l'increment de la quantitat de DNA (que pot ser proporcional o no) a mesura que ascendim en nivells de ploïdia. No obstant això, els canvis en la composició i l'abundància de seqüències repetitives de DNA també tenen un paper crític en l'estructura i en la mida del genoma. Aquestes seqüències es troben pre-

sents en el genoma en múltiples còpies, que varien en ordres de magnitud i que són principalment els retrotransposons, elements transposables de DNA i seqüències repetides en tàndem (o satèl·lits de DNA) (Bennetzen i Wang, 2014). Els transposons de DNA i els retrotransposons són elements mòbils del genoma que s'insereixen a través de mecanismes de tallar-i-enganxar o copiar-i-enganxar, respectivament, que els confereixen la capacitat de colonitzar nous espais de l'espectre genòmic. Entre aquests últims, els retrotransposons de repetició terminal llarga (LTR, de l'anglès *long terminal repeat*) són molt coneguts per la seua capacitat de dominar una fracció substancial del genoma repetitiu de les plantes. Comprenen diferents llinatges organitzats en superfamílies, en què els Ty1/Copia i Ty3/Gypsy són els més comuns entre les plantes (Wicker *et al.*, 2007).

A mesura que creix el nostre coneixement sobre l'organització i el funcionament dels genomes, gràcies en part al desenvolupament de

tècniques de seqüenciació cada vegada més i més potents, podem observar que la fracció repetitiva del genoma augmenta a mesura que ho fa la mida (p. ex., el 80% del genoma de la dacsa, *Zea mays*, és repetitiu [Sofi *et al.*, 2007]). A més, en genomes relativament petits, els canvis en la mida moltes vegades estan governats per l'acció d'un o pocs llinatges d'aquests elements, que arriben a tenir un elevadíssim nombre de còpies (Piegu *et al.*, 2006; Macas *et al.*, 2015). Aquests genomes mostren signatures d'un elevat dinamisme, atès que la majoria d'elements repetitius presenten evidències d'activitat d'amplificació i eliminació freqüent, amb una signatura evolutiva compatible amb un procés de vida recent, d'entre 3 i 5 milions d'anys. D'aquesta manera, les diferències en la capacitat dels elements repetitius per a amplificar-se és el que impacta directament en la grandària final del genoma. La pregunta és, però, si al llarg de l'amplitud de $2.400 \times$ de variació de la grandària del genoma, l'escenari descrit anteriorment per a plantes amb genomes relativament petits continua present. La resposta és que no.

Un dels gèneres model per a l'estudi de l'obesitat genòmica ($1C > 30$ Gb) el trobem en les *Liliaceae*, en el gènere *Fritillaria*. Aquest gènere presenta una diversitat elevada de mides del genoma, que varia des dels 29 Gb/1C a *F. degeniana* fins als 87 Gb/1C a *F. koidzumiana* (Pellicer i Leitch, 2020). L'anàlisi d'alguns d'aquests genomes gegants ha posat de manifest una dinàmica totalment diferent de la que s'havia trobat en genomes petits. Per exemple, el seu genoma repetitiu és heterogeni i està compost per un gran nombre d'elements, amplificats a baixa freqüència, sense presentar evidències d'amplificació o eliminació recents (Kelly *et al.*, 2015). Aquesta dinàmica també s'ha observat en algunes gimnospermes amb genomes grans com *Picea abies* (*Pinaceae*) (Nystedt *et al.*, 2013), i en els genomes gegants d'algunes salamandres (Sun *et al.*, 2012). En conseqüència, la hipòtesi més acceptada per a explicar com han arribat aquests genomes a proporcions gegantesques és la falta d'habilitat per a eliminar seqüències amplificades de DNA repetitiu, fet que n'afavoreix l'acumulació i la fossilització posterior en el genoma. Anàlisis comparatius més recents, incloent-hi una amplitud de diversitat de mides del genoma que va des de 64 Mb fins a 88,5 Gb, indiquen que la fracció repetitiva del genoma canvia en funció de la seva grandària (Novák *et al.*, 2020). De fet, aquesta fracció s'incrementa amb relació a la grandària del genoma fins a arribar

