Bioinformatyka 2 - kurs mały

Adrian Kania[1]

[1]Zakład Biofizyki Obliczeniowej i Bioinformatyki

2025/2026

Dopasowanie sekwencji

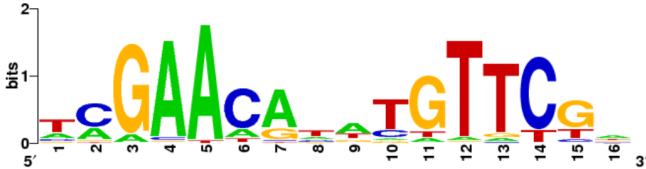
Dopasowanie pomiędzy sekwencjami określa które pozycje odpowiadają sobie.

```
acgtctag   acgtctag    acgtctag
||            |||||    || |||||
actctag-   -actctag    ac-tctag
```

2 matches	5 matches	7 matches
5 mismatches	2 mismatches	0 mismatches
1 not aligned	1 not aligned	1 not aligned

Takie porównania mogą być użyte to:

- szukania relacji ewolucyjnych,
- zidentyfikowania konserwatywnych miejsc,
- zidentyfikowania odpowiadających sobie genów pomiędzy różnymi modelami (np. ludzkimi czy mysimi).



Które dopasowanie jest najlepsze?

Na początku trzeba ustalić punktację za dopasowanie, niedopasowanie i przerwę. Poniżej przykładowe punkty. Łączny wynik dopasowania to suma punktów za każdą pozycję.

```
acgtctag
||
actctag-
```

2 matches
5 mismatches
1 not aligned

S = 2*1 + 5*(-1) + 1*(-2) = -5

```
acgtctag
   |||||
-actctag
```

5 matches
2 mismatches
1 not aligned

S = 5*1 + 2*(-1) + 1*(-2) = 1

```
acgtctag
|| |||||
ac-tctag
```

7 matches
0 mismatches
1 not aligned

S = 7*1 + 0*(-1) + 1*(-2) = 5

Sekwencje aminokwasowe

Do porównywania sekwencji aminokwasowych używamy odpowiednich macierzy podobieństwa (np. BLOSUM)

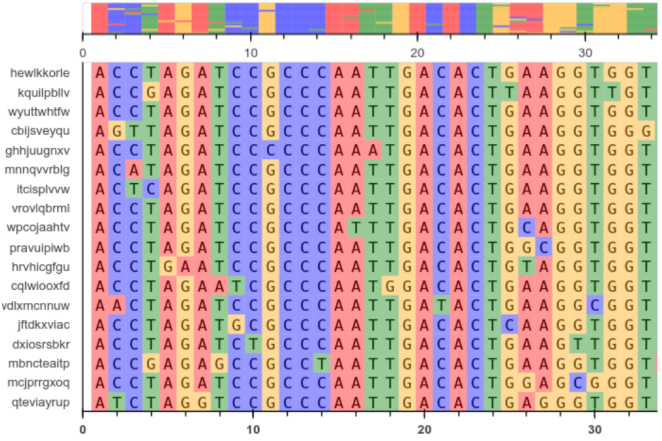
	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	**4**																			
R	-1	**5**																		
N	-2	0	**6**																	
D	-2	-2	**1**	**6**																
C	0	-3	-3	-3	**9**															
Q	-1	**1**	0	0	-3	**5**														
E	-1	0	0	**2**	-4	**2**	**5**													
G	0	-2	0	-1	-3	-2	-2	**6**												
H	-2	0	**1**	-1	-3	0	0	-2	**8**											
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	**4**										
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	**2**	**4**									
K	-1	**2**	0	-1	-3	**1**	**1**	-2	-1	-3	-2	**5**								
M	-1	-1	-2	-3	-1	0	-2	-3	-2	**1**	**2**	-1	**5**							
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	**6**						
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	**7**					
S	**1**	-1	**1**	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	**4**				
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	**1**	**5**			
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	**1**	-4	-3	-2	**11**		
Y	-2	-2	-2	-3	-2	-1	-2	-3	**2**	-1	-1	-2	-1	**3**	-3	-2	-2	**2**	**7**	
V	0	-3	-3	-3	-1	-2	-2	-3	-3	**3**	**1**	-2	**1**	-1	-2	-2	0	-3	-1	**4**

```
AABCDA...BBCDA
DABCDA.A.BBCBB
BBBCDABA.BCCAA
AAACDAC.DCBCDB
CCBADAB.DBBDCC
AAACAA...BBCCC
```

