

Ćwiczenie C05 Geny i genomy

Koncepcja genu Geny różnych organizmów w bazach danych Wielkość genomu i wartość C Genom człowieka

Kornelia Polok

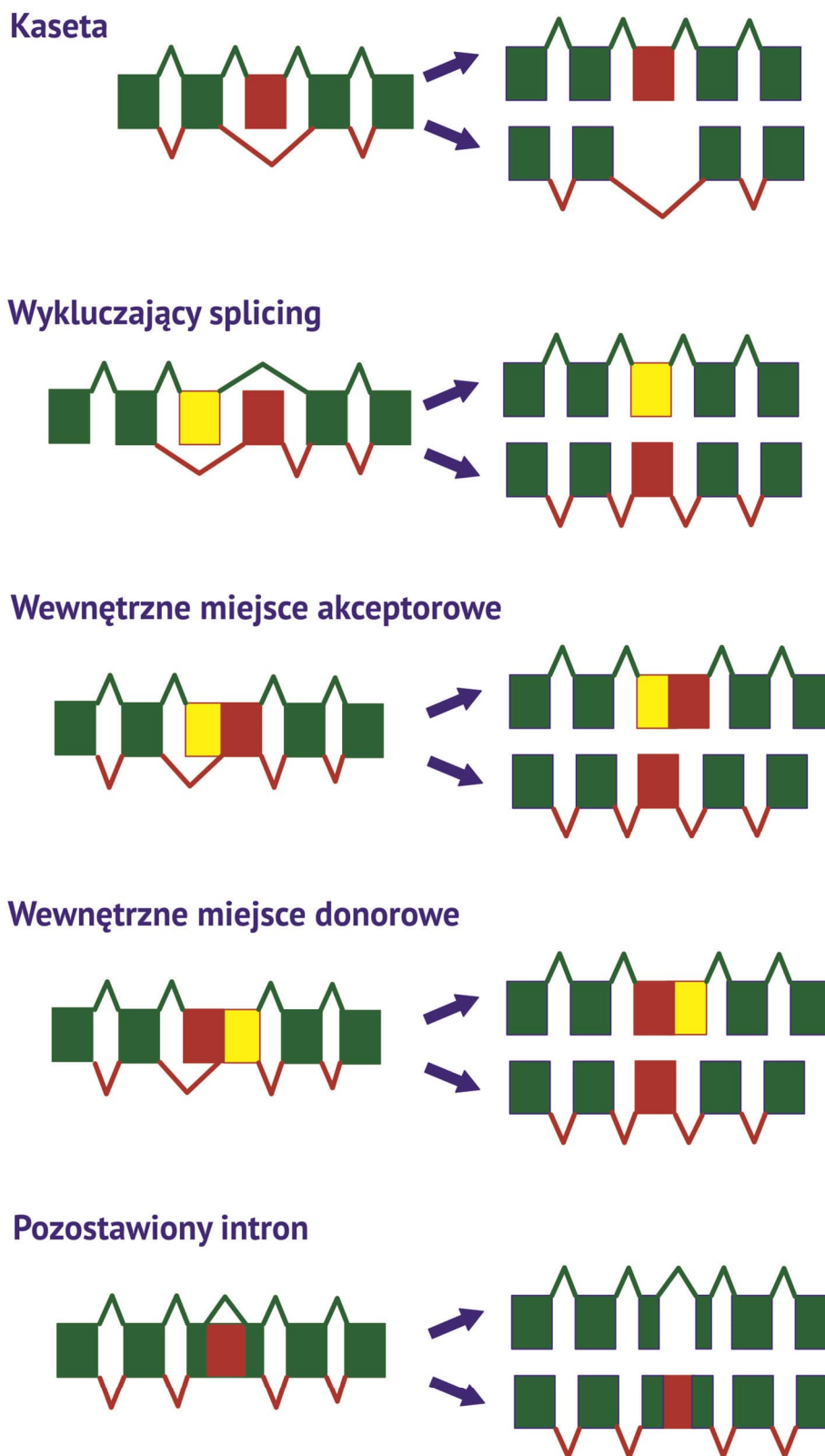
1. Koncepcja genu

1.1. Definicja genu

➔ Pierwotnie gen opisywano jako „jednostkę dziedziczności”, przy czym pojęcie jednostki ewoluowało wraz z postępem badań. Obecnie pojęcie genu odnosi się do fragmentu DNA (lub RNA u wirusów RNA), który odpowiada za powstanie funkcjonalnego produktu.

Gen to fragment lub kilka fragmentów DNA kodujących jeden lub kilka funkcjonalnych produktów będących białkiem lub RNA.

