

Q1.

Berapa panjang (jumlah total pasangan basa dalam) genom mitokondria *Schistosoma mansoni* akses NCBI NC_002545), dan berapa banyak As, Cs, Gs, dan Ts yang dikandungnya?

Anda harus mencari urutan ini melalui situs web NCBI, karena tidak ada dalam database ACNUC. Catatan: *Schistosoma mansoni* adalah cacing parasit yang bertanggung jawab menyebabkan schistosomiasis, yang diklasifikasikan oleh WHO sebagai penyakit tropis terabaikan.

Jawaban:

```
library(seqinr)

schistosoma <- read.fasta(file = "D:/dhil/Tugas/Bioinformatika/Bioinformatics/Tugas3/schistosoma.fasta")

schistosomaseq <- schistosoma[[1]]
length = length(schistosomaseq)
cat("Jumlah pasangan basa genom mitokondria Schistosoma mansoni = ",length,"\n")
print(count(schistosomaseq,1))
```

Output

```
> library(seqinr)
> schistosoma <- read.fasta(file = "D:/dhil/Tugas/Bioinformatika/Bioinformatics/Tugas3/schistosoma.fasta")
> schistosomaseq <- schistosoma[[1]]
> length = length(schistosomaseq)
> cat("Jumlah pasangan basa genom mitokondria Schistosoma mansoni = ",length,"\n")
Jumlah pasangan basa genom mitokondria Schistosoma mansoni = 14415
> print(count(schistosomaseq,1))

  a   c   g   t
3654 1228 3307 6226
```

Jumlah pasangan basa genom mitokondria *Schistosoma* sepanjang **14415** dengan kandungan

As : 3654, Cs : 1228, Gs : 3307, dan Ts : 6226

Q2.

Berapa panjang genom mitokondria *Brugia malayi* (aksesi NCBI NC_004298), dan berapa banyak As, Cs, Gs, dan Ts yang dikandungnya?

Anda harus mencari urutan ini melalui situs web NCBI, karena tidak ada dalam *database* ACNUC. Catatan: *Brugia malayi* adalah cacing parasit yang bertanggung jawab untuk menyebabkan filariasis limfatik, yang diklasifikasikan oleh WHO sebagai penyakit tropis terabaikan.

Jawaban:

```
library(seqinr)

brugia_malayi <- read.fasta(file = "D:/dhil/Tugas/Bioinformatika/Bioinformatics/Tugas3/malayi.fasta")

malayiseq <- brugia_malayi[[1]]

malayi_length = length(malayiseq)
cat("Jumlah pasangan basa genom mitokondria Brugia malayi = ",malayi_length,"\n")
print(count(malayiseq,1))
```

Output

```
> library(seqinr)
> brugia_malayi <- read.fasta(file = "D:/dhil/Tugas/Bioinformatika/Bioinformatics/Tugas3/malayi.fasta")
> malayiseq <- brugia_malayi[[1]]
> malayi_length = length(malayiseq)
> cat("Jumlah pasangan basa genom mitokondria Brugia malayi = ",malayi_length,"\n")
Jumlah pasangan basa genom mitokondria Brugia malayi = 13657
> print(count(malayiseq,1))

  a  c  g  t
2950 1054 2297 7356
```

Jumlah pasangan basa genom mitokondria *Brugia malayi* sepanjang **13657** dengan kandungan

As : 2950, Cs : 1054, Gs : 2297, dan Ts : 7356.

Q3.

Berapa probabilitas urutan genom mitokondria *Brugia malayi* (NCBI NC_004298), sesuai dengan model multinomial di mana probabilitas As, Cs, Gs dan Ts (pA, pC, pG, dan pT) disetel sama dengan fraksi As, Cs, Gs, dan Ts dalam genom mitokondria *Schistosoma mansoni*?

