

## Ćwiczenie 06: Genomy

Miary wielkości genomu. Przeliczanie wielkości genomów podanych w różnych jednostkach. Analiza wielkości genomów na podstawie wartości C (C-value). Gęstość genów.

Kornelia Polok

### 1. Miary wielkości genomu. Przeliczanie wielkości genomów podanych w różnych jednostkach.



#### 1.1. Wielkość genomu

Wielkość genomu jest to zawartość DNA w pojedynczym genomie. Podaje się ją jako całkowitą liczbę par zasad 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 genomie diploidalnym, łatwo określić zawartość w genomie haploidalnym. Podobnie w przypadku sekwencji. Genom diploidalny 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.

## 1.2. Miary wielkości genomów

## 1.2.1. Liczba nukleotydów – liczba par zasad (bp).

- 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)

## 1.2.2. Jednostki masy: najczęściej podajemy w postaci pikogramów.

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

## 1.2.3. Jednostki długości

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

## 1.2.4. Przeliczanie jednostek



Odległość między zasadami w DNA wynosi 3,4 Å. Proszę podać jaką długość ma odcinek DNA. Długość proszę podać w metrach oraz w jednostce odpowiednio mniejszej (np. 0,01 m = 1 cm).

- A. 10 kbp
- B. 1 Mbp
- C. 2 Gbp

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

- A. 100 kbp
- B. 10 Mbp

1.3. Liczba nukleotydów, **długość** oraz masa wybranych chromosomów człowieka

W tabeli 1 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. Proszę przedstawić obliczenia.

Tabela 1. Liczba nukleotydów, długość i masa wybranych chromosomów człowieka.			
Chromosom	Liczba nukleotydów (bp)	Długość (cm)	Masa (pg)
2	242 193 529		
6		5,81	
16			0,09

*Samodzielne wykonanie: 6p*

*Termin: 13.05.2022., 23:59*

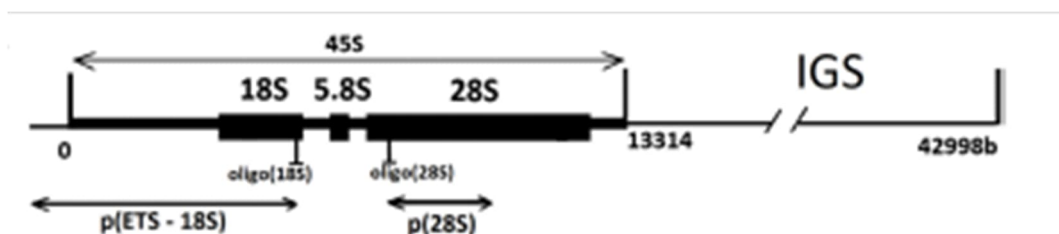
## 2. Analiza wielkości genomów na podstawie wartości C

**Wartość C (C-value); ilość DNA w genomie haploidalnym. Ze względu na różny stopień ploidalności wartość C jest zróżnicowana u różnych organizmów i nie ma związku ze złożonością funkcji tego organizmu. Wysoka wartość C wynika z obecności intronów, sekwencji regulatorowych, sekwencji powtarzalnych, pseudogenów, oraz licznych sekwencji transpozonowych.**

### 2.1. Sekwencje powtarzalne

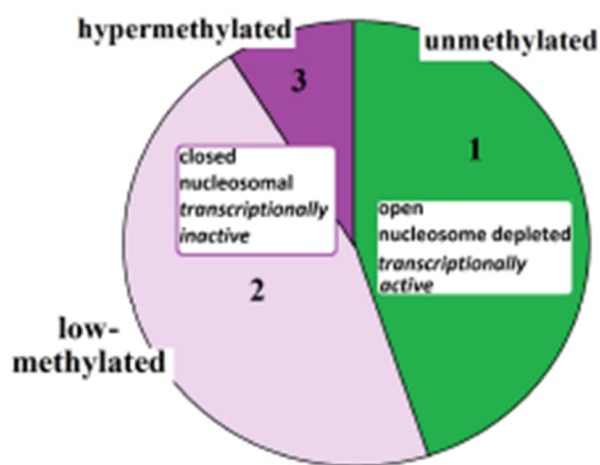
#### ➔ 2.1.1. Sekwencje tandemowo powtórzone, rDNA

