

„Дървото на живота“ и неговите граници. Какво определя и контролира размера и сложността на генома на организмите?

Десислава Нешева

В началото на юни авторитетни научни медии разпространиха като курйоз новината, че елементарната тропическа папрат *Mesipteris oblongeolata* е рекордьор на цялата природа, защото има ДНК с повече от 160 млрд. генетични знака. За сравнение – човешката ДНК има „само“ около 3 млрд. Растението е примитивно, епифит без корени, не може да съществува самостоятелно, а расте върху други растения. Около тази новина се повдигнаха реди въпроси: как и защо се получава толкова дълъг геном, каква информация се кодира в тази верига и защо далеч по-сложните висши бозайници имат по-къса ДНК?

Всъщност това даде добър повод да се върнем отново към базовите генетични понятия. И тук за начало може би трябва да припомним „голямата тайна на живота“ – всички живи форми в природата се обединяват от едно общо нещо – от наличието на нуклеинови киселини (НК) и предаването на носената от тях информация. Независимо дали са ДНК молекули (Дезоксирибонуклеинова киселина) или РНК молекули (Рибонуклеинова киселина при някои вируси), те съдържат градивни структури, наречени нуклеотиди.

Всеки градивен елемент е съставен от три части: 5-въглероден монозахарид-пентоза-дезоксирибоза (ДНК) или рибоза (РНК), фосфатна група и азотна база.

Всеки мономер се свързва със съседния посредством 3', 5' – фосфодиестерни връзки, изграждайки различни по дължина полинуклеотидни вериги. Образоването на двойноверижните молекули ДНК и двойноверижните участъци при РНК става благодарение способността на някои бази да се свързват помежду си. Винаги пуриновите (големите) бази се



Mesipteris oblongeolata е примитивно растение, чийто геном е около 50 пъти по-голям от човешкия

Източник: Fernandez et al., A 160 Gbp fork fern genome shatters size record for eukaryotes iScience 27,109889, June 2024 The Elsevier <https://doi.org/10.1016/j.isci.2024.109889>

сдвояват комплементарно с пиримидиновите (малки) бази.

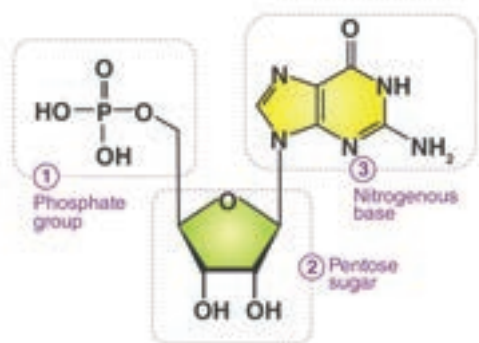
При прокариотите (едноклетъчни организми без ядро в клетката, повечето бактерии) и еукариотите (организми, чиито клетки притежават ядро, включват животни, растения, гъби) ДНК притежава функцията да съхранява и предава в поколенията генетичната информация за синтеза на белтъците и РНК в клетката, докато основната роля на РНК е в реализирането на генетичната информация.

Функцията на организмовия геном е да пренесе надеждно информацията от едно поколение на следващото. Преносът на информацията в живите форми винаги става в една посока ДНК – РНК – Белтък. Като изключение са РНК вирусите, които използват обратна транскриптаза и сложен механизъм, за да синтезират на базата на своята РНК верига ДНК, която ще бъде внедрена в генома на клетката гостоприемник и ще използва неговия механизъм на репликация, транскрипция и транслация за синтеза на нови вирусни частици.

ДНК се реплицира, за да съхрани своя генетичен материал в дъщерните клетки. Тя става и матрица за синтеза на информационна РНК (иРНК), която след процесиране предава информацията за синтеза на полипептидни вериги. Те влизат в състава на ензими и различни белтъчни компоненти, необходими за нормалното функциониране на клетките и организмите. Разнообразието на продуктите, които се получават по време

на транскрипцията и транслацията се дължи на способността на генома да копира множество отделни части от себе си – т.н. транскрипти. Един ген може да даде начало на множество транскрипти чрез алтернативен сплайсинг (смаждане) и чрез алтернативни места за начало и край на транскрипцията.

