

Что такое геном?

Мини-презентация

12 сентября 2025

С.А. Спирин

Что такое геном?

Геном — это совокупность всей наследственной информации, записанной в нуклеиновых кислотах организма.

У большинства организмов носителем генома является ДНК, у некоторых вирусов — РНК.

Определение не вполне чёткое. Например, когда говорят «геном человека», обычно имеют в виду последовательности 22 аутосом, двух половых хромосом и одной митохондриальной хромосомы, итого 25 последовательностей.

Но если речь не об некотором «обобщённом» человеке, а о конкретном В.И.Пупкине 1987 г.р., то имеют в виду его индивидуальные особенности, а это, строго говоря, 46 + 1 последовательность (гомологичные хромосомы немного различаются). Иногда говорят о «геноме раковой опухоли», а там может быть очень много вариантов. И т.д.

Но обычно в каждом конкретном случае интуитивно понятно, что имеется в виду.

Геном бактерии в банке данных

```
ID    CP002302; SV 1; circular; genomic DNA; STD; PR0; 641716 BP.
XX
AC    CP002302;
XX
PR    Project:PRJNA43509;
XX
DT    05-NOV-2010 (Rel. 106, Created)
DT    16-MAY-2014 (Rel. 120, Last updated, Version 4)
XX
DE    Buchnera aphidicola str. JF99 (Acyrtosiphon pisum), complete genome.
XX
KW    .
XX
OS    Buchnera aphidicola str. JF99 (Acyrtosiphon pisum)
OC    Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacterales;
OC    Erwiniaceae; Buchnera.
...
XX
SQ    Sequence 641716 BP; 237982 A; 83872 C; 84812 G; 235050 T; 0 other;
      actactttatc cacagatttg ttctttacta ataataatag taattattat tttttatttt      60
      tttatttttt tgaatttaaa tcttaaagaa aagaaaaaga tctttttttt aagatattat      120
      gtttttaaga ttaacatgtg ttatcttgaa taaaatatta atactattta aatattttta      180
      aatttttaaa aggtttttat atgtttaatt taagaaattt tgatgtaatc gttgttgagg      240
      caggacatgc tggactgaa gcagctatgg cttcatcaag aatgggttgt aaaacattat      300
      tgctcactca aaaaatttca gatttaggtg ctttatcatg taatccagct attggaggga      360
...
      tcaccatgaa tgcaataagc gtctttaata ttacaaatat tcagtttaat taactctaaa      641640
      ataataacat ataaaaaagt tatccgtatg aaaatcacta attgtacaca gttaattaat      641700
      agaataagtt ctttta                                641716
```

