

# Bioinformatyka — ćwiczenie 2 - Pozyskiwanie i wstępna obróbka informacji uzyskanych z bioinformatycznych baz danych

Czas na oddanie finalnego raportu: 1 tydzień

[Zasady obowiązujące na przedmiocie](#), [plik z ocenami](#)

Nazwa pliku z raportem, który trzeba wysłać do nauczyciela: login\_imie\_nazwisko\_pbio\_2.pdf

Cel: Celem ćwiczenia jest zapoznanie studentów z bazami danych wykorzystywanymi w bioinformatyce oraz poznanie sposobów uzyskiwania dostępu do wybranych danych znajdujących się w tych bazach.

## Teoria:

**Jednostkami**, za pomocą których zapisywana jest długość nici DNA lub RNA (generalnie: kwasów nukleinowych) są: kilo par zasad – kpz (ang. kilo base pairs – kbp, w skrócie: kb), mega par zasad – mpz (ang. mega base pairs – Mbp, w skrócie: Mb) i giga par zasad – gpz (ang. giga base pairs – Gbp, w skrócie: Gb).

- 1 kb to 1000 par zasad lub 1000 nukleotydów (1000 nt).
- 1 Mb to 1 000 000 par zasad lub 1 000 000 nukleotydów (1 000 000 nt).
- 1 Gb to 1 000 000 000 par zasad lub 1 000 000 000 nukleotydów (1 000 000 000 nt).

**Scaffold** to pojęcie związane z sekwencjonowaniem DNA, które odnosi się do połączenia kilku krótkich sekwencji DNA w jedną dłuższą sekwencję. Scaffolds są tworzone w procesie asemblacji (łączenia) sekwencji krótkich fragmentów DNA, które zostały uzyskane w laboratoryjnym procesie sekwencjonowania.

**Asemblacja** jest procesem polegającym na scaleniu kilku krótkich sekwencji DNA w jedną dłuższą sekwencję, która będzie odpowiadać sekwencji genomu. Scaffolds są zazwyczaj używane jako punkt wyjścia do dalszej analizy genomu.

**Afiliacja** w publikacji naukowej jest to informacja o instytucji, w której autor publikacji jest zatrudniony lub związany zawodowo. Afiliacja jest zazwyczaj umieszczana na początku artykułu naukowego i wskazuje, w jakiej instytucji autor był zaangażowany w prowadzenie badań i/lub pisanie publikacji. Instytucja ta może być uniwersytetem, instytutem badawczym, firmą lub jakimkolwiek innym miejscem, w którym autor prowadzi swoją działalność. Afiliacja często wskazuje również kraj, w którym instytucja jest zlokalizowana.

**Enzym trawienny** to białko, które jest niezbędne do rozkładania złożonych związków organicznych, takich jak białka, węglowodany i tłuszcze, w prostsze związki, które nasz organizm może wykorzystać jako źródło energii i składników odżywczych.

Niektóre przykłady ludzkich enzymów trawiennych to:

- Pepsyna — enzym występujący w soku żołądkowym, który rozkłada białka.
- Amylaza — enzym, który rozkłada skrobię na prostsze związki, takie jak glukoza.
- Lipaza — enzym, który rozkłada tłuszcze na kwasy tłuszczowe i glicerol.
- Laktaza — enzym, który rozkłada laktozę (cukier mleczny) na glukozę i galaktozę.

**Numer EC** (Enzyme Commission) jest identyfikatorem białek. Każde białko jest przypisane do określonej klasy, podklasy i typu reakcji według kodu EC, który składa się z kilku liczb oddzielonych kropkami. Na przykład, EC 3.2.1.20 oznacza trzecią klasę enzymów (hydrolazy), podklasę 2 (hydrolazy zasadowe), typ 1 (hydrolazy zasadowe złożone z dwóch podjednostek) i 20-ty enzym w tej podklasie.

**Tabela 2. Podział enzymów na klasy**

Nazwa klasy	Inna nazwa klasy	Mechanizm reakcji przez nie katalizowany
EC 1	oksydoreduktazy	katalizują reakcje utleniania i redukcji
EC 2	transferazy	przenoszą grupy funkcyjne
EC 3	hydrolazy	katalizują hydrolizę różnych wiązań
EC 4	liazy	rozcinają różne wiązania na drodze innej niż hydroliza czy utlenianie
EC 5	izomerazy	katalizują zmiany izomeracyjne cząsteczek
EC 6	ligazy	łączą cząsteczki wiązaniami kowalencyjnymi
EC 7	translokazy	katalizują ruch jonów i cząsteczek przez błony lub ich rozdział wewnątrz błon

Źródło: opracowanie własne na podstawie danych z Wikipedii

Źródło: <https://www.forummleczarskie.pl/raporty/1037,nowosci-w-higienizacji-mycie-enzymatyczne>

W biologii molekularnej termin **“ligand”** jest używany do opisu cząsteczki, która wiąże się z białkiem za pomocą interakcji chemicznych. Te interakcje powodują, że ligand i białko tworzą kompleks, który może zmodyfikować funkcję białka. Wiele białek w organizmach żywych pełni funkcje receptorów, które są zaprojektowane do wchodzenia w interakcje z ligandami, takimi jak hormony, czynniki wzrostu i neurotransmitery.