- Współczesna definicja genu uwzględnia sytuacje, gdy na bazie kilku fragmentów DNA może powstać wiele produktów na skutek różnej obróbki potranskrypcyjnej. Ponadto definicja uwzględnia możliwość transkrypcji pseudogenów, oddalenie elementów regulatorowych (Rys 1.1).
- Pseudogeny powstają w wyniku duplikacji funkcjonalnych genów, a następnie ewolucji w kierunku form niefunkcjonalnych. Wiele pseudogenów powstaje jako retrogeny (retrokopie) na skutek działalności retrotranspozonów. U człowieka obecnie jest aktywnych około 100 elementów *LINE* oraz 16 *HERV*. Około 20% genów ludzkich to pseudogeny, które podlegają transkrypcji. Ich funkcja nie jest znana.
- Dane z projektów sekwencjonowania genomów i transkryptomów wykazały, że jeden gen często wpływa na wiele cech (plejotropia), a także, że funkcja genów ulega zmianom pod wpływem czynników środowiskowych, przy czym mogą to być czasowe modulacje. Przykładowo, uruchomienie odpowiedzi immunologicznej może wiązać się z aktywacją retrotranspozonów pod wpływem patogenów. Ekspresja retrotranspozonów z kolei może uruchomić receptory typu Toll, które zainicjują tworzenie przeciwciał.

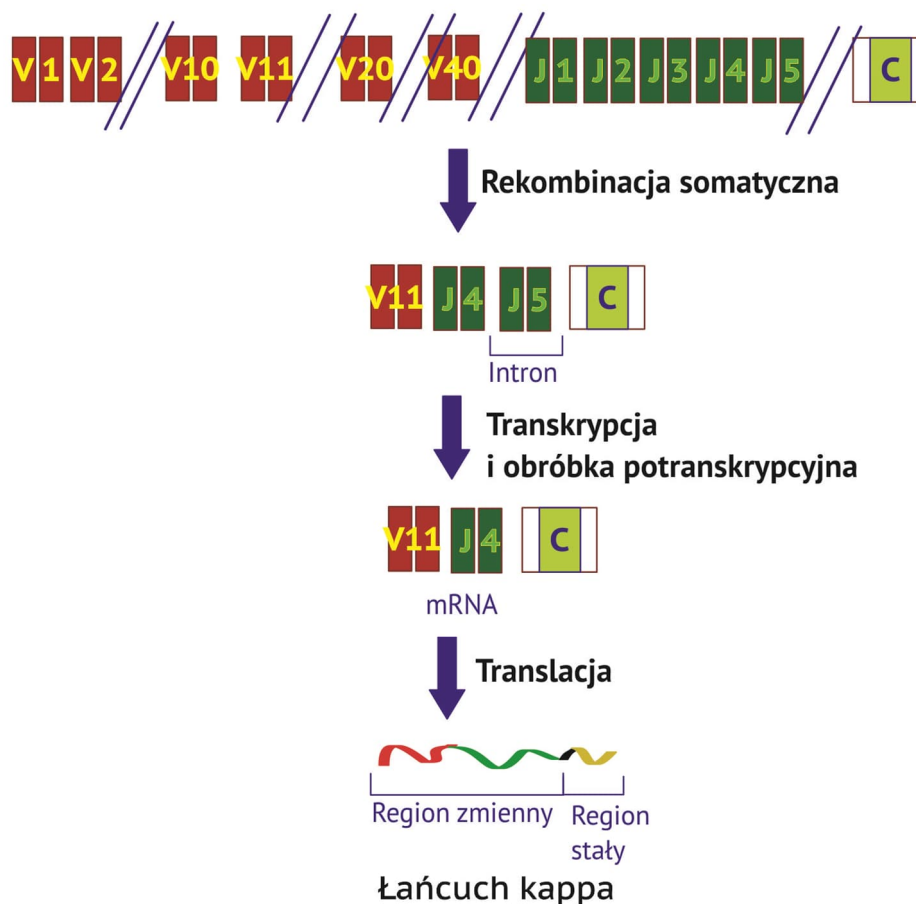


Rys. 1.1. Alternatywny splicing. **Kaseta**: jeden z egzonów jest pominięty w mRNA. **Wykluczający splicing**: egzony, które wykluczają się wzajemnie w mRNA, może występować tylko jeden z egzonów wykluczających się. **Wewnętrzne miejsce akceptorowe**: fragment od końca 5' (zaznaczony na żółto), który leży w obrębie sekwencji kodującej i może być wycięty z mRNA jako intron. **Wewnętrzne miejsce donorowe**: fragment od końca 3' (zaznaczony na żółto), który leży w obrębie sekwencji kodującej i może być wycięty jako intron. **Pozostawiony intron**: intron w obrębie większego egzonu, który może być pozostawiony lub może być wycięty.