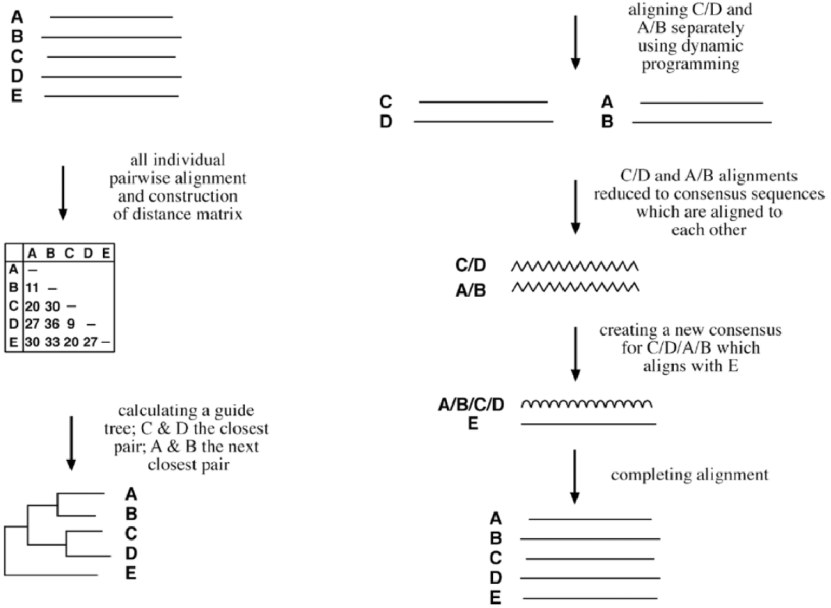
Dopasowanie wielu sekwencji (MSA)

```
ATGGCGAC 8
ATGCCGA- 7
***  ***
```

```
ATGCCGA- 7
ATGGCGAC 8
TTGGCGAC 8
 **  ***
```



Schemat podejścia progresywnego (np. Clustal)



Zielezinski *et al. Genome Biology* (2017) 18:186
DOI 10.1186/s13059-017-1319-7

Genome Biology

REVIEW Open Access

Alignment-free sequence comparison: benefits, applications, and tools

Andrzej Zielezinski[1], Susana Vinga[2], Jonas Almeida[3] and Wojciech M. Karlowski[1*]

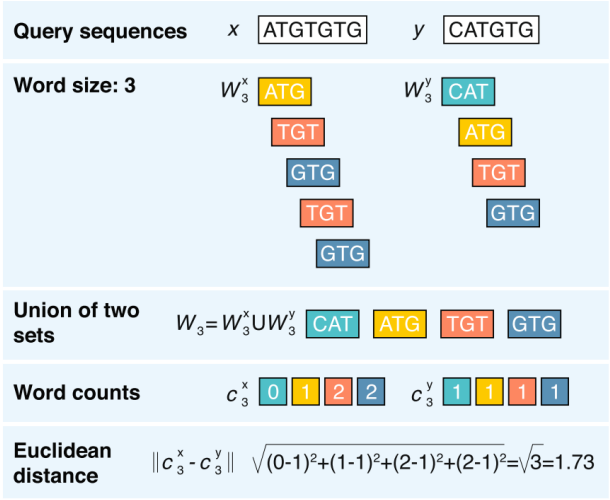
Abstract

Alignment-free sequence analyses have been applied to problems ranging from whole-genome phylogeny to the classification of protein families, identification of horizontally transferred genes, and detection of recombined sequences. The strength of these methods makes them particularly useful for next-generation sequencing data processing and analysis. However, many researchers are unclear about how these methods work, how they compare to alignment-based methods, and what their potential is for use for their research. We address these questions and provide a guide to the currently available alignment-free sequence analysis tools.

Metody bez dopasowania

Pięć przypadków, w których analiza sekwencji oparta na dopasowaniu może być kłopotliwa:

- Programy tworzące dopasowanie zakładają, że sekwencje homologiczne składają się z szeregu ułożonych liniowo i mniej lub bardziej konserwatywnych odcinków sekwencji.
- Dokładność dopasowania sekwencji gwałtownie spada w przypadkach, gdy identyczność sekwencji spada poniżej pewnego punktu krytycznego.
- Podejścia oparte na dopasowaniu są na ogół czasochłonne i pamięciowe, a zatem mają ograniczone zastosowanie w przypadku danych sekwencji w skali wielogenomowej.
- Obliczenie dokładnego dopasowania wielu sekwencji jest problelem NP trudnym, co oznacza, że trudno o rozwiązanie w akceptowalnych ramach czasowych.
- Dopasowanie sekwencji zależy od wielu założeń apriorycznych dotyczących ewolucji porównywanych sekwencji (np. macierzy podstawień, kar za przerwy i wartości progowych parametrów statystycznych).


Query sequences
x ATGTGTG
y CATGTG
Word size: 3
W_3^x ATG TGT GTG TGT GTG
W_3^y CAT ATG TGT GTG
Union of two sets
$W_3 = W_3^x \cup W_3^y$ CAT ATG TGT GTG
Word counts
c_3^x 0 1 2 2
c_3^y 1 1 1 1
Euclidean distance
$\| c_3^x - c_3^y \|$ $\sqrt{(0-1)^2+(1-1)^2+(2-1)^2+(2-1)^2} = \sqrt{3} = 1.73$

Query sequences

x ATGTGTG y CATGTG xy ATGTGTGCATGTG

Lempel-Ziv complexity

$c(x)=4$ $c(y)=5$ $c(xy)=7$

Normalized compression distance

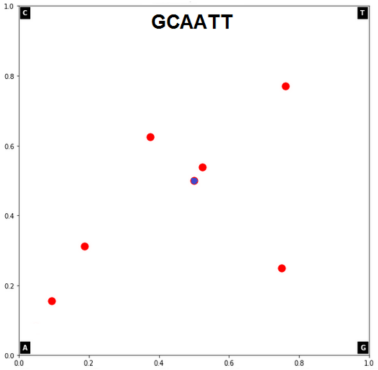
$$\frac{C(xy)\text{-}\min\{C(x), C(y)\}}{\max\{C(x), C(y)\}} \qquad \frac{7\text{-}4}{5}=0.6$$