## Zadania:

Każdego roku w styczniu czasopismo naukowe Nucleic Acids Research (NAR) publikuje artykuł zawierający aktualną i sprawdzoną listę biologicznych baz danych. Znajdź ostatni artykuł i umieść w raporcie zrzut ekranu z pierwszej strony tej publikacji. Podpowiedź: wpisz w google [nucleic acids research review article databases 2023](#) i jednym z pierwszych wyników powinien być odnośnik do tego artykułu [zadanie 1].

OXFORD  
ACADEMIC

Nucleic Acids Research

Article Navigation

JOURNAL ARTICLE

**The 2023 Nucleic Acids Research Database Issue and the online molecular biology database collection**

Daniel J Rigden, Xosé M Fernández

Nucleic Acids Research, Volume 51, Issue D1, 6 January 2023, Pages D1–D8,  
<https://doi.org/10.1093/nar/gkac1186>  
 Published: 06 January 2023

PDF Views Cite Permissions Share

**Abstract**

The 2023 Nucleic Acids Research Database Issue contains 178 papers ranging across biology and related fields. There are 90 papers reporting on new databases and 82 updates from resources previously published in the Issue. Six more papers are updates from databases most recently published elsewhere. Major nucleic acid databases reporting updates include Genbank, ENA, ChIPBase, JASPAR, mirDIP, the Issue's first Breakthrough Article, NACDDB for Circular Dichroism data. Updates from BMRB, RCSB cover experimental protein structural data while AlphaFold 2 computational structure predictions feature widely. STRING and REBASE are stand-out updates in the signalling and enzymes section. Immunology-related databases include CEDAR, the second Breakthrough Article, for cancer epitopes and receptors alongside returning IPD-IMGT/HLA and the new PGG.MHC. Genomics-related resources include Ensembl, GWAS Central and UCSC Genome Browser. Major returning databases for



W treści tego artykułu znajdź odnośnik do wersji on-line tej bazy i podaj ten link (w artykule z 2023 roku jest to ostatni link podany w sekcji o nazwie “Abstract” [zadanie 2]).

<https://www.oxfordjournals.org/nar/database/c/>

W powyższej bazie danych znajdź rekord dotyczący bazy GenBank (np. wybierz “Alphabetical List” i tam znajdź tę bazę) i odnajdź w opisie informację dotyczącą tego, jak często jest ona aktualizowana [zadanie 3].

Aktualizowana jest codziennie.

## GenBank

- ▶ Category/Paper List
- ▶ Search Summary Papers

NAR Molecular Biology Database Collection entry number 7

[www.ncbi.nlm.nih.gov/genbank/](http://www.ncbi.nlm.nih.gov/genbank/)

Sayers, Eric; Cavanaugh, Mark; Clark, Karen; Pruitt, Kim; Sherry, Stephen; Yankie, Linda; Karsch-Mizrachi, Ilene

NCBI

Contact [sayers@ncbi.nlm.nih.gov](mailto:sayers@ncbi.nlm.nih.gov)

## Database Description

GenBank is a comprehensive database that contains publicly available nucleotide sequences for 400 000 formally described species. These sequences are obtained primarily through submissions from individual laboratories and batch submissions from large-scale sequencing projects, including whole genome shotgun and environmental sampling projects. Most submissions are made using BankIt, the National Center for Biotechnology Information (NCBI) Submission Portal, or the tool tbl2asn. GenBank staff assign accession numbers upon data receipt. Daily data exchange with the European Nucleotide Archive and the DNA Data Bank of Japan ensures worldwide coverage. GenBank is accessible through the NCBI Nucleotide database, which links to related information such as taxonomy, genomes, protein sequences and structures, and biomedical journal literature in PubMed. BLAST provides sequence similarity searches of GenBank and other sequence databases. Complete bimonthly releases and **daily updates of the GenBank database are available by FTP**. Recent updates include changes to sequence identifiers, submission wizards for 16S and Influenza sequences, and an Identical Protein Groups resource.

## Recent Developments

Zapoznaj się z inną publikacją <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8022358>, a następnie:

- W jednej z tabel (Tabela S2 - “Cancer databases”) znajdź bazę danych związaną z czerniakiem (sprawdź z użyciem [translate.google.com](https://translate.google.com) jak ta jednostka choroba

określana jest po angielsku jeśli nie wiesz). Podaj nazwę tej konkretnej bazy danych [zadanie 4].

Baza danych: NLM database

- W tabeli S9 “Other databases” znajdź bazę HLA-ADR i wejdź w link “HLA Adverse Drug Reaction Database”, następnie z listy “Drug” wybierz “abacavir” (lek stosowany w terapii antyretrowirusowej czyli m.in. anty HIV) i kliknij “Search”. Z tabeli wynikowej odczytaj z badań jakiej populacji pochodzą te wyniki [zadanie 5]. Podpowiedz poniżej:

Odpowiedz: Caucasian (predominantly)

Line	PubMed Link	Drug	Allele	Old Allele Name	Cohort ethnicity	p-value Pat <sup>1</sup> and ExpCtrl <sup>2</sup>	p-value Pat <sup>1</sup> and Pop <sup>3</sup>	Pa ca
1	18505179	abacavir	B*57:01			0.0005		11
2	22197535	abacavir	B*57:01			0.0001		18
3	15247625	abacavir	B*57:01			0.006		6 /
4	11888582	abacavir	B*57:01			0.0001		14

tutaj podana jest ta nazwa

Wejdź do bazy [NCBI GenBank](#):

- Znajdź wszystkie rekordy zdeponowane pod nazwiskiem “Kuzdraliński” i umieść tutaj listę

gatunków mikroorganizmów, których one dotyczą (widoczne są w ramce po prawej stronie)

[zadanie 6].

