

Что такое геном?

Что такое геном?

Геном – совокупность всей наследственной информации, записанной в ДНК (или РНК в случае РНК-содержащих вирусов).

Геном:

- ▶ одной клетки
- ▶ организма
- ▶ вида (или штамма в случае прокариот)

Как правило, в геном включают только одну из набора *гомологичных* молекул, если их несколько.

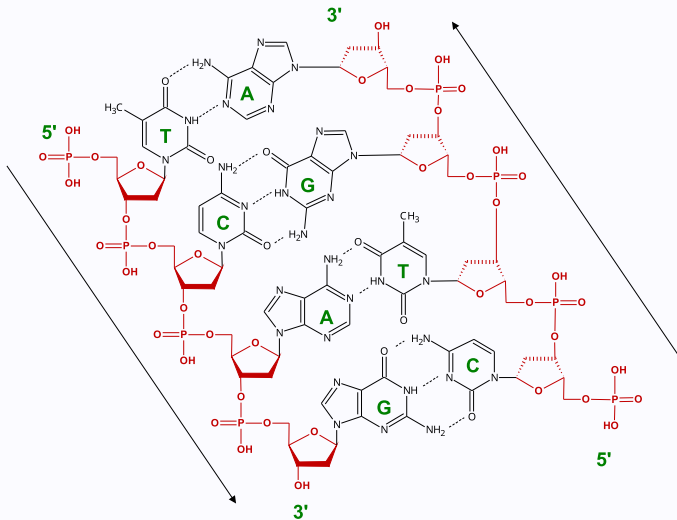
Геномная последовательность

Геном – это текст, т.е. последовательность символов.

Носитель генома:

- ▶ в клетке – молекулы ДНК (или РНК)
- ▶ в компьютере – *текстовый* файл

ДНК – носитель генома



из презентации А.В. Головина для студентов 2 курса

Как записать последовательность в файл?

- ▶ нуклеотиды обозначают латинскими буквами A, C, G и T
- ▶ последовательность **всегда** записывают от 5' к 3'

А еще нужно придумать название последовательности (её идентификатор).

Формат FASTA

```
>NC_000913.3 Escherichia coli str. K-12 substr. MG1655, complete genome
AGCTTTTCATTCTGACTGCAACGGGCAATATGCTCTGTGTGGATTAAAAAAGAGTGTCTGATAGCAGCTTCTGAAGT
GTTACCTGCGGTGAGTAAATTTAATTTTATTGGATTAAGTCACTAAATACTTTAACCBAATATAGGCATAGCGCACAGAC
AGATAAAAATTACAGAGTACACAACATCCATGAAACGCATTAGCACCACCATTACCACCACCATTACCATTACCACAGGT
AACGGTGCAGGGTGCAGCGGTACAGGAAACACAGAAAAAAGCCCGCACCTGACAGTGCAGGGCTTTTTTTTCGACCAAAGG
TAACGAGGTAACAACCATGCGAGTGTGAAAGTTCGGCGGTACATCAGTGGCAAATGCAGAACGTTTTCTGCGTGTGCGCG
ATATTCTGGAAGCAATGCCAGGCAGGGGACCGTCCCTCTCTGCCCCCGCCAAAATCACCACCACTGCTGGTG
GCGATGATTGAAAAAACCATAGCGGCCAGGATGCTTTACCCAATATCAGCGATGCCGAACGTATTTTTGCGGAACCTTT
GACGGGACTCGCCGCGCGCCAGCGGGGTTCGCGTGGCGCAATTGAAAACCTTCGTCGATCAGGAATTTGCCCAAATAA
AACATGTCCTGTCATGGCATTAGTTTGTGGGGCAGTGCCCGGATAGCATCAACGCTGCGCTGATTTGCCGTGGCGAGAAA
ATGTCGATCGCCATTATGGCCGGCGTATGAAGCGCGCGGTCAACAACGTTACTGTTATCGATCCGGTCGAAAACTGCT
GGCAGTGGGGCATTACCTCGAATCTACCTGCGATATTGCTGAGTCCACCCGCGGTATTGCGGCAAGCCGCTTCGGGCTG
ATCAGATGCTGCTGATGGCAGGTTTACCGCCGGTAAATGAAAAAGGCGAACTGGTGGTGTCTGGACGCAACGGTTCCGAC
TACTCTGCTGCGGTGCTGGTGCCTGTTTACGCGCCGATGTTGCGGAGATTTGGACGGACGTTGACGGGGTCTATACCTG
CGACCCGCGTCAAGGTGCCCGATGCGAGGTTGTTGAAGTCGATGTCCTACCAGGAAGCGATGGAGCTTTCCTACTTCCGCG
CTAAAGTTCTTACCCCGCGCACCATTAACCCCATCGCCAGTTCAGATCCCTTGCCCTGATTAATAAATACCGGAAATCCT
CAAGCACCAGGTACGCTCATTGGTGCCAGCCGTGATGAAGACGAATTACCGGTCAAGGGCATTTCGAATCTGAATAACAT
GGCAATGTTACGCGTTTCTGGTCCGGGGATGAAAGGGATGGTCCGCATGGCGGCGCGCGCTTTGACGCGATGTCACGCG
CCCGTATTTCCGTGGTGCTGATTACGCAATCATCTCCGAATACAGCATCAGTTTCTGCGTTCACAAAGCGACTGTGTG
CGAGCTGAACGGGCAATGCAGGAAGATTTCTACCTGGAAGTGAAGAAGGCTTACTGGAGCCGCTGGCAGTGACGGAACG
```