Reprezentacja gry chaosu (CGR)

Jest to metoda służąca do reprezentacji numerycznej sekwencji biologicznych. Reprezentacja gry chaosu jest generowana wg

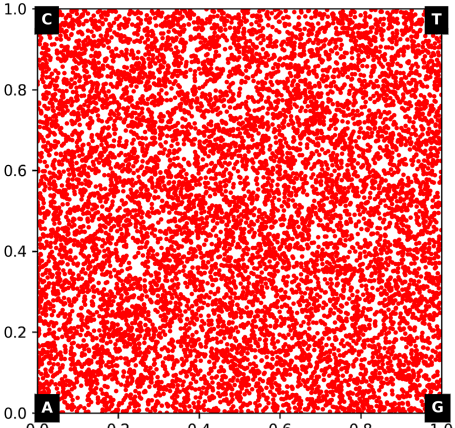
$$(x_{n+1}, y_{n+1}) = (x_n, y_n) - 0.5 \cdot ((x_n, y_n) - N_{n+1})$$

gdzie $N_A = (0, 0)$, $N_T = (1, 1)$, $N_C = (0, 1)$, $N_G = (1, 0)$, a $(x_0, y_0) = (0.5, 0.5)$ jest punktem początkowym.

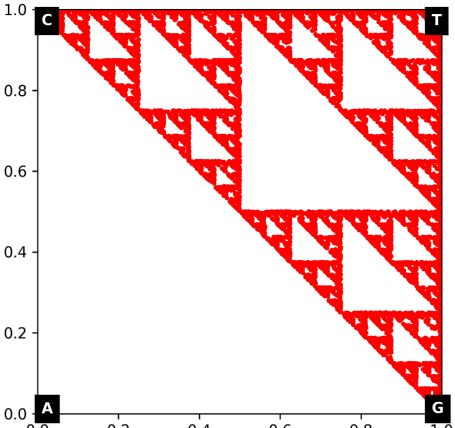


Reprezentacja gry chaosu

Za pomocą takiej reprezentacji możemy łatwo wskazać na pewne wzorce występujące w sekwencjach.



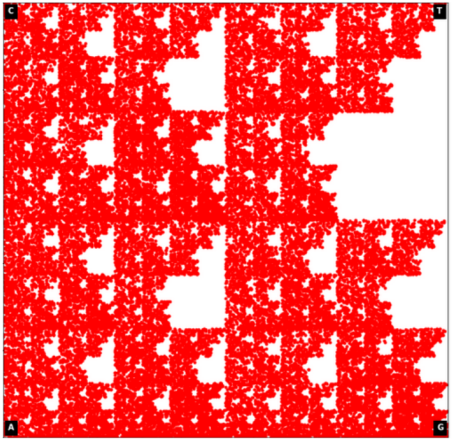
Losowa sekwencja



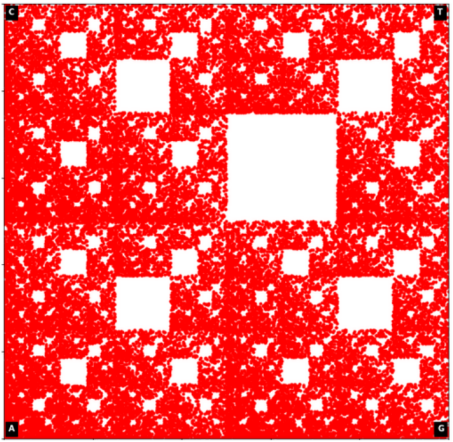
Brak A

Reprezentacja gry chaosu

Za pomocą takiej reprezentacji możemy łatwo wskazać na pewne wzorce występujące w sekwencjach.



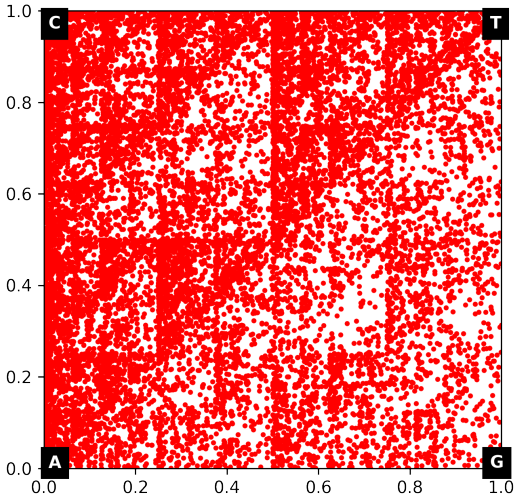
Brak GT w sekwencji



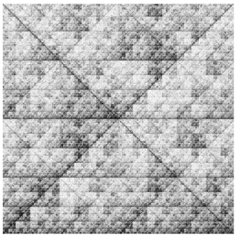
Brak AT w sekwencji

Reprezentacja gry chaosu

Genom mitochondrialny (NCBI ID: FJ986465.1)