Podpowiedz poniżej:

NIH

National Library of Medicine  
National Center for Biotechnology Information

1

Nucleotide

Nucleotide

Kuzdraliński

Search

Baza sekwencji nukleotydowych

W to pole należy wpisać poszukiwane nazwisko

Kliknij “Search”

2

Nucleotide

Kuzdraliński

Search

Summary

20 per page

Sort by Date Released

Send to

Filters: Manage Filters

Items: 1 to 20 of 507

<< First

< Prev

Page 1 of 26

Next >

Last >>

☐ [Fusarium tricinatum isolate 499.2 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene, partial sequence](#)  
 1. [RNA gene and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence](#)  
 520 bp linear DNA  
 Accession: MK342537.1 GI: 1545265855

Results by taxon

Top Organisms

[\[Tree\]](#)  
[Fusarium poae \(155\)](#)  
[Rhizopus arrizus \(96\)](#)  
[Rhizopus stolonifer \(54\)](#)  
[Blumeria graminis f. sp. tritici \(33\)](#)  
[Fusarium sp. \(30\)](#)  
[All other taxa \(139\)](#)  
[More...](#)

Tutaj znajduje się lista gatunków ze zdeponowanych sekwencji. Żeby zobaczyć całą kliknij “More...”

3

Nucleotide Kuzdraliński Search

Create alert Advanced

Summary 20 per page Sort by Date Released Send to: Filters: Manage Filters

Items: 1 to 20 of 507

<< First < Prev Page 1 of 26 Next > Last >>

☐ [Fusarium tricinatum isolate 499.2 internal transcribed spacer 1, partial sequence: 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence](#)  
520 bp linear DNA  
Accession: MK342537.1 GI: 1545265855  
[Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Fusarium avenaceum isolate 522.4 internal transcribed spacer 1, partial sequence: 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence](#)  
520 bp linear DNA  
Accession: MK342538.1 GI: 1545265856  
[Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Fusarium avenaceum isolate 522.5 internal transcribed spacer 1, partial sequence: 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence](#)  
520 bp linear DNA  
Accession: MK342539.1 GI: 1545265857  
[Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

Results by taxon

Top Organisms [Tree]

- Fusarium poae (155)
- Rhizopus arrhizus (96)
- Rhizopus stolonifer (54)
- Blumeria graminis f. sp. tritici (33)
- Fusarium sp. (30)
- Puccinia triticina (26)
- Puccinia striiformis f. sp. tritici (22)
- Epicoccum nigrum (18)
- Fusarium sporotrichioides (16)
- Fusarium avenaceum (14)
- Lactobacillus helveticus (9)
- Fusarium culmorum (6)
- Pichia kudriavzevii (4)
- Fusarium graminearum (3)
- Aureobasidium pullulans (3)
- Rhizopus microsporus (3)
- Fusarium equiseti (2)
- Fusarium oxysporum (2)
- Alternaria sp. (2)
- Rhizoctonia solani (2)
- All other taxa (7)

Less...

Tą listę gatunków umieść w raporcie

Następnie wyświetl tylko rekordy dotyczące drożdży z gatunku *Pichia kudriavzevii* i eksportuj je do jednego pliku w formacie FASTA posortowanego według długości sekwencji i w odpowiedzi do zadania umieść całą zawartość tego pliku [zadanie 7].  
Podpowieź poniżej:

Filters: Manage Filters

Results by taxon

Top Organisms [Tree]

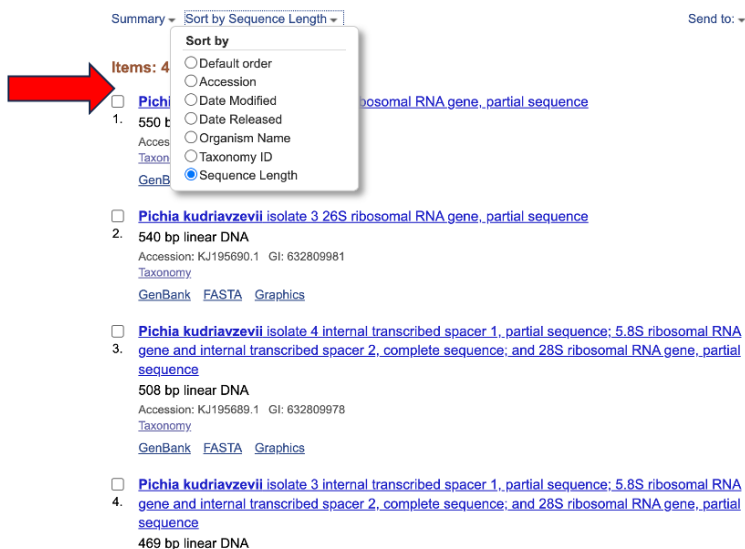
- Fusarium poae (155)
- Rhizopus arrhizus (96)
- Rhizopus stolonifer (54)
- Blumeria graminis f. sp. tritici (33)
- Fusarium sp. (30)
- Puccinia triticina (26)
- Puccinia striiformis f. sp. tritici (22)
- Epicoccum nigrum (18)
- Fusarium sporotrichioides (16)
- Fusarium avenaceum (14)
- Lactobacillus helveticus (9)
- Fusarium culmorum (6)
- Pichia kudriavzevii (4)
- Fusarium graminearum (3)
- Aureobasidium pullulans (3)
- Rhizopus microsporus (3)
- Fusarium equiseti (2)
- Fusarium oxysporum (2)
- Alternaria sp. (2)
- Rhizoctonia solani (2)
- All other taxa (7)