1.2. Składanie genów kodujących lekkie łańcuchy kappa przeciwciał kręgowców

U kręgowców przeciwciała stanowią ochronę przed patogenami i innymi obcymi substancjami atakującymi organizm. Każde przeciwciało składa się z czterech łańcuchów polipeptydowych: dwóch ciężkich i dwóch lekkich. Łańcuchy lekkie występują w dwóch formach: kappa i lambda. Każdy łańcuch peptydowy zawiera region zmienny oraz region stały. W komórkach rozrodczych oraz we wczesnym rozwoju zarodkowym DNA kodujące przeciwciała zbudowane jest z segmentów V, J i C. Są one połączone. Geny warunkujące dane przeciwciało są składane z istniejących segmentów dopiero podczas różnicowania limfocytów B, czyli komórek produkujących przeciwciała.

Składanie genów podczas różnicowania przeciwciał można prześledzić na przykładzie lekkiego łańcucha kappa (Rys. 1.2.). U człowieka za łańcuch ten odpowiada klaster na chromosomie 2, który składa się z 76 segmentów V, z czego co najmniej 40 jest aktywnych. Ponadto w skład klastra wchodzi 5 segmentów J (łączyjących) oraz pojedynczy stały segment C. Podczas różnicowania limfocytów B, w wyniku rekombinacji somatycznej dochodzi do połączenia jednego segmentu V z jednym segmentem J oraz z segmentem C. Wszystkie segmenty między tymi połączonymi stają się intronami i są usuwane podczas dojrzewania. Podobne procesy zachodzą podczas składania genów tworzących łańcuchy ciężkie, łańcuchy lambda oraz receptory białkowe limfocytów T.



Rys. 1.2. Składanie genów kodujących łańcuch lekkie kappa przeciwciał w trakcie dojrzewania limfocytów B u człowieka.



1.2.1. Charakterystyka genu kodującego region zmienny łańcucha kappa

Na podstawie sekwencji genu kodującego region zmienny łańcucha kappa zdeponowanej w bazie NCBI pod numerem akcesyjnym AH002839.2, proszę podać następujące dane.

- A. Długość sekwencji zdeponowanej w NCBI.
- B. Lokalizację na chromosomie.
- C. Tkanę, z której pobrano DNA.
- D. Liczbę, pozycję fragmentów kodujących segmenty J oraz ich symbol.
- E. Jaką długość mają fragmenty kodujące segmenty J?
- F. Proszę podać symbol fragmentu odpowiadającego mRNA dla regionu stałego, C, jego lokalizację i długość.

2. Geny różnych organizmów w bazach danych

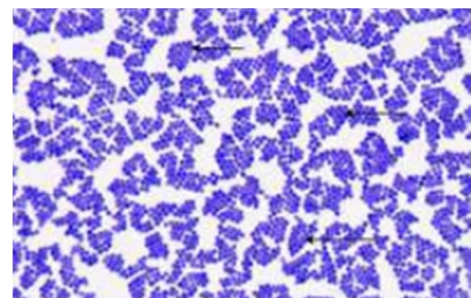
2.1. Geny Prokariota: *Staphylococcus aureus*

➔ Kosmetyki oraz zabiegi kosmetyczne mogą być źródłami zakażenia mikroorganizmami, w tym bakteriami, glonami i grzybami. Generalnie producenci kosmetyków są odpowiedzialni za bezpieczeństwo ich produktów. Z kolei salony kosmetyczne muszą spełniać określone normy higieniczne. Z drugiej strony, kosmetyki stanowią pożywkę dla mikroorganizmów, w tym bakterii, które mogą prowadzić do infekcji skóry. Do najczęściej identyfikowanych bakterii w salonach kosmetycznych należą bakterie z rodzaju *Staphylococcus* i *Pseudomonas*. Ich szczepy często są odporne na antybiotyki.

2.1.1. Charakterystyka *S. aureus*

Staphylococcus aureus (Gronkowiec złoty) to bakteria gram-dodatnia występująca w jamie nosowo-gardłowej oraz na skórze. W warunkach *in vitro* bakteria tworzy żółte lub złote kolonie. Komórki mają sferyczny kształt i tworzą groniaste kolonie (Rys. 2.1.1a). Jest organizmem aerobowym lub czasowo anaerobowym.

Bakteria jest powszechna w środowisku człowieka i szacuje się, że od 15-50% populacji ludzkiej jest jej nosicielem bez wywoływania jakichkolwiek objawów. Szczególnie wysoki poziom kolonizacji przez *S. aureus* (80%) obserwuje się w populacjach pracowników służb medycznych oraz osób często biorących zastrzyki (np. diabetycy). Ryzyko zakażenia pojawia się, gdy dochodzi do przerwania powłok skórnych np. w trakcie zranień, zabiegów. Bakteria najczęściej powoduje zakażenie skóry, ale może ono przenosić się na inne organy, a po wnikięciu do krwi prowadzi do stanów zagrażających życiu. Ewolucja szczepów wielolekoopornych stanowi poważny problem kliniczny.