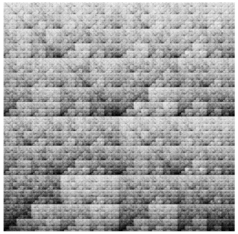


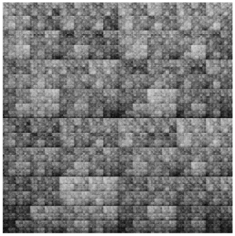
Reprezentacja gry chaosu



pszczoła

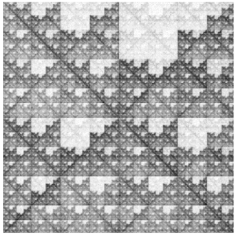
nicień

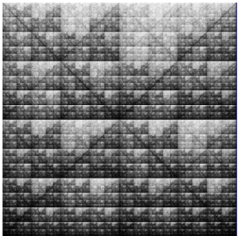




muszka owocowa

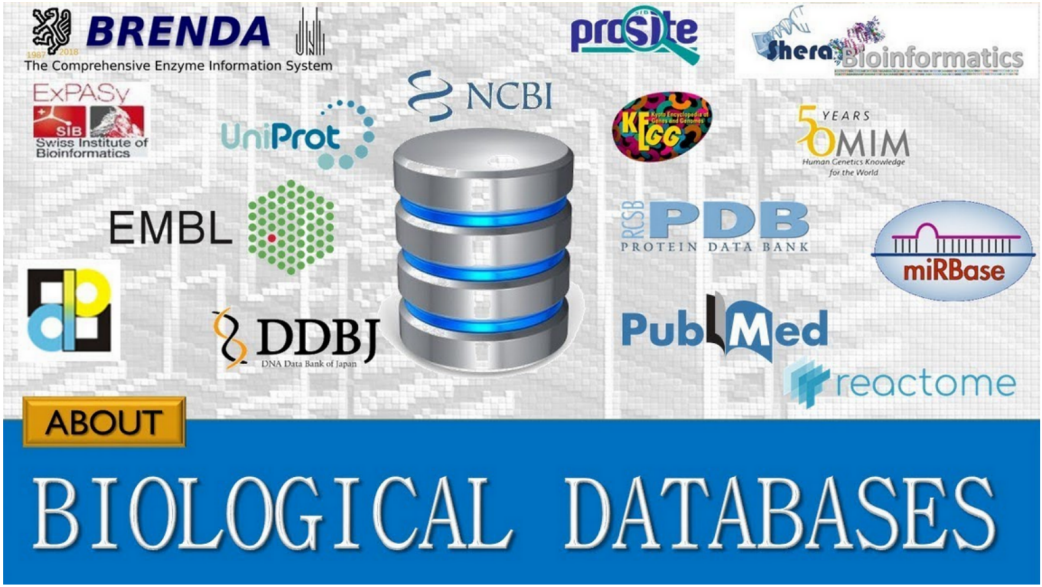
mysz





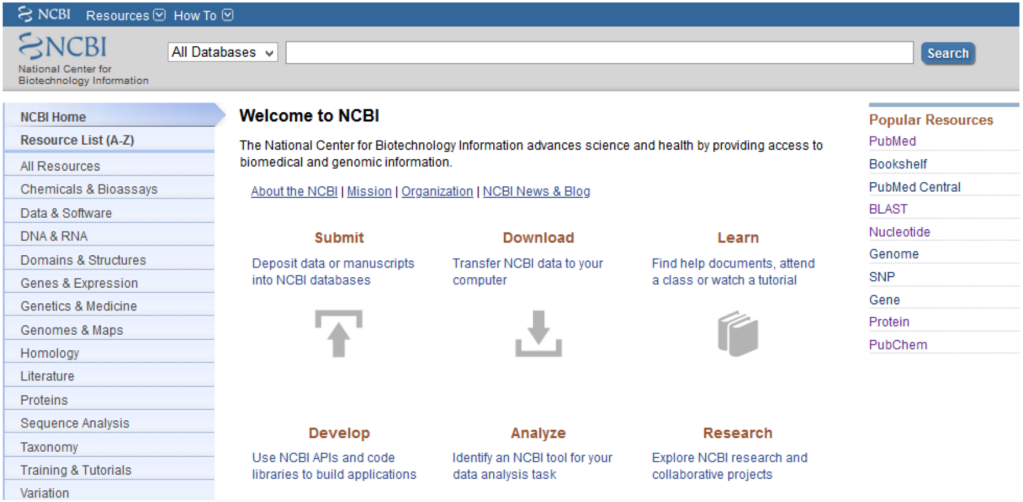
topola kalifornijska

Deschavanne, Patrick, Alain Giron, Joseph Vilain, Guillaume Fagot, and Bernard Fertil. "Genomic Signature: Characterization and Classification of Species Assessed by Chaos Game Representation of Sequences". Mol. Biol. Evol. 16(10): 1391 - 1399. 1999.