Друга особеност, която обединява организмовия свят е генетичният код. Той представлява непрекъсната, неприпокриваща се последователност от по



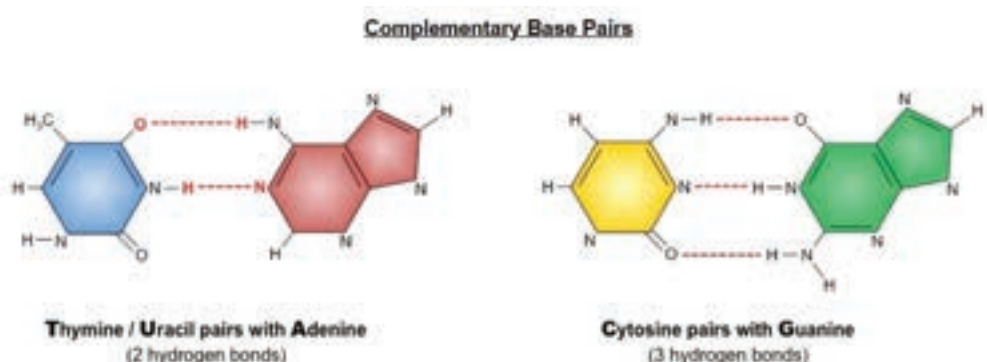
Структура на нуклеотид (база)

Източник: <https://byjus.com/neet/nucleotide/>



Структура на Пуринови (големи) и Пиримидинови (малки) бази

Източник: <https://old-ib.bioninja.com.au/standard-level/topic-2-molecular-biology/26-structure-of-dna-and-rna/nitrogenous-bases.html>



Комплементарно свързване на базите. Аденин се свързва с Тимин/Урацил с две водородни връзки, а Гуанин с Цитозин с три водородни връзки

Източник: <https://old-ib.bioninja.com.au/standard-level/topic-2-molecular-biology/26-structure-of-dna-and-rna/nitrogenous-bases.html>

3 нуклеотида (когони) на зрялата информациялна РНК, които биват разпознавани и свързвани на принципа на комплементарността от тринуклеотидни последователности (антикогони) в мРНК.

Всеки един кодон има определена функция в зависимост от комбинацията на базите в тях. Те са универсални за всеки вид. Кодонът АУГ, който кодира метионин, винаги започва кодиращата протеин част на всяка молекула иРНК. Съществуват когони, свързващи се с

мРНК, отговорни за 22 аминокиселини, а наличието на един от кодоните УАА, УАГ и УГА е сигнал за завършване на протеин-кодиращата част на всяка молекула иРНК.

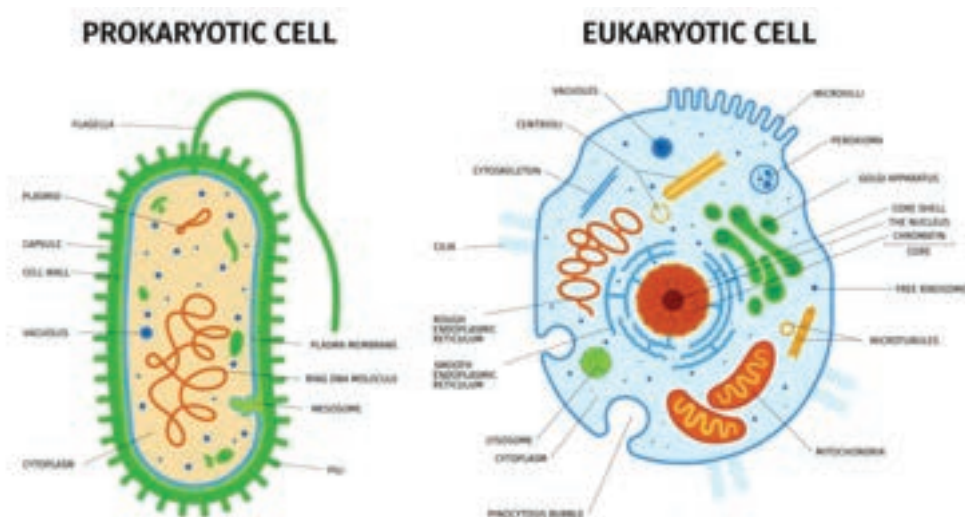
В зависимост от своята функция и видовата принадлежност, всяка клетка има различен по големина геном, който съдържа определен набор от белтък-кодиращи гени. При човека геномът е в размер на 3 милиарда базови двойки. Съдържа кодиращи и не кодиращи участъци.