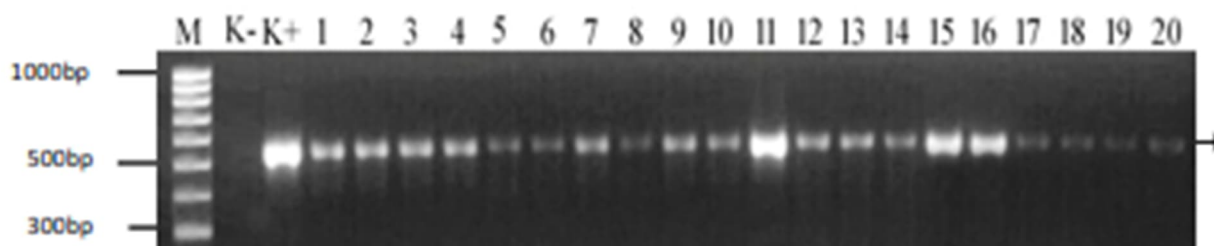
Rys. 2.1.1a. Góra: zabarwiona bakteria *S. aureus*. Dół: typowe, groniaste kolonie bakterii.

2.1.2. Genetyczne uwarunkowania lekooporności *S. aureus*

W diagnostyce klinicznej szczególnie kłopotliwe są szczepy odporne na metacyklinę (MRSA). Cecha ta związana jest z obecnością genu *mecA*, który koduje zmutowane białko wiążące penicylinę (PBP 2a). Gen *mecA* zlokalizowany jest w obrębie kasety *SCCmec*, będącej ruchomym elementem genetycznym. Kaseta zawiera także sekwencję transpozonową Tn554 oraz sekwencje pUB110 i pT181, które zawierają geny warunkujące oporność na nie-metacyklinowe antybiotyki. Sekwencje wchodzące w skład kasety *SCCmec* są rozpowszechniają się w szczepach *S. aureus* na skutek horyzontalnego transferu genów.

Białko PBP 2a jest enzymem ściany komórkowej, który katalizuje produkcję peptydoglikanów. Jego zdolność do łączenia się z beta-laktamami i innymi antybiotykami pochodnymi penicyliny jest ograniczona w porównaniu do formy typowej białka. PBP. Dzięki temu PBP 2a może syntetyzować peptydoglikany w obecności wielu antybiotyków, w tym metacykliny, nafcyliny, oksacyliny i cefalosporyny.

Obecność genu zmutowanego *mecA* w izolatach klinicznych najczęściej potwierdza się przy pomocy reakcji PCR (Rys. 2.1.1b).



Rys. 2.1.1b. Obecność zmutowanego genu *mecA* w izolatach klinicznych *S. aureus* potwierdzona reakcją PCR przy zastosowaniu starterów specyficznych. M: marker masowy. K: kontrola negatywna (-) i pozytywna (+). 1-20: izolaty kliniczne, widoczny prążek o długości ok. 500 bp. odpowiada genowi *mecA*.

2.1.3. Identyfikacja mutacji w genie *mecA*



W pliku *mecA.txt* podano sekwencje genu *mecA* u pięciu szczepów *S. aureus* opornych na metacyklinę. Korzystając ze strony programu Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) proszę określić czy sekwencje genu *mecA* różnią się u poszczególnych szczepów. Proszę uzasadnić.

- A. Proszę wejść na stronę programu i wkleić sekwencje dostarczone w pliku.
- B. Proszę wybrać DNA, a następnie „submit” (Rys. 2.1.3a).

Rys. 2.1.3a. Zrzut ze strony Clustal Omega

C. Po chwili ukaże się uliniowanie sekwencji (Rys. 2.1.3b). Gwiazdka na dole oznacza, że nie występują zmiany. Brak gwiazdki oznacza, że w co najmniej jednej sekwencji zaszła mutacja w stosunku do sekwencji pozostałych szczepów.

Nazwa szczepu		Pozycja nukleotydu dla poszczególnych szczepów
171	----AATCCTAAGTAAAATTGCAGATAAGGGGTACAGAAAATCTAGACTTGATTACAAAA	56
214	--GCAATCCTAAGTAAAATTGCAGATAAGGGGTACAGAAAATCTAGACTTGATTACAAAA	58
346	--GCAATCCTAAGTAAAATTGCAGATAAGGGGTACAGAAAATCTAGACTTGATTACAAAA	58
366	-AGCAATCCTAAGTAAAATTGCAGATAAGGGGTACAGAAAATCTAGACTTGATTACAAAA	59
269	AAGCAATCCTAAGTAAAATTGCAGATAAGGGGTACAGAAAATCTAGACTTGATTACAAAA	60

Gwiazdki wskazują, że w danej pozycji jest ten sam nukleotyd.