Jawaban:

```
brugia_malayi <- read.fasta(file = "D:/dhil/Tugas/Bioinformatika/Bioinformatics/Tugas3/malayi.fasta")
malayiseq <- brugia_malayi[[1]]
length(malayiseq)

table(malayiseq)
tablesch <- table(schistosomaseq)
tablesch
table1 <- tablesch/length(schistosomaseq)

frequencyA <- table1[["a"]]
frequencyC <- table1[["c"]]
frequencyG <- table1[["g"]]
frequencyT <- table1[["t"]]
probability <- c(frequencyA,frequencyC,frequencyG,frequencyT)
probability
multinomialprob <- function(temp_sequence, probability)
{
  nucleotides <- c("A", "C", "G", "T")
  names(probability) <- nucleotides
  temp_sequence <- toupper(temp_sequence)
  seqlength <- length(temp_sequence)
  seqprob <- numeric()
```

Jawaban Lanjutan

```
for (i in 1:seqlength)
{
  nucleotide <- temp_sequence[i]
  nucleotideprob <- probability[nucleotide]
  if (i == 1) { seqprob <- nucleotideprob[[1]] }
  else      { seqprob <- seqprob * nucleotideprob[[1]] }
}
return(seqprob)
}
multinomialprob(malayiseq, probability)
```

Output

```
> brugia_malayi <- read.fasta(file = "D:/dhil/Tugas/Bioinformatika/Bioinformatics/Tugas3/malayi.fasta")
> malayiseq <- brugia_malayi[[1]]
> length(malayiseq)
[1] 13657
> table(malayiseq)
malayiseq
  a  c  g  t
2950 1054 2297 7356
> tablesch <- table(schistosomaseq)
> tablesch
schistosomaseq
  a  c  g  t
3654 1228 3307 6226
> table1 <- tablesch/length(schistosomaseq)
> frequencyA <- table1[["a"]]
> frequencyC <- table1[["c"]]
> frequencyG <- table1[["g"]]
> frequencyT <- table1[["t"]]
> probability <- c(frequencyA,frequencyC,frequencyG,frequencyT)
> probability
[1] 0.25348595 0.08518904 0.22941381 0.43191120
```

```
> multinomialprob <- function(temp_sequence, probability)
+ {
+   nucleotides <- c("A", "C", "G", "T")
+   names(probability) <- nucleotides
+   temp_sequence <- toupper(temp_sequence)
+   seqlength <- length(temp_sequence)
+   seqprob <- numeric()
+   for (i in 1:seqlength)
+   {
+     nucleotide <- temp_sequence[i]
+     nucleotideprob <- probability[nucleotide]
+     if (i == 1) { seqprob <- nucleotideprob[[1]] }
+     else { seqprob <- seqprob * nucleotideprob[[1]] }
+   }
+   return(seqprob)
+ }
> multinomialprob(malayiseq, probability)
```

Berdasarkan output di atas, probabilitas urutan genom mitokondria *Brugia malayi* dalam genom mitokondria *Schistosoma mansoni* adalah **0**.

Q4.

Apa tiga kata 4-bp paling sering (4-mers) dalam genom bakteri *Chlamydia trachomatis* strain D / UW-3 / CX (aksesi NCBI NC_000117), dan berapa kali mereka muncul dalam urutannya?

Catatan: *Chlamydia trachomatis* adalah bakteri yang bertanggung jawab untuk trachoma, yang diklasifikasikan oleh WHO sebagai penyakit tropis terabaikan.

Jawaban:

```
library(seqinr)
chlamydia <- read.fasta(file = "D:/dhil/Tugas/Bioinformatika/Bioinformatics/Tugas3/chlamydia.fasta")
chlamydiaseq <- chlamydia[[1]]
table(chlamydiaseq)

table_count4 <- count(chlamydiaseq,4)
table_count4
cat("urutan dari yang terkecil = \n")
sort(table_count4)
```