Част от некодиращата ДНК в еукариотите се дължи на дълги ДНК последователности, които се намират между гените, а големи количества се намират и в самите гени. Такива гени имат разделена структура, в която сегменти от кодираща последователност (екзони) са разделени от некодиращи последователности (интрони).

Друг фактор, допринасящ за големия размер на еукариотните геноми е, че някои гени се повтарят много пъти. Докато повечето прокариотни гени са представени само веднъж в генома, много еукариотни гени присъстват в множество копия, наречени генни семейства. В някои случаи са необходими множество копия на гени, за да се произведат РНК или протеини, необходими в големи количества. В други случаи различни членове на генно семейство могат да бъдат транскрибирани в различни тъка-

ни или на различни етапи на развитие. Например, α - и β - субединиците на хемоглобина са кодирани от генни семейства в човешкия геном, като различни членове на тези семейства се експресират в ембрионалните тъкани, фетуса и при възрастните. Смята се, че генните семейства са възникнали чрез дублиране на оригинален ген на предците, като различни членове на семейството след това се разминават в резултат на мутации по време на еволюцията. Такава дивергенция може да доведе до еволюцията на свързани протеини, които са оптимизирани да функционират в различни тъкани или на различни етапи на развитие.

Значителна част от еукариотните геноми се състои от силно повтарящи се некодиращи ДНК последователности (от 10^5 до 10^6 пъти). Други повтарящи се ДНК последователности са разпръс-



Структура и характеристика на прокариотна и еукариотна клетка и разликите в организацията на генетичния им материал

Източник: <https://www.biologynotes.site/prokaryotic-vs-eukaryotic-cells-structural-differences-and-cellular-process-comparison/>

нати из целия геном, вместо да бъдат групирани като тандемни повторения. Тези последователности се класифицират като *SINEs* (къси разпръснати елементи) или *LINEs* (дълги разпръснати елементи). Основните *SINEs* в геномите на бозайници са *Alu*-секвенциите, така наречени, защото обикновено съдържат едно място за рестрикционната ендонуклеаза *Alu I*. Те са дълги приблизително 300 базови двойки и около милион такива последователности са разпръснати из целия геном, което представлява почти 10% от общата клетъчна ДНК. Въпреки че *Alu* последователностите се транскрибират в РНК, те не кодират протеини и тяхната функция е неизвестна. Основните човешки линии (които принадлежат към семейството *LINE 1* или *L1*) са с дължина около 6000 базови двойки и се повтарят приблизително 50 000 пъти в генома. *L1* последователностите се транскрибират и поне някои кодират протеини, но подобно на *Alu* последователностите, те нямат известна функция в клетъчната физиология. Наричат ги и *транспозоми*, които са способни да се придвижват към различни места в геномната ДНК. Някои от тях могат да помогнат за регулиране на генната експресия, но повечето *Alu* и *L1* последователности изглежда не дават полезен принос за клетката. Те обаче може да са изиграли важна еволюционна роля.

Схващането, че колкото по-сложен е гаден организъм, толкова по-дълъг е неговият геном се оказва неточно. Вирсният и прокариотният геном по правило е по-малък от еукариотният, но сред еукариотните видове съществуват разлики в съотношението между размера на генома и сложността на организмите. Например, геномите на саламандрите и лилиите съдържат над



Десислава Нешчева е възпитаник на Биологическия факултет на СУ „Св. Климент Охридски“. Бакалавър по молекулярна биология (2006) и магистър по генетика (2008). Работи в Катедра по Медицинска генетика към Медицински университет – София. Доктор по генетика от 2016 г. Специализирала е 2 години популационна, еволюционна и антропогенетика във Флорентинския университет, Италия. Участва в множество проекти и има интерес в области, свързани с изследване на онкологични заболявания, вродени малформации, проблеми на дълголетието и социално значими болести.