Rys. 2.3.3b. Przykład uliniowania sekwencji w programie Clustal Omega.

- D. Proszę wymienić ewentualne różnice pomiędzy szczepami i podać liczbę zmienionych miejsc. Porównujemy do pierwszego szczepu w uliniowaniu.
- E. Proszę określić, czy każdy ze szczepów można jednoznacznie zidentyfikować. Wyniki proszę przedstawić w tabeli.
- F. Proszę podać szczep, u którego wystąpiło najwięcej zmian.

2.2. Geny Eukariota

➔ Genom człowieka jest około 760 razy większy niż genom *E. coli*. Jednakże genów u człowieka jest tylko 5 razy więcej (22 287 u człowieka i 4 300 u *E. coli*). Jednocześnie szacuje się, że liczba aktywnych transkryptów u człowieka wynosi około 150 000, czyli transkryptów jest 7 razy więcej niż genów. Oznacza to, że geny Eukariota, w tym człowieka są bardziej złożone i podlegają różnorodnej regulacji, która prowadzi do różnic w ekspresji w zależności od tkanki. Struktura modułarna genów Eukariota, na którą składają się introny i egzony ułatwia ekspresję specyficzną tkankowo.

- **Egzon** to fragment genu, który wchodzi w skład dojrzałego mRNA i na bazie którego powstaje białko. Termin „egzon” odnosi się do sekwencji DNA genu, a także do sekwencji odpowiedniego transkryptu RNA.
- **Intron** to sekwencja nukleotydowa w obrębie genu, która nie wchodzi w skład dojrzałego mRNA. Wyróżnia się introny tRNA, introny grupy I, grupy II oraz introny spliceomalne.
 - ▶ Introny tRNA: introny w genach kodujących tRNA, najczęściej występują w pętli antykodonowej, usuwane są przez endonukleazę tRNA.
 - ▶ Introny grupy I i grupy II występują w genach kodujących białka, tRNA i rRNA. Introny te tworzą złożone, trójwymiarowe struktury, które umożliwiają im samowycinanie. Innymi słowy intron w pre-mRNA może ulec rearanzacji i precyzyjnie się wyciąć.
 - ▶ Introny spliceomalne charakteryzują się obecnością specyficznych sekwencji na granicy intron-egzon. Są one wycinane przez spliceosom.

Introny są obecne w większości genów Eukariota, ale nie są uniwersalne. Przykładowo, geny kodujące histony nie zawierają intronów. Funkcja większości intronów nie jest znana, Niektóre kodują funkcjonalne cząsteczki RNA oraz białka.

2.2.1. Katalog genów człowieka: baza OMIM

OMIM (ang. Online Mendelian Inheritance in Man) jest publiczną bazą zawierającą informacje o genach człowieka, genotypach oraz chorobach genetycznych. Baza dostępna jest pod adresem: <https://www.omim.org>.

OMIM jest bazą zidentyfikowanych genów ludzkich. Baza zawiera informacje o położeniu genu na chromosomie, funkcji genu oraz charakterystykę chorób związanych z mutacjami w genach. Baza została utworzona w 1960 r., wersja online w 1985 r. Dane z 31 marca 2023 r. wskazują, że w bazie zgromadzono opisy 16 979 genów, w tym:



OMIM®

Online Mendelian Inheritance in Man®

An Online Catalog of Human Genes and Genetic Disorders

Updated December 30, 2021

Search OMIM for clinical features, phenotypes, genes, and more...



Advanced Search : OMIM, Clinical Synopses, Gene Map

Need help? : Example Searches, OMIM Search Help, OMIM Video Tutorials

Mirror site : <https://mirror.omim.org>

Rys. 2.1.1a. Zrzut ekranu dla strony startowej bazy OMIM.

- 16 127 genów zlokalizowanych na autosomach,
- 764 genów sprzężonych z X,
- 51 genów sprzężonych z Y,
- 37 genów zlokalizowanych w mtDNA

Ponadto dostępne są materiały edukacyjne obejmujące filmy YT i przedstawiające sposób przeszukiwania bazy, oraz tutorial na stronach www.

- Proste wyszukiwanie polega na wpisaniu np. określonej choroby np. muscular dystrophy.
- Znak + (np. +muscular + dystrophy) pozwala na zwrot informacji zawierających oba wpisane słowa.
- Zastosowanie cudzysłowu, np. „muscular dystrophy” zwraca rekordy, które zawierają wpisaną frazę.
- Można także wykorzystać standardowe Boolean operators czyli AND, OR i NOT.