Wiele sekwencji powtarzalnych w genomie występuje w postaci tandemowych powtórzeń. Należą do nich między innymi sekwencje rDNA, które kodują rRNA (Rys. 2.1.1a). Zlokalizowane są one na kilku chromosomach (NOR): 13, 14, 15, 21 i 22. Są one silnie skondensowane w postaci heterochromatyny.



Rys. 2.1.1a. Schemat genów rDNA człowieka w postaci tandemowego powtórzenia (Malinovskaya et al. 2019).

Człowiek zawiera około 400 kopii rDNA. Jedynie 30-50% z tych kopii jest aktywna transkrypcyjnie. Frakcja ta występuje w postaci euchromatyny (Rys. 2.1.1b). Histony występują w postaci charakterystycznej dla miejsc aktywnych, większość wysp CpG jest niezmetylowana (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 (fioletowy fragment koła). Zawartość poszczególnych frakcji zmienia się z wiekiem.



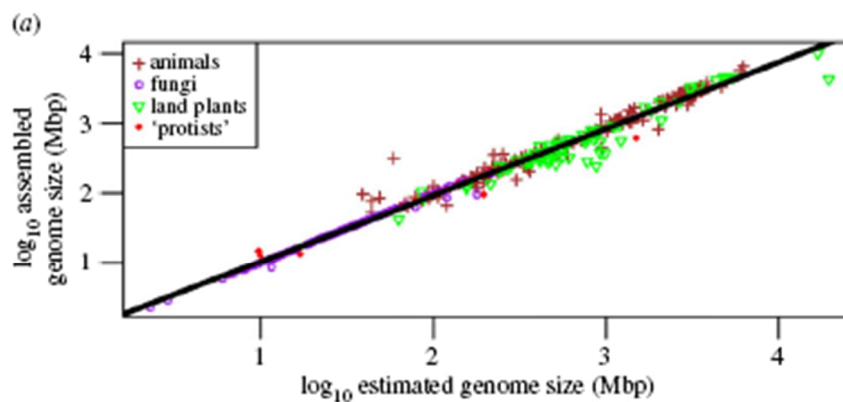
Rys. 2.1.1b. Udział aktywnych transkrypcyjnie i nieaktywnych frakcji genów rDNA u człowieka (Malinovskaya et al. 2019).

### 2.1.2. Rozmieszczenie mRNA

Badania ekspresji genów, a zwłaszcza transkrypcji wykazały, że zaledwie 2,7% sekwencji unikalnych w DNA jest obecna w postaci 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.

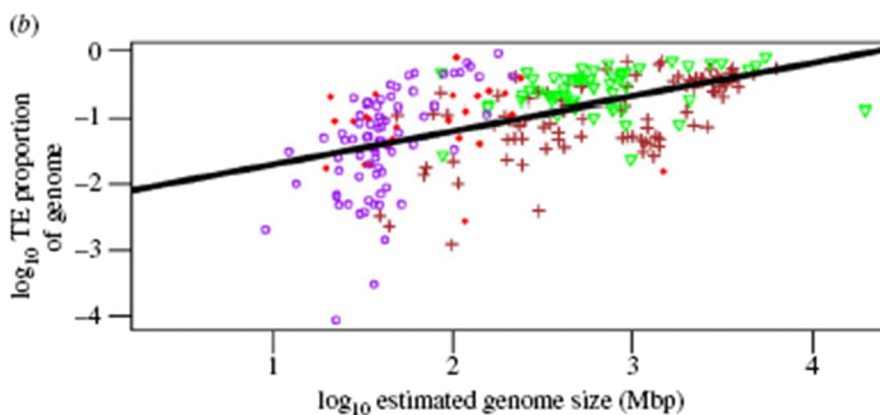
### 2.2. Wartość C (C-value): od czego zależy?