Output

```
> library(seqinr)
> chlamydia <- read.fasta(file = "D:/dhil/Tugas/Bioinformatika/Bioinformatics/Tugas3/chlamydia.fasta")
> chlamydiaseq <- chlamydia[[1]]
> table(chlamydiaseq)
chlamydiaseq
  a   c   g   t
306721 215232 215404 305162
```

```
> table_count4 <- count(chl_amiaseq,4)
> table_count4
```

aaaa	aaac	aaag	aaat	aaca	aacc	aacg	aact	aaga	aagc	aagg	aagt	aata	aatc	aatg	aatt	acaa	acac
14122	6340	9099	8144	6197	3423	3368	4747	10069	6036	5167	4040	6780	5881	4687	6244	5866	2453
acag	acat	acca	accc	accg	acct	acga	acgc	acgg	acgt	acta	actc	actg	actt	agaa	agac	agag	agat
4379	3599	3031	1837	1545	2987	3147	2111	1716	2048	3079	3609	2530	4165	10581	3225	6836	6458
agca	agcc	agcg	agct	agga	aggc	aggg	aggt	agta	agtc	agtg	agtt	ataa	atac	atag	atat	atca	atcc
5679	3126	2890	6860	6516	2861	3372	2973	3420	2755	2428	4591	5944	3680	4967	4914	5153	5136
atcg	atct	atga	atgc	atgg	atgt	atta	attc	attg	attt	caaa	caac	caag	caat	caca	cacc	cacg	cact
3570	6370	4782	3948	3905	3457	5237	5799	4741	8100	7289	4058	4820	4710	3111	1823	1520	2534
caga	cagc	cagg	cagt	cata	catc	catg	catt	ccaa	ccac	ccag	ccat	ccca	cccc	cccg	ccct	ccga	ccgc
5223	4035	3269	2486	3919	4246	3134	4756	4413	2036	2890	3800	2898	1976	1215	3245	2115	1678
ccgg	ccgt	ccta	cctc	cctg	cctt	cgaa	cgac	cgag	cgat	cgca	cgcc	cgcg	cgct	cgga	cggc	cggg	cggt
1180	1802	3436	3691	3209	5135	3484	1606	2557	3617	2701	1407	1287	2882	2580	1321	1198	1481
cgta	cgtc	cgtg	cgtt	ctaa	ctac	ctag	ctat	ctca	ctcc	ctcg	ctct	ctga	ctgc	ctgg	ctgt	ctta	cttc
2544	1786	1541	3234	4925	3107	3009	5205	3607	4903	2607	6937	3589	3992	2913	4301	4078	7234
cttg	cttt	gaaa	gaac	gaag	gaat	gaca	gacc	gacg	gact	gaga	gagc	gagg	gagt	gata	gatc	gatg	gatt
4748	9199	8563	3976	7353	5715	2799	1567	1750	2755	6319	4326	3776	3587	5109	4862	4326	5931
gcaa	gcac	gcag	gcat	gcca	gccc	gccg	gcct	gcga	gcgc	gcgg	gcgt	gcta	gctc	gctg	gctt	ggaa	ggac
5218	1931	4167	3914	2437	1512	1334	2852	2589	1849	1658	2110	4093	4374	4024	5899	6169	1891
ggag	ggat	ggca	ggcc	ggcg	ggct	ggga	gggc	gggg	gggt	ggta	ggtc	ggtg	ggtt	gtaa	gtac	gtag	gtat
5070	5105	2418	1206	1435	3122	4118	1537	2100	1833	2466	1558	1966	3191	3862	1802	3316	3641
gtca	gtcc	gtcg	gtct	gtga	gtgc	gtgg	gtgt	gtta	gttc	gttg	gttt	taaa	taac	taag	taat	taca	tacc
2208	1805	1647	3311	2574	1959	1952	2475	3371	3701	3805	6114	7731	3361	4040	5023	4190	2587
tacg	tact	taga	tagc	tagg	tagt	tata	tatc	tatg	tatt	tcaa	tcac	tcag	tcat	tcca	tccc	tccg	tcct
2384	3347	5489	4158	3510	3081	3697	5240	3945	6946	5380	2568	3577	4742	4773	4009	2681	6387
tcga	tcgc	tcgg	tcgt	tcta	tctc	tctg	tctt	tgaa	tgac	tgag	tgat	tgca	tgcc	tgcg	tgct	tgga	tggc
3413	2639	2026	3145	5638	6380	5032	10060	5373	2149	3545	5048	4432	2396	2593	5526	5021	2462
tggg	tgggt	tgta	tgtc	tgtg	tgtt	ttaa	ttac	ttag	ttat	ttca	ttcc	ttcg	ttct	ttga	ttgc	ttgg	ttgt
2918	2894	4192	2872	3025	5975	5424	3919	4946	6068	5299	6006	3399	10492	5170	5048	4525	5831
ttta	tttc	tttg	tttt														
7671	8462	7280	14021														


```
> cat("urutan dari yang terkecil = \n")
urutan dari yang terkecil =
> sort(table_count4)
```

```