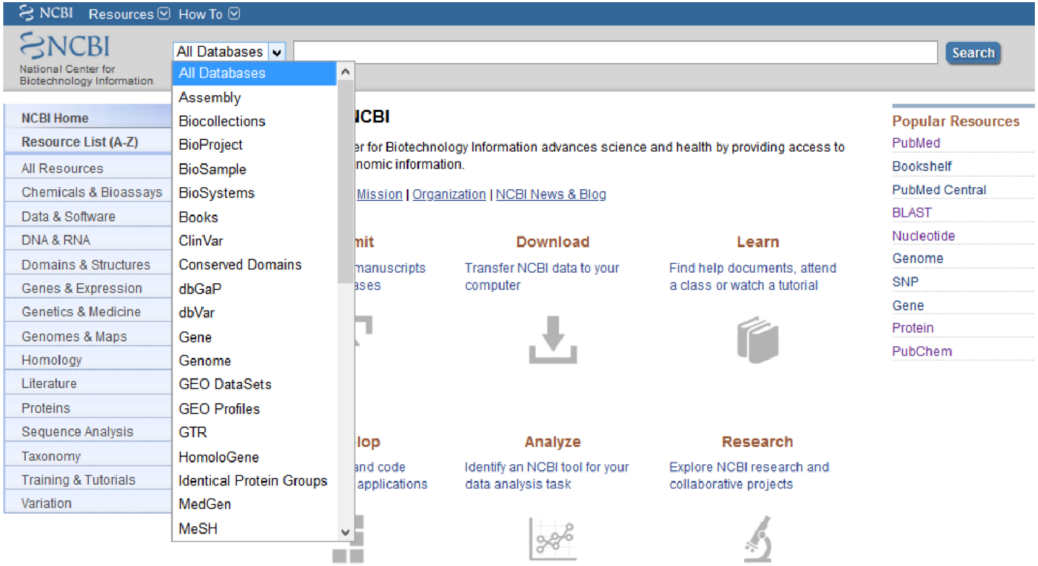

BRENDA
The Comprehensive Enzyme Information System
ExPASy
SIB
Swiss Institute of Bioinformatics
UniProt
NCBI
EMBL
DDBJ
DNA Data Bank of Japan
prosite
KEGG
5 YEARS
OMIM
Human Genetics Knowledge for the World
RCSB
PDB
PROTEIN DATA BANK
miRBase
PubMed
reactome
ABOUT
BIOLOGICAL DATABASES

NCBI

- **NCBI** – National Center for Biotechnology Information
- **NCBI Entrez** – zintegrowany system ponad 30 baz danych gromadzących informację z nauk biomedycznych, głównie w postaci sekwencji i literatury naukowej. Obejmuje m.in Pubmed (streszczenia publikacji), Nucleotide (sekwencje nukleotydowe), Protein (sekwencje aminokwasowe).



NCBI


NCBI Resources How To
NCBI
National Center for Biotechnology Information
All Databases
Search
NCBI Home
Resource List (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis
Taxonomy
Training & Tutorials
Variation
All Databases
Assembly
Biocollections
BioProject
BioSample
BioSystems
Books
ClinVar
Conserved Domains
dbGaP
dbVar
Gene
Genome
GEO DataSets
GEO Profiles
GTR
HomoloGene
Identical Protein Groups
MedGen
MeSH
NCBI
er for Biotechnology Information advances science and health by providing access to nomic information.
Mission | Organization | NCBI News & Blog
mit
manuscripts
ases
Download
Transfer NCBI data to your computer
Learn
Find help documents, attend a class or watch a tutorial
lop
and code
applications
Analyze
Identify an NCBI tool for your data analysis task
Research
Explore NCBI research and collaborative projects
Popular Resources
PubMed
Bookshelf
PubMed Central
BLAST
Nucleotide
Genome
SNP
Gene
Protein
PubChem

