

Ćwiczenie C05B Genomy

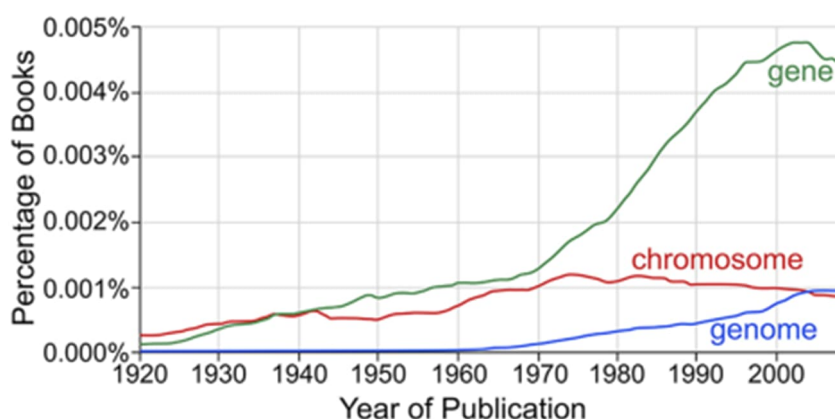
Genom i jego dynamika Wielkość genomu Wartość C Gęstość genów

Prof. dr hab. Roman Zieliński

1. Genom i jego dynamika

➔ 1.1. Definicja genomu

Genom jest często opisywany jako informacja genetyczna organizmu. Taki opis jest nieprecyzyjny i uproszczony, gdyż nie uwzględnia chociażby faktu występowania poliploidów, u których występuje więcej niż jedna kopia informacji genetycznej. Sam termin „genom” jest złożeniem słowa gen i chromosom i początkowo genom odnosił się do haploidalnego zestawu chromosomów. Termin został wprowadzony w 1920 r., ale nie był powszechnie używany i dopiero ostatnie dwie dekady zmieniły tę sytuację (Rys. 1.1).



Rys. 1.1. Procent książek anglojęzycznych, w których pojawiało się słowo genom w latach 1920-2000 (Goldman i Landweber, 2016).

Projekty sekwencjonowania genomów przyczyniły się do znacznego wzrostu ilości danych genomowych. W 2015 r. Bank Genów (NCBI) zawierał 321 miliardów par zasad pochodzących z 2 557 genomów Eukariota, 432 genomów Archaea i 7 474 genomów bakteryjnych. Pomimo tego,

definicja genomu wciąż stwarza problemy i nie jest jednoznaczna. Przykładowo, NIH (National Institute of Health) definiuje genom jako:

„Kompletny zestaw DNA danego organizmu obejmujący wszystkie geny. Każdy genom zawiera pełną informację do budowy i funkcjonowania organizmu.”

Definicja ta zawiera wiele nieścisłości, jest nieprecyzyjna i nie daje wyobrażenia czym jest genom ponieważ:

- Co oznacza kompletny zestaw DNA? Czy u diploida to suma genomów matki i ojca, czy też tylko jeden z nich?
- Genom zawiera nie tylko geny, ale także wiele sekwencji powtarzalnych, w tym transpozony. Czy one nie należą do genomu? Definicja NIH ogranicza genom jedynie do genów.
- Nie jest jasne czy podana definicja odnosi się jedynie do genomu jądrowego, czy obejmuje również genom mitochondrialny.

Z kolei biologia molekularna na ogół definiuje genom jako kompletną informację genetyczną komórki, czyli zawartą w jądrze komórkowym i mitochondriach (lub dodatkowo w chloroplastach w odniesieniu do roślin). W przypadku bakterii, która ma jeden chromosom (materiał genetyczny w nukleoidzie nosi nazwę chromosomu bakteryjnego), definicja jednoznacznie identyfikuje genom. Inaczej jest w przypadku organizmów np. diploidalnych, które mają po dwa chromosomy danego typu (homologiczne), jeden od ojca, drugi od matki. W związku z tym, czy sekwencja genomu dotyczy jednego chromosomu homologicznego czy obu? U organizmów poliploidalnych, mających nawet kilka chromosomów homologicznych sytuacja komplikuje się jeszcze bardziej.

Dlatego definicja genomu powinna odnosić się do podstawowego zestawu chromosomów (x) w jądrze komórkowym. Jednocześnie odniesienie do zestawów chromosomów homologicznych pozwala prawidłowo ocenić ile genomów jest w komórce.

Genom jądrowy to materiał (informacja) genetyczny zawarty w podstawowym zestawie chromosomów, x. Obejmuje on sekwencje kodujące i niekodujące. Genom jest strukturą dynamiczną, która podlega wpływom środowiska i wchodzi w interakcje z czynnikami zewnętrznymi. W komórkach Eukariota występuje także genom mitochondrialny, a u roślin dodatkowo genom chloroplastowy.

1.1.1. Liczba genomów



Proszę podać liczbę genomów jądrowych w komórkach wymienionych w punktach A-D.

- A. Komórka skóry człowieka oraz komórka jajowa człowieka.
- B. Truskawka jest oktaploidem. Ile genomów jest w komórkach liścia truskawki, a ile w komórce generatywnej ziarna pyłku?
- C. Karp królewski jest tetraploidem. Ile genomów jest w komórkach oskrzeli, a ile w plemniku.
- D. Ile genomów jest w komórce skóry człowieka, po zakończeniu fazy S?
- E. Ile genomów jest w spermatocytach człowieka w anafazie II podziału mejyotycznego?

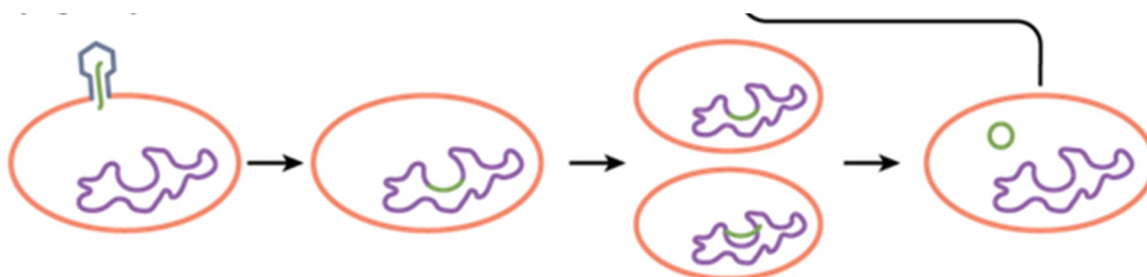
➔ 1.2. Dynamika genomu

Przez wiele lat genomy były postrzegane jako stabilne struktury, które utrzymują swoją integralność w ciągu całego cyklu życiowego. Stabilność genomu zakłada także, że blisko spokrewnione osobniki w obrębie populacji, a nawet gatunku, mają niemalże identyczne genomy. Ten tradycyjny pogląd zmienił się wraz z napływem ogromnej liczby danych z projektów sekwencjonowania genomów. Okazało się, że genomy wykazują znaczne zróżnicowanie wewnątrzgatunkowe, które obejmuje różnice pomiędzy liniami i osobnikami w odniesieniu do substytucji (SNP), niewielkich insercji, delecji, populacji transpozonów w obrębie danego fragmentu DNA, a nawet poliploidalności.

1.2.1. Dynamika genomów wirusów

Cykle życiowe wirusów są najbardziej oczywistym przykładem dynamiki genomu, która może obejmować wbudowanie genomu wirusowego do materiału genetycznego gospodarza, czy zmianę typu kwasu nukleinowego w trakcie cyklu replikacyjnego.

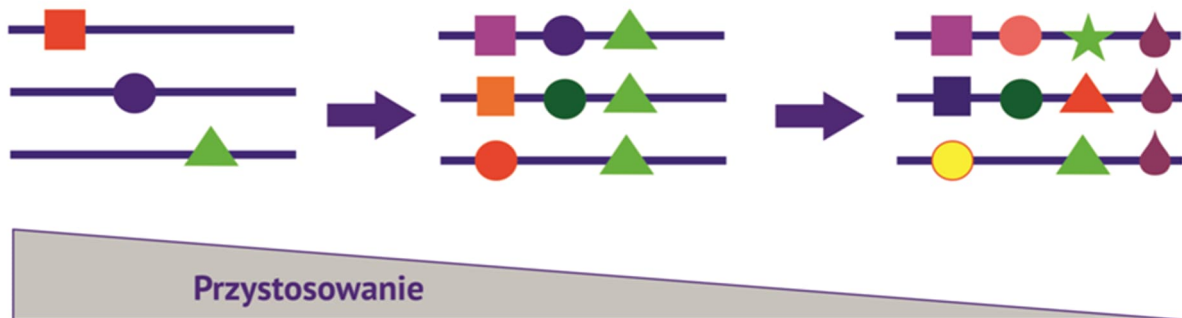
- Wirusy DNA, które przechodzą cykl lizogeniczny wprowadzają swoje DNA do genomu gospodarza i namnażają się wraz z podziałem komórek gospodarza. Wraz z wbudowaniem wirusowego DNA do genomu gospodarza, wirusy przestają być niezależnymi molekułami. Wiele wirusów wchodzi w cykl lityczny, prowadzący do powstania nowych wirionów tylko w ściśle określonych warunkach (Rys. 1.2.1a).



Rys. 1.2.1a. Cykl lizogeniczny wirusa DNA. Wirus może wejść w cykl lityczny (prowadzący do lizy komórek) w warunkach stresowych.

- Retrowirusy: po wnikięciu do komórki, RNA retrowirusów jest kopiowane na cDNA, które następnie jest wstawiane do genomu. Transkrypcja cDNA prowadzi do powstania mRNA i odtworzenia genomu retrowirusa (RNA). W trakcie cyklu replikacyjnego dochodzi do zmiany typu kwasu nukleinowego.
- Wirusy RNA tworzą populacje cząstek, które mają genomy podobne, ale znacznie różniące się między poszczególnymi cząstkami w obrębie jednego gospodarza, a także pomiędzy różnymi gospodarzami. Duże zróżnicowanie genomów wirusów RNA wynika z wysokiej częstości mutacji spowodowanej w głównej mierze brakiem zdolności korekcyjnych polimeraz RNA. Takie silnie zróżnicowane populacje wirusów RNA noszą nazwę kwasigatunków (Rys. 1.2.1b). Klasyczna koncepcja "typu dzikiego" i mutantów nie ma zastosowania dla wirusów, które stanowią roje (chmury) różnych genomów. Ewolucja wirusów polega na selekcji całego spektrum mutantów, a nie indywidualnych genomów (Matinez i inni 2012).

Kwasigatunki: populacje, które stanowią kolekcję blisko spokrewnionych genomów podlegających ciągłemu procesowi mutacji i selekcji, który prowadzi do znacznej plastyczności genetycznej i fenotypowej. Kwasigatunki tworzą roje mutantów.

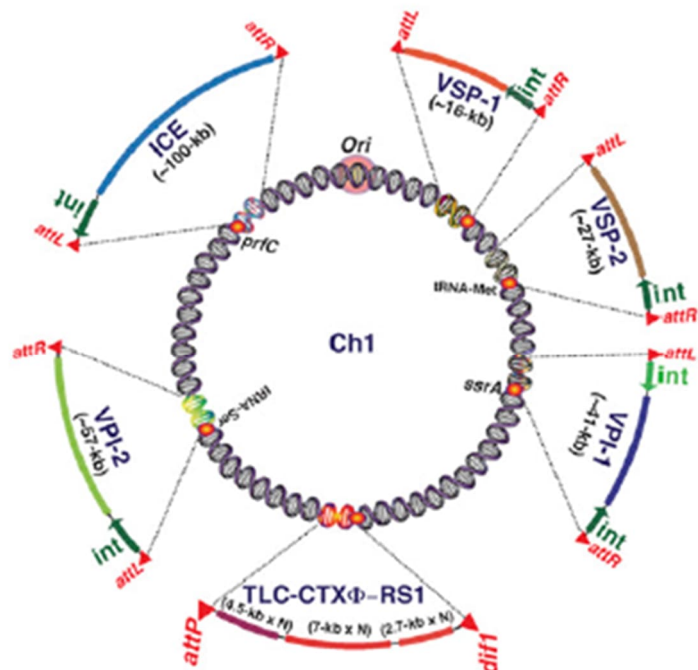


Rys. 1.2.1b. Koncepcja dynamiki kwasigatunków. Linie oznaczają pojedyncze genomy wirusowe. Symbole na liniach oznaczają mutacje. Wraz ze wzrostem liczby mutacji spada zdolność przystosowawcza.