— заголовок: **имя** (до первого пробела) и описание

последовательность

```
...
TAAATGCTAGCTTTCGTCACATTATTTTAATAATCCAAGTGTGATCATACAACATAATAACGTTGGTGAATCCAATTG
TCGAGATTATTTTATATAAAATTATCCTAAGTAAACAGAAAGGATATGTAGCATTTTTTAACAACCTCAACCGTTAGTACA
GTCAGGAAATAGTTTAGCCTTTTTTAAGCTAAGTAAAGGGCTTTTTCTGCGACTTACGTTAAGAATTTGTAAATTCGCAC
CGCGTAATAAGTTGACAGTGATCACCCGGTTCGCGGTTATTGATCAAGAAAGAGTGGCAATATGCGTATAACGATTATTC
TGGTTCGACCCCGCCAGAGCAGAAAAATTTGGGGCAGCGCGCGGGCAATGAAAACGATGGGGTTTAGCGATCTGCGGATT
GTCGATAGTCAAGGCACACTGGAGCCAGGCACCCGCTGGGTTCGACATGGATCTGGTGATATTATTGATAATATTAAGT
TTTCCCGACATTGGCTGAATCGTTACACGATGTCGATTTCACTGTCGCCACCACTGCGCGCAGTCGGGCGAAATATCATT
ACTACGCCACGCAGTTGAACCTGTGCGGCTGTGAGAGAAAAATCTTCATGGATGAGCCATGCCGCGCTGGTGTGTTGGT
CGCGAAGATTCCGGGTTGACTAACGAAGAGTTAGCGTTGGCTGACGTTCTTACTGGTGTGCCGATGGTGGCGGATTATCC
TTCGCTCAATCTGGGGCAGGCGGTGATGGTCTATTGCTATCAATTAGCAACATTAATACAACAACCGGCGAAAAAGTGATG
CAACGGCAGACCAACATCAACTGCAAGCTTTACGCGAACGAGCCATGACATTGCTGACGACTCTGGCAGTGGCAGATGAC
ATAAACTGGTCTGACTGGTTACAACAACGCCCTGGGGCTTTAGAGCAACGAGACACGGCAATGTTGCACCGTTTGTGCA
TGATATTGAAAAAATATACCAAATAAAAAACGCCTTAGTAAGTATTTTTT
```

Произвол при создании записи в базе данных

Отсеквенировали хромосому бактерии, какую последовательность положить в базу данных?

- ▶ где у кольцевой хромосомы начало?
- ▶ где прямая цепь, а где обратная?

Все это на усмотрение создающего запись!

Зачем секвенировать геномы?

Компоненты генома

Репликоны:

- ▶ хромосомы
- ▶ плазмиды
- ▶ ...

Функциональные элементы:

- ▶ гены белков и РНК
- ▶ регуляторные элементы
- ▶ структурные элементы (теломеры, центромеры)
- ▶ ориджины репликации
- ▶ мобильные элементы
- ▶ повторы
- ▶ ...

Аннотированный геном

(на самом деле каждый репликон отдельно)

LOCUS NC_000913 4641652 bp DNA circular CON 17-JUL-2025
DEFINITION Escherichia coli str. K-12 substr. MG1655, complete genome.
ACCESSION NC_000913

...
SOURCE Escherichia coli str. K-12 substr. MG1655
ORGANISM Escherichia coli str. K-12 substr. MG1655
Bacteria; Pseudomonadati; Pseudomonadota; Gammaproteobacteria;
Enterobacterales; Enterobacteriaceae; Escherichia.

...
COMMENT REVIEWED REFSEQ: This record has been curated by NCBI staff. The
reference sequence is identical to U00096.
On Nov 3, 2013 this sequence version replaced NC_000913.2.

