Мобильные элементы генома

Н.Н. Колесников

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

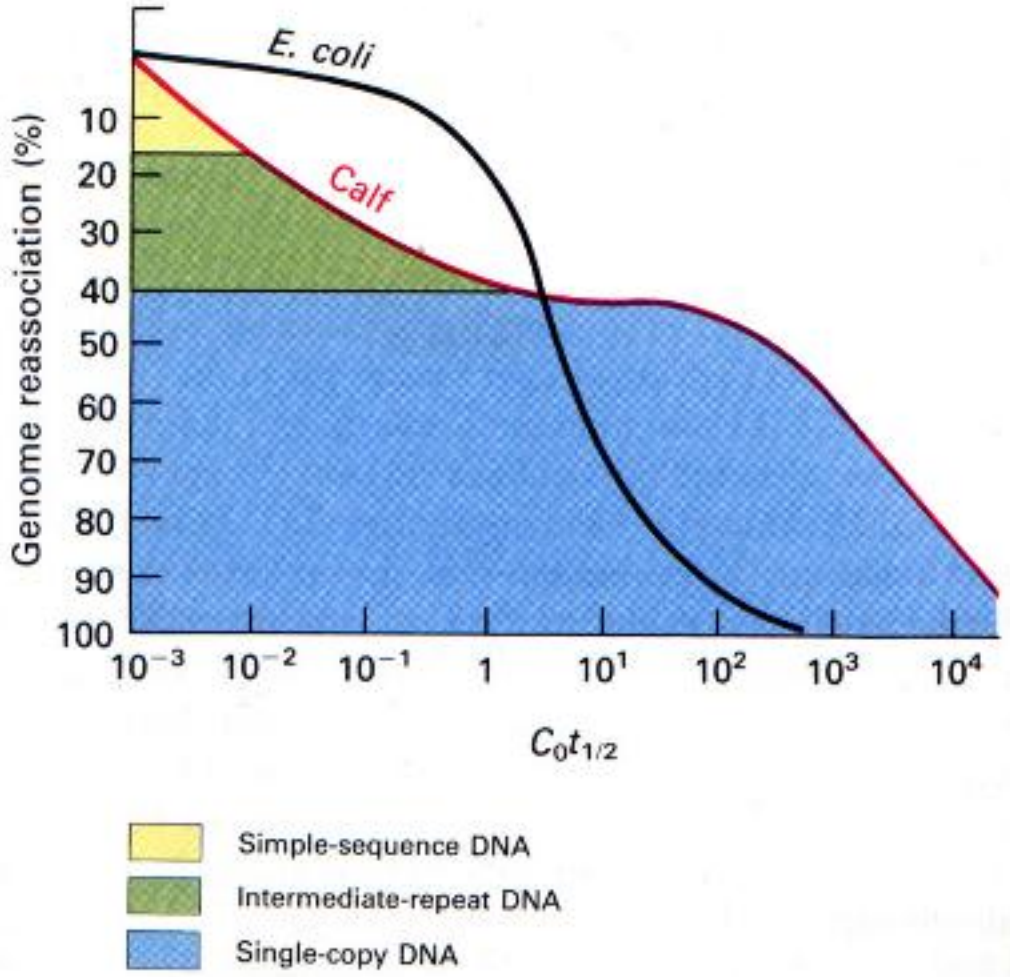
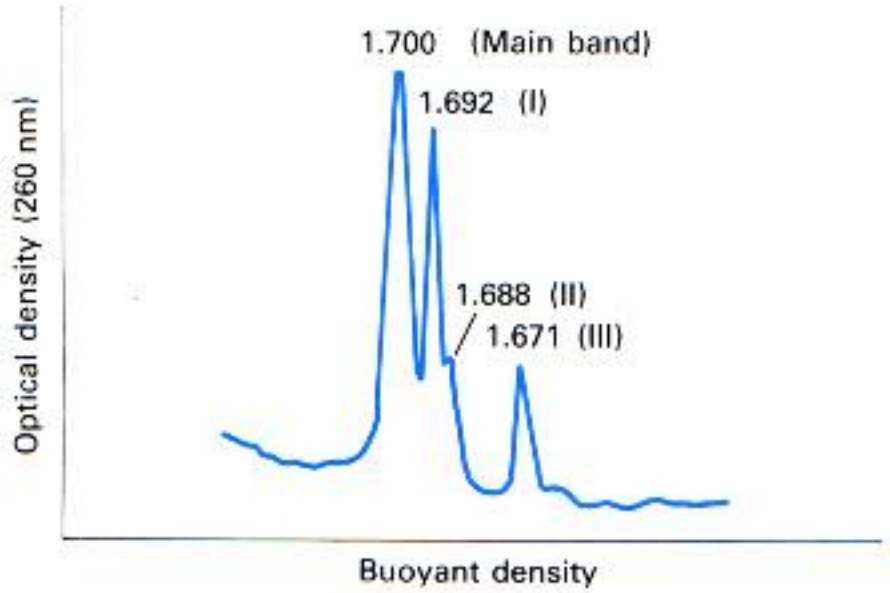

E. coli
Calf
Genome reassociation (%)
10
20
30
40
50
60
70
80
90
100
10^{-3}
10^{-2}
10^{-1}
1
10^{1}
10^{2}
10^{3}
10^{4}
$C_0t_{1/2}$
Simple-sequence DNA
Intermediate-repeat DNA
Single-copy DNA

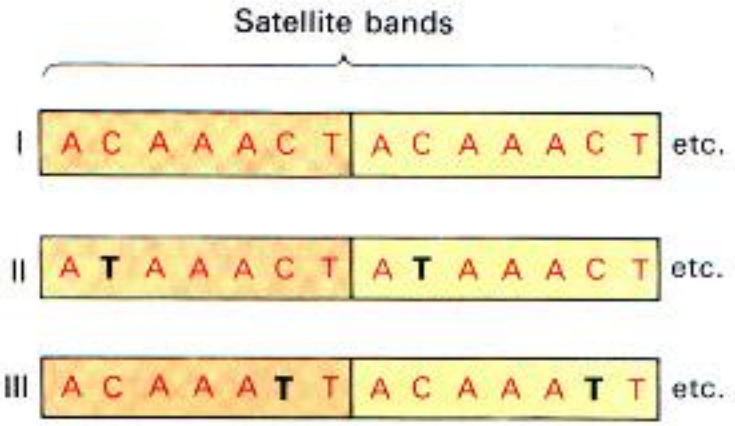
Table 10-6 Examples of simple-sequence DNAs in eukaryotes*

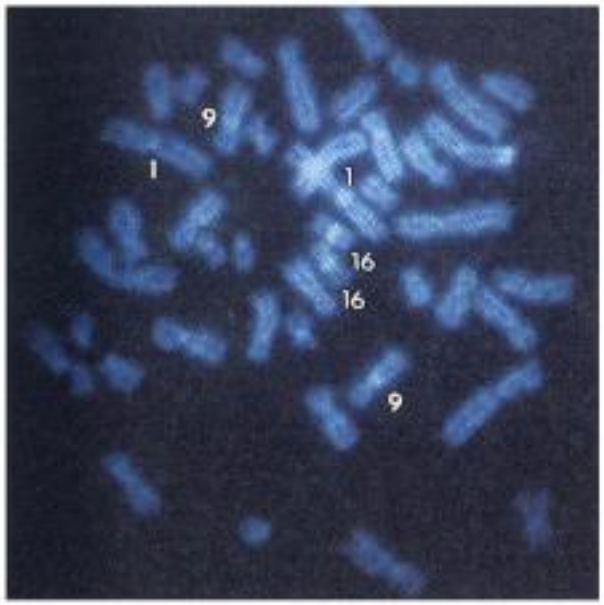
Organism	Base pairs per repeat	Sequence of one repeat unit	Location
Drosophila melanogaster (fruit fly)	5	AGAAG (polypurine) ATAAT	Arms of Y chromosome; centromeric heterochromatin of chromosome 2; long arm of 2, near end
	7	ATATAAT	
	10	AATAACATAG AGAGAAGAAG	Centromeric heterochromatin of all chromosomes; tip of long arm of 2
Drosophila virilis	7	ACAAACT (band I) ATAAACT (band II) ACAAATT (band III)	Centromeric heterochromatin
Cancer borealis (marine crab)	2	AT	?
Pagurus pollicaris (hermit crab)	4	ATCC	?
	3	CTG	?
Cavia porcellus (guinea pig)	6	CCCTAA	Centromeric heterochromatin
Dipodomys ordii (kangaroo rat)	10	ACACAGCGGG	Centromeric heterochromatin
Cercopithecus aethiops (African green monkey; α sequences)*	172	—	Throughout chromosomes
Homo sapiens (human; alphoid sequences)*	171	—	Throughout chromosomes