Less...

Należy kliknąć w tę nazwę gatunku



Sposób  
posortowania  
wyników



Summary ▾ Sort by Sequence Length ▾

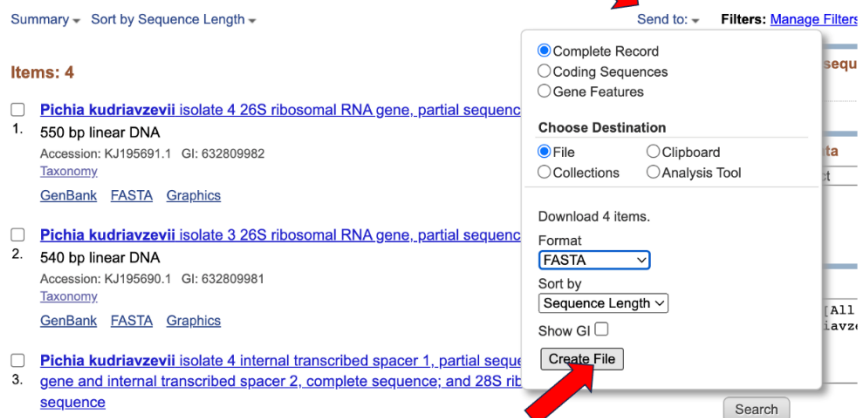
Sort by

- ☐ Default order
- ☐ Accession
- ☐ Date Modified
- ☐ Date Released
- ☐ Organism Name
- ☐ Taxonomy ID
- ☒ Sequence Length

Items: 4

- ☐ [Pichia kudriavzevii isolate 4 26S ribosomal RNA gene, partial sequence](#)  
550 bp linear DNA  
Accession: KJ195691.1 GI: 632809982  
[Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)
- ☐ [Pichia kudriavzevii isolate 3 26S ribosomal RNA gene, partial sequence](#)  
540 bp linear DNA  
Accession: KJ195690.1 GI: 632809981  
[Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)
- ☐ [Pichia kudriavzevii isolate 4 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence](#)  
508 bp linear DNA  
Accession: KJ195689.1 GI: 632809978  
[Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)
- ☐ [Pichia kudriavzevii isolate 3 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence](#)  
469 bp linear DNA

Żeby utworzyć plik w  
formacie FASTA  
klikasz "Send to:"



Summary ▾ Sort by Sequence Length ▾

Items: 4

- ☐ [Pichia kudriavzevii isolate 4 26S ribosomal RNA gene, partial sequence](#)  
550 bp linear DNA  
Accession: KJ195691.1 GI: 632809982  
[Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)
- ☐ [Pichia kudriavzevii isolate 3 26S ribosomal RNA gene, partial sequence](#)  
540 bp linear DNA  
Accession: KJ195690.1 GI: 632809981  
[Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)
- ☐ [Pichia kudriavzevii isolate 4 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence](#)  
508 bp linear DNA

Send to: ▾ Filters: [Manage Filters](#)

☒ Complete Record  
☐ Coding Sequences  
☐ Gene Features

Choose Destination

☒ File ☐ Clipboard  
☐ Collections ☐ Analysis Tool

Download 4 items.

Format: **FASTA**

Sort by: **Sequence Length**

Show GI ☐

[Create File](#)

[Search](#)

Po zastosowaniu  
powyższych ustawień  
tworzysz plik



M11: Alignment Explorer (sequence.fasta)

Data Edit Search Alignment Web Sequencer Display Help

DNA Sequences Translated Protein Sequences


Species/Abbrv	Sequence
1. KJ195691.1 Pichia kudriavzevii isolate 4 26S rRNA	CGGCGAGTGAAGGCGAGAGCTCAAGATTGAAATCGTGCTTTGCGGCACGAGTTGTAGATTGCAAGGTTGGAGTCTGTTGTGGAGGCTGTC
2. KJ195690.1 Pichia kudriavzevii isolate 3 26S rRNA	CGGCGAGAGCTCAAGATTGAAATCGTGCTTTGCGGCACGAGTTGTAGATTGCAAGGTTGGAGTCTGTTGTGGAGGCTGTC
3. KJ195689.1 Pichia kudriavzevii isolate 4 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence	TACACCTGGCTGAGCGGACGAAAAACAACACCTTAAATGTTGGATATAGCATATAGTTCACAAAGAAATCTTACGGACAAAG
4. KJ195688.1 Pichia kudriavzevii isolate 3 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence	AACAAACACCTTAAATGTTGGATATAGCATATAGTTCACAAAGAAATCTTACGGACAAAG

W jednej z baz danych NCBI zdeponowany jest kompletny ludzki genom (całe ludzkie DNA)  
tj.: [ncbi.nlm.nih.gov](http://ncbi.nlm.nih.gov).