Odczytując dane dotyczące fenotypów w OMIM należy pamiętać, że gen nie wywołuje choroby. Gen pełni istotne funkcje w organizmie. Choroba powstaje w wyniku mutacji w genie. Gen ACE (<https://www.omim.org/entry/106180?search=Ace&highlight=ace>) posiada własny promotor zlokalizowany w 12 intronie. Gen koduje enzym konwertujący angiotensynę (kinazę II). Jest to dipeptydylo karboksypeptydaza, która reguluje ciśnienie krwi oraz równowagę elektrolitową. Białko ACE jest integralnym białkiem błonowym. Uwalniane jest ono z powierzchni komórki pod wpływem metalosekretazy cynkowej. Istnieją także dane wskazujące, że allel D tego genu sprzyja wzrostowi mięśni i lepszemu przystosowaniu do długotrwałego wysiłku, co jest korzystne w niektórych dyscyplinach sportowych. Mutacje w genie ACE prowadzą do istotnych zaburzeń.

ANGIOTENSIN I-CONVERTING ENZYME, PLASMA LEVEL OF, INCLUDED

ANGIOTENSIN I-CONVERTING ENZYME, BENIGN SERUM INCREASE, INCLUDED
IgA NEPHROPATHY, PROGRESSION TO RENAL FAILURE IN, SUSCEPTIBILITY TO, INCLUDED

ANGIOTENSIN I-CONVERTING ENZYME, TESTICULAR, INCLUDED

Nazwa genu, opis

HGNC Approved Gene Symbol: **ACE**

Lokalizacja: Chromosom 17, ramię długie, prążek 23.2

Cytogenetic location: 17q23.3 Genomic coordinates (GRCh38): 17:63,477,060-63,498,372 (from NCBI)

Gene-Phenotype Relationships

Położenie w genomie, sekwencja całego genomu GRCh38, dalej koordynaty wskazujące na zakres sekwencji, które obejmują gen.

Location	Phenotype <small>Clinical Synopses</small>	Phenotype MIM number	Inheritance	Phenotype mapping key
17q23.3	Renal tubular dysgenesis	267430	AR	3
	[Angiotensin I-converting enzyme, benign serum increase]			3
	{Microvascular complications of diabetes 3}	612624		3
	{Myocardial infarction, susceptibility to}			3
	{SARS, progression of}			3
	{Stroke, hemorrhagic}	614519	3	

PheneGene Graphics



Fenotypy związane z mutacjami w genie.

Sposób dziedziczenia, tu autosomalny recesywny, dotyczy fenotypu 26430, dla pozostałych fenotypów nie określono.

TEXT

Rys. 2.1.1b. Zrzut ekranu dla wyszukiwań dla genu ACE.



2.2.2. Poszukiwanie informacji o wybranych genach w bazie OMIM

Korzystając z prostego wyszukiwania w bazie OMIM proszę znaleźć informacje na temat genu HIST1H1A.

- Co koduje gen HIST1H1A?
- Proszę podać oficjalny symbol genu.
- Proszę podać lokalizację cytogenetyczną genu. Proszę podać symbole i pełną nazwę.
- Proszę podać funkcję genu.
- Czy analizowany gen występuje u innych grup Eukariota? Proszę uzasadnić.

Samodzielne wykonanie 2.2.2. – 5 punktów

Termin: 30.11.2023. 23:59

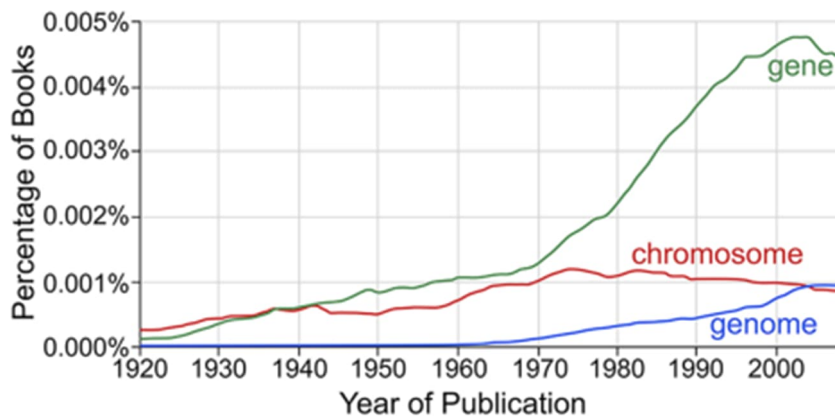
3. Wielkość genomu i wartość C



3.1. Definicja genomu

Genom 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 genom jądrowy, mitochondrialny, a u roślin dodatkowo genom chloroplastowy.

Genom jest często opisywany jako informacja genetyczna organizmu. Taki opis jest nieprecyzyjny i uproszczony, gdyż nie uwzględnia chociażby faktu istnienia 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. 3.1).



Rys. 3.1. Procent książek anglojęzycznych, w których pojawiło się 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.