Wartość C jest najczęściej oceniana na podstawie analiz cytogenetycznych, jednakże w dużej mierze koreluje ona z wielkością genomów określoną w projektach sekwencjonowania (Rys. 2.2a).



Rys. 2.2a. Korelacja wielkości genomu ustalonego na podstawie badań cytogenetycznych (oś odciętych) z wartościami ustalonymi na podstawie sekwencjonowania (oś rzędnych) (Elliot and Gregory 2015).

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. 2.2b).

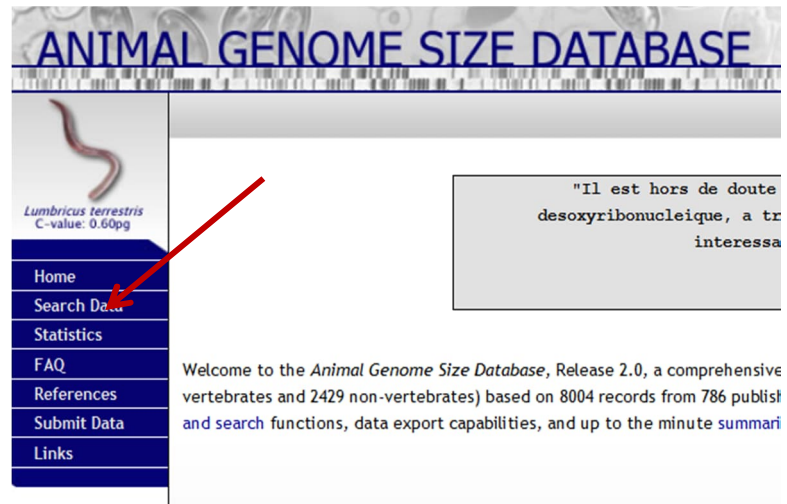


Rys. 2.2b. Korelacja wielkości genomu (oś odciętych) z zawartością ruchomych elementów genetycznych (oś rzędnych) (Elliot and Gregory 2015).

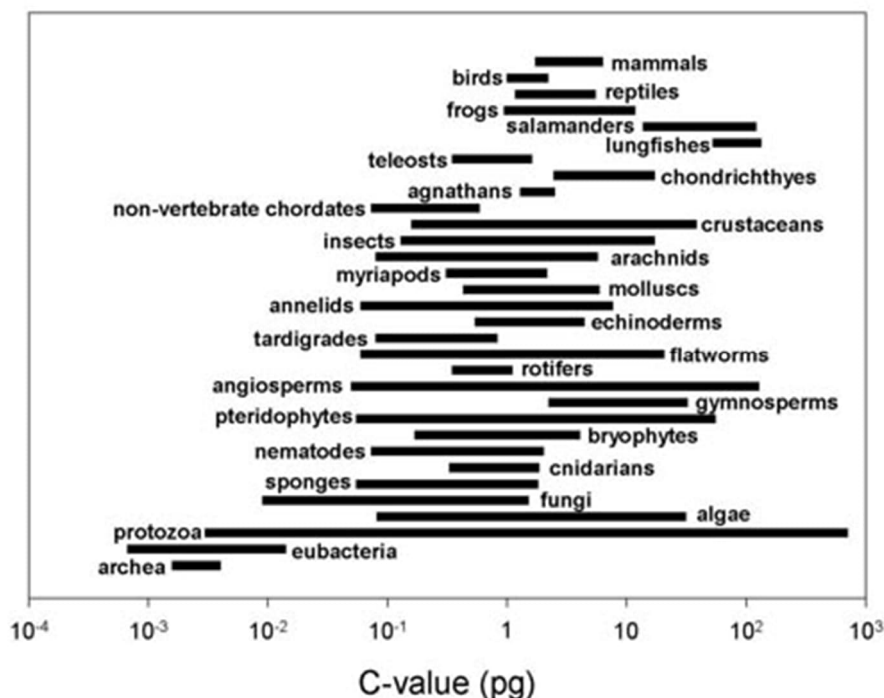
## 2.3. Analiza wartości C u wybranych grup organizmów