aproximadament a uns 10 Gb/1C, mida a partir de la qual s'observa un canvi de dinàmica que suggereix una saturació de l'increment de DNA repetitiu, i fins i tot un declivi en algunes espècies. L'escenari anterior és compatible amb les evidències prèvies descrites en apartats anteriors, que apunten una reducció del dinamisme i una taxa baixa d'eliminació de DNA repetitiu.

Seqüenciant genomes: quins, com i per què és important tenir un coneixement bàsic dels trets genètics?

La publicació del primer genoma vegetal, seqüenciat de l'espècie model *Arabidopsis thaliana* (*Brassicaceae*) (Initiative, 2000), representà el punt de partida de més de dues dècades de recerca en aquest camp que han donat lloc a la publicació de centenars de genomes disponibles en repositoris públics i d'accés obert a la comunitat científica. No obstant això, com s'ha descrit en apartats anteriors, els genomes vegetals són particularment diversos pel que fa a la mida i a la complexitat citotípica, característiques que moltes vegades han descoratjat investigadors i investigadores, ateses la magnitud de la tasca d'assemblatge i les limitacions de la tecnologia disponible durant molts anys. És per això que la distribució taxonòmica dels genomes seqüenciats no ha estat equilibrada, amb una representació esbiaixada (Vallée *et al.*, 2016). Els últims anys, però, han significat una revolució constant pel que fa a les tecnologies disponibles de seqüenciació i d'anàlisi computacional (Suzuki, 2020). Actualment podem analitzar cada vegada fragments més llargs de DNA amb tecnologies com PacBio Iso-Seq i HiFi, Oxford Nanopore i l'ancoratge Hi-C, que permeten assemblar cromosomes amb alta fidelitat i més contigüitat, la qual cosa d'alguna manera obri la possibilitat de poder començar a pensar en la viabilitat de projectes i iniciatives adreçats a seqüenciar qualsevol espècie vegetal independentment dels trets genètics que tinga (Kress *et al.*, 2022). Tanmateix, aquests autors també fan una anàlisi sobre la qualitat dels assemblatges publicats fins al present i posen de rellevància que només una petita fracció compleix els estàndards mínims que s'exigeixen actualment (còntigs amb un N50 d'1 Mb i ancoratges amb un N50 de 10 Mb). Part d'aquestes limitacions rau en el fet que les noves tecnologies de seqüenciació requereixen extraccions de DNA d'alt pes molecular, i aquest punt pot resultar limitant en algunes

espècies, atesa la complexitat química del seu citosol, que pot dificultar el procés d'extracció amb els estàndards de qualitat requerits. En l'era dels grans consorcis i iniciatives de seqüenciació globals, una solució per a aquest problema hauria de raure en la creació de més fòrums de col·laboració, que facilitaren l'accés a protocols i a serveis tecnològics que ajudarien a superar aquests reptes d'una manera més eficient. Igualment, i com era esperable, aquesta «febre de l'or» per la seqüenciació no arriba de manera paritària arreu del planeta. De fet, ja són alguns els autors que alerten sobre la falta de vincles entre l'origen geogràfic de les espècies i dels equips investigadors que les estudien —que moltes vegades descuiden establir i mantenir col·laboracions amb equips de recerca i comunitats locals (Marks *et al.*, 2021)— i que apunten la necessitat d'encoratjar una implicació més gran per part de la comunitat científica.