гесет пъти повече ДНК от човешкия геном, но тези организми очевидно не са гесет пъти по-сложни от хората.

Този очевиден парадокс е решен чрез откритието, че геномите на повечето еукариотни клетки съдържат не само функционални гени, но и големи количества ДНК последователности, които не кодират протеини. По този начин разликата в размерите на генома на саламандъра и на човека отразява по-големи количества некодиреща ДНК, а не повече гени в генома на саламандъра. Следователно хилядократно по-големият размер на човешкия геном в сравнение с този на *Escherichia coli* (*E. coli*) не се

дължи единствено на по-големия брой човешки гени. Смята се, че човешкият геном съдържа приблизително 100 000 гена – само около 25 пъти повече от *E. coli*. Но голяма част от сложността на еукариотните геноми е резултат от изобилието от няколко различни типа некодиращи последователности, които съставляват по-голямата част от ДНК на висшите еукариотни клетки.

Освен това за съществуването на даден организъм са важни определен критичен брой гени. Проучване от 2015 г. при човека, посочва важните 3230 гени. Оказва се, че около 15% от нашите 20 000 гена са толкова критични за съществуването ни, че определени промени могат да ни убият, преди да се родим. Когато един от тези гени мутира, ембрионът обикновено умира или човекът е твърде болен, за да се възпроизвежда – така тези промени се елиминират. Около 20% от човешките гени вече са свързани с някакви патологични изменения, но много от тях все още не са и твърда се изследват. Поради своята важност тези гени могат да бъдат специално наблюдавани по време на разработването на лекарства за потенциални странични ефекти и цитотоксичност. Посоченият брой от 3230 гена не е пълният набор от основни гени в човешкото тяло и чрез изучаване на повече екзони (кодиращите последователности на гена), изследователите ще могат да прецизират това число. Екзоните не покриват междугенната ДНК, която помага за регулирането на генната активност и вариациите там също могат да бъдат важни.

В бактериалните геноми по-голямата част от ДНК кодира протеини. Например, геномът на *E. coli* е дълъг приблизително $4,6 \times 10^6$ базови двойки и съдържа 4288 гена, като почти 90% от ДНК се

използва като протеин-кодираща последователност.

Геномът на дрождите, който се състои от 12×10^6 базови двойки, е около 2,5 пъти по-голям от размера на генома на *E. coli*. Около 4% от гените на *Saccharomyces cerevisiae* (Дрожги) обикновено имат само един малък интрон близо до началото на кодиращата последователност. Това прави техния геном също доста компактен. Приблизително 70% от генома на дрождите се използва като протеин-кодираща последователност, определяйки общо около 6000 протеина.

Геномът на нематода *Caenorhabditis elegans* (*C. elegans*) е 97×10^6 базови

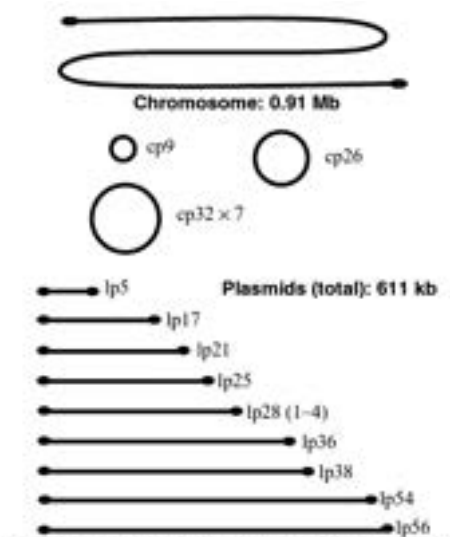


Fig. 3.3. Schematic representation of the total genome of *Borrelia burgdorferi sensu stricto* B31.