### 2.3.1. Baza wartości C w pikogramach dla genomów zwierzęcych

Baza wartości C dla genomów (Rys. 2.3.1a) zwierzęcych zawiera obecnie dane dla 6222 genomów, w tym dla 3793 (61%) genomów kręgowców. Dane podane są jako wartość C dla genomu haploidalnego. Zakres wartości C dla genomów 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.



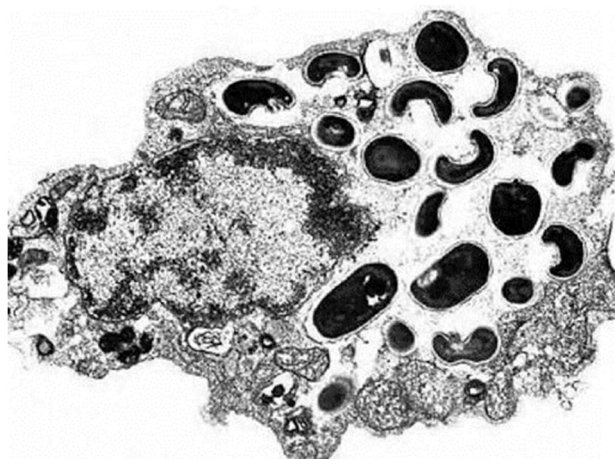
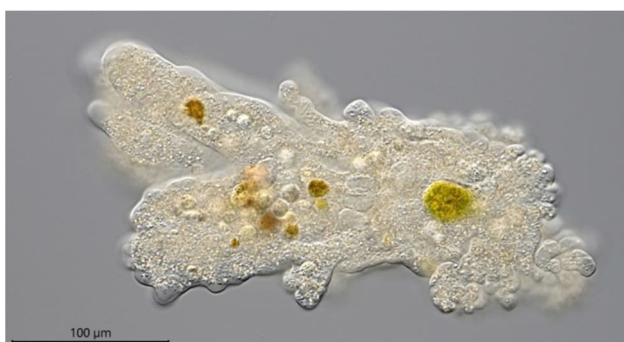
Rys. 2.3.1a. Zrzut ekranu dla bazy genomów zwierzęcych.



Rys. 2.3.1b. Zakres wartości C dla genomów Eukariota.

## 2.3.2. Przykłady organizmów 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. 2.3.2a. *Encephalitozoon intestinalis*.Rys. 2.3.2b. *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.

## 2.3.3. Wartość C u wybranych gatunków.



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 2. Porównanie wielkości genomów u różnych grup kręgowców.

Nazwa gatunkowa	Nazwa polska	Wartość C [pg]	Metoda	Komórka	Standard
<i>Negaprion brevirostris</i>					
<i>Megophrys montana</i>					
<i>Gekko gecko</i>					
<i>Cygnus olor</i>					
<i>Felis lynx</i>					