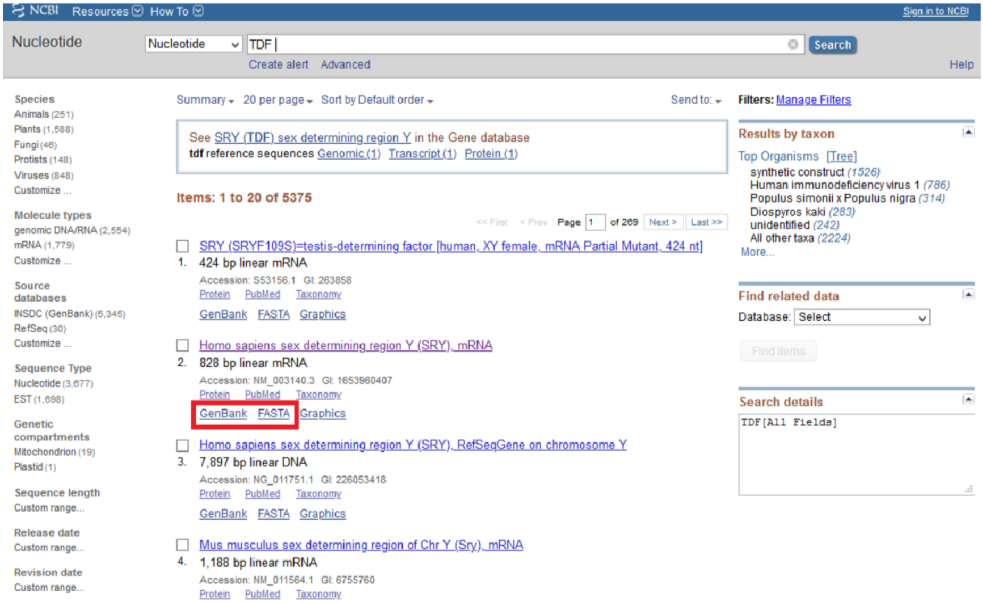
PubMed - tags

TABLE 2.2. Several Selected PubMed Tags and Their Brief Descriptions

Tag	Name	Description
AB	Abstract	Abstract
AD	Affiliation	Institutional affiliation and address of the first author and grant numbers
AID	Article identifier	Article ID values may include the PII (controlled publisher identifier) or doi (digital object identifier)
AU	Author	Authors
DP	Publication date	The date the article was published
JID	Journal ID	Unique journal ID in the National Library of Medicine's catalog of books, journals, and audiovisuals
LA	Language	The language in which the article was published
PL	Place of publication	Journal's country of publication
PT	Publication type	The type of material the article represents
RN	EC/RN number	Number assigned by the Enzyme Commission to designate a particular enzyme or by the Chemical Abstracts Service for Registry Numbers
SO	Source	Composite field containing bibliographic information
TA	Journal title abbreviation	Standard journal title abbreviation
TI	Title	The title of the article
VI	Volume	Journal volume

Source: www.ncbi.nlm.nih.gov/entrez/query/static/help/pmhelp.html.

Format sekwencji



FASTA format

Record name; starts with a “>”

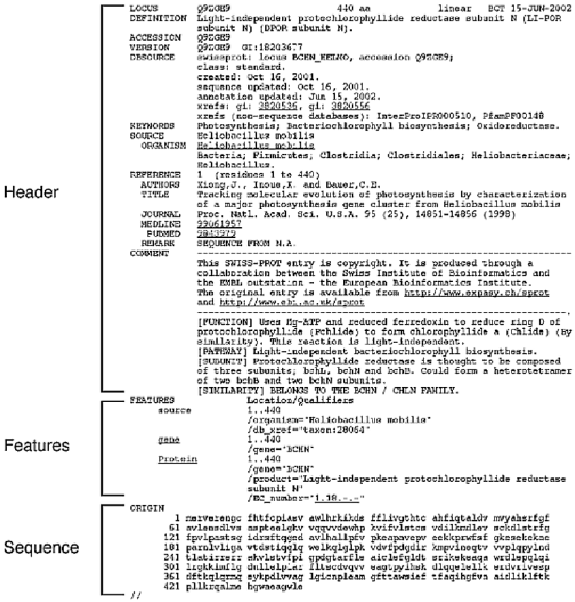
Fasta file:

DNA sequence

```
>sequence_1
TTTCCGGGGCACATAATCTTCAGCCGGGCGC
>sequence_1
TCAGCCGGGCCTTCAGCCGGGGCACATAATA
```

```
>Seq1 [organism=Carpodacus mexicanus] [clone=6b] actin (act) mRNA, partial cds
CCTTTATCTAATCTTTGGAGCATGAGCTGGCATAGTTGGAACCGCCCTCAGCCTCCTCATCCGTGCAGAA
TAATAATTTTCTTTATAGTAATACCAATCATGATCGGTGGTTTCGGAAACTGACTAGTCCCACTCATAAT


>Seq2 [organism=uncultured bacillus sp.] [isolate=A2] corticotropin (CT) gene, complete cds
GGTAGGTACCGCCCTAAGNCTCCTAATCCGAGCAGAACTANGCCAACCCGGAGCCCTTCTGGGAGACGAC
TCAACACCACCTTCTTTGACCCAGCAGGAGGAGGAGACCCAGTACTATACCAGCACCTATTCTGATTCTT
```




```
LOCUS       Q9ZGE9                   440 aa            linear   BCT 15-JUN-2002
DEFINITION  Light-independent protochlorophyllide reductase subunit N (LI-POR
            subunit N) (DPOR subunit N).
ACCESSION   Q9ZGE9
VERSION     Q9ZGE9  GI:18203677
DBSOURCE    swissprot: locus BCHN_HELMO, accession Q9ZGE9;
            class: standard.
            created: Oct 16, 2001.
            sequence updated: Oct 16, 2001.
            annotation updated: Jun 15, 2002.
            xrefs: gi: 3820536, gi: 3820556
            xrefs (non-sequence databases): InterProIPR000510, PfamPF00148
KEYWORDS    Photosynthesis; Bacteriochlorophyll biosynthesis; Oxidoreductase.
SOURCE      Heliobacillus mobilis
  ORGANISM  Heliobacillus mobilis
            Bacteria; Firmicutes; Clostridia; Clostridiales; Heliobacteriaceae;
            Heliobacillus.
REFERENCE   1  (residues 1 to 440)
  AUTHORS   Xiong,J., Inoue,K. and Bauer,C.E.
  TITLE     Tracking molecular evolution of photosynthesis by characterization
            of a major photosynthesis gene cluster from Heliobacillus mobilis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (25), 14851-14856 (1998)
  MEDLINE   99061957
   PUBMED   9843979
  REMARK    SEQUENCE FROM N.A.
COMMENT     -------------------------------------------------------------------
            This SWISS-PROT entry is copyright. It is produced through a
            collaboration between the Swiss Institute of Bioinformatics and
            the EMBL outstation - the European Bioinformatics Institute.
            The original entry is available from http://www.expasy.ch/sprot
            and http://www.ebi.ac.uk/sprot
            -------------------------------------------------------------------.
            [FUNCTION] Uses Mg-ATP and reduced ferredoxin to reduce ring D of
            protochlorophyllide (Pchlide) to form chlorophyllide a (Chlide) (By
            similarity). This reaction is light-independent.
            [PATHWAY] Light-independent bacteriochlorophyll biosynthesis.
            [SUBUNIT] Protochlorophyllide reductase is thought to be composed
            of three subunits; bchL, bchN and bchB. Could form a heterotetramer
            of two bchB and two bchN subunits.
            [SIMILARITY] BELONGS TO THE BCHN / CHLN FAMILY.
FEATURES             Location/Qualifiers
     source          1..440
                     /organism="Heliobacillus mobilis"
                     /db_xref="taxon:28064"
     gene            1..440
                     /gene="BCHN"
     Protein         1..440
                     /gene="BCHN"
                     /product="Light-independent protochlorophyllide reductase
                     subunit N"
                     /EC_number="1.18.-.-"
ORIGIN
        1 msrvetengc fhtfcpiasv awlhrkikds ffliygthtc ahfiqtaldv mvyahsrfgf
       61 avlessdlvs asptealgkv vqqvvdewhp kvifvlstcs vdilkmdlav sckdlstrfg
      121 fpvlpastsg idrsftqged avlhallpfv pkeapavepv eckkprwfsf gkeseksksc
      181 parnlvliga vtdstiqqlq welkqlglpk vdvfpdgdir kmpvineqtv vvplqpylnd
      241 tlatirrerr akvlstvfpi gpdgtarfle aiclsfgldt srikskeaqa wrdlspqlqi
      301 lrgkkimflg dnllelplar fltscdvqvv eagtpyihsk dlqqelellk srdvrivesp
      361 dftkqlgrmq sykpdlvvag lgicnpleam gfttawsief tfaqihgfvs aidliklftk
      421 pllkrqalms hgwaeagvle
//
```