Segons la publicació recent sobre l'estat de la qüestió de l'abast de la seqüenciació de genomes de *Viridiplantae* (Kress *et al.*, 2022), actualment s'han seqüenciat ja 812 espècies, incloent 543 angiospermes, 11 gimnospermes, 5 pteridòfits, 8 briòfits i 249 algues (figura 1b). De fet, entre aquestes últimes hi trobem alguns dels genomes seqüenciats més petits, com és el cas d'*Ostreococcus tauri* (*Prasinophyceae*), amb un genoma de tan sols 12,6 Mb i que representa un dels organismes eucariotes de vida lliure més diminuts que coneixem (Derelle *et al.*, 2006). Un altre exemple, de dimensions similars a l'anterior però que en aquest cas correspon a una alga paràsita, és *Helicosporidium* sp. (*Chlorellaceae*), amb un genoma compacte de només 12,3 Mb (Pombert *et al.*, 2014). Entre els més grans, cal destacar el genoma de *Pinus lambertiana* (*Pinaceae*), que pertany a les gimnospermes i presenta un genoma de 27,6 Gb (Stevens *et al.*, 2016). Més recentment, l'anunci de la seqüenciació del genoma de *Paris polyphylla* (*Melanthiaceae*, angiospermes) ha representat un gran salt quantitatiu, ja que presenta un assemblatge d'aproximadament 80 Gb (Li *et al.*, 2020). Aquest projecte no ha estat exempt de debat, atesa la gran discordança que s'observa entre la mida del genoma assemblat i les dades de què disposem obtingudes amb citometria de flux, les quals indiquen que les espècies diploides del gènere *Paris*, si bé tenen genomes gegants, varien entre 30 i 55 Gb (Pellicer *et al.*, 2014). Certament, la metodologia emprada per a estimar la grandària del genoma ha estat un punt conflictiu, per l'impacte que té en els resultats obtinguts. La

citometria de flux s'ha consolidat durant molts anys com una tècnica altament efectiva i robusta per a estimar la quantitat de DNA (Doležal *et al.*, 2007). Molts projectes de seqüenciació, però, basen estrictament l'estimació de la mida del genoma en l'anàlisi de distribució de *k*-mers o en la representació de seqüències en assemblatges contigus (Sun *et al.*, 2018; Pucker, 2019), que poden produir resultats diferents dels obtinguts amb citometria de flux segons algunes característiques inherents del genoma (p. ex., poliploidia, nivell d'heterozigosi, contingut de DNA repetitiu). D'altres simplement usen valors previs publicats i que potser no són representatius de l'individu analitzat, atesa l'elevada taxa de poliploidització dels genomes vegetals. L'impacte que pot tenir el mètode de quantificació sobre la mida estimada del genoma ha estat estudiat en llinatges d'organismes molt variats, com per exemple en alguns fongs (Kooij i Pellicer, 2020), insectes (Pflug *et al.*, 2020) i en plantes (Al-Qurainy *et al.*, 2021), entre d'altres. Molts d'aquests estudis han trobat diferències entre els valors obtinguts depenent de la metodologia emprada, tot i que no sembla que hi haja un patró comú clar que apunte la sobreestimació i/o infraestimació de la mida del genoma dels mètodes basats en l'anàlisi de seqüències. Per tot això, un punt clau a la base de qualsevol projecte de seqüenciació hauria d'incloure una caracterització citogenètica de l'individu que s'haja d'estudiar que com a mínim incloguera un recompte de cromosomes i la quantificació del contingut de DNA nuclear mitjançant citometria de flux.

Conclusions

El coneixement dels mecanismes responsables de mantenir la biodiversitat requereix, entre d'altres, iniciatives que faciliten l'estudi en profunditat del genoma dels organismes eucariotes, que permeten obtenir-ne una visió crítica i robusta sobre l'estructura i el funcionament. En el cas de les plantes, a més, volem posar de manifest la necessitat de continuar generant dades fonamentals referents al contingut de DNA nuclear i al nombre de cromosomes, que són molt rellevants i, a més a més, aporten informació bàsica, però essencial, a l'hora d'assemblar un genoma; una necessitat que es veu reforçada pel fet que aquests camps estan lluny d'haver estat explorats en profunditat per a molts dels grups taxonòmics (p. ex., només prop del 3 % de les espècies d'angiospermes tenen una mida del genoma coneguda [Pellicer *et al.*, 2018]). Aconseguir