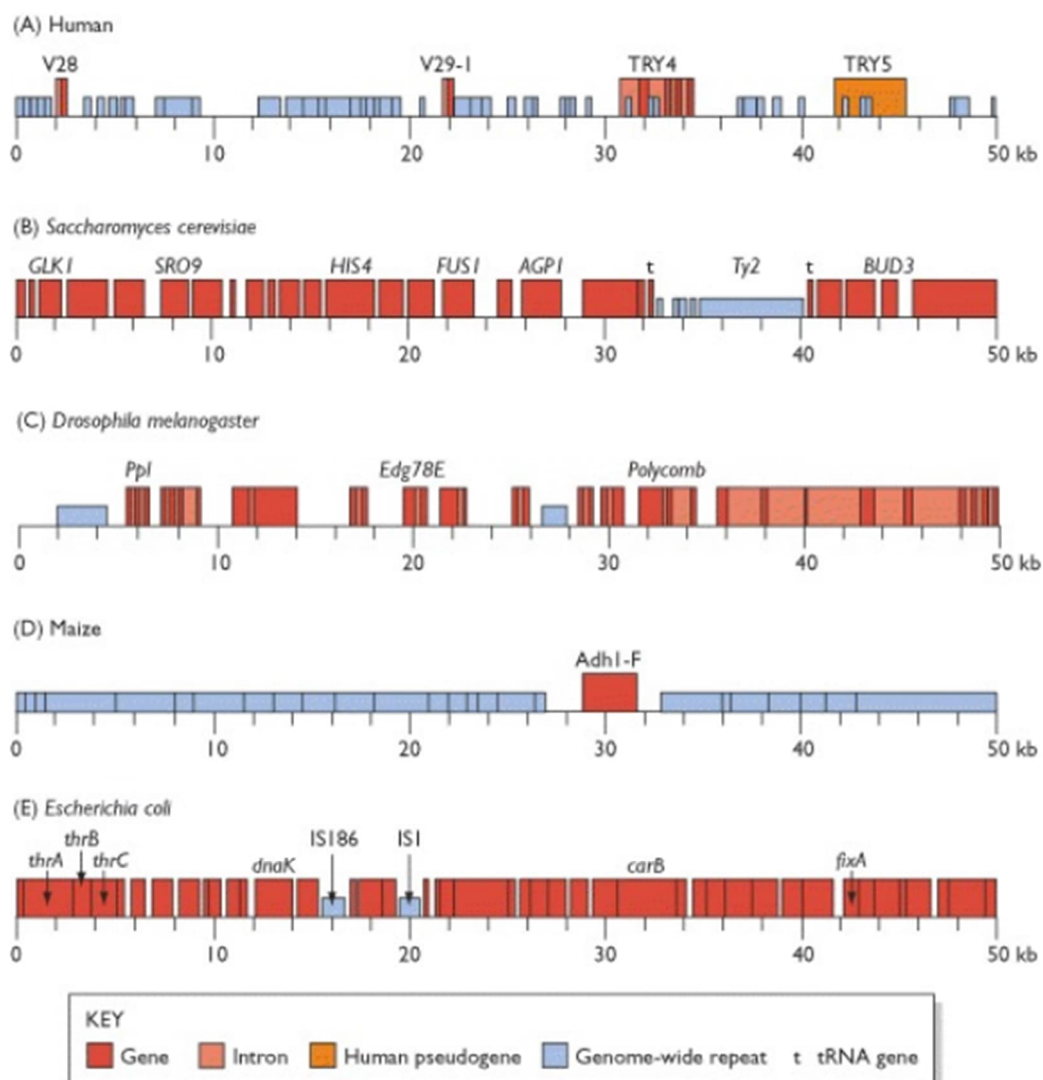


### 3. Gęstość genów

**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.

#### ➔ 3.1. Gęstość genów u różnych 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. 3.1). 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. 3.1. Porównanie gęstości genów (czerwone) u różnych organizmów.

## 3.2. Gęstość genów u człowieka



W tabeli poniżej podano uśrednioną liczbę nukleotydów dla poszczególnych chromosomów człowieka oraz liczbę genów zmapowanych na każdym chromosomie.

A. Czy gęstość genów jest skorelowana z wielkością chromosomu?

B. Proszę podać średnią i medianę gęstości genów.

C. Proszę podać chromosomy o największej i najmniejszej gęstości.

Tabela 3. 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