...
FEATURES Location/Qualifiers
source 1..4641652
/organism="Escherichia coli str. K-12 substr. MG1655"
/mol_type="genomic DNA"
/strain="K-12"
gene 190..255
/gene="thrL"
/locus_tag="b0001"
CDS 190..255
/transl_table=11
/product="thr operon leader peptide"
/translation="MKRISTTITTTITTTGNGAG"

...
ORIGIN
1 agcttttcat tctgactgca acgggcaata tgtctctgtg tggattaaaa aaagagtgtc
61 tgatagcagc ttctgaactg gttacctgcc gtgagtaaata taaaatttta ttgacttagg

...
4641541 acaacaacgc ctggggcttt tagagcaacg agacacggca atgttgacc gtttgcgtga
4641601 tgatattgaa aaaaatatca ccaataaaaa aacgccttag taagtatttt tc

//

Таблица локальных особенностей

# feature	class	...	genomic_acc	start	end	strand	product_acc	...	name	...
gene	protein_coding		NC_000913.3	190	255	+				
CDS	with_protein		NC_000913.3	190	255	+	NP_414542.1		thr operon leader peptide	
gene	protein_coding		NC_000913.3	337	2799	+				
CDS	with_protein		NC_000913.3	337	2799	+	NP_414543.1		fused aspartate kinase/homoserine dehydrogenase 1	
gene	protein_coding		NC_000913.3	2801	3733	+				
CDS	with_protein		NC_000913.3	2801	3733	+	NP_414544.1		homoserine kinase	
gene	protein_coding		NC_000913.3	3734	5020	+				
CDS	with_protein		NC_000913.3	3734	5020	+	NP_414545.1		threonine synthase	
gene	protein_coding		NC_000913.3	5234	5530	+				
CDS	with_protein		NC_000913.3	5234	5530	+	NP_414546.1		DUF2502 domain-containing protein YaaX	
gene	protein_coding		NC_000913.3	5683	6459	-				
CDS	with_protein		NC_000913.3	5683	6459	-	NP_414547.1		DNA binding and peroxide stress response protein YaaA	
gene	protein_coding		NC_000913.3	12163	14079	+				
CDS	with_protein		NC_000913.3	12163	14079	+	NP_414555.1		chaperone protein DnaK	
gene	protein_coding		NC_000913.3	14168	15298	+				
CDS	with_protein		NC_000913.3	14168	15298	+	NP_414556.1		chaperone protein DnaJ	
gene	protein_coding		NC_000913.3	15445	16557	+				
CDS	with_protein		NC_000913.3	15445	16557	+	NP_414557.1		IS186/IS421 family transposase	
gene	protein_coding		NC_000913.3	16751	16960	-				
CDS	with_protein		NC_000913.3	16751	16960	-	NP_414559.1		regulatory protein MokC	
gene	protein_coding		NC_000913.3	16751	16903	-				
CDS	with_protein		NC_000913.3	16751	16903	-	YP_025292.1		small toxic polypeptide HokC	
gene	ncRNA		NC_000913.3	16952	17010	+				
ncRNA	other		NC_000913.3	16952	17010	+			small regulatory RNA antitoxin SokC	
gene	protein_coding		NC_000913.3	17489	18655	+				
CDS	with_protein		NC_000913.3	17489	18655	+	NP_414560.1		Na(+):H(+) antiporter NhaA	
...										