1.2.2. Dynamika genomów bakteryjnych

Genomy bakteryjne są wysoce heterogenne i ich skład zależy od środowiska i stylu życia. Zróżnicowanie genomów bakteryjnych wynika z pozyskiwania DNA przez horyzontalny transfer genów, redukcji genomu w wyniku delecji, mutacji i rearanżacji genomu pod wpływem środowiska. Przykładowo, *Escherichia coli* tworzy kilka patotypów, które różnią się obecnością elementów ruchomych, zawierają różne plazmidy oraz tzw. wyspy patogeniczności.

Podobnie, izolaty *Vibrio cholerae* różnią się głównie składem wysp patogeniczności (Rys. 1.2.2). Wyspy VP1 i VP2 występują we wszystkich szczepach pandemicznych *V. cholerae*. Po wniknięciu do organizmu gospodarza, wyspy patogeniczności integrują się ze specyficznymi miejscami w chromosomie bakteryjnym. Wyspy patogeniczności często związane są z opornością na antybiotyki. W Indiach prawie wszystkie analizowane szczepy *V. cholerae* (99%) wykazują oporność na wiele antybiotyków jednocześnie, przy czym wzór oporności odpowiada antybiotekom rutynowo stosowanym w leczeniu cholery (Shaik i inni 2021).



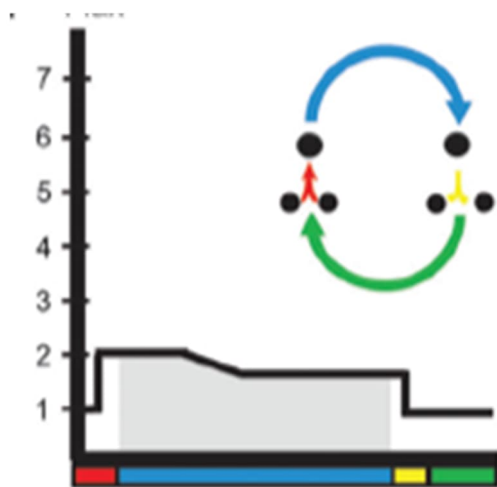
Rys. 1.2.2. Rozmieszczenie wysp patogeniczności (VP) u *Vibrio cholerae*. Wyspy pomagają kolonizować komórki nabłonka jelita (Pant i inni 2020).

Wyspy patogeniczności to segmenty DNA o długości 10-200 kbp, oflankowane prostymi powtórzeniami i rozproszone po całym genomie. Ich skład różni się od składu genomu. Występują u bakterii patogennych i wolnożyjących. Są istotnym elementem warunkującym plastyczność genomu.

1.2.3. Dynamika genomu Eukariota

Genomy Eukariota różnią się znacznie pomiędzy różnymi liniami, osobnikami, a nawet między komórkami tego samego organizmu. Aby zapewnić ciągłość genetyczną pomiędzy pokoleniami, genomy przekazywane następnym pokoleniom są znakowane za pomocą metylacji DNA i acetylacji histonów. Ponadto część genomu jest relatywnie stabilna, podczas gdy inne fragmenty wykazują znaczną zmienność.

Zmienność genomu może dotyczyć różnej zawartości DNA. Typowym przykładem jest poliploidalność niektórych linii, a nawet komórek. Mniej znany jest proces obniżenia zawartości DNA w trakcie życia osobnika w warunkach stresowych. W trakcie mejozy poziom DNA spada z 2C do 1C, a następnie w wyniku zapłodnienia wraca do 2C. U roślin okrytonasiennych poziom DNA w komórkach somatycznych wynosi 2C. U lnu (*Linum usitatissimum*) redukcja zawartości DNA w ciągu życia rośliny może wynosić nawet 15% (Rys. 1.2.3a). Redukcja dotyczy głównie powtarzalnych elementów DNA. Okazuje się, że zmiana zawartości DNA jest odwracalna i po pewnym czasie w nowych komórkach rośliny poziom DNA wraca do stanu wyjściowego (Parfrey i



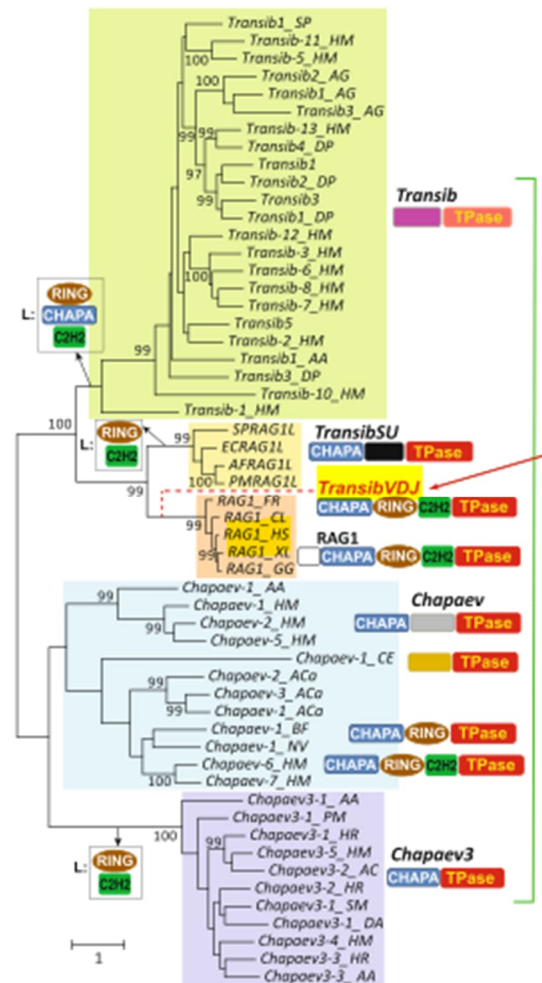
Rys. 1.2.3a. Zmiany poziomu DNA w genomie lnu pod wpływem stresu. Kolory na osi odciętych oznaczają fazę cyklu życiowego. Oś rzędnych przedstawia poziom DNA jako wartość C (Parfrey i inni 2008).

inni 2008).

Zróznicowanie genomu dotyczy także jego składu. Wykazano, że co najmniej 13% genomu człowieka różni się pomiędzy poszczególnymi osobami. Różnice te obserwowane są jako delecje i insercje milionów par zasad w regionach zawierających sekwencje powtarzalne. Ta ogromna zmienność kontrastuje z wysokim podobieństwem genów, których sekwencje nie różnią się znacznie nawet między człowiekiem a szympansem. Występowanie różnych wariantów strukturalnych w genomie człowieka wskazuje, że analiza samych SNP dostarcza jedynie niepełnej informacji o warunkowaniu cech, a także przyczynach chorób.

Istotnym czynnikiem powodującym znaczną zmienność genomów Eukariota są transpozony, czyli elementy ruchome.

- Transpozony DNA: zmieniają pozycję w genomie poprzez wycięcie z jednego miejsca i wstawienie do innego. System rekombinacji V(D)J, podczas tworzenia przeciwciał u kręgowców, wywodzi się z transpozonu DNA *Transib*, należącego do typu CACTA. Z genów kodujących białka w tym transposonie pochodzą współczesne geny *RAG1-RAG2* odpowiedzialne za rekombinację. Prawdopodobnie został on pozyskany w trakcie ewolucji kręgowców około 500 mln lat temu. Transpozony, które są prawdopodobnymi przodkami systemu V(D)J znaleziono u ukwiałów morskich.
- Retrotranspozony: przemieszczają się w genomie za pomocą mRNA. Transpozon podlega transkrypcji, a następnie na matrycy RNA syntetyzowane jest cDNA, które integruje się z genomem. U człowieka do tej grupy należą elementy LINE. Obecnie jest aktywnych około 100 elementów LINE. Retrotranspozony często uaktywniają się pod wpływem stresu, np. oksydacyjnego. Genom przeciwdziała nadmiernej aktywacji poprzez tworzenie tzw. solo LTR. W ten sposób powstały winogrona o różnej barwie owoców.



Rys. 1.2.3b. Pochodzenie systemu V(D)J. Geny *RAG1* kręgowców grupują się z kładem *TransibSU* (Kapitonov i Koonin 2015).



1.2.4. Analiza wariantów strukturalnych w genomie człowieka

Warianty strukturalne to regiony w genomie o długości co najmniej 1 kbp, które zawierają inwersje, translokacje, większe insercje i delecje.

- Proszę wejść na stronę <https://www.ncbi.nlm.nih.gov/dbvar/browse/>
- W tabeli proszę wybrać poniższe pozycje i zaakceptować enterem (Rys. 1.2.4.a):

- ▶ *Homo sapiens*,
- ▶ GRCH38,
- ▶ Chr. 10.

- Po wyświetleniu się widoku chromosomu 10 proszę rozwinąć koło zębate w prawym dolnym rogu i zaznaczyć pozycje zgodnie z Rys. 1.2.4b.

View data across studies on the genome

Select an Organism: Select an Assembly:

Select a region using search or your uploaded data

Search	User Data and Track Hubs
<input type="text" value="Chr10"/> ▶ Enter a location, gene name or phenotype	<input type="text" value="- no added tracks or track ..."/> <input type="button" value="Options"/>
	<input type="button" value="Supported File/Data types"/>

Rys. 1.2.4a. Zrzut ekranu dla bazy dbVar - wprowadzanie danych.

Należy „kliknąć” lewym klawiszem myszy na koło zębate

Configure Page

Available Tracks Find Tracks Custom Data

Active Tracks

Click on track to display settings. To re-order tracks, drag and drop track names.

Active	Track name
<input checked="" type="checkbox"/>	Sequence
<input checked="" type="checkbox"/>	Genes, NCBI Homo sapiens Annotation Release 110, 2022-04-08
<input checked="" type="checkbox"/>	Common Variations (MAF >= 0.01), dbSNP b155 v2
<input checked="" type="checkbox"/>	Live RefSNPs, dbSNP b155 v2
<input checked="" type="checkbox"/>	dbVar NCBI Curated Common Structural Variants (nstd186)

Tracks shown: 5/4698

Rys. 1.2.4b. Pozycje, które należy zaznaczyć po wyświetleniu się widoku chromosomu. Odpowiednie pozycje znajdujemy klikając na nazwy po lewej stronie zgodnie ze zaznaczonymi strzałkami.

Otrzymany obraz pokazuje warianty strukturalne w obrębie chromosomu 10. Warianty strukturalne oznaczone są kolorami.

- ▶ Barwa ciemnofioletowa: zmiany liczby kopii różnego typu.
- ▶ Barwa granatowa: zwiększenie liczby kopii, najczęściej są to duplikacje.
- ▶ Barwa czerwona: utrata kopii, najczęściej delecje.
- ▶ Barwa ciemnobrązowa: tandemowe powtórzenia.
- ▶ Barwa jasnofioletowa: inwersje.
- ▶ Barwa fioletowa, paskowana: translokacje.
- ▶ Barwa błękitna, lazurowa: złożone zmiany strukturalne różnego typu.