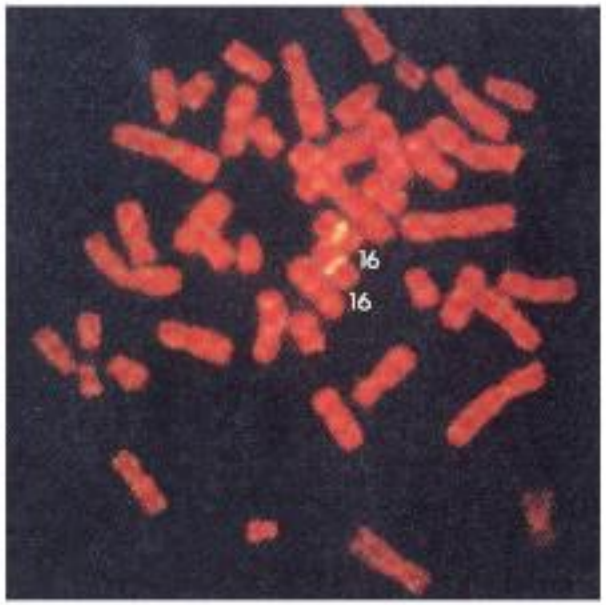
*All eukaryotic species have more than one type of simple-sequence DNA, characterized by the sequence repeat; the table includes only selected examples. Many repeats are 10 bp or less in length, but several longer repeats, such as the primate α and alphoid sequences and human minisatellites are now known. The α and alphoid sequences are not shown because of their length.

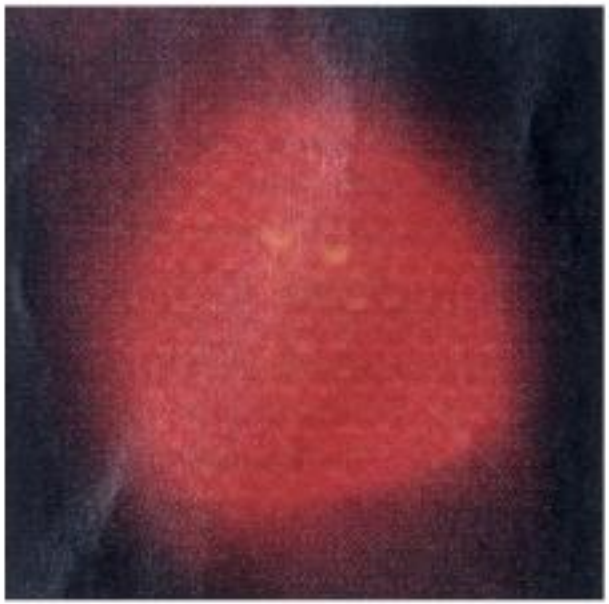
SOURCE: K. Tartoff, 1975, *Ann. Rev. Genet.* 9:355; and A. J. Jeffreys et al., 1985, *Nature* 314:69.


1.700 (Main band)
1.692 (I)
1.688 (II)
1.671 (III)
Optical density (260 nm)
Buoyant density


Satellite bands
I A C A A A C T A C A A A C T etc.
II A T A A A C T A T A A A C T etc.
III A C A A A T T A C A A A T T etc.


9
1
1
16
16
9


16
16



λ 33.1 minisatellite

Consensus sequence: AAGGGTGGGCAGGAAGTGGAGTGTGTGCCTGCTTCCCTTCCCTGTCTTGTCCTGGAAACTCA

Changes in individual repeats	
1–24	(no changes)
25	G
26	G

λ 33.5 minisatellite

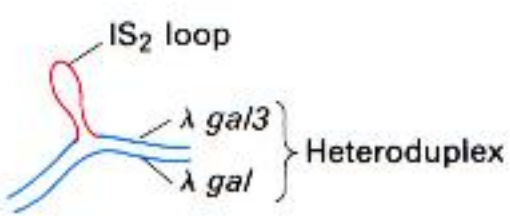
Consensus sequence:

T
C GGGCAGG•AGGGGGAGG

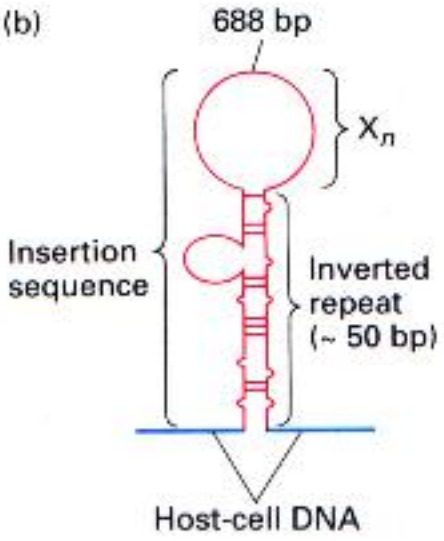
Changes in individual repeats								
1	T			CAG				
2	C			•	A			
3	C	A		•			•	
4	T			•	A			A
5	T			•	A	T		
6	T	A		•		A	G	
7	T			•	A			
8	T			G				

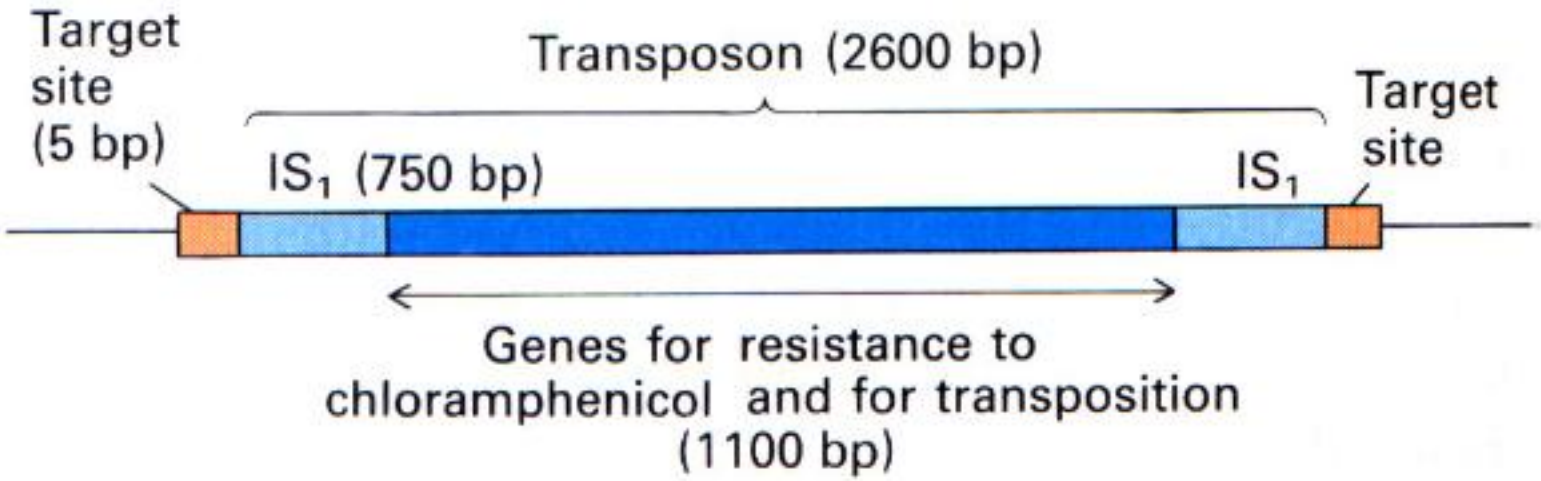
(a)