//

Геном — это текст, некоторая порция информации

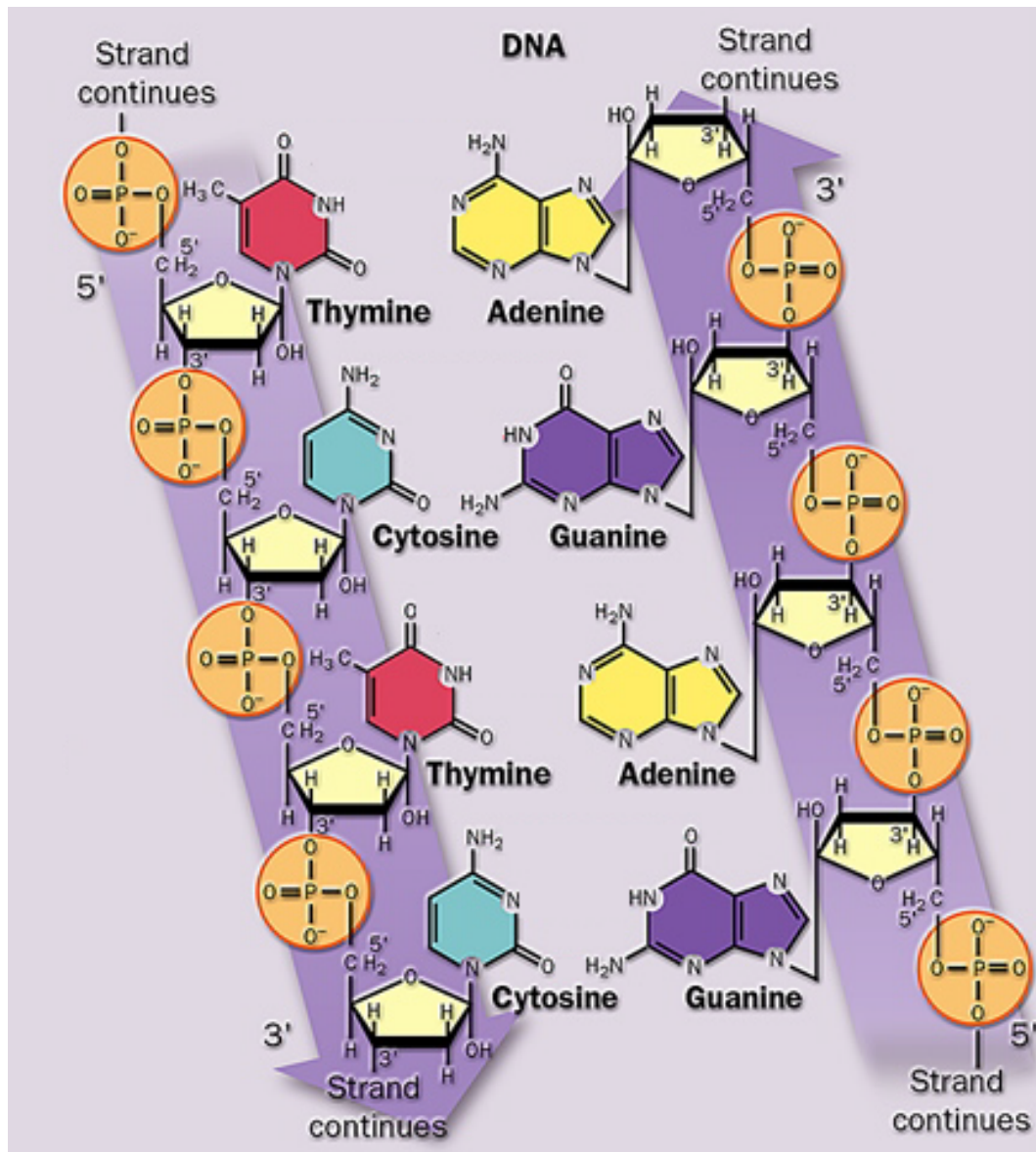
Носителем генома в организме являются молекулы ДНК или РНК.

Носителем генома для биоинформатика является текстовый файл.

Текстовые файлы хранятся в банках данных. Например, в БД Genomes на сайте NCBI

Это одна и та же информация, т.к. последовательность оснований ДНК однозначно определяет химическую формулу одной молекулы ДНК, а последовательность одной цепочки однозначно определяет двухцепочечную ДНК