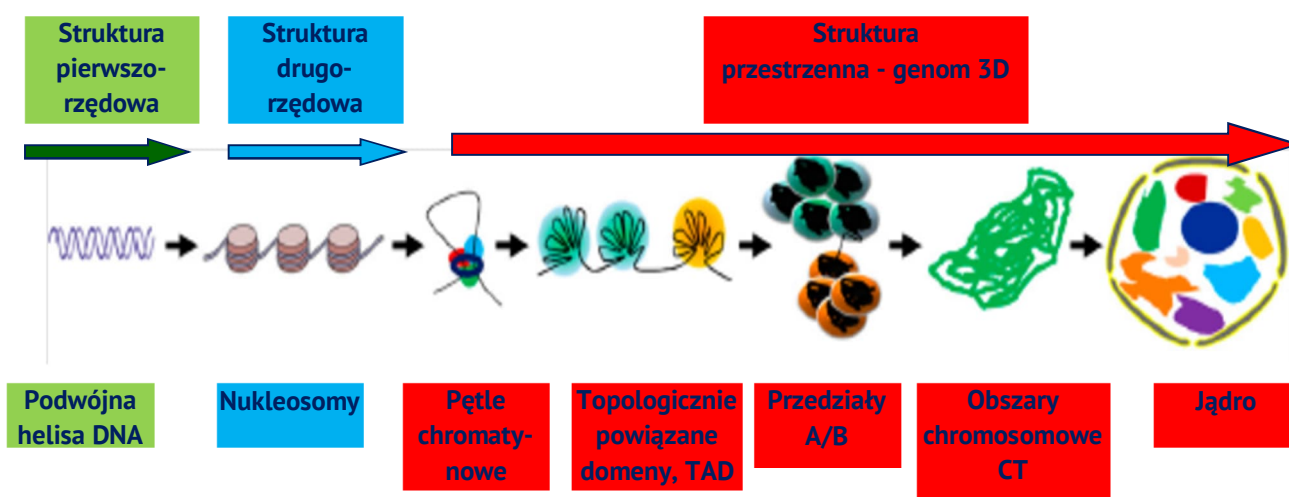
- A. Na podstawie otrzymanego obrazu chromosomu 10 oraz podanych powyżej wyjaśnień, proszę podać jakiego typu warianty strukturalne możemy wyróżnić.
- B. Za pomocą suwaka +/- proszę zawęzić obserwowany zakres do obszaru 45-48 mln bp i z otrzymanego rysunku odczytać numer i długość najdłuższego wariantu związanego ze zmiennością liczby kopii różnego typu.
- C. Proszę podać symbole oraz funkcje genów, które występują w regionie objętym wariantem strukturalnym opisanym w punkcie B.

➔ 1.3. Genom 3D u Eukariota

1.3.1. Struktura przestrzenna genomu Eukariota (genom 3D)

Genomy Eukariota tworzą złożoną strukturę, która obejmuje trzy poziomy organizacji:

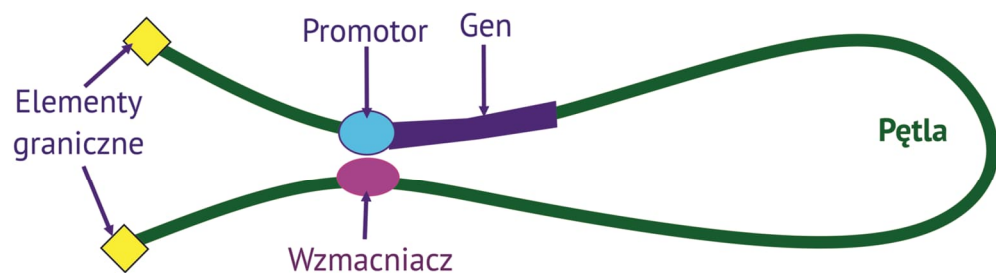
- strukturę pierwszorzędową, czyli sekwencję nukleotydów;
- strukturę drugorzędową, czyli upakowanie nukleotydów na nukleosomach;
- strukturę przestrzenną zwaną także genomem 3D, który obejmuje rozmieszczenie przestrzenne chromatyny w jądrze (Rys 1.3.1a).



Rys. 1.3.1a. Hierarchiczna organizacja genomu jądrowego Eukariota (Krumm i Duan 2018).

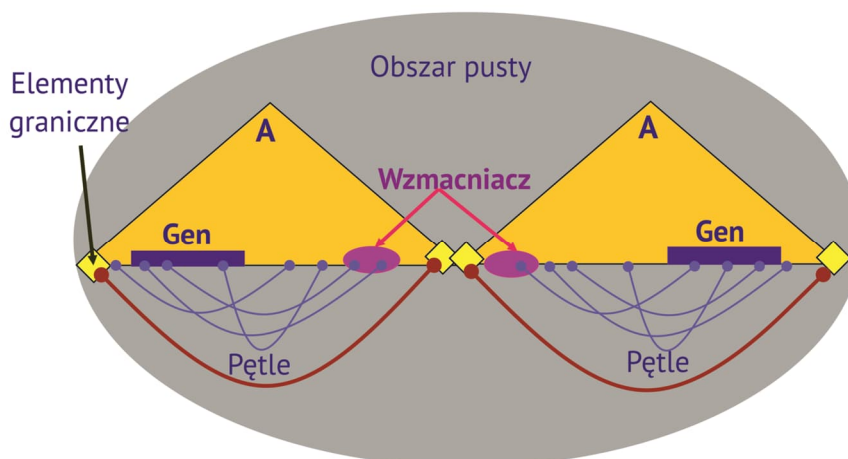
Interakcje między chromosomami i zmiany topologiczne, które zachodzą pod wpływem środowiska wpływają na replikację, ekspresję genów i utrzymanie integralności genomu. Struktura przestrzenna chromatyну obejmuje pętle chromatynowe, topologicznie powiązane domeny (TAD), przedziały (A/B) oraz obszary chromosomowe (CT).

- Pętla chromosomowa jest definiowana jako dwa loci (geny), które znajdują się w bliskiej odległości w przestrzeni jądrowej, pomimo iż oba geny oddzielone są sekwencjami nukleotydowymi o znacznej długości (Rys. 1.3.1b). Pętle umożliwiają komunikację między elementami regulatorowymi (promotorami) położonymi w pobliżu genu oraz tymi położonymi w znacznej odległości (np. wzmacniaczami). U myszy gen sekwencji regulatorowej, *LCR* dla genu beta-globiny oddalony jest od tego genu o 50 kbp. Jednakże w komórkach wątroby płodu sekwencja *LCR* oraz gen beta-globiny znajdują się w bezpośrednim sąsiedztwie w przestrzeni jądra. Z kolei w komórkach mózgu płodu gen *LCR* i gen beta-globiny są oddalone. Różnice w przestrzennym rozmieszczeniu związane są ze specyficzną ekspresją beta-globiny płodowej u myszy. Klasycznym przykładem regulacji na znaczną odległość jest gen *Shh*, który koduje białko sygnałowe kluczowe w morfogenezie wszystkich zwierząt. W szczególności *Shh* kontroluje rozwój układu nerwowego. Wzmacniacz regulujący ekspresję *Shh* położony jest w odległości około 1 Mbp od genu.



Rys. 1.3.1b. Pętla chromosomowa umożliwiająca komunikację między oddalonymi elementami regulatorowymi.

- Topologicznie powiązane domeny (TAD, ang. topologically associated domains) to regiony, w których sekwencje DNA fizycznie są powiązane i wchodzi w interakcję (Rys. 1.3.1c). Rozmiary TAD to około 880 kbp. Regiony TAD regulują ekspresję genów przez

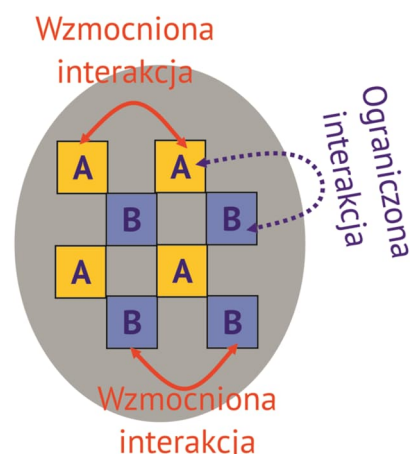


Rys. 1.3.1c. Topologicznie powiązane domeny, TAD (czerwone półkola) na tle przedziału A i obszarów chromosomowych, CT (barwa szara). W obrębie TAD pokazane są pętle.

zmianę interakcji między wzmacniaczami i promotorami. W obrębie locus *Hox* domeny

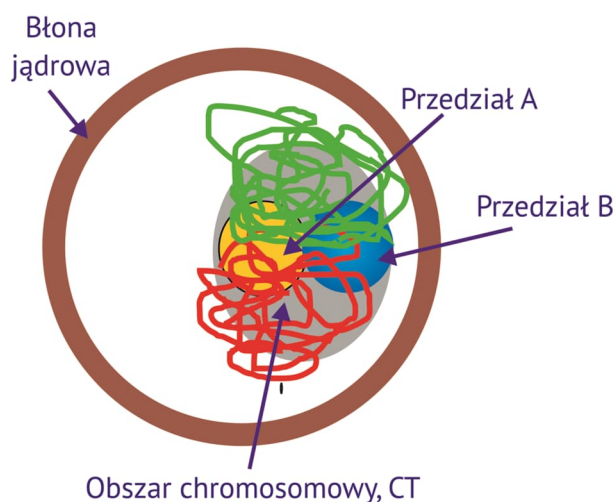
TAD różnią się w zależności od tkanki, co prowadzi do różnych interakcji między promotorem i wzmacniaczem transkrypcji (Chakraborty i Ay 2019). Zaburzenia w interakcjach w obrębie TAD mogą prowadzić do chorób nowotworowych i neurodegeneracyjnych. TAD są uważane za podstawową jednostkę strukturalną chromosomów i organizacji przestrzennej genomu. Wykazują one pewien stopień konserwatywności u poszczególnych gatunków. Ponadto wyróżnia się domeny powiązane z blaszką jądrową (LAD, ang. lamina associated domains) oraz z jąderkiem (NAD, ang. nucleolus associated domains).

- Przedziały A/B (ang. A/B compartments) to dwa regiony w obrębie jądra, które różnią się rozmieszczeniem i aktywnością transkrypcyjną. Przedziały A/B tworzą wzór zbliżony do szachownicy (Rys. 1.3.1d). Regiony chromosomów w danym obszarze preferencyjnie wchodzi w interakcje z innym regionem chromosomów zlokalizowanych w tym samym obszarze. Region A na ogół obejmuje chromatynę aktywną transkrypcyjnie, która jest bogata w geny oraz charakteryzuje się wysoką zawartością par GC. Region A zlokalizowany jest na ogół we wnętrzu jądra. Region B jest nieaktywny transkrypcyjnie. Jest on ubogi w geny i jest zlokalizowany peryferyjnie. Przedział B jest połączony z domenami powiązanymi z blaszką jądrową (LAD). Przedziały A/B są dynamiczne. Przykładowo, podczas różnicowania komórek macierzystych dochodzi do przemieszczenia 36-60% DNA pomiędzy przedziałami (Chakraborty i Ay 2019). Korelacja przedziałów A/B z cytogenetycznie wyróżnialną euchromatyną i heterochromatyną nie jest jednoznaczna. Przedziały A/B zbudowane są z topologicznie powiązanych domen, TAD.



Rys. 1.3.1d. Koncepcja przedziałów A/B w jądrze.

- Obszary chromosomowe (CT, ang. chromosome territories) to regiony w jądrze, które preferencyjnie zajmowane są przez określone chromosomy. Obszary CT mają kształt sferoidalny o średnicy od jednego do kilku mikrometrów. Przykładowo, duże chromosomy ubogie w geny na ogół zlokalizowane są na peryferiach jądra w pobliżu blaszki jądrowej. Poszczególne chromosomy w obrębie obszarów chromosomowych CT są rozmieszczone w przedziałach A lub B (Rys. 1.3.1e).



Rys. 1.3.1e. Koncepcja obszarów chromosomowych, CT na tle przedziałów A/B.

Organizacja przestrzenna genomu jest dynamiczna, co prowadzi do zmienności genomu 3D między komórkami, a nawet w obrębie danej komórki. Struktura zmienia się także w trakcie cyklu życiowego komórki. Silna kondensacja genomu w plemnikach ssaków związana jest ze znacznie większą liczbą interakcji między domenami chromatynowymi w porównaniu z fibroblastami. Z kolei struktura 3D genomu oocytów w stadium metafazy II jest podobna do struktury genomu w komórkach dzielących się mitotycznie, która charakteryzuje się brakiem TAD i przedziałów A/B. Podczas różnicowania komórek zmienia się liczba i wielkość TAD oraz może dochodzić do przelączenia między przedziałami A/B.