Организация и размер на генома на *Borrelia burgdorferi* – причинител на лаймската болест. Хромозомата при Борелията е линейна двойноверижна ДНК, подобна на тази на еукариотите. Геномът ѝ е съставен от линейна хромозома от 910 725 базови двойки и най-малко 17 линейни и кръгови плазмиди с общ размер повече от 533 000 базови двойки
Източник: https://microbewiki.kenyon.edu/index.php/Borrelia_burgdorferi_NEU2011

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

Генетичният код. Таблица, представяща организацията на кодоните (тринуклеотидите) за всяка една аминокиселина, стартовият кодон АУГ за Метионин, както и Стоп кодоните УАА, УАГ и УГА, които са сигнал за прекратяване на транслацията

Източник: <https://byjus.com/biology/genetic-code/>

двойки и съдържа приблизително 19 000 гени, кодиращи протеини, което го изпраща на междинно място по размер и сложност между геномите на дрождите и бозайниците. Докато геномът на *C. elegans* е 8 пъти по-голям от този на дрождите, той съдържа само около три пъти повече гени. Това корелира с наличието на значителен брой интрони в *C. elegans*. Всеки ген в *C. elegans* съдържа средно пет интрона и обхваща средно 5000 бази. В съответствие с това, само около 25% от последователността на *C. elegans* съответства на екзони, срещу 70% протеин-кодираща последователност в генома на дрождите. Въпреки че геномът на *Drosophila melanogaster* (*D. melanogaster*, Винена мушица) е 180×10^6 базови двойки, *D. melanogaster* съдържа по-малко гени от *C. elegans* (около 13 600). Следователно протеин-кодираща-

та последователност съответства само на около 13% от генома на *D. melanogaster*.

Геномите на висшите животни (включително на човека) са още по-сложни и съдържат големи количества некодираща ДНК. По този начин се очаква само малка част от 3×10^9 базови двойки на човешкия геном да съответстват на протеин-кодираща последователност. Приблизително една трета от генома съответства на силно повтарящи се последователности, оставяйки приблизително 2×10^9 базови двойки за функционални гени, псевдогени и неповтарящи се спейсерни последователности. Ако средният ген обхваща 10 000 – 20 000 базови двойки (включително интрони), може да се очаква човеш-

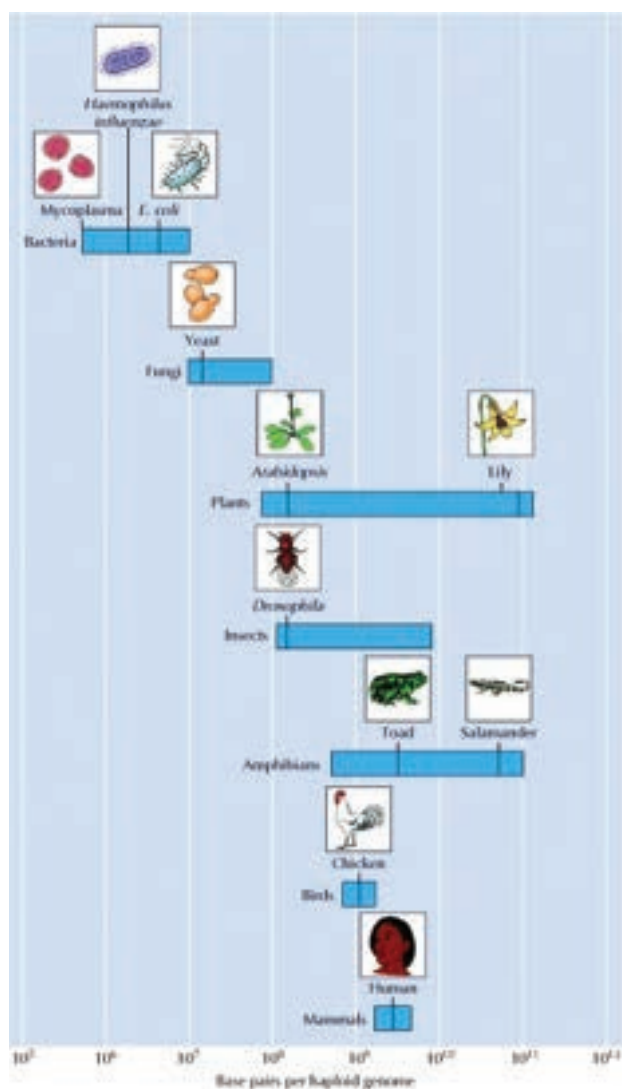
кият геном да се състои от около 100 000 гена, като последователностите, кодиращи протеини, съответстват само на около 3 – 5% от човешката ДНК.