- Co oznacza kompletny zestaw DNA? Czy u diploida (np. człowiek) 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 genomy danego typu (homologiczne), jeden od ojca, drugi od matki. W związku z tym, czy sekwencja genomu dotyczy jednego genomu homologicznego czy obu? U organizmów poliploidalnych, mających nawet kilka genomów 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.

3.2. 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.



3.2.1. Liczba nukleotydów, długość oraz masa wybranych chromosomów człowieka

W tabeli 1.3 podano wielkość wybranych chromosomów człowieka w różnych jednostkach. Przyjmując odległość między zasadami $3,4 \text{ \AA}$ oraz masę 1 pg dla 978 Mbp proszę uzupełnić brakujące wartości w tabeli. Długość i masę proszę zaokrąglić do części setnych.

Tabela 3.2.1. Liczba nukleotydów, długość i masa wybranych chromosomów człowieka.			
Chromosom	Liczba par zasad (bp)	Długość (cm)	Masa (pg)
3	198 295 560		
6		3,89	
16			0,05

➔ 3.2.2. Analiza wartości C u wybranych grup organizmów

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.



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>					
<i>Homo sapiens</i>					

4. Genom człowieka

➔ Jądrowy genom człowieka obejmuje 3,2 mld par zasad. Wielkość genomu człowieka oceniono metodami cytogenetycznymi. Dane te potwierdzono w Projekcie Poznania Genomu Ludzkiego. Około 55-60% genomu człowieka stanowią sekwencje transpozonowe. Ponadto genom człowieka zawiera liczne ślady rearanżacji, które zaszły w trakcie ewolucji.

HGP (Human Genome Project) – projekt poznania genomu ludzkiego rozpoczął się w 1990 r. Jego celem pierwotnym było utworzenie biblioteki liniowych fragmentów DNA specyficznych dla ludzkich chromosomów, poprawa efektywności sekwencjonowania oraz zwiększenie mocy obliczeniowej komputerów. Dalszym celem było poznanie pełnej sekwencji ludzkiego genomu aby zrozumieć funkcję genów oraz przyczyny chorób genetycznych

4.1. Rodziny genowe u człowieka

Genom człowieka zawiera liczne, szybko ewoluujące rodziny genowe, które podlegają reorganizacjom strukturalnym, w tym duplikacjom segmentalnym w zakresie od jednego tysiąca do kilku tysięcy par zasad. Około 400 bloków w genomie człowieka zawiera duplikacje, które zaszły podczas ewolucji hominidów. Dlatego badając rodziny genowe należy uwzględnić:

- obecność homologów u innych gatunków;
- rozkład genów należących do rodziny na chromosomach;
- stopień konserwacji sekwencji,
- strukturę domen w obrębie rodziny.

Wielkość rodzin genowych u człowieka, podobnie jak u innych organizmów jest różna. Niektóre rodziny składają się z 2-3 genów, inne obejmują kilkanaście, a nawet kilkaset genów. W zależności od wielkości możemy wyróżnić rodziny wielogenowe oraz superrodziny.

- **Rodziny wielogenowe:** grupa genów charakteryzujących się homologicznymi sekwencjami oraz podobnymi, czasami pokrywającymi się funkcjami.
- **Superrodziny genowe:** grupa genów mających wspólne pochodzenie, ale różniących się funkcją. Jeżeli geny zawierają większą liczbę domen to mogą należeć do różnych superrodzin. Cechą genów należących do super rodzin jest różny wzór ekspresji. Superrodziny mogą obejmować kilka rodzin oraz pojedyncze geny.

Wiele rodzin genowych tworzy klaster. Najlepiej poznanym przykładem jest rodzina genów globinowych. Rodziny takie najczęściej powstają przez niesymetryczny crossing-over. Może on być promowany insercją transpozonu, czego przykładem jest rodzina immunoglobulin, której segmenty pochodzą od transpozonu transib.

Grupa rodzin genowych rozproszonych: powstaje najczęściej poprzez retroduplikację czyli odwrotną transkrypcję na bazie RNA i następnie integrację do genomu. Przykładem retrogenów u człowieka są kinaza fosfoglicerylowa, która powstała przez odwrotną transkrypcję genu na chromosomie X, syntetaza arginosukcenylowa, która obecnie tworzy pseudogen, α -tubulina, cytochrom c, dehydrogenaza 3-fosfogliceraldehydowa, białko rybosomowe L32.

Tabela 4.1. Przykłady dużych superrodzin u człowieka.

