

APPENDIX 4

Useful Facts

COMPILED BY JOAN KOBORI

Appendix 4 contains information commonly used in the analysis of genomes. The following are included in this appendix:

- Conversion factors for calculations involving nucleic acids and proteins
- Genome comparisons
- Average sizes of DNA fragments generated by cleavage with restriction enzymes
- Single-letter abbreviations for amino acids
- Genetic code
- Codon facts
- Isotope information

CONVERSION FACTORS

Nucleic Acid Weight Conversions

1000 bases = 6.5×10^5 daltons of double-stranded DNA (sodium salt)

1000 bases = 3.3×10^5 daltons of single-stranded DNA (sodium salt)

1000 bases = 3.4×10^5 daltons of single-stranded RNA (sodium salt)

Average molecular weight of a dNMP = 324.5 daltons

Average molecular weight of a base pair of DNA = 649 daltons

Nucleic Acid Molar Conversions

1 $\mu\text{g/ml}$ solution of DNA = 3.08 μM phosphate

1 $\mu\text{g/ml}$ solution of 1-kb DNA = 3.08 nM 5' ends

1 μg of 1-kb DNA = 1.54 pmole (3.08 pmole of ends)

Nucleic Acid Spectrophotometric Conversions

In a quartz cuvette with a 1-cm path length, an OD_{260} of 1 corresponds to:

- a 33 $\mu\text{g/ml}$ solution of single-stranded oligonucleotide (unmodified)
- a 36 $\mu\text{g/ml}$ solution of single-stranded DNA
- a 50 $\mu\text{g/ml}$ solution of double-stranded DNA
- a 40 $\mu\text{g/ml}$ solution of single-stranded RNA

Protein Conversions

1 kb of DNA can encode a 333-amino-acid protein with a m.w. of 37,000.

2.7 kb of DNA can encode a 900-amino-acid protein with a m.w. of 100,000.

GENOME COMPARISONS

Genome	Number of haploid chromosomes	Approximate size of haploid genome		Approximate amount of DNA per cell (pg)	Ploidy
		kb	kD		
<i>Homo sapiens</i> (human being)	23	3,300,000	2.1×10^{12}	6	diploid
<i>Mus musculus</i> (mouse)	20	2,700,000	1.75×10^{12}	5.4	diploid
<i>D. melanogaster</i> (fruit fly)	4	165,000	1.1×10^{11}	0.33	diploid
<i>C. elegans</i> (nematode)	6	100,000	6.5×10^{10}	0.2	diploid
<i>Arabidopsis thaliana</i> (plant)	5	100,000	6.5×10^{10}	0.2	diploid
<i>S. cerevisiae</i> (yeast)	16	14,000	9.1×10^9	0.014	haploid
<i>E. coli</i> (bacteria)	1	4700	3.05×10^9	0.0047	haploid
Bacteriophage λ (bacteriophage)	1	48.5	3.1×10^7	n.r.	n.r.

(n.r.) Not relevant.

AVERAGE SIZES OF DNA FRAGMENTS GENERATED BY CLEAVAGE WITH RESTRICTION ENZYMES

The table below lists those restriction enzymes that are known or predicted to cleave infrequently in ten commonly studied genomes (*Arabidopsis thaliana*, *C. elegans*, *D. melanogaster*, *E. coli*, human being, mouse, *Rhodobacter sphaeroides*, *S. cerevisiae*, *Staphylococcus aureus*, and *Xenopus laevis*). Factors affecting the ability of restriction enzymes to cleave a particular genome include (1) percentage G + C content, (2) specific dinucleotide, trinucleotide, and/or tetranucleotide frequencies, and (3) methylation. Using available information on percentage G + C content, dinucleotide frequencies, and a few kilobases of DNA sequence, predictions can be made about potential cleavage with restriction enzymes. (Reprinted, with permission, from New England Biolabs.)

