## 2. Genomic information

- proteins:
  - amino acid chain; typically ~300 a.a. in length
  - each protein encoded by a "gene" (~1000 bases)
- bacterial genomes: circular chromosome(s)
  - <u>E. coli K-12 MG1655</u>: 4.6Mb, ~4400 genes; 2/3 "identified"
  - <u>4 different strains of *E. coli*</u>: 4.6 5.4 Mb (common: 3.8Mb)
  - <u>range for 150+ bacteria</u>: 0.5 10 Mb



- approx 85% of the E. coli genome codes for proteins
- operon structure
  - genes coding for proteins that function in the same pathway <u>may</u> be located adjacent to one another and controlled as a single unit that is transcribed into a polycistronic mRNA called "operon"



- regulator of an operon is often located nearby (and divergently transcribed)

- different operons involved in the same function can be located far away

• other genomic elements and features

- transfer RNA: 61 different types, 74-95 nt in length



 → codons not equally represented: 86 tRNA genes in *E. coli MG1655* → codons not equally used in coding sequences [= codon bias] (strong bias especially in genes involved in transcription and translation) <u>ribosomal RNA</u>: 5S, 120 nt; 16S, 1542 nt; 23S, 2904 nt
 (22 genes encoding rRNA in MG1655)

- small (regulatory) RNA: ~50 discovered so far
- prophages: ~10 (5,000 ~ 50,000 nt)

- insertion sequences (transposons)



	Target Inverted repeat repeat		Transposon	Inverted Target repeat repeat
			Overall length	Target selection
IS1	9 bp	23 bp	768 bp	random
IS2	5 bp	41 bp	1327 bp	hotspots
IS4	11-13 bp	18 bp	1428 bp	AAAN <sub>20</sub> TTT
IS5	4 bp	16 bp	1195 bp	hotspots
IS10R	9 bp	22 bp	1329 bp	NGCTNAGCN
IS50R	9 bp	9 bp	1531 bp	hotspots
IS903	9 bp	18 bp	1057 bp	random

- replication origin (oriC)
- termination of replication (terC)

## • plasmids:

types:

plasmids	F, P1	R1	pSC101	CoIE1
copy #	1-2	3-5	~10	~30
length	90 kb	102 kb	6.5 kb	7.2 kb

- maintenance: provides a service to the host
  - e.g., poison-antidote system, antibiotic marker (genetic engineering)





• plasticity of the genome

## mutation ≠ passive decay of genome

- o *Deinococcus radiodurans:* Withstands 1.5 Mrads, dessication, starvation, UV light, hydrogen peroxide; grows well at 6 krad/h
  o mutator strain: can increase mutation rate 1000-fold
- base-pair substitution (BPS): 10<sup>-9</sup>/base/replication
- insertions and deletions: ~1/3 of BPS rate
- horizontal transfer
  - competence
  - conjugation
  - recombination







requires reasonably long regions of homology [expansion of repeats; transposable elements]

> site-specific recombination



[Watt et al, 85]



only occurs for specific sequences; requires special proteins

Question: how is homology recognized?

## > consequence of recombination



direct repeat (e.g., insertion sequences)

- recombination releases material between repeats as circular molecule [may be transferred between cells]
- reverse process: chromosomal integration



inverted repeat:

- sequence inversion (important mechanism of gene regulation)