ccgg  cggg  ggcc  cccg  cgcg  cggc  gccg  cgcc  ggcg  cggg  gccc  cacg  gggc  cgtg  accg  gggt  gacc  cgac
1180  1198  1206  1215  1287  1321  1334  1407  1435  1481  1512  1520  1537  1541  1545  1558  1567  1606
gtcg  gcgg  ccgc  acgg  gacg  cgtc  ccgt  gtac  gtcc  cacc  gggg  accc  gcgc  ggac  gcac  gtgg  gtgc  ggtg
1647  1658  1678  1716  1750  1786  1802  1802  1805  1823  1833  1837  1849  1891  1931  1952  1959  1966
cccc  tcgg  ccac  acgt  gggg  gcgt  acgc  ccga  tgac  gtca  tacg  tgcc  ggca  agtg  gcca  acac  tggc  ggta
1976  2026  2036  2048  2100  2110  2111  2115  2149  2208  2384  2396  2418  2428  2437  2453  2462  2466
gtgt  cagt  actg  cact  cgta  cgag  tcac  gtga  cgga  tacc  gcga  tgcg  ctcg  tcgc  tccg  cgca  agtc  gact
2475  2486  2530  2534  2544  2557  2568  2574  2580  2587  2589  2593  2607  2639  2681  2701  2755  2755
gaca  gcct  aggc  tgct  cgct  agcg  ccag  tggg  ccca  ctgg  tggg  aggt  acct  ctag  tgtg  acca  acta  tagt
2799  2852  2861  2872  2882  2890  2890  2894  2898  2913  2918  2973  2987  3009  3025  3031  3079  3081
ctac  caca  ggct  agcc  catg  tcgt  acga  gggt  cctg  agac  cggt  ccct  cagg  gtct  gtag  tact  taac  aacg
3107  3111  3122  3126  3134  3145  3147  3191  3209  3225  3234  3245  3269  3311  3316  3347  3361  3368
gtta  aggg  ttcg  tcga  agta  aacc  ccta  atgt  cgaa  tagg  tgag  atcg  tcag  gagt  ctga  acat  ctca  actc
3371  3372  3399  3413  3420  3423  3436  3457  3484  3510  3545  3570  3577  3587  3589  3599  3607  3609
cgat  gtat  atac  cctc  tata  gttc  gagg  ccat  gttg  gtaa  atgg  gcat  cata  ttac  tatg  atgc  gaac  ctgc
3617  3641  3680  3691  3697  3701  3776  3800  3805  3862  3905  3914  3919  3919  3945  3948  3976  3992
tccc  gctg  cagc  aagt  taag  caac  ctta  gcta  ggga  tagc  actt  gcag  taca  tgta  catc  ctgt  gagc  gatg
4009  4024  4035  4040  4040  4058  4078  4093  4118  4158  4165  4167  4190  4192  4246  4301  4326  4326
gttc  acag  ccaa  tgca  ttgg  agtt  aatg  caat  attg  tcat  aact  cttg  catt  tcca  atga  caag  gatc  ctcc
4374  4379  4413  4432  4525  4591  4687  4710  4741  4742  4747  4748  4756  4773  4782  4820  4862  4903
atat  ctaa  ttag  atag  tgga  taat  tctg  tgat  ttgc  ggag  ggat  gata  cctt  atcc  atca  aagg  ttga  ctat
4914  4925  4946  4967  5021  5023  5032  5048  5048  5070  5105  5109  5135  5136  5153  5167  5170  5205
gcaa  caga  atta  tatc  ttca  tgaa  tcaa  ttaa  taga  tgct  tcta  agca  gaat  attc  ttgt  acaa  aatc  gctt
5218  5223  5237  5240  5299  5373  5380  5424  5489  5526  5638  5679  5715  5799  5831  5866  5881  5899
gatt  ataa  tggt  ttcc  aagc  ttat  gttt  ggaa  aaca  aatt  gaga  aaac  atct  tctc  tcct  agat  agga  aata
5931  5944  5975  6006  6036  6068  6114  6169  6197  6244  6319  6340  6370  6380  6387  6458  6516  6780
agag  agct  ctct  tatt  cttc  ttgt  caaa  gaag  ttaa  taaa  attt  aaat  ttct  gaaa  aaag  cttt  tctt  aaga
6836  6860  6937  6946  7234  7280  7289  7353  7671  7731  8100  8144  8462  8563  9099  9199  10060  10069
ttct  agaa  tttt  aaaa
10492 10581 14021 14122
```