1 μm


IS2 loop
λ gal3
λ gal
Heteroduplex

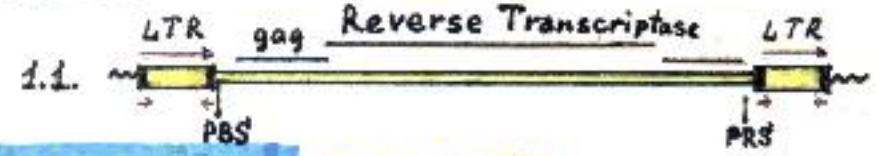
(b)


688 bp
Xn
Insertion sequence
Inverted repeat (~ 50 bp)
Host-cell DNA

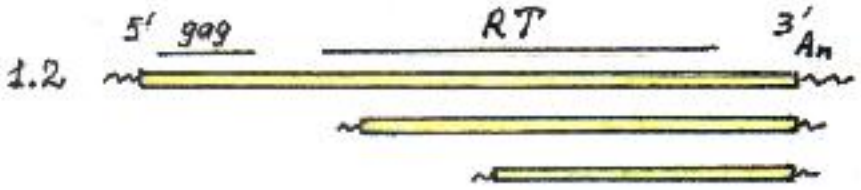

Target
site
(5 bp)
Transposon (2600 bp)
IS1 (750 bp)
IS1
Target
site
Genes for resistance to
chloramphenicol and for transposition
(1100 bp)