Опять FASTA

```
>lcl|NC_000913.3_cds_NP_414542.1_1 [gene=thrL] [locus_tag=b0001] [db_xref=UniProtKB/Swiss-Prot:P0AD86] [protein=thr operon leader p]>>
ATGAAACGCATTAGCACCACCATTACCACCACCATCACCATTACCACAGGTAACGGTGCGGGCTGA
>lcl|NC_000913.3_cds_NP_414543.1_2 [gene=thrA] [locus_tag=b0002] [db_xref=UniProtKB/Swiss-Prot:P00561] [protein=fused aspartate kin]>>
ATGCGAGTGTGTAAGTTCGGCGGTACATCAGTGGCAAAATGCAGAACGTTTTCTGCGTGTGCGGATATTCTGGAAGCAA
TGCCAGGCAGGGGCAGGTGGCCACGCTCTCTGTCGCCCGCCAAAATCACCAACCACCTGGTGGCGATGATTGAAAAAA
CCATTAGCGGCCAGGATGCTTTACCCAATATCAGCGATGCCGAACGTATTTTTGCCGAACCTTTGACGGGACTCGCCGCC
GCCAGCGCGGGTTCCCGCTGGCGCAATTGAAAACTTTCTGTCGATCAGGAATTTGCCCAATAAAACATGTCCTGCAATGG
CATTAGTTTTGTTGGGGCAGTGCCCGGATAGCATAACCGCTGCGCTGATTGCGCGTGGCGAGAAAAATGTCGATCGCCATTA
TTACTACCATCAGTTGCGTTATGCGCGGAGAAAATCGCGGCGTAAATTCCTCTATGACACCAACGTTGGGGCTGGATTAC
CGGTTATTGAGAACCTGCAAAATCTGCTCAATGCAGGTGATGAATTGATGAAGTTCTCCGGCATTCTTTCTGGTTGCGCTT
TCTTATATCTTCGGCAAGTTAGACGAAGGCATGAGTTTCTCCGAGGCGACACGCTGGCGCGGGAAATGGGTATACCGA
ACCGGACCCGCGAGATGATCTTTCTGGTATGGATGTGGCGCGTAAACTATTGATTCTGCTCGTGAACGGGACGTGAAC
TGGAGCTGGCGGATATTGAAATTGAACCTGAGTGTGCTGCGCAGAGTTTAAACGCCGAGGGGTATGTTGCGCGCTTTTATGGCG
AATCTGTCACAACTCGACGATCTCTTTGCCGCGCGCTGGCGAAGGCCGTGATGAAGGAAAAGTTTTGCGCTATGTTGG
CAATATTGATGAAGATGGCGTCTGCCGCGTGAAGATTGCCGAAGTGATGGTAATGATCCGCTGTTCAAAGTGAAAAATG
GCGAAAACGCCCTGGCCTTCTATAGCCACTATTATCAGCCGCTGCCGTTGGTACTGCGCGGATATGGTGGCGGCAATGAC
GTTACAGCTGCCGCTGTCTTTGCTGATGCTACGTAACCTCTCATGGAAGTTAGGAGTCTGA
>lcl|NC_000913.3_cds_NP_414544.1_3 [gene=thrB] [locus_tag=b0003] [db_xref=UniProtKB/Swiss-Prot:P00547] [protein=homoserine kinase] >>
ATGGTTAAAGTTTATGCCCGGGCTTCCAGTGCCAAATATGAGCGTCGGGTTTGATGTGCTCGGGGCGGCGGTGACACCTGT
TGATGGTGCATTGCTCGGAGATGTAGTCACGGTTGAGGCGGCAGAGACATTAGTCTCAACAACCTCGGACGCTTTGCCG
ATAAGCTGCCGTACAGAACCCAGGGAATAATCTGTTTATCAGTGTGCGGAGCGTTTTTGGCAGGAATGGGTAAAGCAAATT
CCAGTGGCGATGACCTTGGAAAAGAATATGCCGATCGGTTCCGGCTTAGGCTCCAGTGCCTGTTCCGGTGGTTCGCGGCGCT
GATGGCGATGAATGAACACTGCGGCAAGCCGCTTAATGACACTCGTTTGCTGGCTTTGATGGGCGAGCTGGAAGGCCGTA
TCTCCGGCAGCATTATTACGACAACGTGGCACCGCTGTTTTCTCGGTGGTATGCAGTTGATGATCGAAGAAAACGACATC
ATCAGCCAGCAAGTGCCAGGGTTTGATGAGTGGCTGTGGGTGCTGCGGCGTATCCGGGGATTAAAGTCTCGACGGCAGAAGC
CAGGGCTATTTTACCGGCGCAGTATCGCCGCCAGGATTGCATTGCGCACGGGCGACATCTGGCAGGCTTCATTACGCGCT
GCTATTCGCGTCAGCCTGAGCTTGCCGCGAAGCTGATGAAGATGTTATCGTGAACCTTACCCTGAACGGTTACTGCCA
GGCTTCGGGCGAGGCGCGGCGAGCGGTGCGGGAATCGGCGCGGTAGCGAGCGGTATCTCCGGCTCCGGCCCCGACCTTGTT
CGCTCTGTGTGACAAGCCGGAACCGCCACGCGCGTTGCGCACTGGTTGGGTGAAGAACTACCTGCAAAATCAGGAAGGTT
TTGTTTCATATTTGCCGGCTGGATACGGCGGGCGCACGAGTACTGGAAAACTAA
>lcl|NC_000913.3_cds_NP_414545.1_4 [gene=thrC] [locus_tag=b0004] [db_xref=UniProtKB/Swiss-Prot:P00934] [protein=threonine synthase]>>
ATGAAACTCTACAATCTGAAAGATCACACGAGCAGGTACGCTTTGCGCAAGCCGTAACCCAGGGGTGGGCAAAAAATCA
CCGGTGATACCGGAGCGGCAGTGGCTCATGCTTTCTACGGTTTACCGAATGTGAAAGTGTTATCCTCTATCCACGAGGC
```

...