Enzyme	Sequence	ATH	CEL	DRO	ECO	HUM	MUS	RSS	YSC	STA	XEL
<i>Apa</i> I	GGGCC	25,000	40,000	6,000	15,000	2,000	3,000	8,000	20,000	70,000	5,000
<i>Asc</i> I	GGCGGCC	400,000	400,000	60,000	20,000	80,000	100,000	4,000,000	500,000	600,000	200,000
<i>Avr</i> II	CCTAGG	15,000	20,000	20,000	150,000	8,000	7,000	10,000	20,000	20,000	15,000
<i>Bam</i> HI	GGATTC	6,000	9,000	4,000	5,000	5,000	4,000	5,000	9,000	15,000	5,000
<i>Bgl</i> I	GCCN ₅ GGC	20,000	25,000	4,000	3,000	3,000	4,000	10,000	15,000	30,000	6,000
<i>Bgl</i> II	GCGCGC	2,000	4,000	4,000	6,000	3,000	3,000	3,000	4,000	6,000	3,000
<i>Bss</i> HI	GCGCGC	50,000	30,000	6,000	2,000	10,000	15,000	80,000	30,000	40,000	20,000
<i>Dra</i> I	TTTAAA	2,000	1,000	1,000	2,000	2,000	3,000	2,000	1,000	1,000	2,000
<i>Eag</i> I	CGGCCG	10,000	20,000	3,000	4,000	10,000	15,000	60,000	20,000	50,000	15,000
<i>Eco</i> R I	GAATTC	4,000	2,000	4,000	5,000	5,000	5,000	3,000	3,000	4,000	4,000
<i>Hind</i> III	AAGCTT	1,000	3,000	4,000	5,000	4,000	3,000	6,000	3,000	2,000	3,000
<i>Nae</i> I	GCCGGC	6,000	15,000	3,000	2,000	4,000	6,000	40,000	15,000	20,000	6,000
<i>Nar</i> I	GGCGCC	10,000	15,000	3,000	2,000	4,000	6,000	50,000	15,000	15,000	7,000
<i>Nhe</i> I	GCTAGC	10,000	30,000	10,000	25,000	10,000	10,000	15,000	10,000	10,000	10,000
<i>Not</i> I	GCGGCCG	200,000	600,000	30,000	200,000	100,000	200,000	2,000,000	450,000	1,000,000	200,000
<i>Pac</i> I	TTAATTAA	70,000	20,000	25,000	50,000	60,000	100,000	100,000	15,000	9,000	50,000
<i>Pme</i> I	GTTTAAAC	60,000	40,000	40,000	40,000	70,000	80,000	50,000	50,000	25,000	50,000
<i>Rsr</i> II	CGGWCCG	25,000	50,000	15,000	10,000	60,000	60,000	50,000	60,000	150,000	70,000
<i>Sac</i> I	GAGCTC	3,000	4,000	4,000	10,000	3,000	3,000	3,000	9,000	10,000	4,000
<i>Sac</i> II	CCGCGG	10,000	20,000	5,000	3,000	6,000	8,000	70,000	20,000	40,000	15,000
<i>Sal</i> I	GTCGAC	6,000	8,000	5,000	5,000	20,000	20,000	25,000	10,000	15,000	15,000
<i>Sfi</i> I	GGCCN ₅ GGCC	400,000	1,000,000	60,000	150,000	30,000	40,000	150,000	350,000	2,000,000	100,000
<i>Sgr</i> A I	CXCCGGXG	30,000	100,000	20,000	8,000	70,000	80,000	100,000	90,000	150,000	90,000
<i>Sma</i> I	CCCGGG	10,000	30,000	10,000	6,000	4,000	5,000	30,000	50,000	50,000	5,000
<i>Spe</i> I	ACTAGT	8,000	8,000	9,000	60,000	10,000	15,000	6,000	6,000	6,000	8,000
<i>Sph</i> I	GCATGC	10,000	15,000	5,000	4,000	6,000	6,000	7,000	10,000	9,000	6,000
<i>Srf</i> I	GCCCGGGC	400,000	1,000,000	90,000	50,000	50,000	90,000	1,000,000	600,000	2,000,000	100,000
<i>Sse</i> I	CCTGCAGG	100,000	200,000	50,000	40,000	15,000	15,000	60,000	150,000	200,000	30,000
<i>Ssp</i> I	AATATT	3,000	1,000	1,000	2,000	2,000	3,000	2,000	1,000	1,000	2,000
<i>Swa</i> I	ATTTAAAT	50,000	9,000	15,000	40,000	30,000	60,000	25,000	15,000	6,000	30,000
<i>Xba</i> I	TCTAGA	5,000	4,000	9,000	70,000	5,000	8,000	7,000	4,000	5,000	6,000
<i>Xho</i> I	CTCGAG	4,000	5,000	4,000	15,000	7,000	7,000	6,000	15,000	25,000	10,000