Podaj, na ile obecnie ocenia się liczbę genów kodujących białka w ludzkim genomie (tabela  
"Feature counts", na przecięciu "GRCh38.p14" i "protein-coding")



#### Feature counts

Feature	GRCh38.p14	GRCh38.p14 Primary Assembly	GRCh38.p14 All Alt Loci	GRCh38.p14 PATCHES	T2T-CHM13v2.0
Genes and pseudogenes 	59,265	58,654	2,487	2,863	57,771
protein-coding	20,024	19,853	845	1,069	20,011
non-coding	21,894	21,575	718	861	20,716
Transcribed pseudogenes	1,222	1,219	101	148	1,195
Non-transcribed pseudogenes	15,654	15,547	655	750	15,394

[zadanie 8]. Posługując się wyszukiwarką Google, znajdź dane z około 1999-2001 roku na temat naszych wyobrażeń dotyczących liczby genów kodujących białka w ludzkim genomie

Odpowiedz: „w ludzkim genomie jest od 60-ciu do 80-ciu tysięcy genów. „

[zadanie 9]. Dzisiaj już wiemy, że jeszcze niedawne wyobrażenia były błędne.

## Genome assembly GRCh38.p14

reference

Download

 datasets

curl

Actions

NCBI RefSeq assembly GCF\_000001405.40



Submitted GenBank  
assembly GCA\_000001405.29



Taxon *Homo sapiens* (human)

Synonym hg38

Assembly type haploid with alt loci

Submitter Genome Reference Consortium

Date Feb 3, 2022

View the [legacy Assembly page](#)

W bazie danych [NCBI](#) posługując się wyszukiwarką znajdź rekord dla chromosomu 22 z GRCh38.p14

(wpisz w wyszukiwarce <https://www.ncbi.nlm.nih.gov/>: “Homo sapiens chromosome 22, GRCh38.p14 Primary Assembly” uprzednio zaznaczając żeby przeszukana została baza

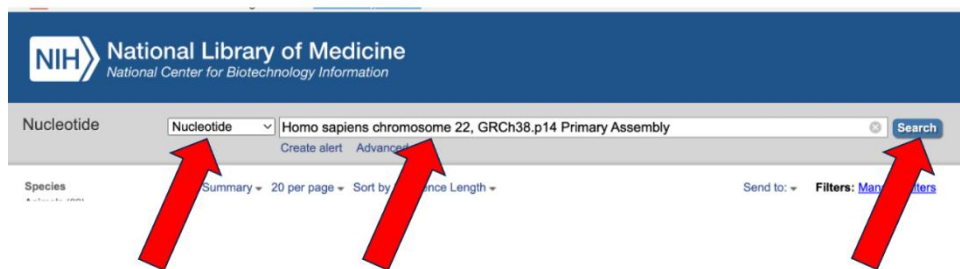
# Assembly statistics

	RefSeq	GenBank
Genome size	3.1 Gb	3.1 Gb
Total ungapped length	2.9 Gb	2.9 Gb
Gaps between scaffolds	349	349
Number of chromosomes	24	24
Number of organelles	1	1
Number of scaffolds	470	473
Scaffold N50	67.8 Mb	67.8 Mb
Scaffold L50	16	16
Number of contigs	996	999
Contig N50	57.9 Mb	57.9 Mb
Contig L50	18	18
GC percent	40.5	40.5
Assembly level	Chromosome	Chromosome

“Nucleotide”) i podaj jakiej wielkości jest ten chromosom (podpowiedź: ponad 50 mln par zasad) [zadanie 10].

Podpowiedź poniżej:





Items: 1 to 20 of 62

<< First < Prev Page 1 of 3

☐ [Homo sapiens chromosome 22, GRCh38.p14 Primary Assembly](#)

1. XXXXXXXXXX bp linear DNA  
Accession: NC\_000022.11 GI: 568815576  
[Assembly](#) [BioProject](#) [Protein](#) [PubMed](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

tutaj podana jest  
wielkość  
chromosomu 22  
u człowieka

Następnie wejdź do tego rekordu klikając jego nazwę. Przejdź do bazy danych BioProject klikając w link znajdujący się obok "BioProject:". Tam w sekcji Project Data, w tabeli Assembly Details kliknij w kolumnie Chrs pozycję "25". Wygenerowaną listę uszereguj według wielkości (Sort by ->Sequence Length). Jakiej mają oznaczenia (numery) trzy największe ludzkie chromosomy [zadanie 11]?