Struktura przestrzenna genomu jest utrzymywana przy pomocy białek i RNA. Do najważniejszych białek wpływających na genom 3D zalicza się czynniki wiążące sekwencję CCCTC, takie jak represor transkrypcji, CTCF zawierający palec cynkowy, kohezyny i kondensyny należące do białek SMC (białka utrzymujące strukturę chromosomu), białka macierzy jądrowej, w tym białka blaszki jądrowej i czynniki transkrypcyjne. Przykładowo, czynnik YY1 przyczynia się do tworzenia pętli pomiędzy promotorami i wzmacniaczami transkrypcji.

Istotną rolę w utrzymaniu struktury 3D genomu ma lncRNA (long non-coding RNA), którego cząsteczki wiążą się z chromatyną. Cząsteczki RNA typu Firre umożliwiają kolokalizację regionów genomowych zlokalizowanych na różnych chromosomach. Z kolei RNA typu Xist umożliwia tworzenie struktury 3D w nieaktywnym chromosomie X ssaków. Transkrypcja ThymoD RNA jest niezbędna do zwijania chromatyny i tworzenia przedziałów umożliwiających interakcję między promotorami i wzmacniaczami w limfocytach T (Krumm i Duan 2018).

1.3.2. Znaczenie genomu 3D w warunkowaniu chorób

Struktura przestrzenna genomu odgrywa istotną rolę w regulacji ekspresji genów, a także w replikacji i naprawie DNA. Funkcja struktury przestrzennej jest komplementarna do mechanizmów epigenetycznych (metylacja DNA i modyfikacja histonów przez acetylację), a także regulacji za pomocą niekodującego RNA. W wielu stanach chorobowych obserwuje się różnice w rozmieszczeniu genów i chromosomów w jądrze. Wykazano także, że struktura przestrzenna chromatyny ma wpływ na częstość mutacji (Krumm i Duan 2018). Topologicznie powiązane domeny (TAD) stanowią podstawowy element ułatwiający interakcję między wzmacniaczami i promotorami. Zaburzenie struktury TAD prowadzi do zmian w ekspresji wielu genów. Błędnie utworzone pętle mogą doprowadzić do połączenia egzonów z różnych genów i powstania chimerycznego mRNA. Zjawisko to jest często obserwowane w komórkach nowotworowych.

- Przelączenie 12% przedziałów A/B zaobserwowano w komórkach nowotworu piersi. Większość przypadków przelączenia z przedziału B na A dotyczyło małych chromosomów (16-22). Przelączenie doprowadziło do zmiany wzoru ekspresji, aktywacji szlaków związanych z nowotworzeniem.
- Zaburzenia w formowaniu domen TAD prowadzą do chorób przejawiających się w deformacji kończyn. Gen *Epha4* koduje receptor efryn (efryny należą do rodziny receptorowych kinaz tyrozynowych), który uczestniczy w tworzeniu granic między tkankami i segmentacji. Inwersja fragmentu o długości 1,1 Mbp prowadzi do brachydaktylii (krótkopalczaści dłoni). Duplikacja fragmentu genu o długości 1,1 Mbp powoduje powstanie syndromu F polegającego na zespoleniu kości nadgarstka i stępu, któremu towarzyszą złożone anomalie palców rąk i nóg. Z kolei duplikacja fragmentu o długości 0,9 Mbp prowadzi do polidaktylii. Wszystkie te deformacje związane są z

zaburzeniami w strukturze TAD w obrębie locus *Epha4*, które powodują zmianę interakcji między promotorem a wzmacniaczem. W efekcie dochodzi do zaburzonej ekspresji genu *Epha4*, której towarzyszy ekspresja otaczających loci. Zmiany są dominujące.

- Gen *Sox9* zawiera dwie domeny TAD, jedna obejmuje cały gen *Sox9* (intra-TAD), druga dodatkowo zawiera geny kanałów potasowych, *Kcnj2* i *Kcnj16* (inter-TAD). Duplikacja TAD w obrębie *Sox9* (intra-TAD) prowadzi do zamiany fenotypu żeńskiego na męski. Z kolei duplikacja obejmująca inter-TAD powoduje powstanie nowej domeny, neo-TAD, która łączy nowy zestaw elementów regulatorowych oraz geny *Kcnj2* i *Kcnj16*. Zmiany te prowadzą do powstania syndromu Cook'a - dominującej choroby dziedzicznej, która charakteryzuje się obustronną hipoplazją paznokci kciuka, palca wskazującego i środkowego, brakiem paznokci (anonychia) palca serdecznego i małego, wydłużeniem kciuków i bulwiastością palców rąk.
- Alfa talasemia jest chorobą dziedziczną autosomalnie recesywnie. Polega na zaburzeniu syntezy alfa-globiny na skutek defektów biosyntezy globin. Przyczyną są mutacje punktowe w genach alfa-globin. Mutacje te prowadzą do zaburzenia pętli chromosomowych łączących promotory genów alfa-globin ze wzmacniaczami transkrypcji.
- Gen *CFTR* reguluje własną ekspresję za pomocą sekwencji regulatorowej zlokalizowanej w intronie dzięki utworzeniu pętli między intronem a wzmacniaczem. Mutacja w intronie likwidująca pętlę chromosomową prowadzi do obniżonej ekspresji genu *CFTR* i objawów mukowiscydozy.

➔ 2. Wielkość genomu

Wielkość genomu jest to zawartość DNA w pojedynczym genomie (x). Podaje się ją jako całkowitą liczbę par zasad (bp = pz) lub w pikogramach. Zawartość DNA w pikogramach w haploidalnych jądrach określana jest mianem wartości C (C-value).

Zgodnie z definicją wielkość genomu powinna być zawsze odnoszona do pojedynczego genomu. W przypadku organizmów diploidalnych nie następuje to problemów interpretacyjnych. Pojedynczy genom odpowiada zawartości DNA w haploidalnym zestawie chromosomów. Tym samym znając zawartość DNA w pikogramach w jądrze diploidalnym, łatwo określić zawartość w genomie haploidalnym. Podobnie jest w przypadku określenia liczby sekwencji przypadających na genom. Jądro diploidalne zawiera po dwie kopie każdej sekwencji, a więc łatwo określić liczbę nukleotydów pojedynczego genomu. Kłopoty zaczynają się w odniesieniu do allopoliploidów, czyli form powstałych w wyniku krzyżowania różnych gatunków. Wówczas genomy pochodzące od różnych gatunków różnią się nieznacznie i sekwencjonowanie prowadzi do uzyskania sekwencji, które odzwierciedlają zestaw haploidalny. Tymczasem np. u pszenicy taki haploidalny zestaw zawiera trzy genomy, a więc liczbę nukleotydów należałoby podzielić przez 3. Niestety w publikacjach często pomija się aspekt ploidalności podając wielkość genomu bez określenia czy wartość ta dotyczy pojedynczego zestawu, czy też zestawu haploidalnego zawierającego więcej genomów. W efekcie dla gatunków poliploidalnych możemy spotkać się z dużą rozbieżnością w ocenie wielkości genomu na podstawie liczby nukleotydów. Dlatego wydaje się rozsądniejsze korzystanie z wartości C, która zawsze odnosi się do haploidalnych jąder.

2.1. Miary wielkości genomu

2.1.1. Liczba par zasad

- 1 000 nukleotydów: 10^3 bp = 1 kbp (tysiąc)
- 1 000 000 nukleotydów: 10^6 bp = 1 Mbp (milion)
- 1 000 000 000 nukleotydów: 10^9 bp = 1 Gbp (miliard)

2.1.2. Jednostki masy: najczęściej w postaci pikogramów

- **1 pikogram (pg) to 978 Mbp (w przybliżeniu 1 Gbp)**
- 1 pg = 10^{-12} g.

2.1.3. Jednostki długości

- Odległość pomiędzy zasadami w DNA wynosi 3,4 Å czyli $3,4 \times 10^{-10}$ m



2.2. Przeliczanie jednostek

2.2.1. Proszę podać długość odcinków DNA w metrach i centymetrach.

- A. 15 kbp
- B. 1,2 Mbp
- C. 3 Gbp

2.2.2. Przyjmując, że 1 Gbp waży 1 pg proszę podać masę następujących genomów w gramach.

- A. 80 kbp
- B. 25 Mbp

2.2.3. W tabeli 2.2.3 podano wielkość wybranych chromosomów człowieka w różnych jednostkach. Przyjmując odległość między zasadami 3,4 Å oraz masę 1 pg dla 978 Mbp proszę uzupełnić brakujące wartości w tabeli. Długość i masę proszę zaokrąglić do części setnych.

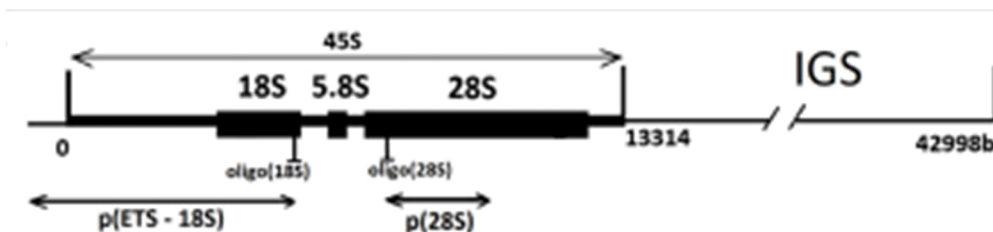
Tabela 2.2.3. Liczba nukleotydów, długość i masa wybranych chromosomów człowieka.			
Chromosom	Liczba nukleotydów (bp)	Długość (cm)	Masa (pg)
1	248 956 422		
3		6,48	
20			0,07

3. Wartość C

➔ 3.1. Sekwencje powtarzalne i unikalne

3.1.1. Sekwencje tandemowo powtórzone, rDNA

Wiele sekwencji powtarzalnych w genomie występuje w postaci tandemowych powtórzeń. Należą do nich między innymi sekwencje rDNA, które kodują rRNA (Rys. 3.1.1a). Zlokalizowane są one na kilku chromosomach (NOR): 13, 14, 15, 21 i 22. Większość jest silnie skondensowana w postaci heterochromatyny.



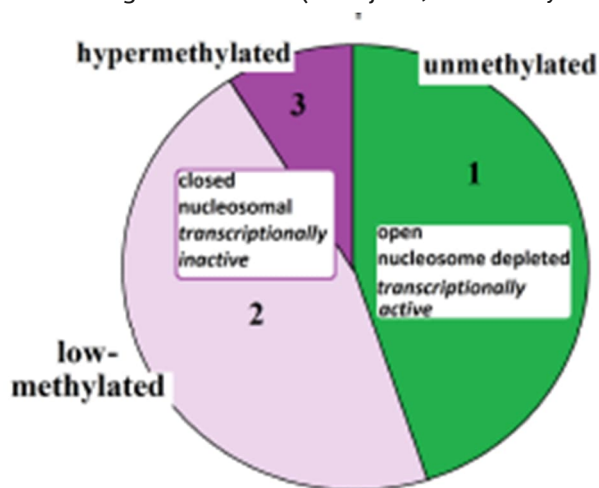
Rys. 3.1.1a. Schemat genów rDNA człowieka w postaci tandemowego powtórzenia (Malinovskaya et al. 2018).

Genom człowieka zawiera około 400 kopii rDNA. Jedynie 30-50% z tych kopii jest aktywna transkrypcyjnie. Frakcja ta występuje w postaci euchromatyny (Rys. 3.1.1b). Histony występują w postaci charakterystycznej dla miejsc aktywnych, większość wysp CpG jest niezmetylowana (frakcja 1, zielony fragment koła). Kolejna frakcja (frakcja 2) charakteryzuje się niską metylacją, jest zlokalizowana w jąderkach, w regionach aktywnych transkrypcyjnie. Jednakże jest ona nieaktywna ze względu na metylację wybranych wysp CpG w regionach promotorowych. Frakcja ta stanowi 30–50% (jasno różowy fragment koła). Ponadto część genów rDNA jest silnie zmetylowana, występuje w regionach peryferyjnych jąderka i stanowi pas heterochromatyny otaczający jąderko. Frakcja ta stanowi około 20% genów rDNA (frakcja 3, fioletowy fragment koła). Zawartość poszczególnych frakcji zmienia się z wiekiem.