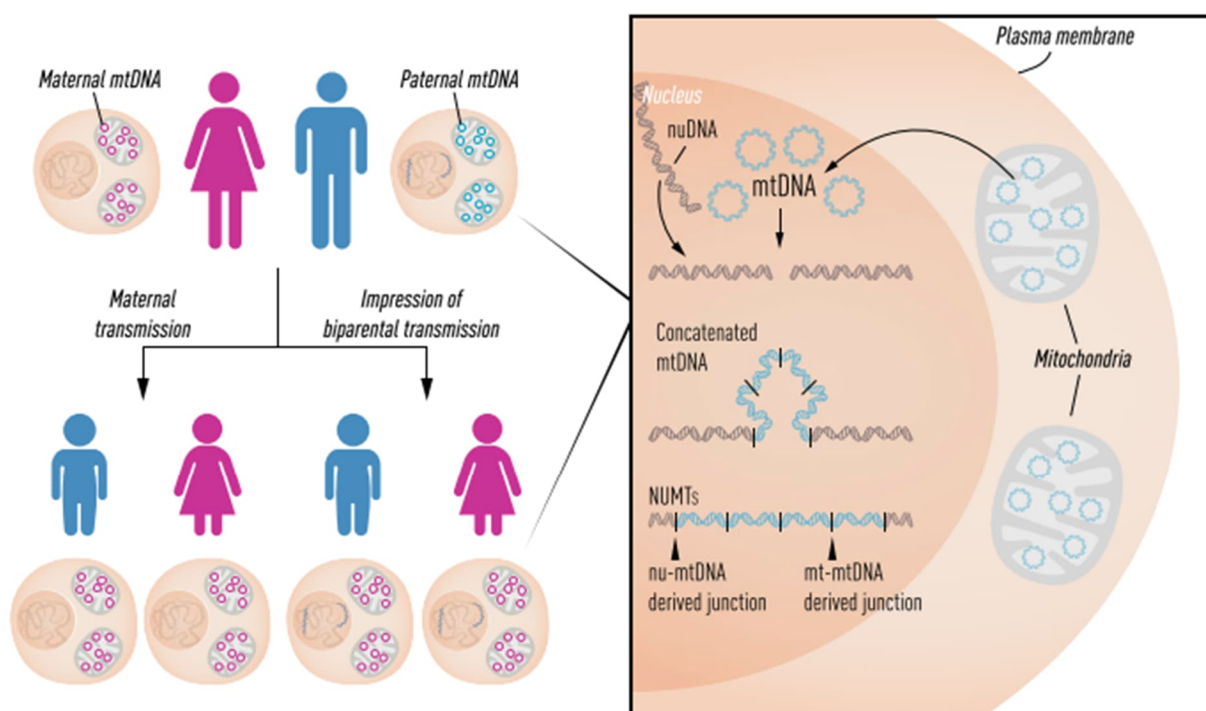
Domena	Liczba białek
Palce cynkowe typu C2H2564	564
Immunoglobuliny	381
Motywy rozpoznające RNA	224
Rodzina KRAB-box	204
Domena PH	193
Domena homeobox	160
Domena WD40	136
Rodzina Ras	126

4.2. Dziedziczenie mitochondrialne

Genom mitochondrialny człowieka składa się z 16,6 kbp. Zawiera on 37 genów rozproszonych na dwóch niciach: H (łańcuch ciężki), L (łańcuch lekki). Geny mitochondrialne nie zawierają intronów. Kodują one głównie enzymy łańcucha oddechowego, geny dla mitochondrialnego rRNA oraz tRNA. W trakcie ewolucji geny mitochondrialne są przenoszone do jądra (sekwencje NUMT: sekwencje jądrowe pochodzenia mitochondrialnego). U człowieka występuje co najmniej 211 takich sekwencji. Ostatnie insercje NUMT miały miejsce 6-4 mln lat temu.

Mitochondria u różnych gatunków dziedziczą się na ogół od jednego rodzica. U człowieka dziedziczą się one w linii maceknej, podobnie jak u większości ssaków. Dziedziczenie macekne powoduje, że cechy zlokalizowane w mitochondrium matki przekazywane są całemu potomstwu. Ta właściwość umożliwia wykorzystanie mtDNA w analizach filogenetycznych.

Sporadyczne przypadki dziedziczenia ojcowskiego obserwowano u kóz, myszy oraz bydła domowego. Również u człowieka zdarzają się przypadki dziedziczenia mitochondriów po obojgu rodziców. Jedną z przyczyn obserwacji takiego zjawiska może być obecność mtDNA w genomie jądrowym. Kopie jądrowe ojcowskiego DNA mogą sprawiać wrażenie ojcowskiego dziedziczenia mitochondriów (Rys. 4.2). Z drugiej strony istnieją przykłady heteroplazmii, która wskazuje na dziedziczenie od obojga rodziców. Zjawisko to nie jest wyjaśnione w literaturze.



Rys. 4.2. „Falszywe” ojcowskie dziedziczenie mitochondriów jako efekt obecności kopii mtDNA w jądrze (NUMT).