D. melanogaster

Класс I



ретротранспозоны COPIA-LIKE

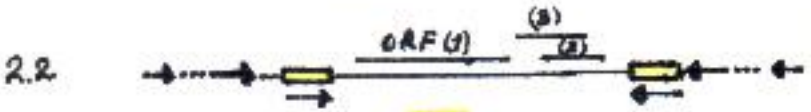


I, F, G, Jockey, LINE

Клade II

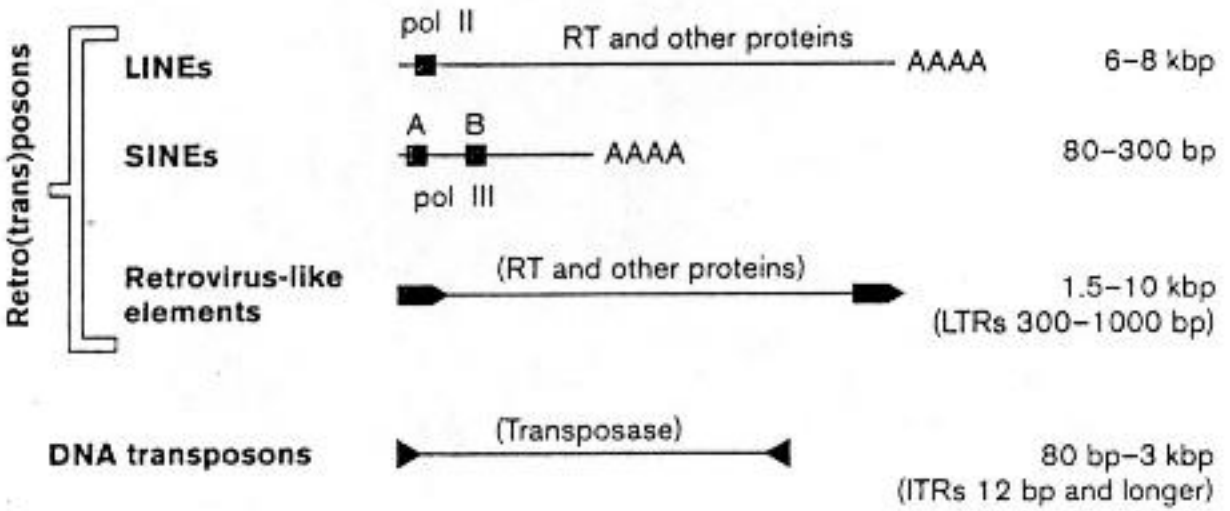
ТРАНСПОЗОНЫ

P, Hobo

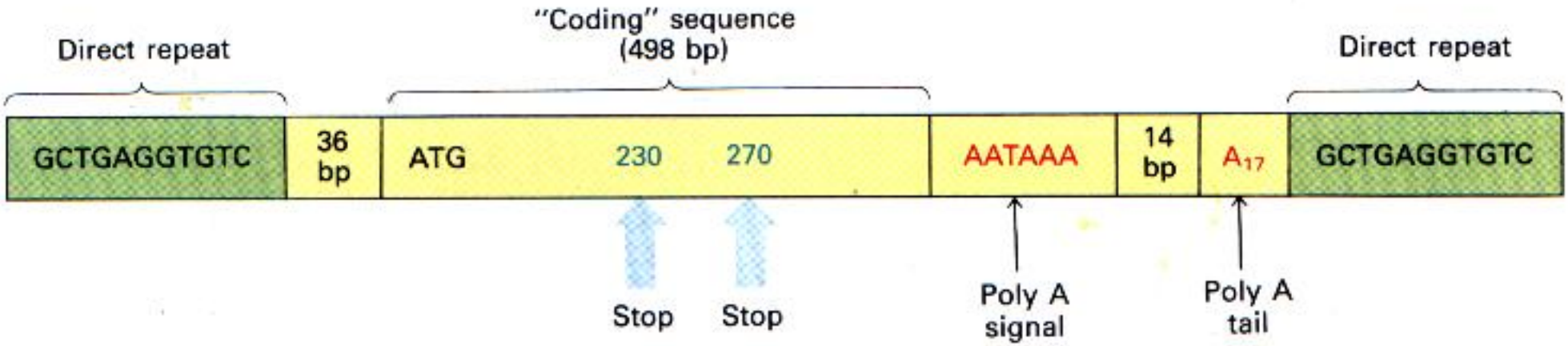


FB

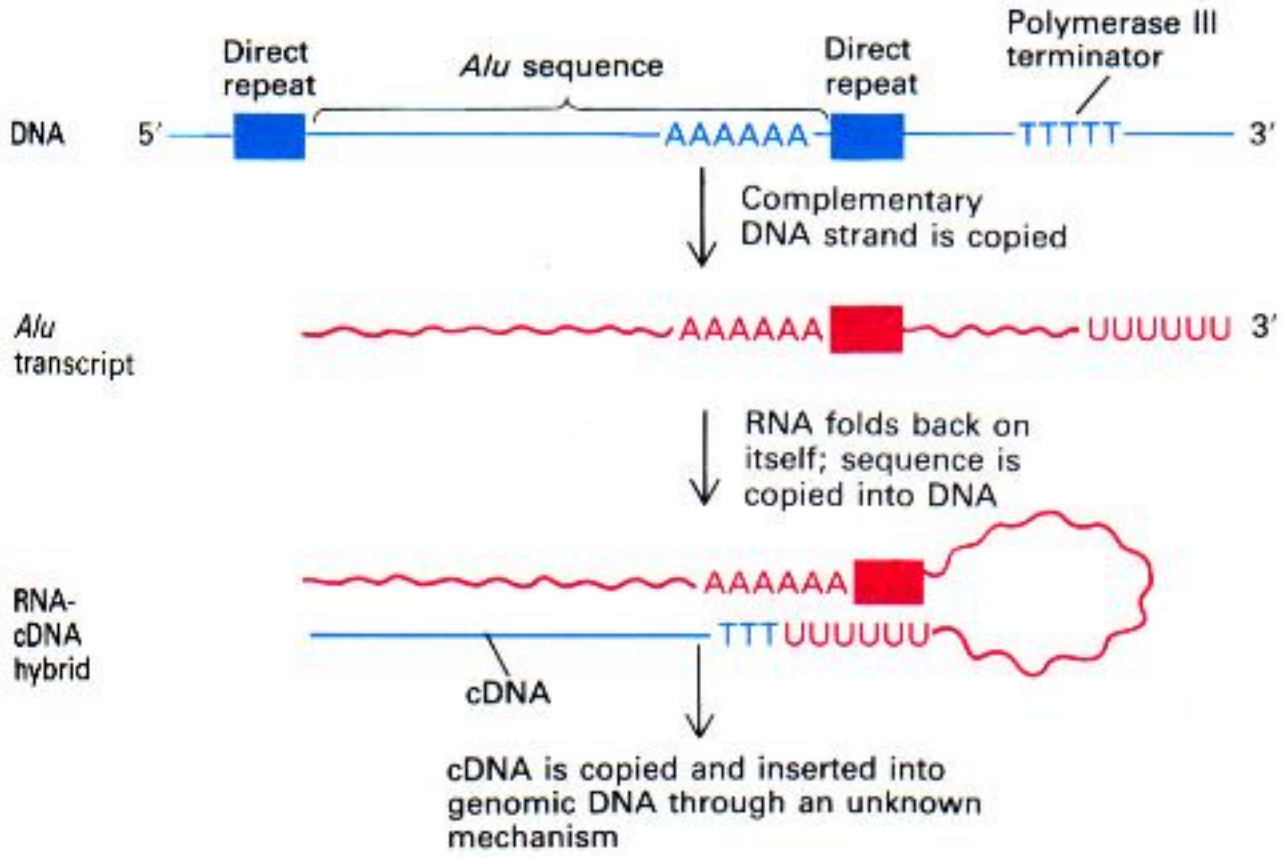
TE

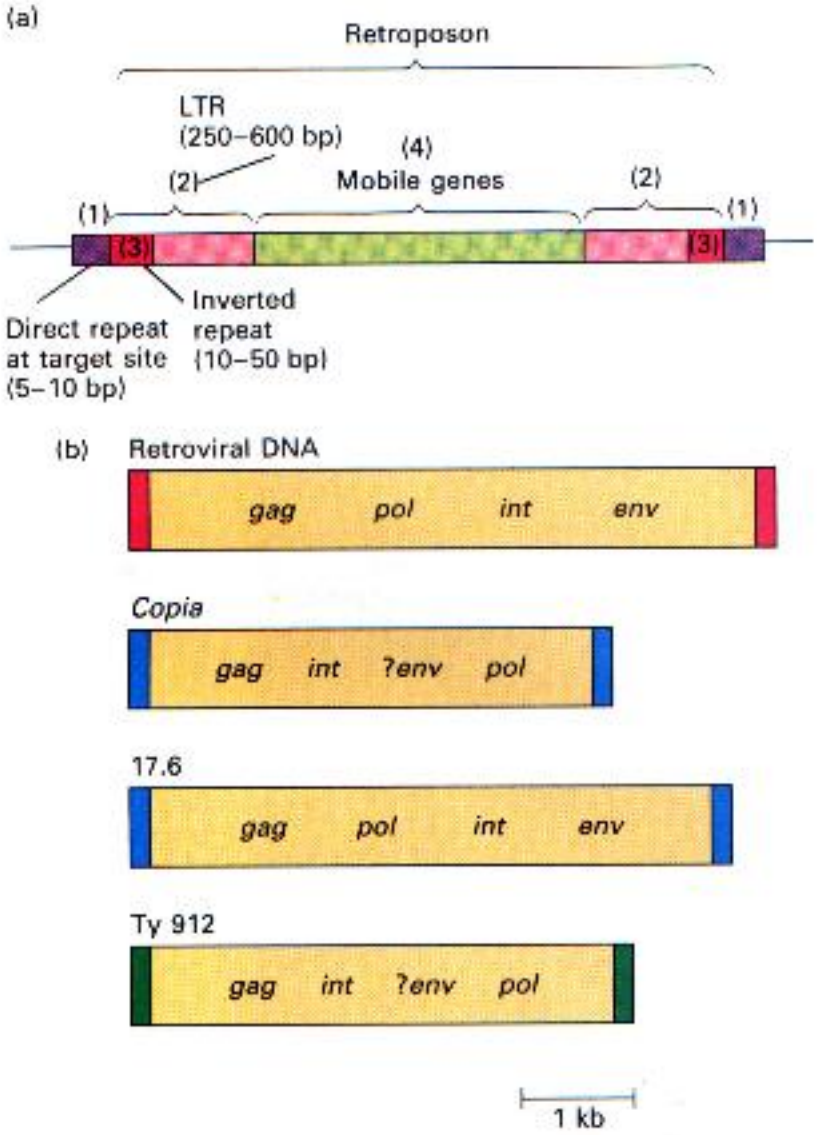


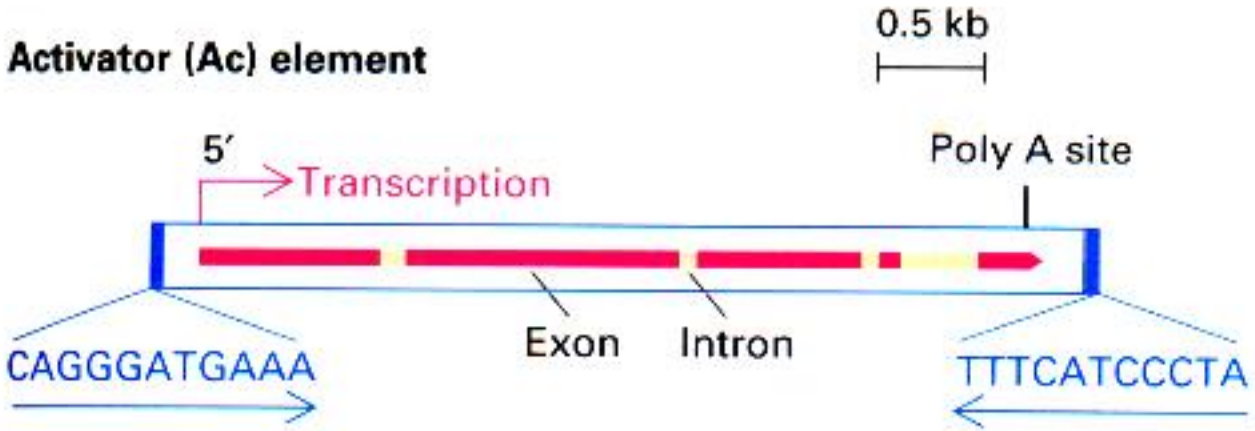

1 15 16 30 31 45
7SL •GCCGGGCGCGGTGG CGCGTGCCTGTAGTC CCAGCTACT•CGGGAG
Alu-cons GGCCGGGCGCGGTGG CTCACGCCTGTAATC CCAGC•ACTTTGGGAG
46 60 61 75 76 90
7SL GCTGAGGCTGGAGGA TCGCTTGAGTCCAGG AGTTC••••CCAGCC
Alu-cons GCCGAGGCGGGCGGA TCACCTGAGGTCAGG AGTTCGAGACCAGCC
91 105 106 120 121 135
7SL TGGGCAACATAGCGA GACCCCGTCTCT••• ••••••••••••••••
Alu-cons TGGCCAACATGGTGA AACCCCGTCTCTACT AAAAATACAAAAATT
136 150 151 165 166 180
7SL •GCCGGGCGCGGTGG CGCGTGCCTGTAGTC CCAGCTACTCGGGAG
Alu-cons AGCCGGGCGTGGTGG CGCGCGCCTGTAATC CCAGCTACTCGGGAG
181 195 196 210 211 225
7SL GCTGAGGCTGGAGGA TCGCTTGAGTCCAGG AGTTCTGGGCTGTAG
Alu-cons GCTGAGGCAGGAGAA TCGCTTGAACCCGGG AGGCGGAGGTTGCAG
226 240 241 255 256 270
7SL TGCGCCTGTGA••••G CCAGTGCACTCCAGC CTGGGCAACATAGCG
Alu-cons TGAGCC•GAGATCGCG CCAGTGCACTCCAGC CTGGGCGACAGAGCG
271 285
7SL AGACCCCGTCTCT
Alu-cons AGACTCCGTCTCAAA AAAAA