3.1.2. Sekwencje unikalne

Badania ekspresji genów, a zwłaszcza transkrypcji wykazały, że zaledwie 2,7% sekwencji unikalnych w DNA jest obecna jako mRNA na terenie cytoplazmy. Z kolei w jądrze, aż 28% sekwencji unikalnych ma swoje odpowiedniki na poziomie RNA. Frakcja RNA odpowiadająca sekwencjom unikalnym jest zróżnicowana pod względem rozmiarów i tworzy tzw. heterogenny RNA (hnRNA). **Dane te pokazują, że**

znaczna porcja genów podlega transkrypcji, ale niewiele z tych transkrybowanych sekwencji opuszcza jądro i podlega translacji. Tym samym zróżnicowanie RNA na terenie jądra jest około 10-krotnie wyższe niż w cytoplazmie.

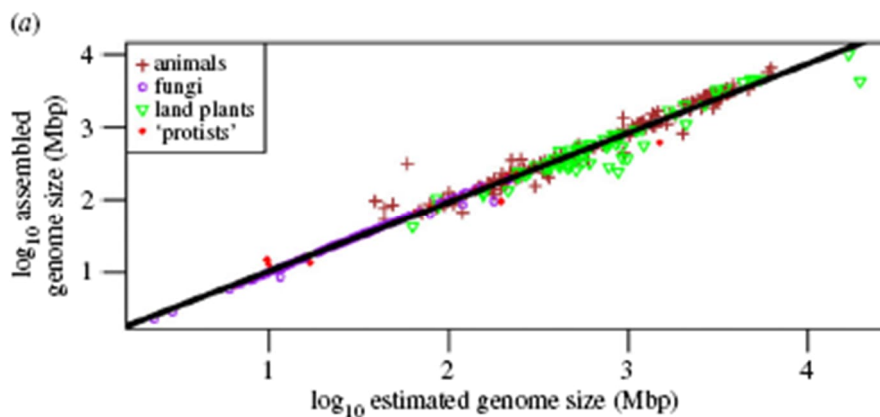


Rys. 3.1.1b. Udział aktywnych transkrypcyjnie i nieaktywnych frakcji genów rDNA u człowieka (Malinovskaya et al. 2018).

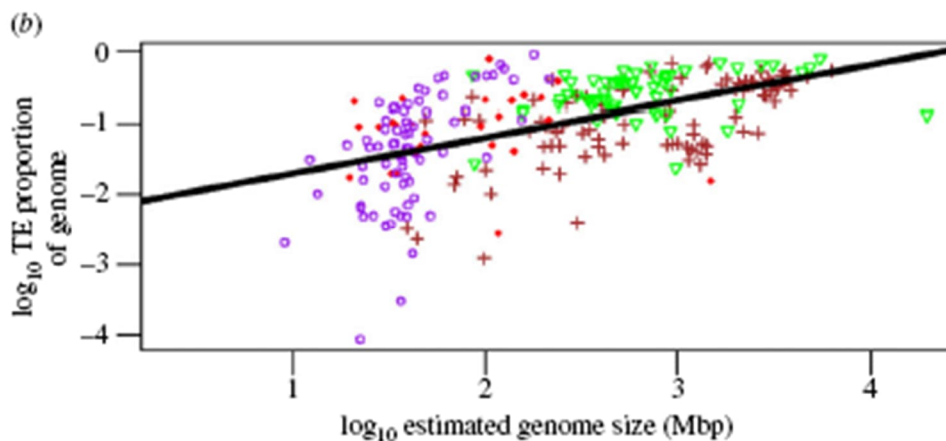
➔ 3.2. Od czego zależy wartość C

Wartość C jest najczęściej oceniana na podstawie analiz cytogenetycznych, jednakże w dużej mierze koreluje ona z wielkością genomów określoną w projektach sekwencjonowania (Rys. 3.2a).

Wartość C koreluje z wielkością komórek, z tempem podziału komórek. Oznacza to, że duże genomy najczęściej występują w dużych, wolno dzielących się komórkach. Ponadto istotna korelacja występuje pomiędzy wartością C a zawartością transpozonów (Rys. 3.2b).



Rys. 3.2a. Korelacja wielkości genomu ustalonego na podstawie badań cytogenetycznych (oś odciętych) z wartościami ustalonymi na podstawie sekwencjonowania (oś rzędnych) (Elliot and Gregory 2015).



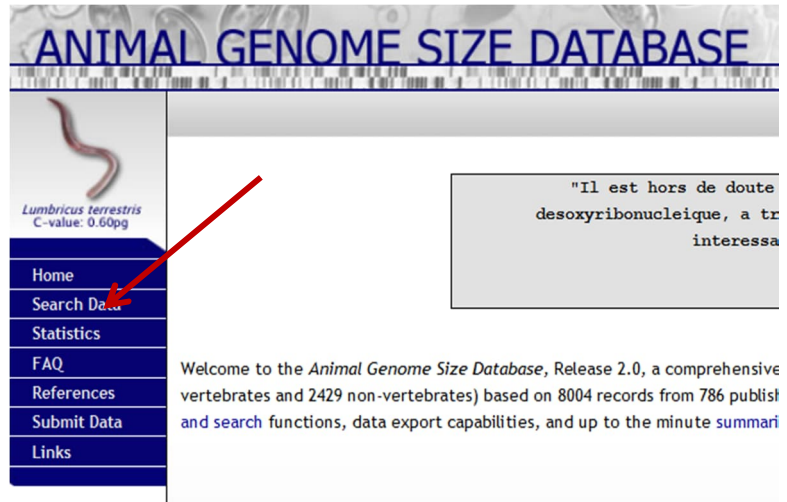
Rys. 3.2b. Korelacja wielkości genomu (oś odciętych) z zawartością ruchomych elementów genetycznych (oś rzędnych) (Elliot and Gregory 2015).

➔ 3.3. Analiza wartości C u różnych grup organizmów

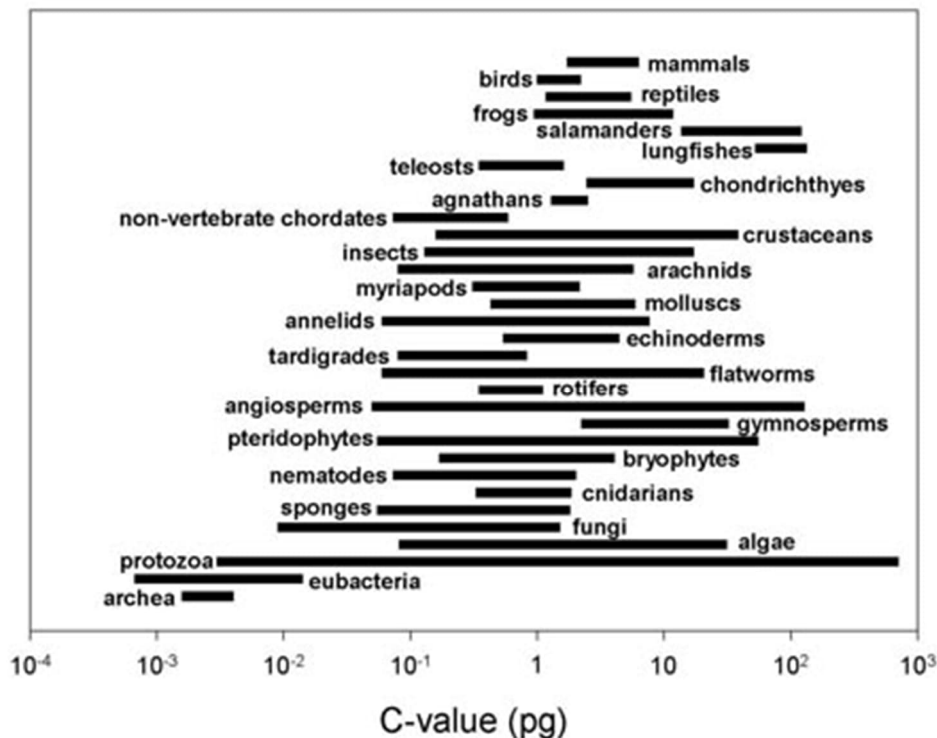
3.3.1. Baza wartości C w pikogramach dla genomów zwierzęcych

Baza wartości C dla zwierząt (Rys. 3.3.1a) zawiera obecnie dane dla 6222 genomów, w tym dla 3793 (61%) genomów kręgowców. Dane w bazie podane są jako **wartość C dla jądra haploidalnego**. Zakres wartości C dla Eukariota waha się od 0,0023 pg dla pasożytniczego pierwotniaka *Encephalitozoon intestinalis* do 1 400 pg dla wolno żyjącego pierwotniaka *Chaos chaos*. W literaturze często podaje się, że największy genom ma *Amoeba dubia* (*Polychaos dubia*) (700 pg).

Nie mniej jednak wszystkie te wartości stanowią przybliżenie, gdyż zostały ocenione metodami biochemicznymi, które także mierzą zawartość DNA w mitochondriach. Ponadto ameby często zawierają liczne (500-1000), małe chromosomy. Prawdopodobnie są one też poliploidami.



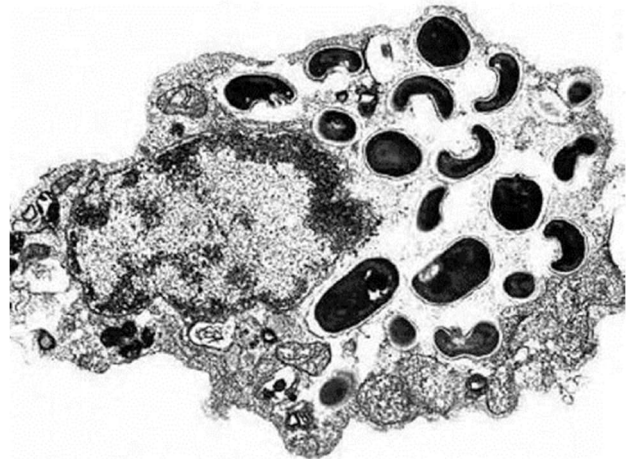
Rys. 3.3.1a. Zrzut ekranu dla bazy genomów zwierzęcych.



Rys. 3.3.1b. Zakres wartości C dla Eukariota.

3.3.2. Przykłady organizmów zwierzęcych o dużym i małym genomie

- Encephalitozoon intestinalis:** pierwotniak, pasożyt przewodu pokarmowego kręgowców. Przenoszony za pośrednictwem skażonej wody. U człowieka najczęściej powoduje zakażenia oportunistyczne. Ma najmniejszy genom u Eukariota, 2,25 Mbp. W porównaniu z gatunkami wolnożyjącymi ma zredukowany genom, co związane jest z silną zależnością od gospodarza.

Rys. 3.3.2a. *Encephalitozoon intestinalis*.Rys. 3.3.2b. *Chaos chaos*.

- Chaos chaos:** wolno żyjący pierwotniak z rodziny Amoebidae. Jest to jeden z największych pierwotniaków dochodzący do 5 mm. Wytwarza liczne nibynóżki, za pomocą których się przemieszcza. Posiada wiele jąder.



3.3.3. Wartość C u wybranych gatunków zwierząt

W bazie Animal Genome Size Database (<http://www.genomesize.com/index.php>) proszę przejść do sekcji „search data” i wprowadzić nazwę gatunkową w polu przeszukiwania „species”. Na podstawie danych podanych w sekcji proszę uzupełnić tabelę.

Tabela 3.3.3. Porównanie wielkości genomów u różnych grup kręgowców.