aquests reptes requereix estratègies de recerca coordinades a escala global com l'EBP i les seues filials regionals, que han de facilitar l'accés a la tecnologia, a l'intercanvi d'informació

i al desenvolupament de plans de formació. Prioritzar l'estudi comparatiu d'espècies amb un mostratge equilibrat i inclusiu, representatiu de la diversitat taxonòmica i geogràfica de

les plantes, així com dels seus trets genòmics, morfològics i ecològics serà, llavors, el gran repte d'aquestes iniciatives en els pròxims anys.

Bibliografia

- AL-QURAINY, F. [et al.] (2021). «Estimation of genome size in the endemic species *Reseda pentagyna* and the locally rare species *Reseda lutea* using comparative analyses of flow cytometry and *k*-mer approaches». *Plants*, 10: 1362.
- BANYAI, W. [et al.] (2010). «Overproduction of artemisinin in tetraploid *Artemisia annua* L.»». *Plant Biotechnol.*, 27: 427-433.
- BENNETZEN, J. L.; WANG, H. (2014). «The contributions of transposable elements to the structure, function, and evolution of plant genomes». *Ann. Rev. Plant Biol.*, 65: 505-530.
- BREINHOLT, J. W. [et al.] (2021). «A target enrichment probe set for resolving the flagellate land plant tree of life». *App. Plant Sci.*, 9: e11406.
- CASTIGLIONE, M.; CREMONINI, R. (2012). «A fascinating island: $2n = 4$ ». *Plant Biosyst.*, 146: 711-726.
- CLARK, J. [et al.] (2016). «Genome evolution of ferns: Evidence for relative stasis of genome size across the fern phylogeny». *New Phytol.*, 210: 1072-1082.
- CROSLAND, M. W. J.; CROZIER, R. H. (1986). «*Myrmecia pilosula*, an ant with only one pair of chromosomes». *Science*, 231: 1278.
- DERELLE, E. [et al.] (2006). «Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features». *Proc. Nat. Acad. Sci. USA*, 103: 11647-11652.
- DOLEŽEL, J. [et al.] (2007). «Estimation of nuclear DNA content in plants using flow cytometry». *Nat. Protoc.*, 2: 2233-2244.
- FLEISCHMANN, A. [et al.] (2014). «Evolution of genome size and chromosome number in the carnivorous plant genus *Genlisea* (Lentibulariaceae), with a new estimate of the minimum genome size in angiosperms». *Ann. Bot.*, 114: 1651-1663.
- GREGORY, T. R.; MABLE, B. K. (2005). «Polyploidy in animals». A: GREGORY, T.; RYAN, B. T. (ed.). *The evolution of the genome*. Burlington (MA): Academic Press, 427-517.
- GREILHUBER, J. [et al.] (2005). «The origin, evolution and proposed stabilization of the terms "genome size" and "C-value" to describe nuclear DNA contents». *Ann. Bot.*, 95: 255-260.
- HAUFLER, C. H. [et al.] (2014). «Ever since Klekowski: Testing a set of radical hypotheses revises the genetics of ferns and lycophytes». *Am. J. Bot.*, 101: 2036-2042.
- HEGARTY, M. J.; HISCOCK, S. J. (2008). «Genomic clues to the evolutionary success of polyploid plants». *Curr. Biol.*, 18: R435-R444.
- HIDALGO, O. [et al.] (2017). «Genomic gigantism in the whisk-fern family (*Psilotaceae*): *Tmesipteris obliqua* challenges record holder *Paris japonica*». *Bot. J. Linn. Soc.*, 183: 509-514.
- (2021). «La botànica: entre art i ciència». A: REDONDO, M.; FIGUERAS, M. (ed.). *HerbArt: Confluències entre art i ciència*. Barcelona: Universitat de Barcelona, 226-243.
- INITIATIVE, T. A. G. (2000). «Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*». *Nature*, 408: 796-815.