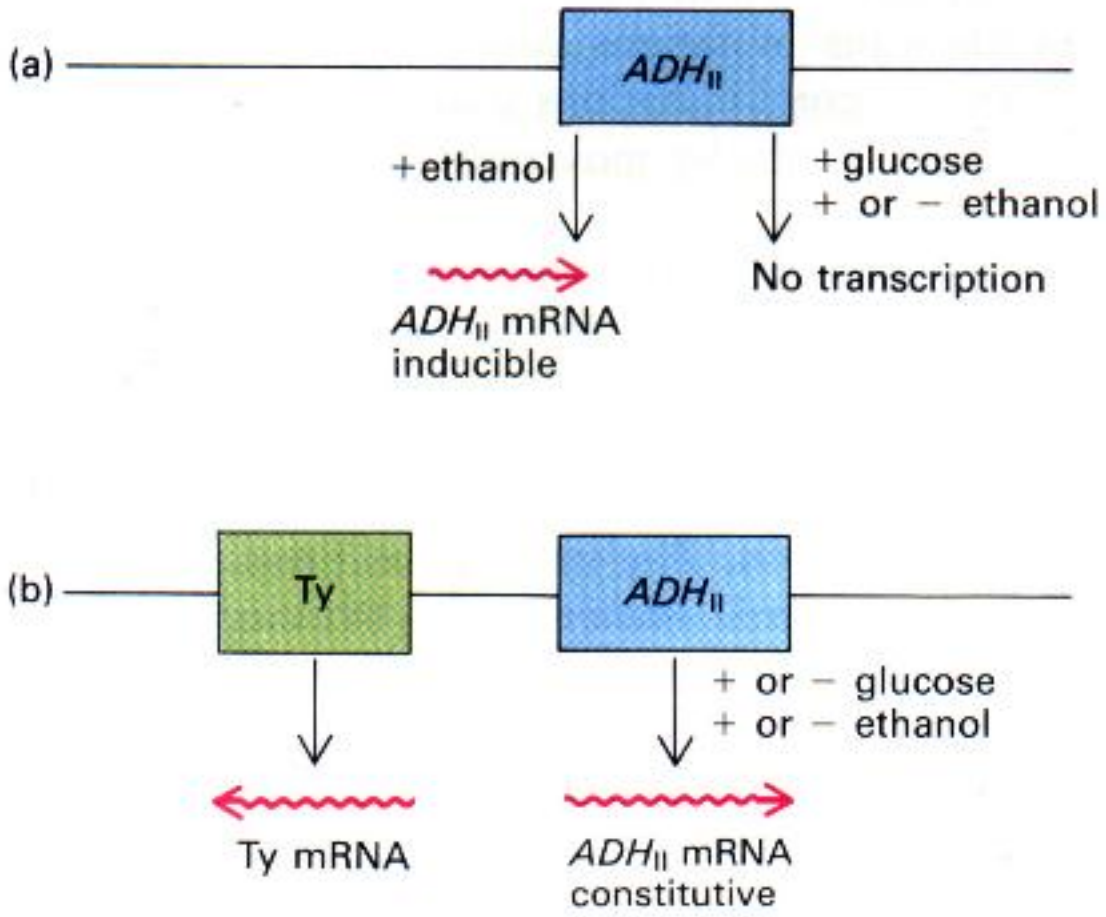

Direct repeat
"Coding" sequence
(498 bp)
Direct repeat
GCTGAGGTGTC
36
bp
ATG
230
270
AATAAA
14
bp
A17
GCTGAGGTGTC
Stop
Stop
Poly A
signal
Poly A
tail

5′ Direct repeat	*Alu* sequence		3′ Direct repeat
		3′ A-rich sequence	
AAACAAGCAGGAGAGGCT	Human *Alu* . . .	$A_7CA_5CA_7TCA_4CA_2TCA_3$	AAAACAAGCAGGAGGGGCT
AAGATTCACTTGTTTAG	. . . . .	$A_{12}GAGAGATTGATTGA_2$	AAGATTCACTTGTTTAG
AAAGAAATGG	. . . . .	$A_{14}GA_3GA_3GA_4GA_5GA_6GA_3$	AAATAAATGG
GTTTAGATAAG	. . . . .	A_{25}	GTTTAGATAAA
AAAAGAAACTTGGAAAGAG	Cho *Alu*	$A_2TA_3TA_3TA_4TCTTA_7$	AAAAGGAAACTTGGAAAGGA
AACATACTAATTTTG	. . . . .	A_4CA_2	AACTATAATTTTTG
GTCAGCC	. . . . .	$TGA_5CCA_5GA_7GA_5GA_5GA_3GTTCCAGGCCA$	GTCAGCC
AGCTCATGAATGAAG	. . . . .	$CCA_5CA_3TCA_4CCAGACAGGCACAGCCCC$	AGCCCAT
GAGACAACAAATCAGAG	Mouse *Alu* . . .	$A_7CCA_3CCA_3CCA_3CCA_6CC$	GAGACAACAAATCAAAT


Direct repeat
Alu sequence
Direct repeat
Polymerase III terminator
DNA
5′
AAAAAA
TTTTT
3′
Complementary DNA strand is copied
Alu transcript
AAAAAA
UUUUUU
3′
RNA folds back on itself; sequence is copied into DNA
RNA-cDNA hybrid
AAAAAA
TTT UUUUUU
cDNA
cDNA is copied and inserted into genomic DNA through an unknown mechanism


(a)
Retroposon
LTR
(250–600 bp)
(4)
(2)
Mobile genes
(2)
(1)
(1)
(3)
(3)
Direct repeat
at target site
(5–10 bp)
Inverted
repeat
(10–50 bp)
(b)
Retroviral DNA
gag
pol
int
env
Copia
gag
int
?env
pol
17.6
gag
pol
int
env
Ty 912
gag
int
?env
pol
1 kb


Activator (Ac) element
0.5 kb
5′
Transcription
Poly A site
Exon
Intron
CAGGGATGAAA
TTTCATCCCTA
Dissociation (Ds) elements
Ds9
Ds2d1
Ds2d2
Ds6


(a)
ADH_{II}
+ethanol
+glucose
+ or − ethanol
No transcription
ADH_{II} mRNA
inducible
(b)
Ty
ADH_{II}
+ or − glucose
+ or − ethanol
Ty mRNA
ADH_{II} mRNA
constitutive