ДНК — носитель генома

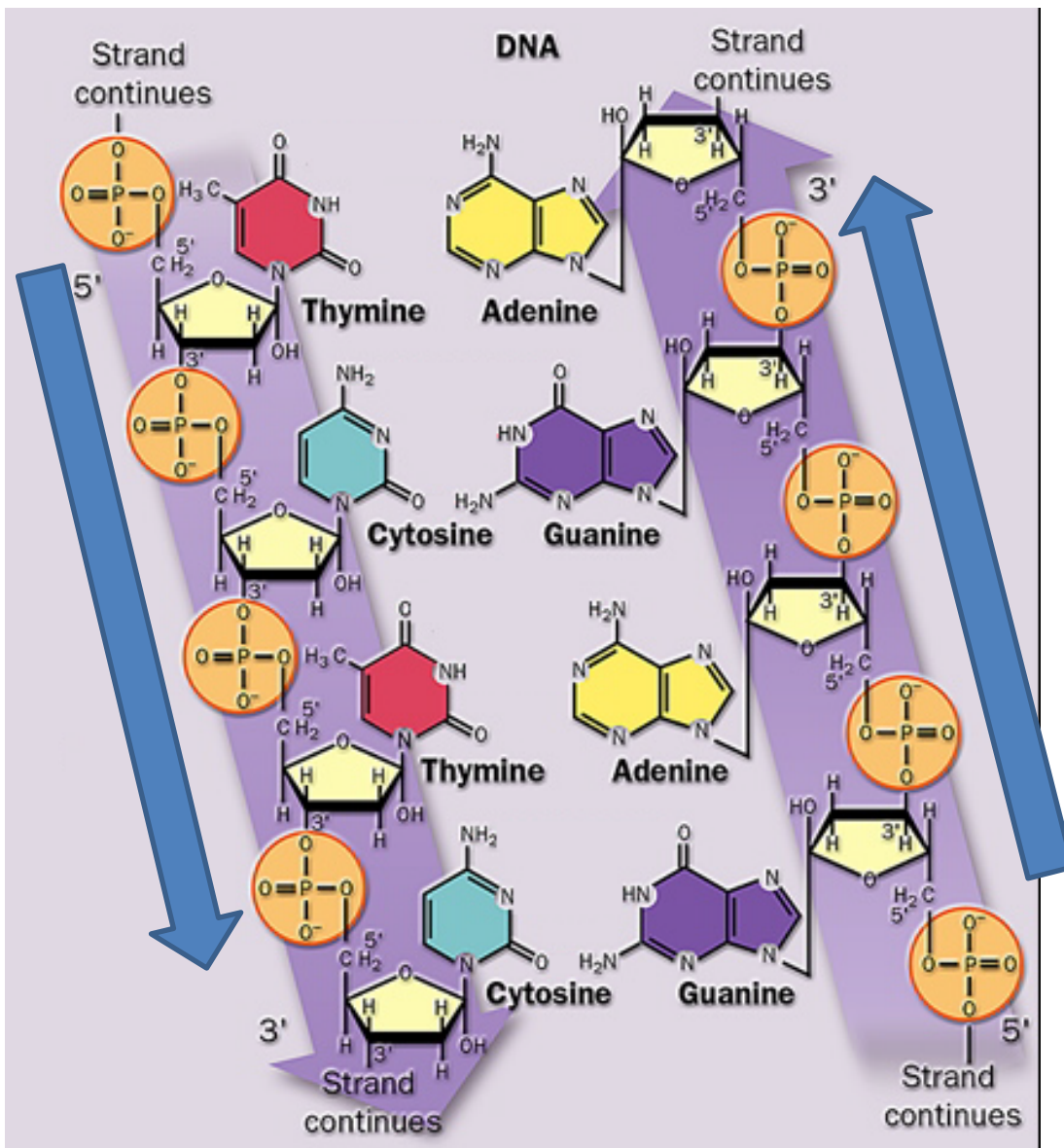


Последовательность этого участка ДНК можно записать **двумя** способами.

Две цепи ДНК равноправны, можно выбрать любую.

Получатся два разных текста (разные последовательности букв), но они отвечают **одной и той же** двухцепочечной ДНК.

ДНК — носитель генома



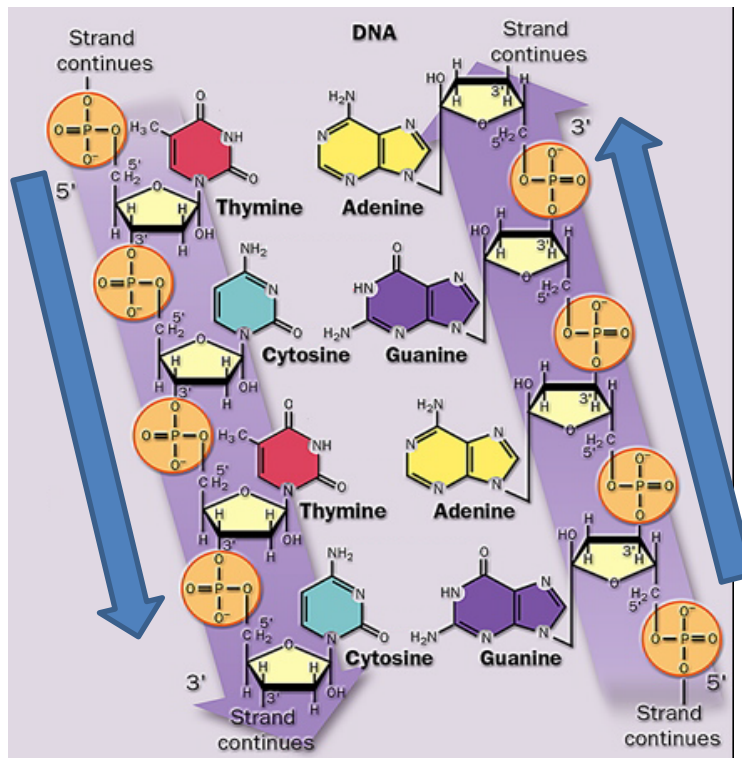
Последовательность этого участка ДНК можно записать **двумя** способами.

Две цепи ДНК равноправны, можно выбрать любую. Получатся два разных текста (разные последовательности букв), но они отвечают **одной и той же** двухцепочечной ДНК.