Figure 2.3: NCBI GenBank/GenPept format showing the three major components of a sequence file.

GenBank

```
LOCUS       X56734 1859 bp mRNA linear PLN 25-NOV-2005
DEFINITION  Trifolium repens mRNA for non-cyanogenic beta-glucosidase.
ACCESSION   X56734 S46826
VERSION     X56734.1  GI:21954
KEYWORDS    beta-glucosidase.
SOURCE      Trifolium repens (white clover)
  ORGANISM  Trifolium repens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Trifolium.
REFERENCE   1  (bases 1 to 1859)
  AUTHORS   Oxtoby,E., Dunn,M.A., Pancoro,A. and Hughes,M.A.
  TITLE     Nucleotide and derived amino acid sequence of the cyanogenic
            beta-glucosidase (linamarase) from white clover (Trifolium repens
            L.)
  JOURNAL   Plant Mol. Biol. 17 (2), 209-219 (1991)
   PUBMED   1907511
REFERENCE   2  (bases 1 to 1859)
  AUTHORS   Hughes,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-1990) Hughes M.A., University of Newcastle Upon
            Tyne, Medical School, Newcastle Upon Tyne, NE2 4HH, UK
COMMENT     On Jun 10, 2005 this sequence version replaced gi:233395.
```

```
FEATURES             Location/Qualifiers
     source          1..1859
                     /organism="Trifolium repens"
                     /mol_type="mRNA"
                     /db_xref="taxon:3899"
                     /clone="TRE361"
                     /tissue_type="leaves"
                     /clone_lib="lambda gt10"
     mRNA            1..1859
                     /experiment="experimental evidence, no additional details
                     recorded"
     CDS             14..1496
                     /EC_number="3.2.1.21"
                     /note="non-cyanogenic"
                     /codon_start=1
                     /product="beta-glucosidase"
                     /protein_id="CAA40058.1"
                     /db_xref="GI:21955"
                     /db_xref="GOA:P26204"
                     /db_xref="InterPro:IPR001360"
                     /db_xref="UniProtKB/Swiss-Prot:P26204"
                     /translation="MDFIVAIFALFVISSFTITSTNAVEASTLLDIGNLSRSSFPRGF
                     IFGAGSSAYQFEGAVNEDGRGPSIWDTFTHKYPEKIRDGSNADITVDQYHRYKEDVGI
                     MKDQNMDSYRFSISWPRILPKGKLSGGINHEGIKYYNNLINELLANGIQPFVTLFHWD
                     LPQVLEDEYGGFLNSGVINDFRDYTDLCFKEFGDRVRYWSTLNEPWVFSNSGYALGTN
                     APGRCSASNVAKPGDSGTGPYIVTHNQILAHAEAVHVYKTKYQAYQKGKIGITLVSNW
                     LMPLDDNSIPDIKAAERSLDFQFGLFMEQLTTGDYSKSMRKIVKNRLPKFSKFESSLV
                     NGSFDFIGINYYSSSYISNAPSHGNAKPSYSTNPMTNISFEKHGIPLGPRAASIWIYV
                     YPYMFIQEDFEIFCYILKINITILQFSITENGMNEFNDATLPVEEALLNTYRIDYYYR
                     HLYYIRSAIRAGSNVKGFYAWSFLDCNEWFAGFTVRFGLHFVD"
```

```
ORIGIN
        1 aaacaaacca aatatggatt ttattgtagc catatttgct ctgtttgtta ttagctcatt
       61 cacaattact tccacaaatg cagttgaagc ttctacttct cttgacatag gtaacctgag
      121 tcggagcagt tttcctcgtg gcttcatctt tggtgctgga tcttcagcat accaatttga
      181 aggtgcagta aacgaaggcg gtagaggacc aagtatttgg gataccttca cccataaata
      241 tccagaaaaa ataagggatg gaagcaatgc agacatcacg gttgaccaat atcaccgcta
                ...
     1741 agaagctatg atcataacta taggttgatc cttcatgtat cagtttgatg ttgagaatac
     1801 tttgaattaa aagtcttttt ttatttttct aaaaaaaaaa aaaaaaaaaa aaaaaaaaa
//
```