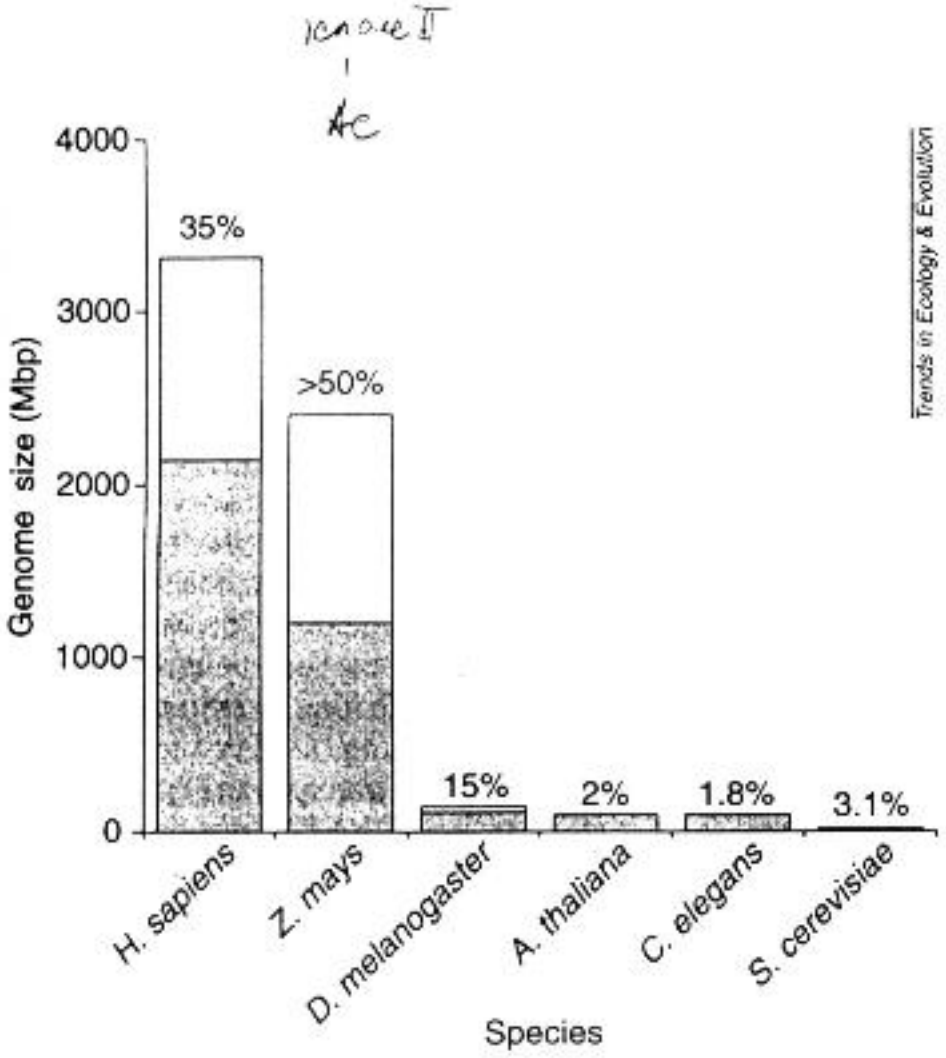
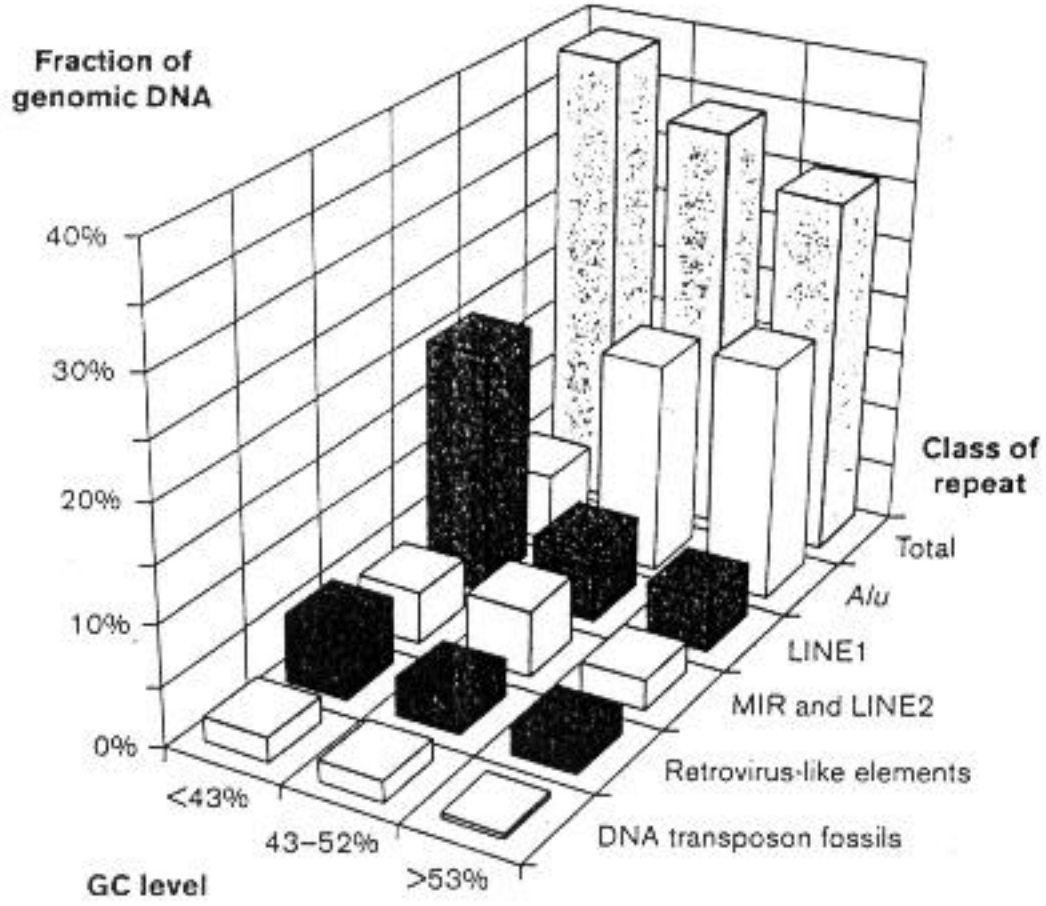

4000
3000
2000
1000
0
Genome size (Mbp)
35%
>50%
15%
2%
1.8%
3.1%
H. sapiens
Z. mays
D. melanogaster
A. thaliana
C. elegans
S. cerevisiae
Species
Trends in Ecology & Evolution

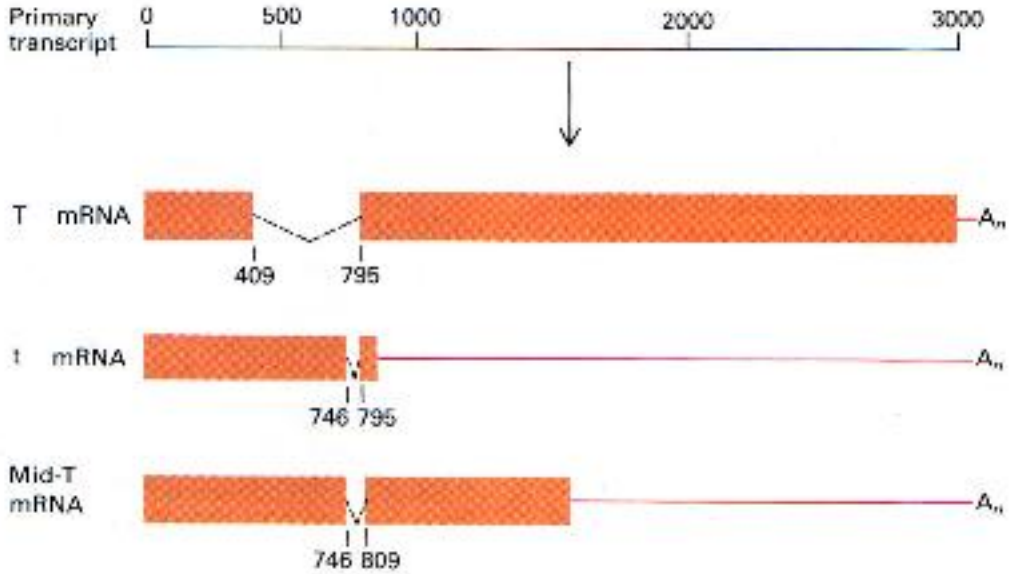
Table 1

Interspersed repeat composition of the human genome.