Nazwa gatunkowa	Nazwa polska	Wartość C [pg]	Metoda	Komórka	Standard
<i>Latimeria chalumnae</i>					
<i>Pelodytes punctatus</i>					
<i>Laudakia caucasia</i>					
<i>Apus caffer</i>					
<i>Capra hircus</i>					



3.3.4. Analiza wielkości genomu u różnych grup roślin.

- Proszę wejść na stronę: <https://www.asteraceaegenomesize.com>
- Jeżeli nie można wejść z pliku, proszę skopiować adres do okna przeglądarki.
- W polu „search”, proszę przejść do „advanced” i w polu „Genus” wybrać rodzaje, do których należą gatunki podane w tabeli. Proszę nie dokonywać wyboru w pozostałych polach, gdyż błędny wpis uniemożliwi wyszukanie danych.
- Dla każdego gatunku z tabeli poniżej proszę podać średnią wartość parametru i wpisać do tabeli. Dane można pobrać w pliku Excel.
- Proszę porównać wartość C dla genomów u różnych gatunków.



Search...

 Author Infrasp. author Chrom
Optional identification of the user

Advanced search options

Subfamily Tribe Genus Specific

2C value - Ploidy - Chrom. number

Year - Infrspecific category Show

Welcome to the GSAD: Genome Size in Asteraceae



The GSAD is an exhaustive catalog of genome size data for Asteraceae species. Data are now available for 1,555 species as of July 2018. The current update release is 3.0, released July 2018. There are 337 species (2

The main goal of the database is to provide genome size information accessible to scientists in a variety of fields (e.g. plant molecular systematics, breeding, races or raw material identification).

The database holds data from previous updates and future updates are planned.

Rys. 3.3.4a. Zrzut ekranu bazy GSAD (Genome size in Asteraceae database).

Advanced search options

Subfamily Tribe Genus

2C value - Ploidy - Chrom. number

Year - Infrspecific category

Rys. 3.3.4b. Zrzut ekranu pola „search” w bazie GSAD.

Tabela 3.3.4. Porównanie wielkości genomów u różnych grup roślin.

Nazwa gatunkowa	Nazwa polska	1C [pg]	2C [pg]	2C [Mbp]
<i>Cichorium intybus</i>				
<i>Centaurea solstitialis</i>				
<i>Matricaria chamomilla</i> subsp. <i>chamomilla</i>				
<i>Solidago virgaurea</i>				
<i>Tussilago farfara</i>				

3.4. Podawanie wielkości genomu w różnych jednostkach



Przyjmując, że odległość między zasadami w DNA wynosi 3,4 Å oraz 978 Mbp waży 1 pg proszę uzupełnić tabelę 3.4.

Tabela 3.4. Wielkość genomów oraz zawartość DNA w komórkach różnych grup organizmów.

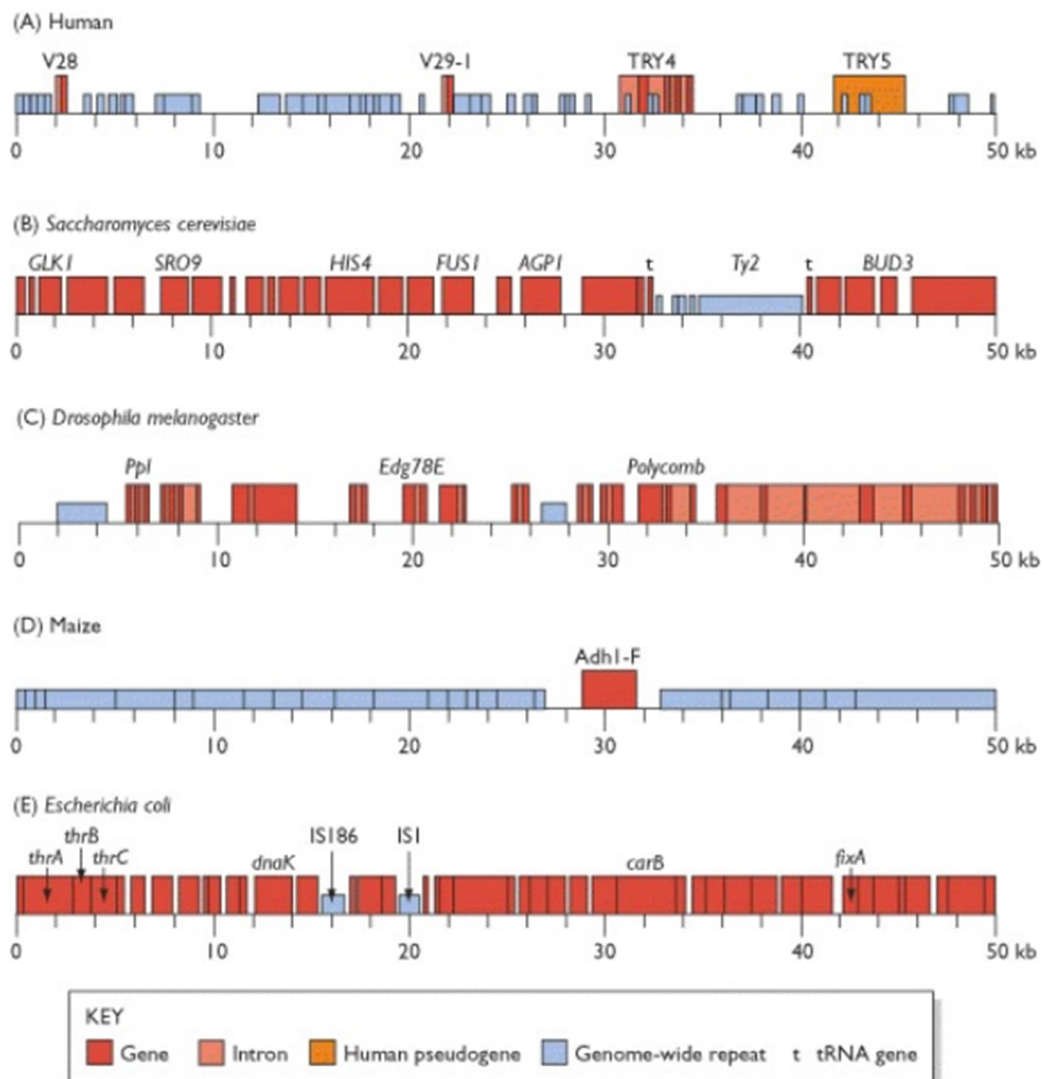
Gatunek	Wielkość genomu [Mbp]	Wielkość genomu w metrach [m]	Wielkość genomu w pikogramach [pg]	Wartość C [C-value]	Zawartość DNA w komórkach somatycznych (2n)
<i>Escherichia coli</i>	4,6				
<i>Drosophila melanogaster</i>	140,0				
<i>Physcomitrella patens</i>	510,0				
<i>Triticum aestivum</i>	17 000,0				
<i>Homo sapiens</i>	3 200,0				

4. Gęstość genów

Gęstość genów: liczba genów przypadająca na milion par zasad (Mbp). U Prokariota obserwujemy wysoką gęstość genów, u Eukariota, zwłaszcza w przypadku dużych genomów gęstość genów jest niska.

➔ 4.1. Gęstość genów u różnych grup organizmów

Porównując organizację genomu prostych Eukariota, jak np. genom *Saccharomyces cerevisiae* (drożdże) z genomem *Homo sapiens*, widoczne jest bardziej ekonomiczne upakowanie genów u *S. cerevisiae*. Podobnie, geny u *D. melanogaster* są gęściej upakowane w porównaniu z *H. sapiens* czy *Z. mays* (Rys. 4.1). Głównym czynnikiem, który przyczynia się do niższej gęstości genów u człowieka i kukurydzy jest duży udział sekwencji powtarzalnych, głównie pochodzenia retrotranspozonowego. Gęstość genów w chromosomach ma wpływ na ułożenie chromosomów w jądrze oraz na parowanie podczas mejozy. Chromosomy bogate w geny na ogół zlokalizowane są w wewnętrznej części obszaru chromosomowego jądra.



Rys. 4.1. Porównanie gęstości genów (czerwone) u różnych organizmów.



4.1.1. Gęstość genów u różnych grup organizmów

Proszę uzupełnić poniższą tabelę.

Tabela 4.1.1. Gęstość genów u różnych organizmów.				
Obiekt	Wielkość genomu	Liczba genów, która przypadłaby na 1 m długości genomu	Liczba genów	Gęstość genów
<i>A</i>	4,3 Mbp		4 000	
<i>B</i>	120 Mbp		20 000	
<i>C</i>	15 Gbp		30 000	
<i>D</i>	120 Gbp		60 000	



4.1.2. Gęstość genów u człowieka

W tabeli poniżej podano uśrednioną liczbę nukleotydów dla poszczególnych chromosomów człowieka oraz liczbę genów zmapowanych na każdym chromosomie. Proszę obliczyć gęstość genów dla chromosomów. Uzyskane wyniki proszę przedstawić na wykresie i na jego podstawie dokonać analizy rozkładu gęstości genów u człowieka.

Tabela 4.2. Gęstość genów na chromosomach człowieka.

Chromosom	Liczba nukleotydów (bp)	Liczba genów kodujących białka	Średnia gęstość
1	248 956 422	2 058	
2	242 193 529	1 309	
3	198 295 559	1 078	
4	190 241 555	752	
5	181 538 259	876	
6	170 805 979	1 048	
7	159 345 973	989	
8	145 138 636	677	
9	138 394 717	786	
10	133 797 422	733	
11	135 086 622	1 298	
12	133 275 309	1 034	
13	114 364 328	327	
14	107 043 718	830	
15	101 991 189	613	
16	90 338 345	873	
17	83 257 441	1 197	
18	80 373 285	270	
19	58 617 616	1 472	
20	64 444 167	544	
21	46 709 983	234	
22	50 818 468	488	
X	156 040 895	842	
Y	57 227 415	71	

5. Literatura

- Chakraborty A, Ay F. 2019. The role of 3D genome organisation in disease: from compartments to single nucleotides. *Semin Cell Dev Biol* 90:104-113. Dostęp: doi:10.1016/j.semcdb.2018.07.005.
- Elliott TA, Gregory TR. 2015. What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. *Phil Trans R Soc B* 370:20140331. Dostęp: <http://dx.doi.org/10.1098/rstb.2014.0331>
- Goldman AD, Landweber LF. 2016. What is a genome? *PLoS Genet* 12:e1006181. Dostęp doi:10.1371/journal.pgen.1006181
- Kapitonov VV, Koonin EV. 2015. Evolution of the *RAG1-RAG2* locus: both proteins came from the same transposon. *Biology Direct* 10:20. Dostęp: DOI10.1186/s13062-015-0055-8
- Krumm A, Duan Z. 2018. Understanding the 3D genome: emerging impacts on human disease. *Semin Cell Dev Biol* 90:62-77. Dostęp: <https://doi.org/10.1016/j.semcdb.2018.07.004>
- Malinovskaya EM, Ershova ES, Golimbet VE, Porokhovnik LN, Lyapunova NA, Kutsev SI, Veiko NN, Kostyuk SV. 2018. Copy number of human ribosomal genes with aging: unchanged mean but narrowed range and decreased variance in elderly group. *Front Genet* 9:306. Dostęp: doi:10.3389/fgene.2018.00306
- Martínez MA, Martrus G, Capel E, Parera M, Franco S, Nevot M. 2012. Quasispecies dynamics of RNA viruses. W Witzany G (ed.) *Viruses: essential agents of life*. Dordrecht: Springer Science. p. 21-42. Dostęp: DOI 10.1007/978-94-007-4899-6_2
- Pant A, Bag S, Saha B, Verma J, Kumar P, Banerjee S, Kumar B, Kumar Y, Desigamani A, Maitie S. 2021. Molecular insights into the genome dynamics and interactions between core and acquired genomes of *Vibrio cholerae*. *PNAS* 117:23762-23773. Dostęp: www.pnas.org/cgi/doi/10.1073/pnas.
- Parfrey LW, Lahr DJG, Katz LA. 2008. The dynamic nature of Eukariotic genome. *Mol Biol Evol* 25:787-794. Dostęp: doi:10.1093/molbev/msn032
- Shaik S, Suresh A, Ahmed N. 2021. Resistant bacteria: implications for global infection control priorities. *JID* 224:876-S882. Dostęp: <https://doi.org/10.1093/infdis/jiab456>

Odpowiedzi

1. Genom i jego dynamika

1.1. Definicja genomu

1.1.1. Liczba genomów

Proszę podać liczbę genomów jądrowych w komórkach wymienionych w punktach A-E.