Berdasarkan tabel di atas, tiga kata 4-bp paling sering muncul (4-mers) dalam genom bakteri *Chlamydia trachomatis* adalah **AAAA** sebanyak **14122**, **TTTT** sebanyak **14021**, dan **AGAA** yaitu sebanyak 10581

Q5.

Tulis fungsi R untuk menghasilkan urutan DNA acak yang panjangnya n huruf (yaitu, panjang n basis) menggunakan model multinomial di mana probabilitas pA, pC, pG, dan pT ditetapkan sama dengan fraksi As, Cs, Gs dan Ts dalam genom mitokondria *Schistosoma mansoni* (di sini pA berarti probabilitas As, pC adalah probabilitas Cs, dll.)

Petunjuk: lihat halaman bantuan untuk fungsi "sample ()" dalam R, karena mungkin berguna untuk digunakan dalam fungsi R Anda.

Jawaban:

```
library(seqinr)
schistosomaseq <- read.fasta(file = "D:/dhil/Tugas/Bioinformatika/Bioinformatics/Tugas3/schistosoma.fasta")[[1]]

#probabilitas mschitosoma
schistosoma_length = length(schistosomaseq)

pA = table(schistosomaseq)['a']/schistosoma_length
pC = table(schistosomaseq)['c']/schistosoma_length
pG = table(schistosomaseq)['g']/schistosoma_length
pT = table(schistosomaseq)['t']/schistosoma_length
probability = c(pA,pC,pG,pT)

#fungsi
randomSeq = function(n, p){
  sample(c("A", "C", "G", "T"),n,rep=TRUE,prob=p)
}

#sampel
print(randomSeq(10,probability))
print(randomSeq(10,probability))
print(randomSeq(10,probability))
```

Output

```
> library(seqinr)
> schistosomaseq <- read.fasta(file = "D:/dhil/Tugas/Bioinformatika/Bioinformatics/Tugas3/schistosoma.fasta")[[1]]
> #probabilitas mschitosoma
> schistosoma_length = length(schistosomaseq)
> pA = table(schistosomaseq)['a']/schistosoma_length
> pC = table(schistosomaseq)['c']/schistosoma_length
> pG = table(schistosomaseq)['g']/schistosoma_length
> pT = table(schistosomaseq)['t']/schistosoma_length
> probability = c(pA,pC,pG,pT)
> #fungsi
> randomSeq = function(n, p){
+   sample(c("A", "C", "G", "T"),n,rep=TRUE,prob=p)
+ }
> #sampel
> print(randomSeq(10,probability))
[1] "T" "T" "T" "G" "T" "T" "T" "G" "T" "T"
> print(randomSeq(10,probability))
[1] "T" "C" "T" "A" "T" "A" "G" "T" "A" "G"
> print(randomSeq(10,probability))
[1] "T" "G" "G" "T" "T" "T" "G" "T" "T" "T"
```

Dengan menggunakan fungsi R tersebut kita dapat menghasilkan urutan DNA acak yang panjangnya sebanyak n huruf sesuai yang kita ingin tampilkan.

Q6.

Berikan contoh penggunaan fungsi Anda dari **Q5** untuk menghitung urutan acak yang panjangnya 20 huruf, menggunakan model multinomial dengan $p_A = 0,28$, $p_C = 0,21$, $p_G = 0,22$, dan $p_T = 0,29$.