Обратите внимание на синие стрелки: цепи ДНК ориентированы противоположно!

Правила записи последовательности ДНК в файл

- Записывается одна из двух цепей
- Последовательность записывается в направлении от 5' к 3'-концу



**TCTC или GAGA
(но не AGAG!)**

Произвол в записи генома

Когда лаборатория завершает работы (секвенирование и сборку) по чтению генома, его нужно положить в банк. При этом для каждой прочитанной молекулы ДНК нужно выбрать, какую из двух цепей записывать в файл.

Этот выбор произволен!!!

И это не всё. У большинства бактерий и архей носитель генома — **кольцевые** молекулы ДНК. Текст в компьютере всегда линейный. Значит, надо выбрать не только одну из цепей, но и где её начать. Этот выбор тоже произволен.

(ну почти... Для бактерий есть некоторый «лучший» вариант, но на практике далеко не всегда выбирают именно его).

Гены на «прямой» и «обратной» цепи

(в записи банка данных)

```

FT   gene           221499..221873 
FT               /locus_tag="CWS_01060"
FT   CDS           221499..221873
FT               /codon_start=1
FT               /transl_table=11
FT               /locus_tag="CWS_01060"
FT               /product="mutator MutT protein"
FT               /note="COG0494 NTP pyrophosphohydrolases including
FT               oxidative damage repair enzymes"
FT               /db_xref="EnsemblGenomes-Gn:CWS_01060"
FT               /db_xref="EnsemblGenomes-Tr:ADP67177"
FT               /protein_id="ADP67177.1"
FT               /translation="MELLSKKKVYITRGKYKKNIWEFPGGKVKKHENIVHALKRELLEE
FT               VGIIVLKINFFQYIEYIYPEKKIKLYFFLIKKWKGRPYSIEGYTYLWKRLCHLRALDFP
FT               LANHSVINALKKNNILIKFR"
FT   gene           complement(221856..222509) 
FT               /locus_tag="CWS_01065"
FT   CDS           complement(221856..222509)
FT               /codon_start=1
FT               /transl_table=11
FT               /locus_tag="CWS_01065"
FT               /product="dephospho-CoA kinase"
FT               /note="COG0237 Dephospho-CoA kinase"
FT               /db_xref="EnsemblGenomes-Gn:CWS_01065"
FT               /db_xref="EnsemblGenomes-Tr:ADP67178"
FT               /protein_id="ADP67178.1"
FT               /translation="MTYIVALTTGGISSGKTTISNGFKKIGINVIDTDIIAKNIIIEKNLQ
FT               VSFSIKRKFGKKILNIDNSINRLLLRQYVFNNHHHRLWLENLLHPKIYQESKHQIKMTQ
FT               SNWCLWVVP LLVEKKLEKKAHRILLIDTPVKEQIKRTVRRDKISFLEAKKIIALQSSRK
FT               TRISLSDDTIIFNKKNFKKINLYIYYFNLLYSHLSRTYNNKNTINIKKNYLTKEY"

```

Таблица особенностей генома

(маленький фрагмент)

#	feature	class	start	end	strand	product_accession	name
gene	protein_coding		1	1524	+		dnaA
CDS	with_protein		1	1524	+	NP_214515.1	chromosomal replication initiator protein DnaA
gene	protein_coding		2052	3260	+		dnaN
CDS	with_protein		2052	3260	+	NP_214516.1	DNA polymerase III subunit beta
gene	protein_coding		3280	4437	+		recF
CDS	with_protein		3280	4437	+	NP_214517.1	DNA replication/repair protein RecF
gene	protein_coding		4434	4997	+		
CDS	with_protein		4434	4997	+	NP_214518.1	hypothetical protein
gene	protein_coding		5240	7267	+		gyrB
CDS	with_protein		5240	7267	+	NP_214519.2	DNA gyrase subunit B
gene	protein_coding		7302	9818	+		gyrA
CDS	with_protein		7302	9818	+	NP_214520.1	DNA gyrase subunit A
gene	protein_coding		9914	10828	+		
CDS	with_protein		9914	10828	+	NP_214521.1	membrane protein
gene	tRNA		10887	10960	+		ileT
tRNA			10887	10960	+		tRNA-Ile
gene	tRNA		11112	11184	+		alaT
tRNA			11112	11184	+		tRNA-Ala
gene	protein_coding		11874	12311	-		
CDS	with_protein		11874	12311	-	NP_214522.1	cell wall synthesis protein CwsA
gene	protein_coding		12468	13016	+		ppiA
CDS	with_protein		12468	13016	+	NP_214523.1	iron-regulated peptidyl-prolyl cis-trans isomerase PpiA
gene	protein_coding		13133	13558	-		
CDS	with_protein		13133	13558	-	NP_214524.1	membrane protein