A. Komórka skóry człowieka oraz komórka jajowa człowieka.

- Człowiek jest diploidem. Komórka skóry zawiera dwa genomy, komórka jajowa – jeden genom.

B. Truskawka jest oktaploidem. Ile genomów jest w komórkach liścia truskawki, a ile w komórce generatywnej ziarna pyłku?

- Komórki liścia truskawki zawierają 8 genomów, komórka generatywna ziarna pyłku jest haploidem i zawiera o połowę mniej genomów niż komórka somatyczna, zatem zawiera ona 4 genomy.

C. Karp królewski jest tetraploidem. Ile genomów jest w komórkach oskrzeli, a ile w plemniku.

- Komórki oskrzeli zawierają 4 genomy. Plemniki są haploidalne, a więc zawierają o połowę mniej genomów niż komórka somatyczna, zatem zawierają dwa genomy.

D. Ile genomów jest w komórce skóry człowieka, po zakończeniu fazy S?

- 4 genomy. W komórce skóry, która jest komórką somatyczną, a człowiek jest diploidem, są dwa genomy. W fazie S dochodzi do syntezy DNA, każdy chromosom składa się z dwóch chromatyd, które odpowiadają dwóm cząsteczkom DNA. Zatem po zakończeniu fazy S występują 4 genomy.

E. Ile genomów jest w spermatocytach człowieka w anafazie II podziału meiotycznego?

- 2 genomy. Człowiek jest diploidem, zatem przed mejozą ma 2 genomy w komórkach, po fazie syntezy poprzedzającej mejozę, w spermatocytach są 4 genomy. I podział jest podziałem redukcyjnym, zmniejsza się liczba chromosomów i tym samym genomów o połowę. Spermatocyty po I podziale meiotycznym mają 2 genomy. W anafazie II chromosomy są na przeciwległych biegunach, ale nie doszło jeszcze do podziału jądra, zatem są 2 genomy.

1.2. Dynamika genomu

1.2.4. Analiza wariantów strukturalnych w genomie człowieka

Warianty strukturalne to regiony w genomie o długości co najmniej 1 kbp, które zawierają inwersje, translokacje, większe insercje i delecje.

- Proszę wejść na stronę <https://www.ncbi.nlm.nih.gov/dbvar/browse/>
- W tabeli proszę wybrać poniższe pozycje i zaakceptować enterem (Rys. 1.2.4.a):

- ▶ *Homo sapiens*,
- ▶ GRCH38,
- ▶ Chr. 10.

- Po wyświetleniu się widoku chromosomu 10 proszę rozwinąć koło zębate w prawym dolnym rogu i zaznaczyć pozycje zgodnie z Rys. 1.2.4b.

Otrzymany obraz pokazuje warianty strukturalne w obrębie chromosomu 10. Warianty strukturalne oznaczone są kolorami.

- ▶ Barwa ciemnofioletowa: zmiany liczby kopii różnego typu.
- ▶ Barwa granatowa: zwiększenie liczby kopii, najczęściej są to duplikacje.
- ▶ Barwa czerwona: utrata kopii, najczęściej delecje.
- ▶ Barwa ciemnobrązowa: tandemowe powtórzenia.
- ▶ Barwa jasnofioletowa: inwersje.
- ▶ Barwa fioletowa, paskowana: translokacje.
- ▶ Barwa błękitna, lazurowa: złożone zmiany strukturalne różnego typu.

A. Na podstawie otrzymanego obrazu chromosomu 10 oraz podanych powyżej wyjaśnień, proszę podać, jakiego typu warianty strukturalne możemy wyróżnić.

- W obrębie chromosomu 10 najczęściej występują warianty dotyczące (Rys. 1.2.4c):
 - ▶ zwiększenia liczby kopii przez duplikacje (barwa granatowa);
 - ▶ utraty liczby kopii na skutek delecji (barwa czerwona).

B. Za pomocą suwaka +/- proszę zawęzić obserwowany zakres do obszaru 45-48 mln bp i z otrzymanego rysunku odczytać numer i długość najdłuższego wariantu związanego ze zmiennością liczby kopii różnego typu.

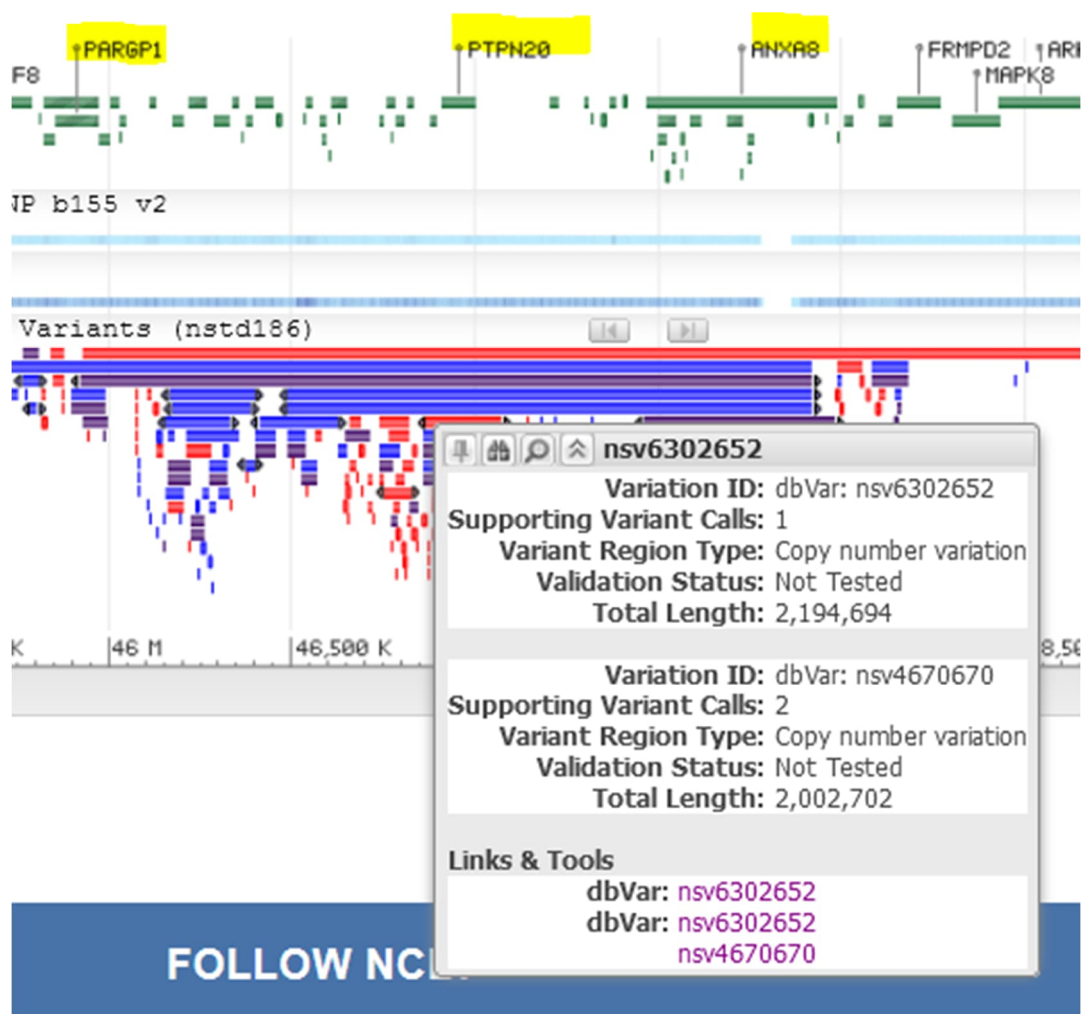
- Odczytujemy dane dla ciemnofioletowego fragmentu (Rys 1.2.4d).
- Wariant 1: numer – nsv6302652, długość obszaru – 2 194 694 bp.
- Wariant 2: numer – nsv4670670, długość obszaru – 2 002 702 bp.

C. Proszę podać symbole oraz funkcje genów, które występują w regionie objętym wariantem strukturalnym opisanym w punkcie B.

- PARGP1-AGAP4: pseudogen.
- PTPN20: gen fosfatazy tyrozynowej, typ niereceptorowy 20.
- ANXA8: gen aneksyny A8.



Rys. 1.2.4c. Warianty strukturalne w chromosomie 10 człowieka. Widok ogólny.



Rys. 1.2.4d. Warianty strukturalne w chromosomie 10 człowieka, w regionie 45-48 mln bp.

2. Wielkość genomu

2.2. Przeliczanie jednostek

2.2.1. Proszę podać długość odcinków DNA w metrach, centymetrach i milimetrach.

A. 15 kbp

- Odległość między zasadami w DNA wynosi 3,4 Å, czyli $3,4 \times 10^{-10}$ m.
- 15 kbp to 15×10^3 bp.
- **$15 \times 10^3 \text{ bp} \times 3,4 \times 10^{-10} \text{ m} = 51 \times 10^{-7} \text{ m} = 5,1 \times 10^{-6} \text{ m} = 5,1 \times 10^{-4} \text{ cm} (0,00051 \text{ cm}) = 5,1 \times 10^{-3} \text{ mm} (0,0051 \text{ mm}) = 5,1 \mu\text{m}.$**

B. 1,2 Mbp

- Odległość między zasadami w DNA wynosi 3,4 Å, czyli $3,4 \times 10^{-10}$ m.
- 1,2 Mbp = $1,2 \times 10^6$ bp.
- **$1,2 \times 10^6 \times 3,4 \times 10^{-10} \text{ m} = 4,08 \times 10^{-4} \text{ m} (0,000408 \text{ m}) = 4,08 \times 10^{-2} \text{ cm} (0,0408 \text{ cm}) = 4,08 \times 10^{-1} \text{ mm} = 0,408 \text{ mm}.$**

C. 3 Gbp.

- Odległość między zasadami w DNA wynosi 3,4 Å, czyli $3,4 \times 10^{-10}$ m.
- 3 Gbp = 3×10^9 bp.
- **$3 \times 10^9 \times 3,4 \times 10^{-10} \text{ m} = 10,2 \times 10^{-1} \text{ m} = 1,02 \text{ m} = 102 \text{ cm} = 1020 \text{ mm}.$**

2.2.2. Przyjmując, że 1 Gbp waży 1 pg proszę podać masę następujących genomów w gramach.

A. 80 kbp

- 10^9 bp waży 1 pg.
- 80 kbp = 80×10^3 bp, zatem
- 80 kbp waży $80 \times 10^3 / 10^9 = 80 \times 10^3 \times 10^{-9} = 8 \times 10^4 \times 10^{-9} = 8 \times 10^{-5}$ pg.
- **$1 \text{ pg} = 10^{-12} \text{ g}$ zatem $8 \times 10^{-5} \text{ pg} = 8 \times 10^{-17} \text{ g} = 80 \text{ attogramów}.$**

B. 25 Mbp

- 10^9 bp waży 1 pg.
- 25 Mbp = 25×10^6 bp, zatem
- 25 Mbp waży $25 \times 10^6 / 10^9 = 25 \times 10^6 \times 10^{-9} \text{ pg} = 25 \times 10^{-3} \text{ pg}.$
- **$1 \text{ pg} = 10^{-12} \text{ g}$ zatem $25 \times 10^{-3} \text{ pg} = 25 \times 10^{-15} \text{ g} = 25 \text{ femtogramów}.$**

2.2.3. W tabeli 2.2.3 podano wielkość wybranych chromosomów człowieka w różnych jednostkach. Przyjmując odległość między zasadami 3,4 Å oraz masę 1 pg dla 978 Mbp, proszę uzupełnić brakujące wartości w tabeli. Długość i masę proszę zaokrąglić do części setnych.