- JOHNSON, M. G. [et al.] (2019). «A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using *k*-medoids clustering». *Syst. Biol.*, 68: 594-606.
- KELLY, L. J. [et al.] (2012). «Why size really matters when sequencing plant genomes». *Plant Ecol. Div.*, 5: 415-425.
- (2015). «Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size». *New Phytol.*, 208: 596-607.
- KHANDELWAL, S. (1990). «Chromosome evolution in the genus *Ophioglossum* L.»». *Bot. J. Linn. Soc.*, 102: 205-217.
- KOOIJ, P. W.; PELLICER, J. (2020). «Genome size versus genome assemblies: Are the genomes truly expanded in polyploid fungal symbionts?»». *Genome Biol. Evol.*, 12: 2384-2390.
- KRESS, W. J. [et al.] (2022). «Green plant genomes: What we know in an era of rapidly expanding opportunities». *Proc. Nat. Acad. Sci. USA*, 119: e2115640118.
- LANDIS, J. B. [et al.] (2018). «Impact of whole-genome duplication events on diversification rates in angiosperms». *Amer. J. Bot.*, 105: 348-363.
- LESSE, H. (1970). «Les nombres de chromosomes dans le groupe de *Lysandra argester* et leur incidence sur sa taxonomie [*Lep. Lycaenidae*]». *Bull. Soc. Entomol. France*, 75 (3-4): 64-68.
- LI, J. [et al.] (2020). «An enormous *Paris polyphylla* genome sheds light on genome size evolution and polyploid biogenesis». *BioRxiv*, 2020.06.01.126920.
- LI, Z. [et al.] (2018). «Multiple large-scale gene and genome duplications during the evolution of hexapods». *Proc. Nat. Acad. Sci. USA*, 115: 4713-4718.
- LUGHADHA, E. N. [et al.] (2016). «Counting counts: Revised estimates of numbers of accepted species of flowering plants, seed plants, vascular plants and land plants with a review of other recent estimates». *Phytotaxa*, 272: 82-88.
- MACAS, J. [et al.] (2015). «In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe *Fabeae*». *PLOS ONE*, 10: 1-23.
- MARKS, R. A. [et al.] (2021). «Representation and participation across 20 years of plant genome sequencing». *Nat. Plants*, 7: 1571-1578.
- NOVÁK, P. [et al.] (2020). «Repeat-sequence turnover shifts fundamentally in species with large genomes». *Nat. Plants*, 6: 1325-1329.
- NYSTEDT, B. [et al.] (2013). «The Norway spruce genome sequence and conifer genome evolution». *Nature*, 497: 579-584.
- OGUR, M. [et al.] (1951). «Nucleic acids in relation to cell division in *Lilium longiflorum*». *Exp. Cell. Research*, 2: 73-89.
- OTTO, S. P.; WHITTON, J. (2000). «Polyploid incidence and evolution». *Annu. Rev. Genet.*, 34: 401-437.
- PANCHY, N. [et al.] (2016). «Evolution of gene duplication in plants». *Plant Physiol.*, 171: 2294-2316.
- PARISOD, C. [et al.] (2010). «Evolutionary consequences of autopolyploidy». *New Phytol.*, 186: 5-17.
- PELLICER, J. [et al.] (2010). «The largest eukaryotic genome of them all?»». *Bot. J. Linn. Soc.*, 164: 10-15.
- (2014). «A universe of dwarfs and giants: Genome size and chromosome evolution in the monocot family *Melanthiaceae*». *New Phytol.*, 201: 1484-1497.
- (2018). «Genome size diversity and its impact on the evolution of land plants». *Genes*, 9: 88.
- (2022). «Morphological and genome-wide evidence of homoploid hybridisation in *Urospermum* (Asteraceae)». *Plants*, 11: 182.
