**Figure 22.7** Introns in vertebrate genes range from short to long.

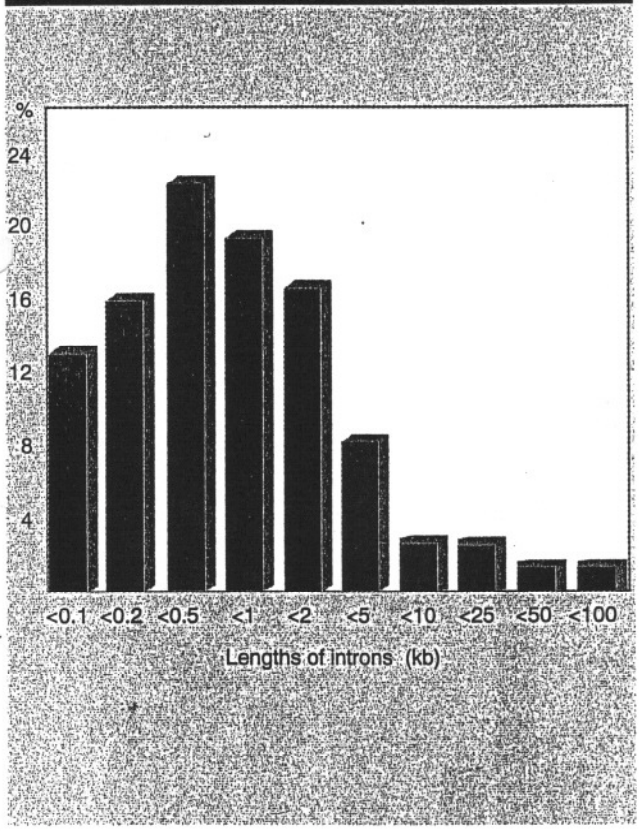



**Figure 22.6** Exons coding for proteins are usually short.

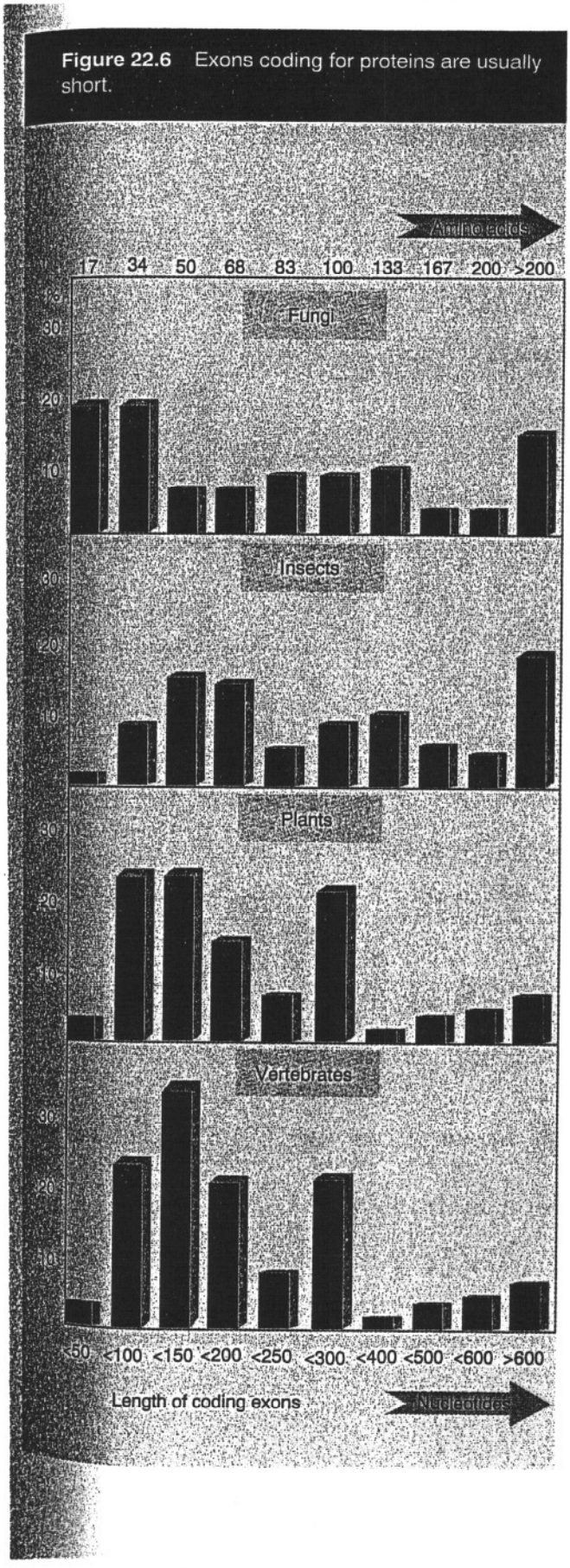


**Figure 21.7** The proportions of different sequence components vary in eukaryotic genomes. The absolute content of nonrepetitive DNA increases with genome size, but reaches a plateau at ~2 x $10^9$ bp.