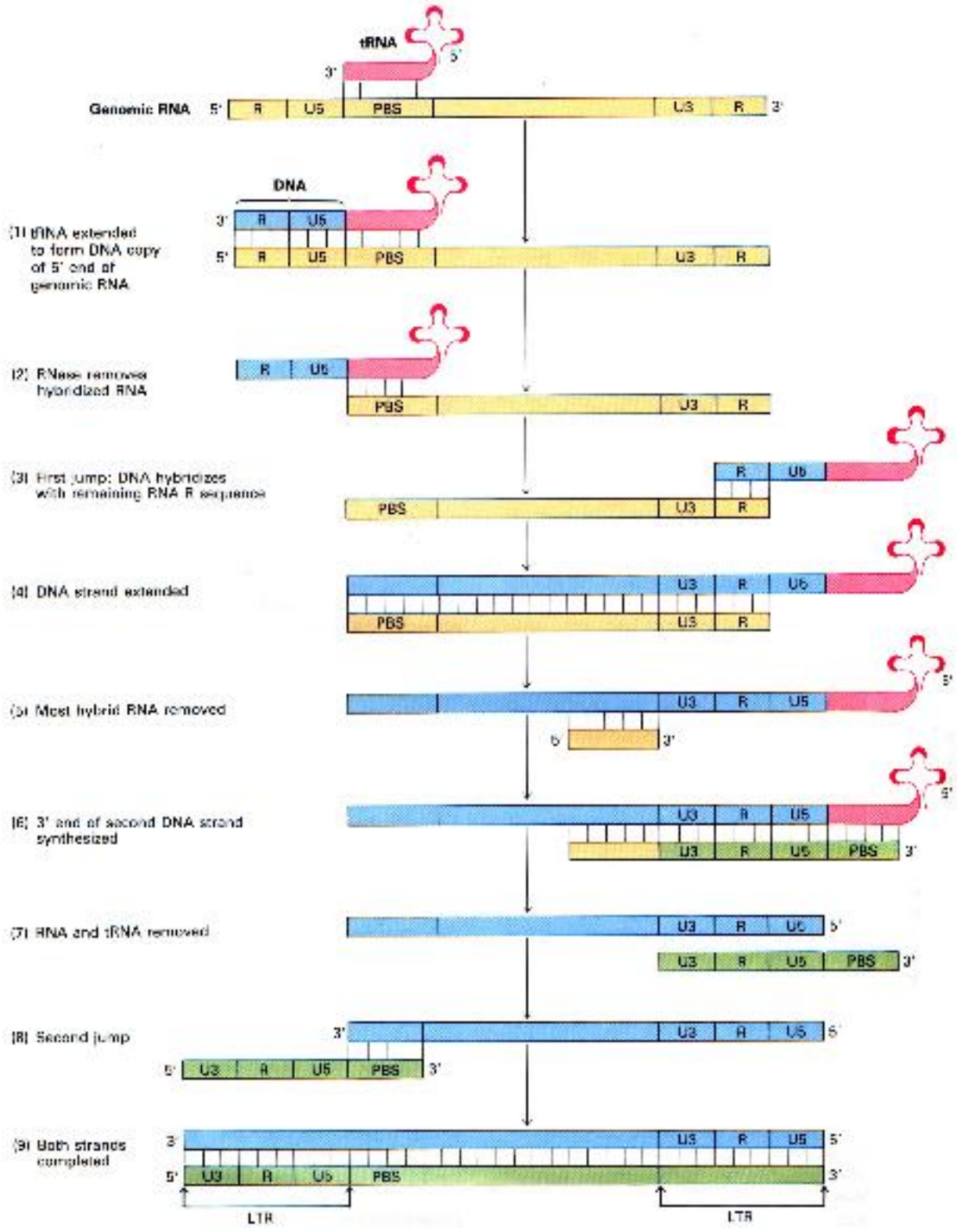
CG level and total size database entries		SINEs		LINEs		Elements with LTRs			DNA transposons		Unclassified elements	Total
		Alu	MIR	LINE1	LINE2	HERVs	MaIRs	others	*mariner*	others		
36–43% GC	Number:	885	494	939	351	80	261	48	15	283	80	
(4102 kb)	Fraction:	5.7%	1.6%	20.5%	2.1%	1.7%	3.1%	0.8%	0.16%	1.7%	1.0%	38.5%
43–52% GC	Number:	1182	290	252	178	21	75	19	2	101	38	
(1724 kb)	Fraction:	17.9%	2.1%	6.1%	2.6%	0.7%	1.9%	0.7%	0.7%	1.3%	0.7%	34.1%
52–63% GC	Number:	996	110	168	90	12	42	13	0	44	15	
(1225 kb)	Fraction:	20.2%	0.9%	4.6%	1.4%	0.5%	1.1%	0.7%	0.0%	0.5%	0.4%	30.3%
Extrapolation to a 3 billion bp genome												
Copy number (in thousands)		1188	402	593	271	50	167	34	8	192	60	2969
Fraction of total genome		10.0%	1.7%	14.6%	2.1%	1.3%	2.6%	0.7%	0.1%	1.5%	0.8%	35.5%

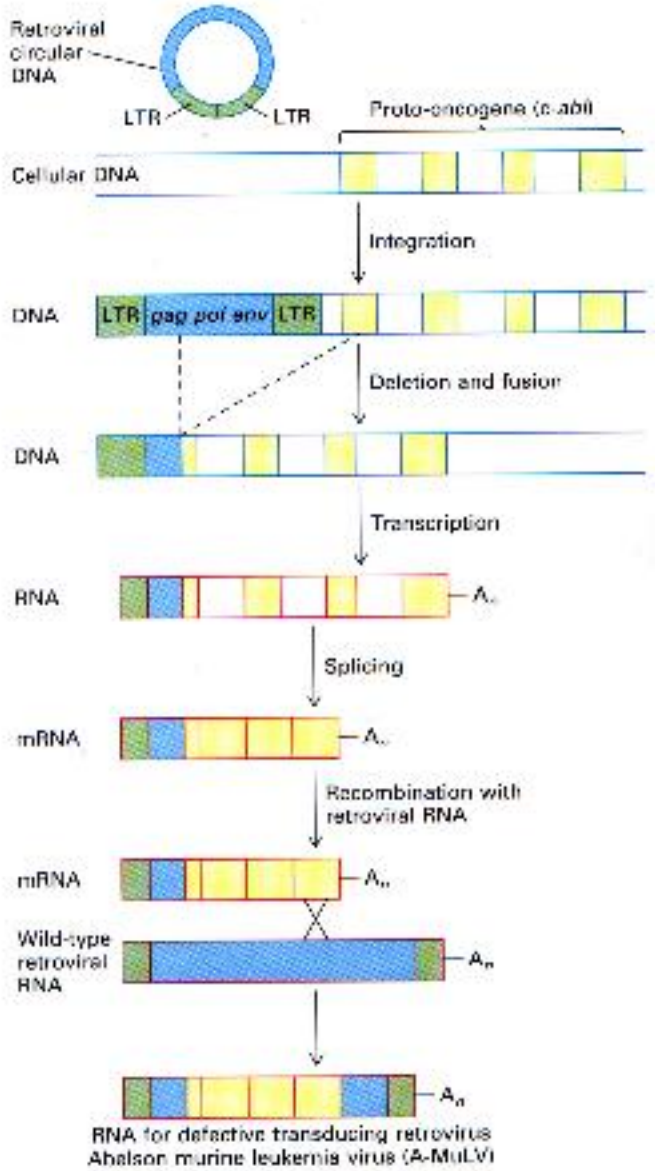
The data in this table are based on the analysis of all unique human genomic GenBank entries >40 kb as of June 1996; a total of 7051 kb derived from 40 distinct loci. For this we used an extended version of the database of human interspersed repeat consensus sequences [50] and the program RepeatMasker [51]. In calculating the number of repeats, fragmented sequences were counted as one. The sequences have been pooled by the GC content to show the differential repeat distribution in AT- and GC-rich DNA (see Fig. 2). As ~60% of human DNA is <43% GC, 30% is 43–52% GC, and 3–5% is >52% GC [52] (the rest is satellite DNA relatively devoid of interspersed repeats) and these fractions comprise 58%, 24.5% and 17.5% of the analyzed sequences respectively, adjustments were made to calculate the genome-wide numbers (GC-rich DNA is probably overrepresented in the database as a consequence of its higher gene density [52]). Except for *Alu*, probably all numbers are underestimates – especially those of MIR and LINE2 – as very old copies/members probably escape detection.