Items: 1 to 20 of 25

<< First < Prev Page 1 of 2 Next > Last >>

☐ [Homo sapiens chromosome 1, GRCh38.p14 Primary Assembly](#)

1. 248,956,422 bp linear DNA  
Accession: NC\_000001.11 GI: 568815597  
[Assembly](#) [BioProject](#) [Protein](#) [PubMed](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Homo sapiens chromosome 2, GRCh38.p14 Primary Assembly](#)

2. 242,193,529 bp linear DNA  
Accession: NC\_000002.12 GI: 568815596  
[Assembly](#) [BioProject](#) [Protein](#) [PubMed](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Homo sapiens chromosome 3, GRCh38.p14 Primary Assembly](#)

3. 198,295,559 bp linear DNA  
Accession: NC\_000003.12 GI: 568815595  
[Assembly](#) [BioProject](#) [Protein](#) [PubMed](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

Podpowiedź poniżej:

tu znajduje się oznaczenie  
chromosomu w postaci  
jego numeru



- ☐ [Homo sapiens chromosome 1, GRCh38.p14 Primary Assembly](#)
1. 248,956,422 bp linear DNA  
Accession: NC\_000001.11 GI: 568815597  
[Assembly](#) [BioProject](#) [Protein](#) [PubMed](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

Trzy pierwsze elementy tej listy (a więc trzy największe, ludzkie chromosomy) przenieś do pliku "Summary", a jego treść (w całości) umieść jako odpowiedź do zadania [zadanie 12].

1. Homo sapiens chromosome 1, GRCh38.p14 Primary Assembly  
248,956,422 bp linear DNA  
NC\_000001.11 GI:568815597

2. Homo sapiens chromosome 2, GRCh38.p14 Primary Assembly  
242,193,529 bp linear DNA  
NC\_000002.12 GI:568815596

3. Homo sapiens chromosome 3, GRCh38.p14 Primary Assembly  
198,295,559 bp linear DNA  
NC\_000003.12 GI:568815595

Podpowieź poniżej:

2. Kliknij "Send to"

1. Zaznacz rekordy, które mają znaleźć się w pliku

W bazie danych [PubMed](#) wyszukaj wszystkie publikacje afiliowane obecnie na Polskę. Podaj ich liczbę jako odpowiedź na to zadanie [zadanie 13].

249.833

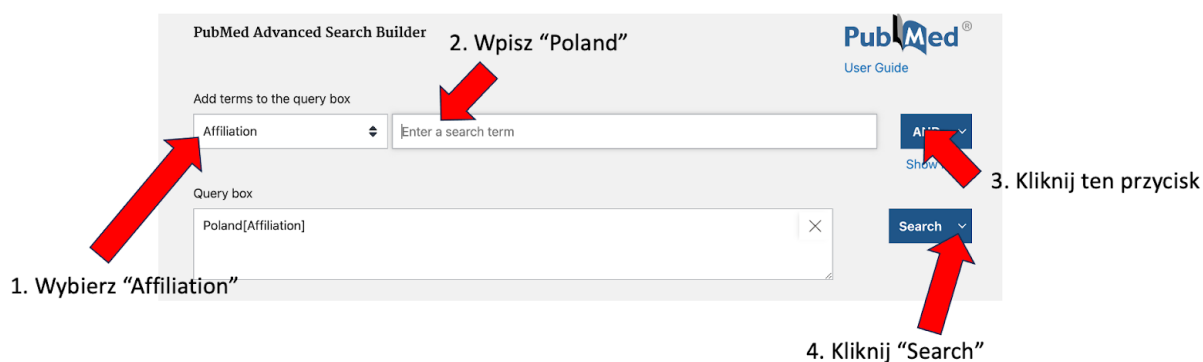
249,833 results

☐ Pattern of human  
1 Ożańska A, Szymczak I  
Cite Scand J Immunol. 2020  
PMID: 32243617 Fi  
Share

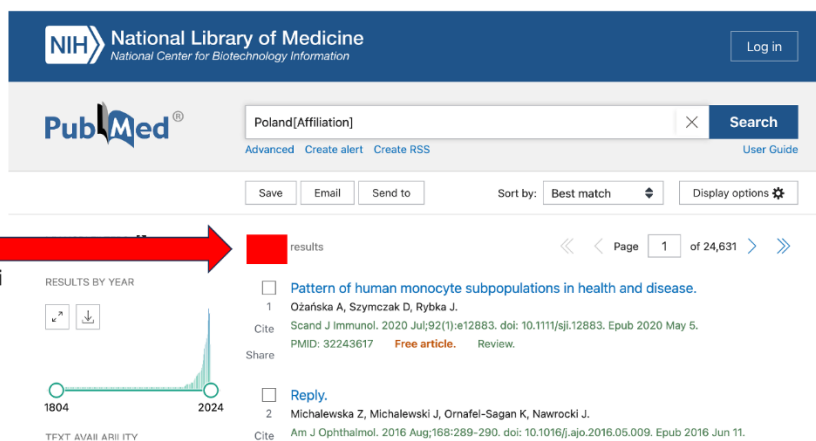
Podpowieź poniżej:



Kliknij "Advanced"



Tutaj jest  
liczba  
publikacji



W polu “results by year” możesz eksportować plik csv z liczbą publikacji z każdego roku (przycisk ze strzałką w dół). Skopiuj jako odpowiedź do zadania dane od roku 2015 do roku 2022 z rozdzielczością co do roku **[zadanie 14]**.