Неопровержимо доказателство, че размерът и сложността на даден организъм не отговаря пропорционално на размера на неговият геном се дава от изследването, за което стана дума в началото на статията. През 2024 г. бе съобщено, че геномът на *Tmesipteris oblancoolata* е 50 пъти по-дълъг от този на човека и със своите 160 млрд. базови двойки ($160.45 \text{ Gbp}/1\text{C} - 1\text{C}$ е съдържание на ядрена ДНК в гаметиночно ядро) е рекордьор на досега проучените организми. Това откритие отваря нови пътища за изследване на еволюционната динамика на геномния гигантизъм.

Васкуларните растения са изключително интересни сред еукариотите по-

ради отличителното им разнообразие в размера на генома, което варира около 2400 пъти. Проведените изследвания на близо 20 000 еукариотни генома разкрива удивително количество видове с големи размери на генома, разпределени в еукариотното „Дърво на живота“. Тези екстремни размери повлияват размерите на клетките, жизнените цикли, физиологията и морфологията и в крайна сметка оказват влияние върху екологията и еволюцията на видовете. Геномите на някои представители на гъбичките с миниатюрен размер, включително най-малкият еукариотен геном, открит в микроспоридиите *Encephalitozoon intestinalis* (2,6 Mbp/1C), контрастират с тези, открити в други групи, където геномите са достигнали размери до пет порядъка. В рамките на няколко животински и растителни вида са установени екстремни геномни размери над 100 Gbp/1C. Сред животните с големи геноми са представители на хордови – двайнодошащи, белодробни риби (клас *Dipnoi*), където достига до 129,90 Gbp /1C при *Protopterus aethiopicus* (африканска двайнодошаща риба) и протеидни саламандри (клас *Amphibia*), с геном до 117,47 Gbp/1C при *Necturus lewisi* (Водно куче от река Неусе).

От растителните представители васкуларни растения включват шест от първите десет най-големи известни еукариотни генома. Въпреки че повече



Диапазонът от размери на геномите на представителни групи организми е показан в логаритмична скала
Източник: The Cell, A Molecular Approach. Sunderland (MA): Sinauer Associates; 2000

мо от тях са покритосеменни, особено в семействата едносеменни *Liliaceae* и *Melanthiaceae* (*Paris japonica*, бяло цъфтящо растение от Япония), екстремни геноми също са докладвани в паразитния евдикот *Viscum album* (Бял имел,

Organism	Genome size (Mb) ^a	Protein-coding sequence (percent) ^b	Number of genes ^b
<i>E. coli</i>	4.6	90	4,288
<i>S. cerevisiae</i>	12	70	5,885
<i>C. elegans</i>	97	25	19,099
<i>Drosophila</i>	180	13	13,600
Human	3,000	3	100,000

a Mb = millions of base pairs.

b Determined from the *E. coli*, *S. cerevisiae*, *C. elegans*, and *Drosophila* genomic sequences and estimated for the human genome.

Големината на кодона, броя на гените и процентът на протеин-кодиращите гени в представителна извадка организми.

Източник: *The Cell, A Molecular Approach*. Sunderland (MA): Sinauer Associates; 2000