1. **Entrez Programming Utilities** – oprogramowanie umożliwiające dostęp do danych zintegrowanych w ramach systemu Entrez bez konieczności obsługi formularzy na stronach WWW
   1. **ESearch** – wyszukiwanie identyfikatorów/kodów dostępu rekordów wybranej bazy danych,
   2. **EFetch** – pobieranie rekordów o wskazanych identyfikatorach lub wyników wskazanego wyszukiwania

EFetch

EFetch (efetch.fcgi) returns full data records for a list of unique identifiers (UIDs) in a format specified in the parameters. The list of UIDs is either provided in the parameters, or is retrieved from the History server.

EFetch Parameters

EFetch Required Parameters

- `db` (required): Database containing the unique identifiers (UIDs) for which you wish to retrieve records. You can see NCBI's table of Entrez Unique Identifiers (UIDs) for a complete list of allowable database names, but some example values include:
  - `pubmed`: PubMed
  - `pmc`: PubMed Central
  - `nlmcatalog`: NLM Catalog
- `id` (required): Either a single unique identifier (UID) or a comma-delimited list of UIDs. All of the UIDs must be from the database specified by the `db` parameter.

EFetch Optional Parameters

- `retstart` (optional): Setting this parameter helps limit which records will be shown in the output, as it determines whether the record for the first input unique identifier (UID) is retrieved, or whether to skip to a later UID in the input list. For example, if `retstart` is set to `10`, the output will begin with the record for the tenth UID. The default of this parameter is `1`, corresponding to the first UID in the input list. This parameter can be used in conjunction with `retmax` to download an arbitrary subset of records.
- `retmax` (optional): Total number of records to be shown in the output, up to a maximum of 10,000. If the set of records you are trying to retrieve is larger than 10,000, you can submit multiple EFetch requests, and increase the `retstart` parameter each time.
- `retmode`/`rettype`: These two parameters determine how your results will be displayed. `retmode` determines the data format your records will be returned in (e.g. XML, plain text, etc.). `rettype` determines the specific view your records will be returned in (e.g. MEDLINE, Abstract, list of PMIDs, etc.). Different databases have different allowable data formats and record views, and not all `retmode` data formats are compatible with all `rettype` record views, and vice versa.

  The table below shows the allowable combinations of `retmode` and `rettype` for some of the databases. **Bold** `retmode` values are the default data format for the specified database. **Bold** `rettype` parameters are the default record view for the specified data format and database.

EFetch Examples

- Retrieve the abstract view (text format) of two PubMed records.
  - https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=pubmed&id=17284678,9997&retmode=text&rettype=abstract
- Retrieve two PubMed records in XML format.
  - https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=pubmed&id=11748933,11700088&retmode=xml

ESearch Formatting Parameters

- `retstart` (optional): Setting this parameter helps limit which of the unique identifiers (UIDs) in the results set will be shown in the output, as it determines whether the output begins at the first retrieved UID, or with a UID that is later in the results set. For example, if `retstart` is set to `10`, the first ten UIDs in the results set will be skipped, and the output will begin with the eleventh UID. The default of this parameter is `0`, corresponding to the first record in the entire set. This parameter can be used in conjunction with `retmax` to download an arbitrary subset of UIDs retrieved from a search.
- `retmax` (optional): Total number of unique identifiers (UIDs) from the retrieved set to be shown in the output (default=20). Increasing `retmax` allows more of the retrieved UIDs to be included in the output, up to a maximum of 100,000 UIDs. If you need to retrieve more than 100,000 UIDs, you can submit multiple ESearch requests, and increase the `retstart` parameter each time. For example:
  - https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=pubmed&term=cancer&retmax=100: This URL will return results 1 through 100 of a search for "Cancer".
  - https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=pubmed&term=cancer&retstart=100&retmax=100: This URL will return results 101 through 200 of the same search for "Cancer".
- `rettype` (optional): Retrieval type. There are two supported values:
  - `uilist` (default): Displays the standard XML output, including a list of unique identifiers (UIDs), the total number of results, and the query translation for the search.
  - `count`: Displays only the total number of results, without the list of UIDs or query translation.
- `retmode` (optional): Determines the format of the returned output. The default value is `xml`, but `json` is also supported.