Search query: Poland[Affiliation]	
Year,Count	
2022,26203	
2021,27941	
2020,23662	
2019,18677	
2018,17296	
2017,15962	
2016,15123	
2015,13743	

Przykładowo dane dla lat 2003-2006 prezentują się następująco:

<b>2006</b>	4948
<b>2005</b>	4548
<b>2004</b>	4101
<b>2003</b>	3791

[Zapoznaj się z informacjami](#) dotyczącymi markera genetycznego typu SNP o oznaczeniu rs53576. Znajdź rekord dotyczący tego markera genetycznego w tej bazie: [NCBI SNP](#). Jaki jest MAF dla tego markera w populacji ludzkiej **[zadanie 15]**?

A=0.326459/24474 ([ALFA](#))

□ rs53576 [Homo sapiens]

1.

Variant type: SNV  
Alleles: A>G,T [Show Flanks]  
Chromosome: 3:8762685 (GRCh38)  
3:8804371 (GRCh37)  
Canonical SPDI: NC\_000003.12:8762684:A:G,NC\_000003.12:  
Gene: OXTR (Varview)  
Functional Consequence: intron\_variant  
Validated: by frequency,by alfa,by cluster  
MAF: A=0.326459/24474 (ALFA)  
A=0.28169/120 (SGDP\_PRJ)  
A=0.306872/81226 (TOPMED)  
A=0.309437/43359 (GnomAD)  
G=0.310185/67 (Vietnamese)  
G=0.313452/247 (PRJEB37584)  
A=0.322276/1195 (TWINSUK)  
A=0.323819/1248 (ALSPAC)  
A=0.324074/70 (Qatari)  
G=0.337201/988 (KOREAN)  
A=0.34/204 (NorthernSweden)  
A=0.342685/342 (GoNL)  
A=0.35/14 (GENOME\_DK)  
A=0.361111/13 (Siberian)  
A=0.364063/1631 (Estonian)  
G=0.373416/6258 (TOMMO)  
A=0.378357/1895 (1000Genomes)  
A=0.5/14 (PRJEB36033)  
...less

Podpowiedź 1: wartość MAF pobierz stąd: "MAF: A=N.NNNNN/24474 (ALFA) z dokładnością do pierwszego miejsca po przecinku.

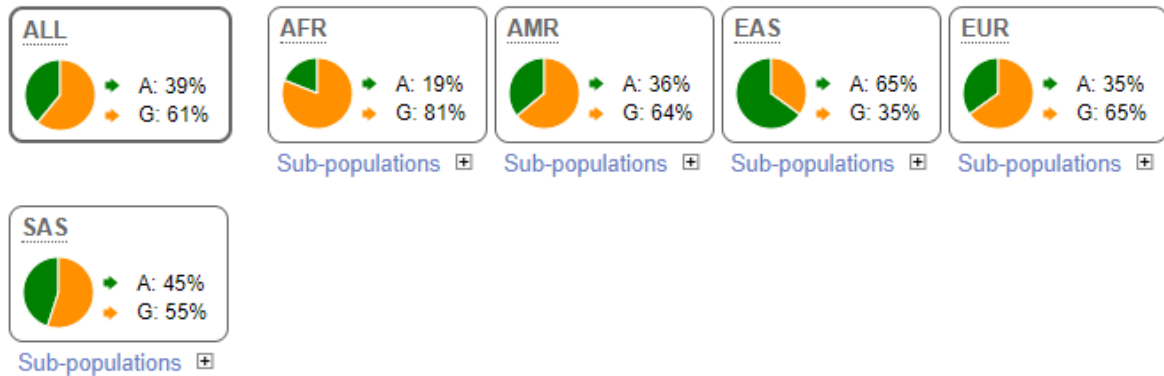
Podpowiedź 2: MAF to z ang. minor allele frequency czyli częstość rzadszego allele. Marker genetyczny rs53576 może mieć dwa allele: A lub G, w tym przypadku A występuje rzadziej i to jego dotyczy ten parametr.

Wejdź do rekordu tego markera w bazie danych [ENSEMBL](#) i odpowiedz na pytanie: w jakiej populacji (odnośnik: "Population genetics") allele związany z fenotypem predysponującym do bycia bardziej empatycznym występuje w największej frekwencji tj. 81% [zadanie 16]?

Odpowiedz: Afryka

## Population genetics ?

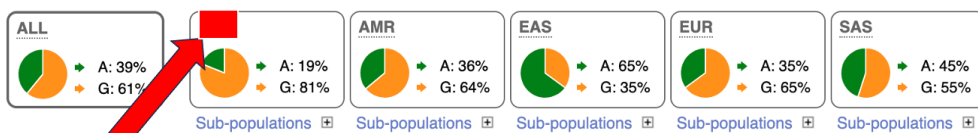
### 1000 Genomes Project Phase 3 allele frequencies



Podpowiedź poniżej:

Population genetics ?

### 1000 Genomes Project Phase 3 allele frequencies



Tutaj znajduje się odpowiedź na to pytanie

Znajdź w bazie danych [UNIPROT](#) rekord o numerze P0DTE8 reprezentujący ludzki enzym trawienny. W odpowiedzi do zadania wpisz nazwę enzymu [zadanie 17].