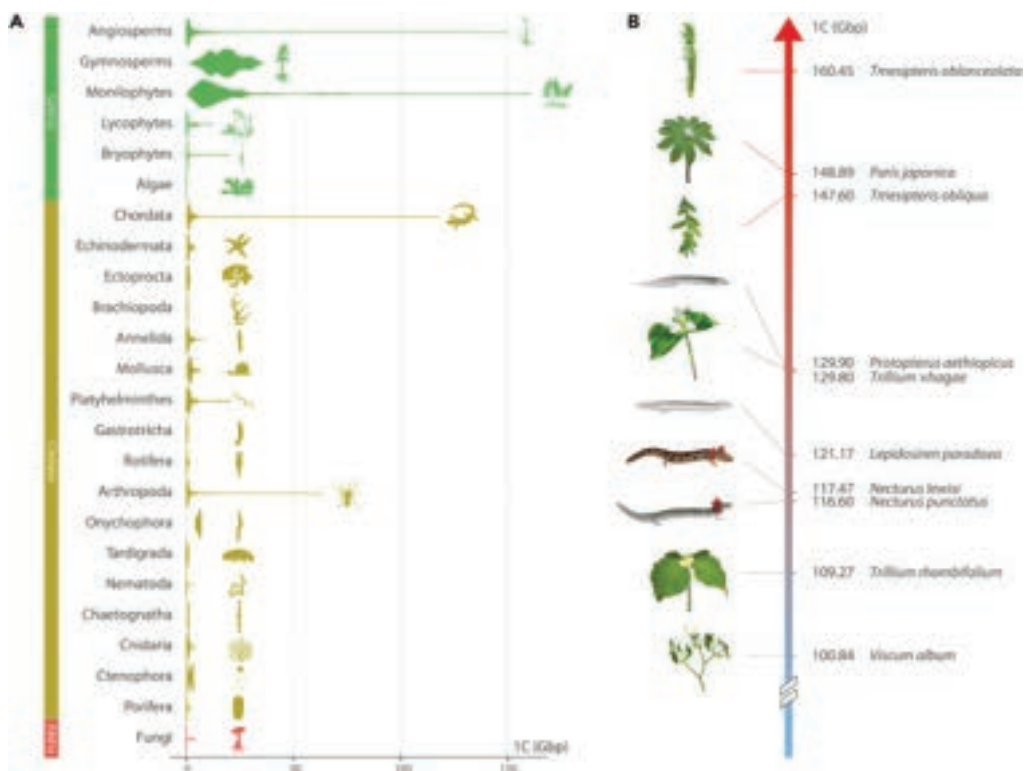
сем. *Santalaceae*, 100,84 Gbp/1C) и в направоудните в семейство *Psilotaceae* (*Tmesipteris obliqua*, дълга вилицоподобна напрам, 147.60 Gbp/1C). Този огромен обхват от вариации в размера, привидно несвързани със сложността на организма и известни като „парадокса на С-стойността“ или „енигмата на С-стойността“, интригуват изследователите повече от половин век. Бързият напредък в методите за секвенирането на ДНК, предпоставя убедителни доказателства за това, че вариациите в количеството на ДНК възникват предимно от разликите в честотата на полиплоидизацията и множество некодираща, повтаряща се ДНК. Въпреки това, възниква въпросът дали сме разкрили пълната степен на разнообразието на размера на генома.

Наличието на голям брой хромозоми е по-често при напрамите, където дори се пише за „синдром на натрупване на хромозоми“, който е основно следствие от многобройните цикли на удвояване на целия геном.

Сред животните разширяването на геномите до гигантски мащаби досега се приписва на натрупването на повта-

рящи се елементи и запазването на некодираща последователност (въпреки че се съобщава за полиплоидизация при някои саламангри и белогробни риби с по-малки геноми). От друга страна, най-малките еукариотни геноми, както се срещат при виговете *Encephalitozoon* (*микроспоридии*), се открояват с това, че са изключително компактни, с намаляване на броя и размера на гените и с минимални количества на некодиращи, повтарящи се ДНК последователности. Като се имат предвид тези наблюдения, перспективата за намиране на много по-малки размери на еукариотния геном изглежда невероятна.

Възниква въпросът „Какво контролира размера и сложността на генома на организмите?“ Това е критичен въпрос за еукариотите организми като цяло защото изследванията показват, че има ясни биохимични, метаболитни и регулаторни механизми, свързани с увеличаване размера на ДНК. Тези механизми предизвикват каскадни ефекти, оказващи влияние върху биологията на организмите, тяхната екология и потенциала им да се запазят при еволюционните процеси.