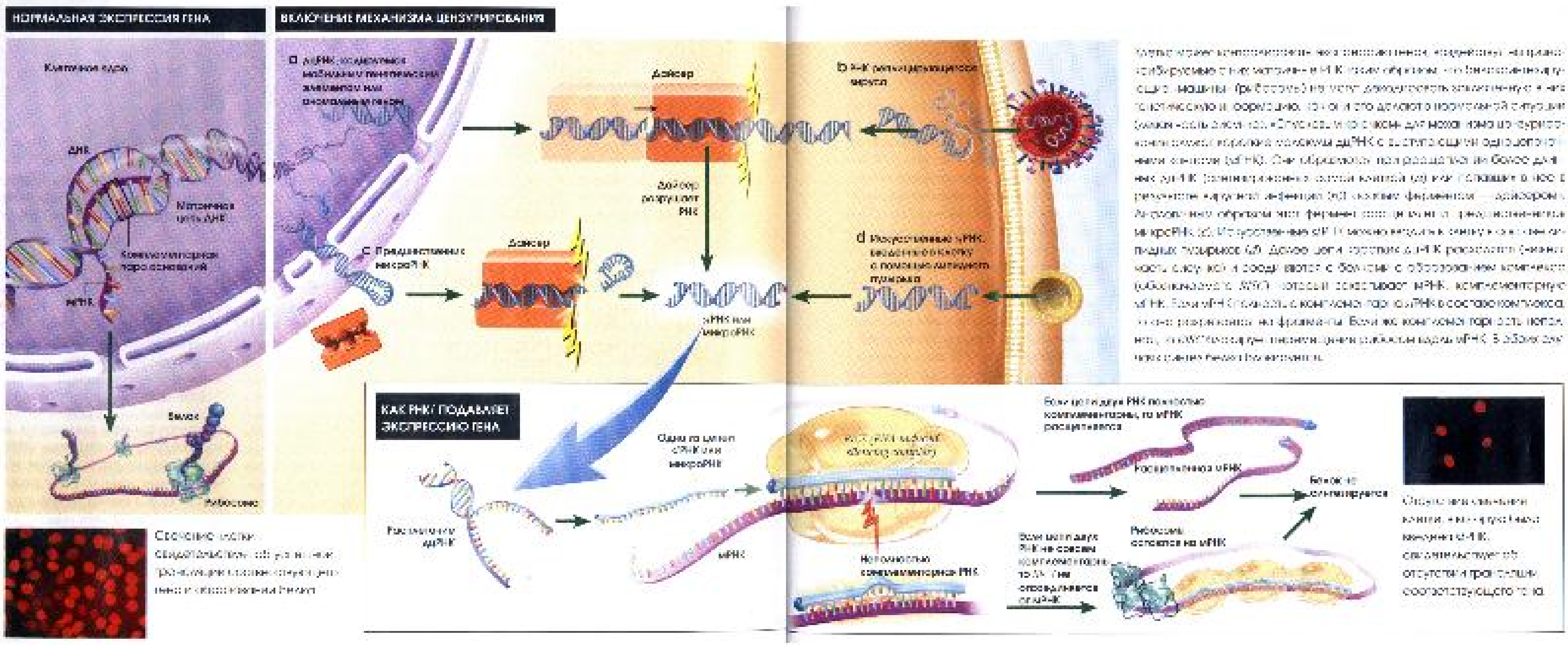


▲ **Figure 24-11** The early region of polyoma virus is expressed in one 3000-base-long nuclear RNA that can be spliced in three ways to yield three different mRNAs and therefore three early proteins. The thick orange bars denote protein-coding regions; the thin red lines, noncoding regions. [See M. Rassoulzadegan et al., 1982, *Nature* 300:713.]


tRNA
Genomic RNA
R
U5
PBS
U3
DNA
(1) tRNA extended to form DNA copy of 5' end of genomic RNA
(2) RNase removes hybridized RNA
(3) First jump: DNA hybridizes with remaining RNA R sequence
(4) DNA strand extended
(5) Most hybrid RNA removed
(6) 3' end of second DNA strand synthesized
(7) RNA and tRNA removed
(8) Second jump
(9) Both strands completed
LTR
LTR



▲ **Figure 24-14** The formation of a transducing retrovirus. Although the events that generate a transducing retrovirus occur at such a low frequency that they have not been studied in the laboratory, the structure of the transducing viruses strongly suggests that they are formed as depicted here. Formation of the Abelson virus is depicted because it has many features found in other retroviruses. The *c-abl* gene is shown as a series of exons in cellular DNA. A wild-type retrovirus upstream of *c-abl* probably integrates randomly once in about 10^6 integrations. A subsequent deletion-fusion event fuses a *c-abl* exon into the *gag* region of the retrovirus. Transcription of the fused DNA and subsequent splicing produces an mRNA for the *gag-abl* fusion protein. If the cell is also infected by a wild-type retrovirus, it will package the *gag-abl* mRNA along with wild-type RNA. In the next generation, the two RNAs can recombine (presumably by a switch of templates during reverse transcription) to generate the RNA of the final transducing retrovirus (now called Abelson murine leukemia virus).



НОРМАЛЬНАЯ ЭКСПРЕССИЯ ГЕНА
ВКЛЮЧЕНИЕ МЕХАНИЗМА ЦЕНЗУРИРОВАНИЯ
Дайсер
ДНК
Дайсер
КАК РНКi ПОДАВЛЯЕТ ЭКСПРЕССИЮ ГЕНА

ПРИМЕНЕНИЕ РНКi В МЕДИЦИНЕ

В клетках млекопитающих механизм геномного цензурирования был обнаружен два года назад, но уже сейчас несколько компаний пытаются использовать его для лечения или предотвращения заболеваний у человека.

КОМПАНИЯ	ПЛАНЫ	ПОЛОЖЕНИЕ
Alnylam Pharmaceuticals Кембридж шт. Массачусетс, США	Исследование возможности применения РНКi в медицине; конкретные заболевания пока не названы	Основана в 2002 г. Получила начальное финансирование и несколько патентов
Cenix Biosciences Дрезден, Германия	Исследование возможности применения РНКi для лечения рака и вирусных инфекций	Создает библиотеку *si*РНК, охватывающую весь геном человека
Ribopharma Кульмбах, Германия	Химическая модификация *si*РНК с целью создания препаратов для лечения глиобластомы, рака поджелудочной железы и гепатита С	Предполагается начать клинические испытания на больных, страдающих раком головного мозга
Sirna Therapeutics Боулдер, шт. Колорадо, США	Тестирование препаратов на основе РНК, обладающей ферментативной активностью, которые предназначены для лечения рака прямой кишки; разработка методов лечения с использованием РНКi	Под таким названием компания появилась в апреле 2003 г. (прежнее название *Ribozyme Pharmaceuticals*)