Tabela 2.2.3. Liczba nukleotydów, długość i masa wybranych chromosomów człowieka.			
Chromosom	Liczba nukleotydów (bp)	Długość (cm)	Masa (pg)
1	248 956 422	8,46	0,25
3	190 588 235	6,48	0,19
20	68 460 000	2,33	0,07

● Chromosom 1

- ▶ Odległość między zasadami wynosi 3,4 Å, czyli $3,4 \times 10^{-10}$ m.
- ▶ $248\,956\,422 \text{ bp} \times 3,4 \times 10^{-10} \text{ m} = 248\,956\,422 \text{ bp} \times 3,4 \times 10^{-8} \text{ cm} = 2,489\,564\,220 \times 3,4 \text{ cm} = \mathbf{8,46 \text{ cm}}$.
- ▶ 978 Mbp czyli 978×10^6 waży 1 pg, zatem
- ▶ $248\,956\,422 \text{ bp}$ waży $248\,956\,422/978 \times 10^6 = 254\,556,67 \times 10^{-6} \text{ pg} = \mathbf{0,25 \text{ pg}}$.

● Chromosom 3

- ▶ Odległość między zasadami wynosi 3,4 Å, czyli $3,4 \times 10^{-10}$ m czyli $3,4 \times 10^{-8}$ cm.
- ▶ Zatem liczba zasad odpowiadająca 6,48 cm to $6,48/3,4 \times 10^{-8} = 1,905\,882\,353 \times 10^8$ czyli **190 588 235 bp**.
- ▶ 978 Mbp czyli 978×10^6 waży 1 pg, zatem
- ▶ $190\,588\,235 \text{ bp}$ waży $190\,588\,235/978 \times 10^6 = 194\,875,50 \times 10^{-6} \text{ pg} = \mathbf{0,19 \text{ pg}}$.

● Chromosom 20

- ▶ 978 Mbp czyli 978×10^6 waży 1 pg, zatem
- ▶ $0,07 \text{ pg}$ to $978 \times 10^6 \times 0,07 = 68,46 \times 10^6 \text{ bp}$, czyli **68 460 000 bp**.
- ▶ Odległość między zasadami wynosi 3,4 Å, czyli $3,4 \times 10^{-10}$ m czyli $3,4 \times 10^{-8}$ cm.
- ▶ $3,4 \times 10^{-8} \text{ cm} \times 68\,460\,000 = \mathbf{2,33 \text{ cm}}$.

3. Wartość C

3.3. Analiza wartości C u różnych grup organizmów

3.3.3. Wartość C u wybranych gatunków zwierząt

W bazie Animal Genome Size Database (<http://www.genomesize.com/index.php>) proszę przejść do sekcji „search data” i wprowadzić nazwę gatunkową w polu przeszukiwania „species”. Na podstawie danych podanych w sekcji proszę uzupełnić tabelę.

Tabela 3.3.3. Porównanie wielkości genomów u różnych grup kręgowców.					
Nazwa gatunkowa	Nazwa polska	Wartość C [pg]	Metoda	Komórka	Standard
<i>Latimeria chalumnae</i>	Latimeria	2,75	Cytometria przepływ.	Krwinka czerwona	<i>Gallus domesticus</i>
<i>Pelodytes punctatus</i>	Nurzaniec błotny	2,44	Cytometria przepływ.	Krwinka czerwona	<i>Homo sapiens</i> , <i>Rana temporaria</i>
<i>Laudakia caucasia</i>	Agama kaukaska	1,87	Cytometria przepływ.	Krwinka czerwona	<i>Bufo fowleri</i>
<i>Apus caffer</i>	Jerzyk widłosterny	1,18	Feulgen	Krwinka czerwona	<i>Gallus domesticus</i>
<i>Capra hircus</i>	Koza domowa	3,24	Feulgen	Leukocyty	<i>Bos taurus</i>

3.3.4. Analiza wielkości genomu u różnych grup roślin

Tabela 3.3.4. Porównanie wielkości genomów u różnych grup roślin.				
Nazwa gatunkowa	Nazwa polska	1C [pg]	2C [pg]	2C [Mbp]
<i>Cichorium intybus</i>	Cykoria podróżnik	1,145	2,285	2 235
<i>Centaurea solstitialis</i>	Chaber wetnisty	1,283	2,561	2 506
<i>Matricaria chamomilla</i> subsp. <i>chamomilla</i>	Rumianek pospolity	2,613	5,218	5 103
<i>Solidago virgaurea</i>	Nawłóć pospolita	0,860	1,716	1 678
<i>Tussilago farfara</i>	Podbiał pospolity	1,700	3,470	3 321

3.4. Podawanie wielkości genomu w różnych jednostkach

Przyjmując, że odległość między zasadami w DNA wynosi 3,4 Å oraz 978 Mbp waży 1 pg, proszę uzupełnić tabelę 3.4.

Tabela 3.4. Wielkość genomów oraz zawartość DNA w komórkach różnych grup organizmów.

Gatunek	Wielkość genomu [Mbp]	Wielkość genomu w metrach [m]	Wielkość genomu w pikogramach [pg]	Wartość C [C-value]	Zawartość DNA w komórkach somatycznych (2n)
<i>Escherichia coli</i>	4,6	$15,64 \times 10^{-4} = 1,564 \text{ mm}$	$4,74 \times 10^{-3}$	$4,74 \times 10^{-3}$	$4,74 \times 10^{-3}$
<i>Drosophila melanogaster</i>	140,0	$476 \times 10^{-4} \text{ m} = 47,6 \text{ mm} = 4,76 \text{ cm}$	0,143 pg	0,143	0,286
<i>Physcomitrella patens</i>	510,0	$1734 \times 10^{-4} \text{ m} = 173,4 \text{ mm} = 17,34 \text{ cm}$	0,521 pg	0,521	0,521
<i>Triticum aestivum</i>	17 000,0	5,78 m	17,38	52,14	104,28
<i>Homo sapiens</i>	3 200,0	1,088 m	3,27 pg	3,27	6,54

4. Gęstość genów

4.1. Gęstość genów u różnych grup organizmów

4.1.1. Gęstość genów u różnych grup organizmów Proszę uzupełnić poniższą tabelę

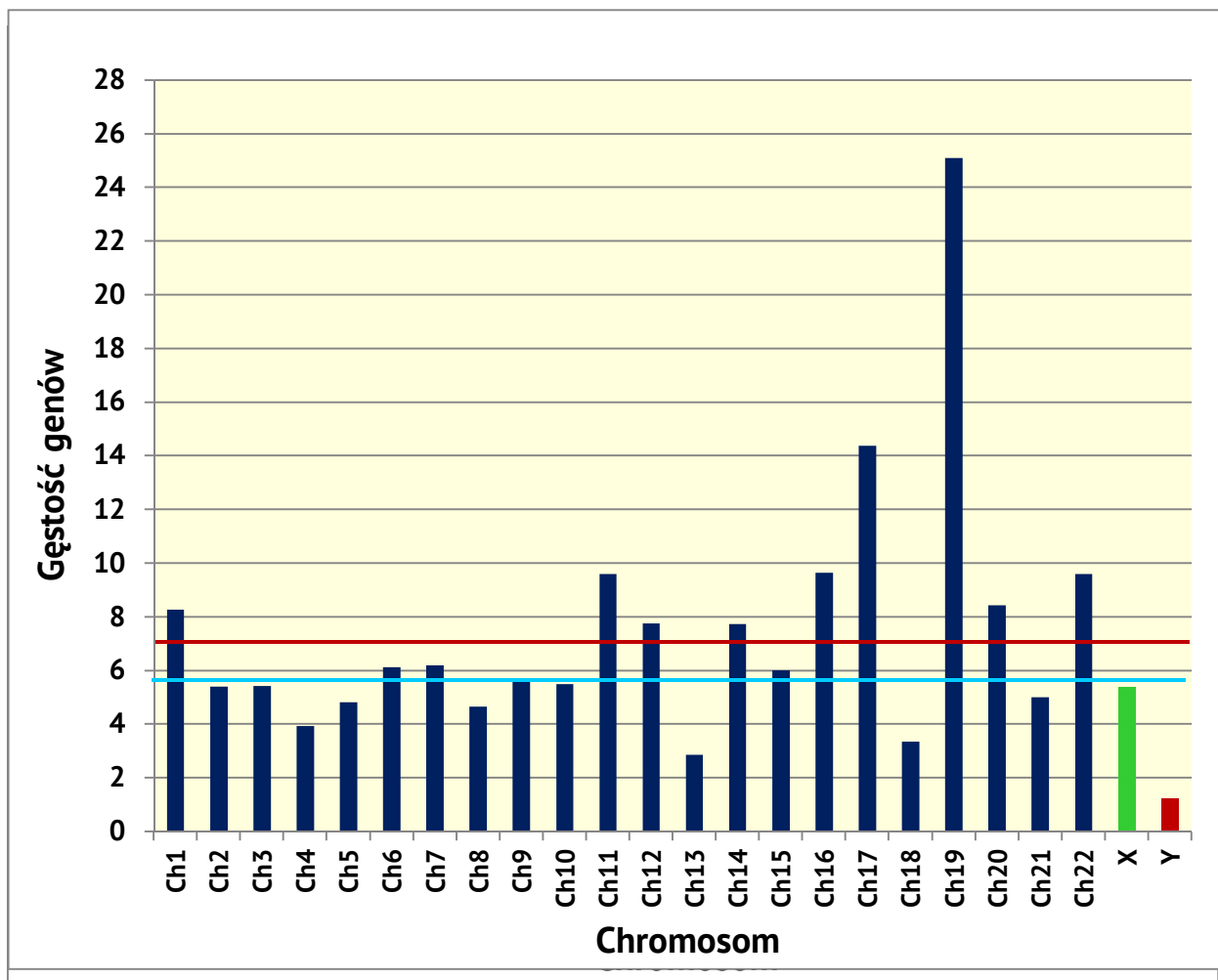
Tabela 4.1.1. Gęstość genów u różnych organizmów.

Obiekt	Wielkość genomu	Liczba genów, która przypadłaby na 1 m długości genomu	Liczba genów	Gęstość genów
A	4,3 Mbp	2 735 971	4 000	930,23
B	120 Mbp	490 206	20 000	166,67
C	15 Gbp	5 883	30 000	2,00
D	120 Gbp	1 471	60 000	0,50

4.2. Gęstość genów u człowieka

W tabeli poniżej podano uśrednioną liczbę nukleotydów dla poszczególnych chromosomów człowieka oraz liczbę genów zmapowanych na każdym chromosomie. Proszę obliczyć gęstość genów dla chromosomów. Uzyskane wyniki proszę przedstawić na wykresie i na jego podstawie dokonać analizy rozkładu gęstości genów u człowieka.

Tabela 4.2. Gęstość genów na chromosomach człowieka.			
Chromosom	Liczba nukleotydów (bp)	Liczba genów kodujących białka	Średnia gęstość
1	248 956 422	2 058	8,27
2	242 193 529	1 309	5,40
3	198 295 559	1 078	5,44
4	190 241 555	752	3,95
5	181 538 259	876	4,83
6	170 805 979	1 048	6,14
7	159 345 973	989	6,21
8	145 138 636	677	4,66
9	138 394 717	786	5,68
10	133 797 422	733	5,49
11	135 086 622	1 298	9,61
12	133 275 309	1 034	7,76
13	114 364 328	327	2,86
14	107 043 718	830	7,75
15	101 991 189	613	6,01
16	90 338 345	873	9,66
17	83 257 441	1 197	14,38
18	80 373 285	270	3,36
19	58 617 616	1 472	25,11
20	64 444 167	544	8,44
21	46 709 983	234	5,01
22	50 818 468	488	9,60
X	156 040 895	842	5,40
Y	57 227 415	71	1,24



- Linia czerwona oznacza gęstość średnią (7,2), linia niebieska: mediana (5,8).
- Około połowa chromosomów ma gęstość poniżej średniej, oscylującą wokół mediany.
- Gęstość genów wyższa niż średnia występuje w chromosomach 1, 11, 12, 14, 16, 17, 19, 20, 22.
- Najniższą gęstość obserwuje się dla chromosomu Y, najwyższą dla chromosomu 19.
- Generalnie gęstość genów u człowieka jest niska.