Геном в формате Fasta

>CP002302.1 *Buchnera aphidicola* str. JF99 (*Acyrtosiphon pisum*), complete genome.

```
ACTACTTATCCACAGATTTGTTCTTTACTAATAATAATAGTAATTATTATTTTTTATTTT
TTTATTTTTTTTGAATTTAAATCTTAAAGAAAAGAAAAAGATCTTTTTTTTAAAGATATTAT
GTTTTTAAGATTAACATGTGTTATCTTGAATAAAATATTAATACTATTTAAATATTTTTTA
AATTTTTTAAAGGTTTTTATATGTTTAATTTAAGAAATTTTGATGTAATCGTTGTTGGAG
CAGGACATGCTGGTACTGAAGCAGCTATGGCTTCATCAAGAATGGGTTGTAAAACATTAT
TGCTCACTCAAAAAATTTTCAAGATTTAGGTGCTTTATCATGTAATCCAGCTATTGGAGGGA
TAGGAAAAAGTCATTTAGTAAAAGAAATAGATGCATTAGGTGGTATGATGGCTAAAGCTA
TTGATTACTCTGGTATTCAGTTTAGAATTCTGAATTCTAGCAAAGGTCCTGCTGTAAGAT
CTACTAGAGCTCAAGCCGATAAAATCCTGTATCATGAGACAGTAAAAAAAATATTAAAAA
AGCAAAATAATTTGTTAATTTTAGAAGCAGAAGTTAAGGATTTAATTTTTTAAAAATTACT
CTGTTGTTGGTGTGTTAACACAAAATGAAATTAATTTCTACTCTCGTTCAGTTGTCTTAG
CAGCAGGCACTTTTTTAGGAGGTAAAATACATATAGGGTTAAAAGTTATTCTGCTGGTA
GAATAGGAGATAAGTCTGCTATAGATTTATCTGTTTCGTTTAAAGAGAATTGTCTTTGCGAG
TTAATAGATTAAAAACAGGAACGCCGCTCGAATTGATATCAACACTGTTAATTTTAACA
ATTTATTAATTCAAATAGCGATACTCCTGTTCCAGTTTTTTTCATTTATGGGAAACGTTT
CACATCATCCTAAACAGATACCATGTTACCTCACACATACTAATGAAAAAACACATGAAA
TAATACGTAAAAATTTAGATAAAAGTCCAATATATACGGGTTTTTTTAAAAGGTTTAGGAC
CTCGATACTGTCCTTCTATTGAAGATAAAATTGTACGTTTTTCCAGATAGAAAATCACATC
AAGTATTTTTTAGAACCTGAAGGATTATCTAGTATTAAAGTATATCCTAATGGTATTTTCGA
CCAGTCTTCCAATAGAAGTTCAAGAACAAATAGTAGCGTCTATAAAAGGTTTAGAAAAAT
CCAAAATTATCAGACCTGGATATGCCATTGAATATGATTTTTTTTGATCCAAAAGACTTAA
ATTTAACTTTAGAAAGCAAATTAATTAAGGGTTATTTTTTTGCCGGGCAAATTAACGGAA
CTACTGGTTATGAAGAAGCGGCTTCACAAGGATTGTTGGCAGGATTAAATGCAGCATTAA
GTTCAAAGAATACTGAAGGATGGTTTCCTAGACGTGATCAAGCATATTTGGGTGTTTTAA
TAGATGATTTAACTACTCAAGGCACAGAGGAACCATACCGCATGTTTACTTCACGTGCCG
AATATCGACTTTCTTTAAGAGAAGATAATGCAGATTTACGTTTAACTGAAATTGGTCGTA
AACTTGGTTTAGTAAATGATTCAAGATGGATTGTTATAATCAAAAAGTATTGAATATTC
AAACAGAAATGAATCGTTTAAAAAAAATAAAATAAGTCCTATATCACCAGATGCTGATA
TTTTAAAAAAATTATATAATATTAATTTAATTAAGAAATAAGTATGTCTGAGTTATTAA
```