Jawaban:

```
library(seqinr)

pA = 0.28
pC = 0.21
pG = 0.22
pT = 0.29
probability <- c(pA,pC,pG,pT)

#contoh
print(randomSeq(20,probability))
print(randomSeq(20,probability))
print(randomSeq(20,probability))
```

Output

```
> library(seqinr)
> pA = 0.28
> pC = 0.21
> pG = 0.22
> pT = 0.29
> probability <- c(pA,pC,pG,pT)
> #contoh
> print(randomSeq(20,probability))
[1] "T" "C" "C" "T" "A" "G" "A" "G" "A" "G" "T" "C" "A" "A" "A" "C" "A" "C" "T" "A"
> print(randomSeq(20,probability))
[1] "A" "C" "T" "C" "A" "G" "A" "A" "A" "A" "C" "T" "T" "T" "A" "T" "A" "G" "A" "T"
> print(randomSeq(20,probability))
[1] "T" "A" "G" "G" "C" "C" "T" "G" "C" "G" "T" "G" "A" "T" "C" "C" "A" "G" "G" "C"
```

Q7.

Berapa banyak urutan protein dari virus rabies yang ada di *database* NCBI Protein?

Anda harus mencari urutan ini melalui situs web NCBI, karena tidak mungkin melakukan pencarian ini menggunakan SeqinR. Catatan: virus rabies adalah virus yang bertanggung jawab atas rabies, yang diklasifikasikan oleh WHO sebagai penyakit tropis terabaikan.

Jawaban:

The screenshot shows the NCBI Protein search interface. The search bar contains 'rabies virus[All Fields]' and the search button is labeled 'Search'. Below the search bar, there are links for 'Create alert' and 'Advanced'. The results section shows a message: 'See the results of this search (10415 items) in our new Identical Protein Groups database.' Below this, it says 'Items: 1 to 20 of 34350'. On the right side, there is a 'Results by taxon' section with a list of top organisms: Rabies lyssavirus (31459), Homo sapiens (375), Eptesicus fuscus (170), and synthetic construct (90). The page also includes navigation links like '<< First', '< Prev', 'Page 1 of 1718', 'Next >', and 'Last >>'. The left sidebar shows 'Species' and 'Source databases' with counts for various categories.

Berdasarkan hasil pencarian di *database* NCBI yang telah difilter khusus protein dengan pencarian *rabies virus*, dapat diketahui ada sebanyak **34350** item virus rabies yang telah diklarifikasi oleh WHO.

Q8.

Apa akses NCBI untuk genom virus Mokola?

Catatan: Virus Mokola dan virus rabies adalah virus terkait erat yang keduanya termasuk dalam kelompok virus yang disebut *Lyssavirus*. Virus mokola menyebabkan infeksi seperti rabies pada mamalia termasuk manusia.

Jawaban:

NCBI Resources How To Sign in to NCBI

Genome Genome mokola virus Search

Create alert Limits Advanced Help

Display Settings: Overview

Send to: ID: 5153

Organism Overview

Mokola lyssavirus

Lineage: Viruses[17266]; Riboviria[3925]; Negarnaviricota[688]; Haploviricota[342]; Monjiviricetes[329]; Mononegavirales[321]; Rhabdoviridae[189]; Lyssavirus[17]; Mokola lyssavirus[1]

Summary

Assembly level: Complete Genome
Assembly: GCA_000856485.1 ViralProj15013 scaffolds: 1 contigs: 1 N50: 0 L50: 0
Statistics: total length (Mb): 0.01194
protein count: 5
GC%: 44.4

Replicon Info

Type	Name	RefSeq	INSDC	Size (Kb)	GC%	Protein	Gene
Chr	Unknown	NC_006429.1	Y09762.1	11.94	44.4	5	5

Tools

Genome Resource

Related information

Assembly
Gene
Other genomes for species
Components
Protein
Taxonomy

Search details

"Mokola lyssavirus"[Organism]

Akses NCBI untuk genom virus Mokola yaitu **NC_006429**.