Nazwa: Alpha-amylase 1C

## P0DTE8 · AMY1C\_HUMAN

The screenshot shows the UniProt entry for P0DTE8 (AMY1C\_HUMAN). The 'Gene' field is highlighted with a red arrow. The entry includes the following information:

- Protein: **AMY1C**
- Gene: **AMY1C**
- Status: UniProtKB reviewed (Swiss-Prot)
- Organism: **Homo sapiens (Human)**
- Amino acids: 511 (go to sequence)
- Protein existence: Evidence at protein level
- Annotation score: 5/5

Tutaj znajduje się nazwa enzymu

Znajdź na tej samej podstronie numer EC tego enzymu [zadanie 18].

Numer EC: EC:3.2.1.1



# Names & Taxonomy<sup>i</sup>

## Protein names<sup>i</sup>

Recommended name	EC 3.2.1.1 (UniProtKB   ENZYME <a href="#">↗</a>   Rhea <a href="#">↗</a> ) <span>1 Publication</span>
EC number	EC 3.2.1.1

Numer EC enzymu

W bazie danych [RCSB](#) wyszukaj rekordy, które posiadają numer EC ten sam co enzym, który wytypowałeś w poleceniu powyżej.

RCSB PDB Deposit Search Visualize Analyze Download Learn About Documentation Careers COVID-19 MyPDB Contact us

Search Query History Browse Annotations MyPDB

QUERY: Full Text = "EC:3.2.1.1" MyPDB Login Search API

Advanced Search Query Builder [Help](#)

Full Text  Count  
Add Term Add Subquery Remove Subquery  
Add Subquery

Structure Attributes ☐  
Chemical Attributes ☐  
Sequence Similarity ☐  
Sequence Motif ☐  
Structure Similarity ☐  
Structure Motif ☐  
Chemical Similarity ☐

Return Structures grouped by No Grouping Include Computed Structure Models (CSM) ☐ Count Clear Search

Search Summary This query matches 18,856 Structures.

Refinements [?](#)

Structure Determination Methodology  
☐ experimental (18,856)

Scientific Name of Source Organism  
☐ Homo sapiens (8,805)  
☐ Mus musculus (1,193)  
☐ Severe acute respiratory syndrome coronavirus 2 (1,134)  
☐ Rattus norvegicus (590)  
☐ Human immunodeficiency virus 1 (473)  
☐ Escherichia coli (457)  
☐ Bos taurus (454)  
☐ synthetic construct (395)  
☐ Escherichia coli K-12 (251)  
☐ Sus scrofa (212)  
[More...](#)

Taxonomy  
☐ Eukaryota (12,454)  
☐ Bacteria (4,517)  
☐ Riboviria (3,183)  
☐ other sequences (401)  
☐ Archaea (282)

1 to 25 of 18,856 Structures Page 1 of 755 25 Sort by Score

**8IM8** [Download File](#) [View File](#)  
Crystal structure of Periplasmic alpha-amylase (MalS) from E.coli  
An, Y., Park, J.T., Park, K.H., Woo, E.J.  
(2023) Molecules 28:  
Released 2023-05-24  
Method X-RAY DIFFRACTION 2.7 Å  
Organisms Escherichia coli K-12  
Macromolecule Periplasmic alpha-amylase (protein)  
Unique Ligands CA

**1JFH** [Download File](#) [View File](#)  
STRUCTURE OF A PANCREATIC ALPHA-AMYLASE BOUND TO A SUBSTRATE ANALOGUE AT 2.03 ANGSTROM RESOLUTION  
Qian, M., Payan, F.  
(1997) Protein Sci 6: 2285-2296  
Released 1998-12-02  
Method X-RAY DIFFRACTION 2.03 Å  
Organisms Sus scrofa

Dlaczego wyszukiwarka pokazuje rekordy również z innych

gatunków niż człowiek (np. *Triticum aestivum*, *Synechocystis* sp. i inne) [zadanie 19]?  
Podpowiedź: [kliknij](#).

Odpowiedz: skategoryzowane są na podstawie schematów enzymów i bazują na reakcjach chemicznych.

W bazie danych <https://www.ncbi.nlm.nih.gov/genome/viruses/> znajdź sekwencję referencyjną genu wirusa SARS-CoV-2 oznaczoną jako Wuhan-Hu-1 (wpisz w wyszukiwarce "Search by sequence": "NC\_045512.2" i wejdź w pierwszy znaleziony rekord), następnie w oknie "Nucleotide details" kliknij ponownie w "NC\_045512". Jaką ma wielkość i z jakiego rodzaju kwasu nukleinowego jest zbudowany ten genom [zadanie 20]?

29903 bp ss-RNA

nucleocapsid phosphoprotein

#### Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

NCBI Reference Sequence: NC\_045512.2

[FASTA](#) [Graphics](#)

[Go to:](#) ☺

LOCUS	NC_045512	bp	linear	VRL 18-JUL-2020
DEFINITION	Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome.			
ACCESSION	NC_045512			
VERSION	NC_045512.2			
DBLINK	BioProject: <a href="#">PRJNA485481</a>			
KEYWORDS	RefSeq.			

wielkość  
genomu

rodzaj kwasu  
nukleinowego

Znajdź datę pozyskania tej próbki ("collection\_date") [zadanie 21].

Dec-2019