- PELLICER, J.; LEITCH, I. J. (2020). «The Plant DNA C-values Database (release 7.1): An updated online repository of plant genome size data for comparative studies». *New Phytol.*, 226: 301-305.
- PFLUG, J. M. [et al.] (2020). «Measuring genome sizes using read-depth, *k*-mers, and flow cytometry: Methodological comparisons in beetles (*Coleoptera*)». *G3 Genes|Genomes|Genetics*, 10: 3047-3060.
- PIEGU, B. [et al.] (2006). «Doubling genome size without polyploidization: Dynamics of retrotransposon-driven genomic expansions in *Oryza australiensis*, a wild relative of rice». *Genome Research*, 16: 1262-1269.
- POMBERT, J. F. [et al.] (2014). «A lack of parasitic reduction in the obligate parasitic green alga *Helicosporidium*». *PLOS Genetics*, 10: e1004355.
- PUCKER, B. (2019). «Mapping-based genome size estimation». *BioRxiv*, 607390.
- RASTOGI, S.; OHRI, D. (2020). «Chromosome numbers in Gymnosperms – an update». *Silvae Genet.*, 69: 13-19.
- RICE, A. [et al.] (2019). «The global biogeography of polyploid plants». *Nat. Ecol. Evol.*, 3: 265-273.
- SCHAAL, B. (2019). «Plants and people: Our shared history and future». *Plants, People, Planet*, 1: 14-19.
- SOFI, P. [et al.] (2007). «Sequencing the maize genome: Rationale, current status and future prospects». *Curr. Sci.*, 92: 1702-1708.
- SOLTIS, P. S. [et al.] (2016). «Ancient WGD events as drivers of key innovations in angiosperms». *Curr. Opin. Plant Biol.*, 30: 159-165.
- SONNLEITNER, M. [et al.] (2016). «Ecological differentiation of diploid and polyploid cytotypes of *Senecio carniolicus sensu lato* (Asteraceae) is stronger in areas of sympatry». *Ann. Bot.*, 117: 269-276.
- STEVENS, K. A. [et al.] (2016). «Sequence of the sugar pine megagenome». *Genetics*, 204: 1613-1626.
- STUESSY, T. F. (2009). *Plant taxonomy: The systematic evaluation of comparative data*. Nova York: Columbia University Press.
- (2011). «Multiple sources of comparative data for creative monography». A: STUESSY, T. F.; LACK, H. (ed.). *Monographic plant systematics: Fundamental assessment of plant biodiversity*. Ruggell: A. R. G. Gantner Verlag K. G., 33-47.
- SUN, C. [et al.] (2012). «Slow DNA loss in the gigantic genomes of salamanders». *Genome Biol. Evol.*, 4: 1340-1348.
- SUN, H. [et al.] (2018). «*findGSE*: Estimating genome size variation within human and *Arabidopsis* using *k*-mer frequencies». *Bioinformatics*, 34: 550-557.
- SUZUKI, Y. (2020). «Advent of a new sequencing era: Long-read and on-site sequencing». *J. Hum. Genet.*, 65: 1.
- TWYFORD, A. D. (2018). «The road to 10,000 plant genomes». *Nat. Plants*, 4: 312-313.
- VALLÉE, G. C. [et al.] (2016). «Economic importance, taxonomic representation and scientific priority as drivers of genome sequencing projects». *BMC Genomics*, 17: 782.
- WANG, X. [et al.] (2021). «Genome downsizing after polyploidy: Mechanisms, rates and selection pressures». *Plant J.*, 107: 1003-1015.
- WENDEL, J. F. (2015). «The wondrous cycles of polyploidy in plants». *Amer. J. Bot.*, 102: 1753-1756.
- WICKER, T. [et al.] (2007). «A unified classification system for eukaryotic transposable elements». *Nature Rev. Genet.*, 8: 973-982.
- WOOD, T. E. [et al.] (2009). «The frequency of polyploid speciation in vascular plants». *Proc. Nat. Acad. Sci. USA*, 106: 13875-13879.