Разнообразието в размера на генома при еукариотите

Източник: Ferná'ndez et al., A 160 Gbp fern genome shatters size record for eukaryotes *iScience* 27,109889, June 2024 The Elsevier <https://doi.org/10.1016/j.isci.2024.109889>

Повече от десетилетие се провеждаха изследвания, свързани с границите в размера на генома на еукариотите, за да се открие по-голям геном от този на *Paris japonica*, като не може напълно да се изключи възможността в бъдеще да бъдат открити още по-големи геноми. Независимо от това, множе-

ството физиологични, екологични и еволюционни механизми, свързани с геномни разширения в такива гигантски мащаби, най-вероятно предполагат, че ако горната граница все още не е достигната, то тази на *Tmesipteris oblongeolata* трябва да е много близо до нея. ■

НОВИНИ ОТ НАУКАТА

SOS! – опасна инвазивна мравка прониква в Европа

През м. септември, 2023 г. научни списания и публични медии в Европа разпространиха новината, че в Италия е открита за първи път активна колония от един опасен световен нашественик – Червената огнена мравка, чието научно име е *Solenopsis invicta*. Тя е позната и с името *RIFA* – съкращение на популярното английско име на мравката – *Red Imported Fire Ant*. Този опасен инвазивен вид отгава е включен и в списъка на Световната база данни за „100-те най-инвазивни видове в света“. Обявен е и за петият най-скъп инвазивен вид в страните на Южна Америка, Карибите, Австралия и САЩ, където ежегодно причинява около 6 милиарда долара загуби.

И в миналото червената огнена мравка е намирана в Европа и по-конкретно в Испания, Холандия и даже в по-северната Финландия, но

това са находки на единични мъртви екземпляри главно в трюмовете на търговски кораби с транспортирани стоки от други неевропейски страни. Тревожното съобщение на 7 италиански учени, начело с д-р Матиа Мачети (Dr Mattia Machetti), този път се отнася за откриването на няколко активни колонии на този опасен нашественик, намерени от научния екип на о-в Сицилия, в околностите на гр. Сиракуза. Кога и как е пристигнал на острова този силно инвазивен вид е неизвестно, но вече са открити 88 гнезда на мравките в района около гр. Сиракуза. Първите генетични анализи на намерените колонии, публикувани в научното списание *Current Biology* показали, че в Италия новите нашественици са внесени най-вероятно от Китай или САЩ с търговски стоки.

Червените огнени мравки не се различават по форма, размери и външни особености от останалите видове мравки, но са изключително активни, адаптивни към условията на средата и нанасят значителни вреди както на човека и човешката дейност, така и в природата. Те са топлолюбив вид и се очаква да се разпространят предимно в средиземноморските страни в Европа. На дължина варират между 2.4 – 6 мм., но се отличават със сравнително голяма глава, здрави хитинени челюсти, бързина на придвижванията, способност за бързо размножа-



Solenopsis invicta – общ вид на работна мравка



Мравка в готовност за ужилване на неприятеля

ване. Живеят обикновено в колонии с една царица, но са наблюдавани и „супер колонии“ с множество царици, които бързо умножават популацията на вида, когато условията на средата са благоприятни. Колониите на мравките са главно в почвата, но често образуват и малки струпвания от почва на повърхността.

Вредите, които червените огнени мравки нанасят, са многостранни, защото те са почти „всеядни“ и нападат както растения и гъби, така и много животински обитатели. Те са опасни хищници за гребните почвени безгръбначни животни – червеи, охлюви, многоножки, ларви и възрастни насекоми и гр. и сериозни вредители в силози, складове на различни хранителни и нехранителни продукти, посевите на полезни растения и гр. Като защитно средство срещу неприятели притежават отровни жлези с остро жило на коремчето и

причиняват болезнени ужилвания на човека и животните.

Борбата с червените огнени мравки е много трудна. В някои от южните щати на САЩ са правени много опити за унищожаването на мравките на техните територии, но безуспешно. Досега само Нова Зеландия е успяла и напълно е ликвидирала през 2015 г. присъствието на инвазивния вид на територията на острова, а проникването му отново е под строга карантина. Италианските учени считат, че крайбрежните средиземноморски градове и пристанища могат да улеснят най-бързо разпространението на този опасен нашественик не само в Италия, но и в много топли райони и страни в Европа и призовават експертите и обществеността за строга бдителност и разработване на мерки за контрол на популацията в Сицилия .

По Current Biology и The Gardian