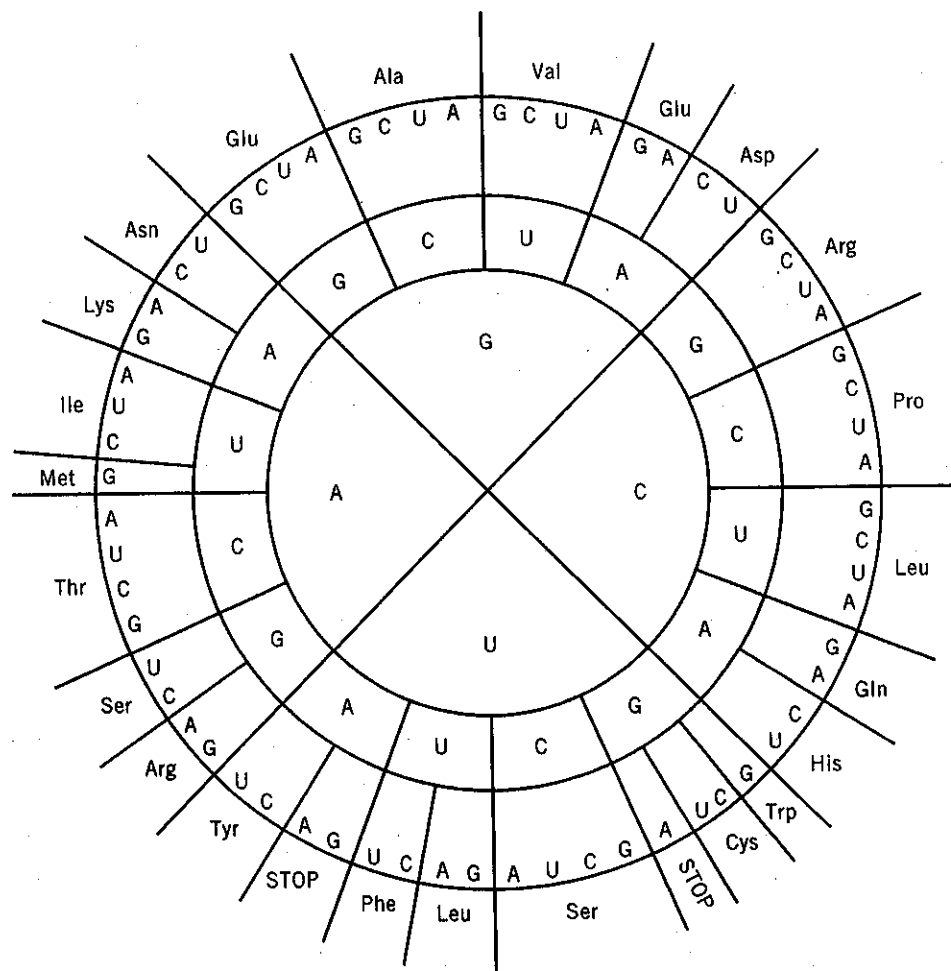
Average size fragments predicted for ATH *Arabidopsis thaliana*, CEL *Caenorhabditis elegans*, DRO *Drosophila melanogaster*, ECO *Escherichia coli*, HUM Human, MUS Mouse, RSS *Rhodobacter sphaeroides*, YSC *Saccharomyces cerevisiae*, STA *Staphylococcus aureus*, and XEL *Xenopus laevis*.

SINGLE-LETTER ABBREVIATIONS FOR AMINO ACIDS

Amino acid	One-letter code
Alanine	A
Arginine	R
Aspartic acid	D
Asparagine	N
Cysteine	C
Glycine	G
Glutamic acid	E
Glutamine	Q
Histidine	H
Isoleucine	I
Leucine	L
Lysine	K
Methionine	M
Phenylalanine	F
Proline	P
Serine	S
Threonine	T
Tryptophan	W
Tyrosine	Y
Valine	V

GENETIC CODE

The first base in each mRNA codon is in the center of the circle, the second base is in the second circle from the center, and the third base is in the third circle from the center. The standard three-letter abbreviation for the amino acid encoded by each codon is on the outside. UAG, UAA, and UGA stop codons are designated by the word STOP.



CODON FACTS

Codon Redundancy

Amino acid	Number of synonymous codons
W, M	1
N, D, C, Q, E, H, K, F, Y	2
I	3
A, G, P, T, V	4
R, L, S	6

Termination Codons

UAA	ochre
UAG	amber
UGA	opal

ISOTOPE INFORMATION

Isotope Half-lives

Isotope	Emitted particle	Half-life
^{14}C	beta	5730 years
^3H	beta	12.4 years
^{125}I	gamma	60 days
^{32}P	beta	14.3 days
^{35}S	beta	87.4 days

Radioactivity Conversions

$$1 \text{ Ci} = 1000 \text{ mCi}$$

$$1 \text{ mCi} = 1000 \mu\text{Ci}$$

$$1 \mu\text{Ci} = 2.2 \times 10^6 \text{ dpm}$$

$$1 \text{ Bq} = 1 \text{ disintegration/second}$$

$$1 \mu\text{Ci} = 3.7 \times 10^4 \text{ Bq}$$

$$1 \text{ Bq} = 2.7 \times 10^{-5} \mu\text{Ci}$$