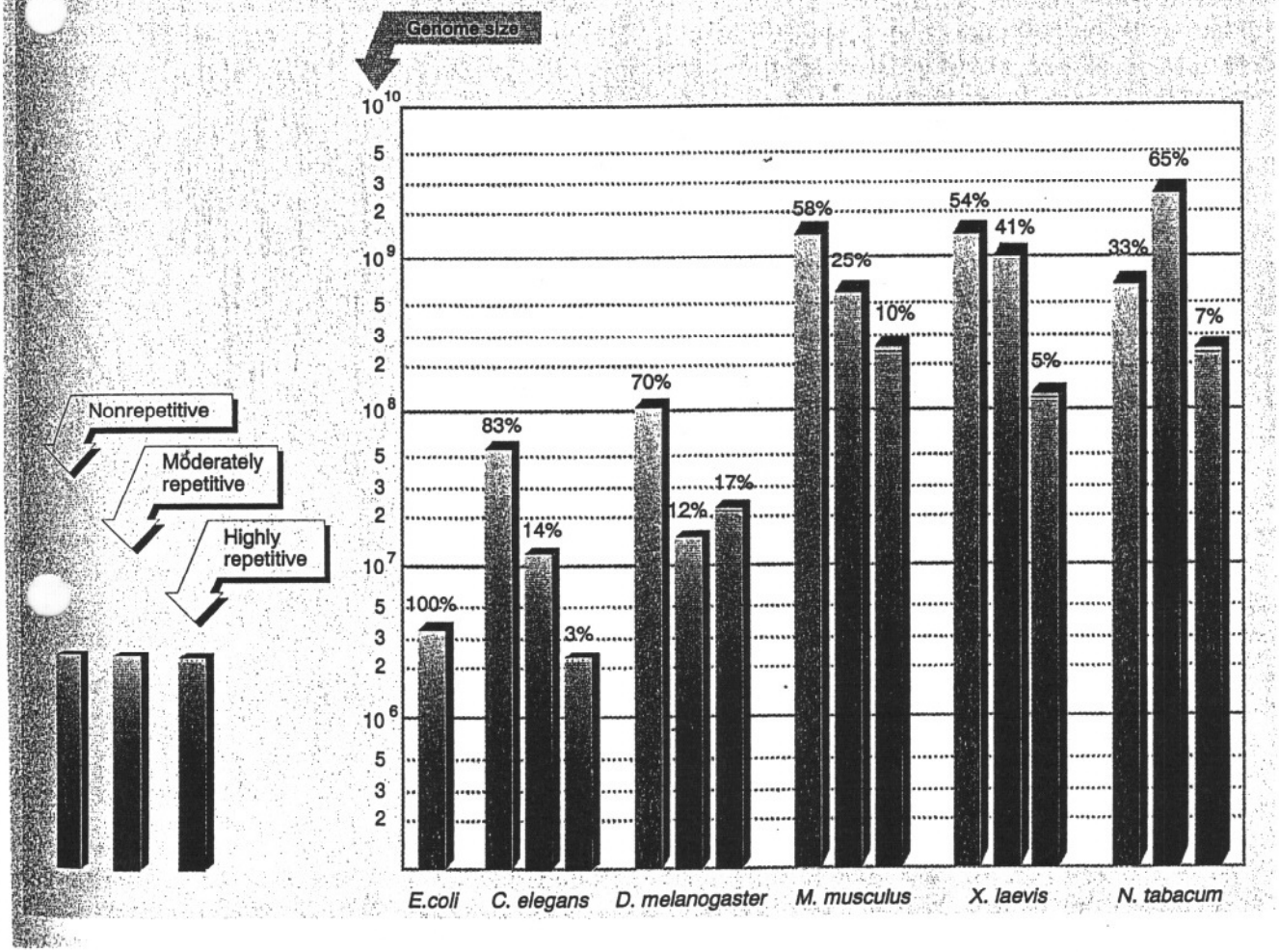
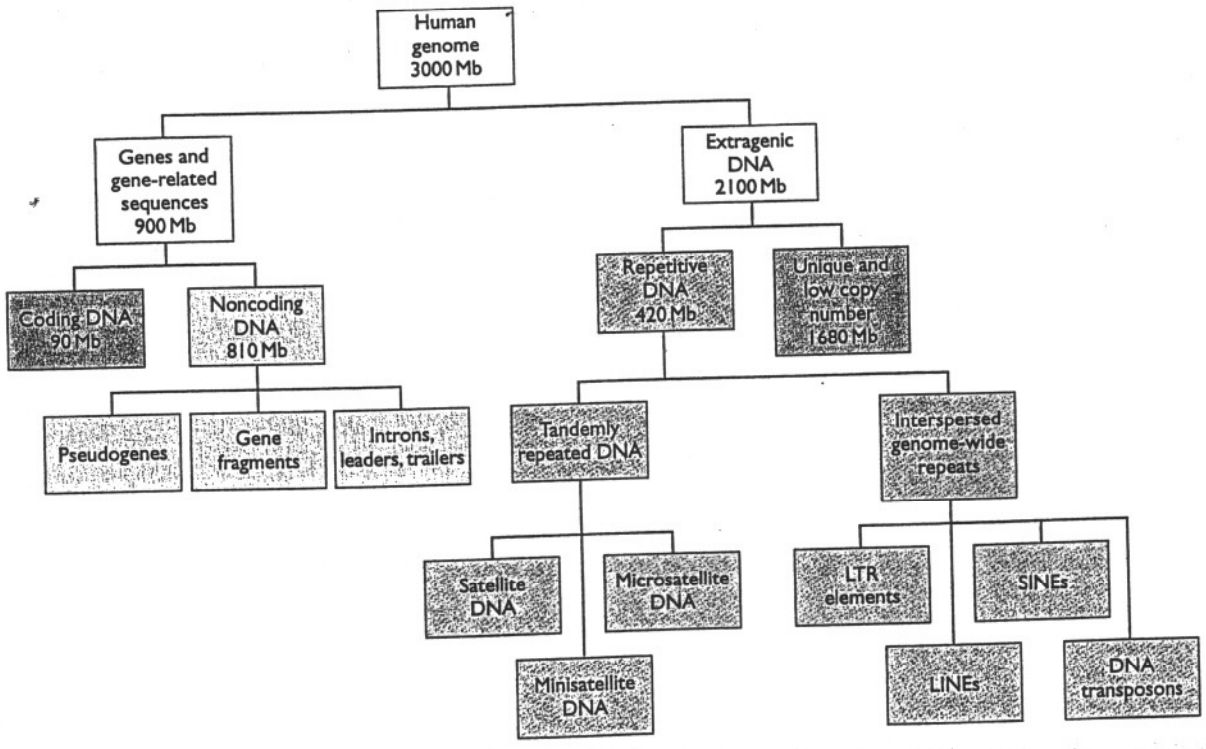


Table 6.6 Partial gene catalogs for *Escherichia coli*, *Haemophilus influenzae* and *Mycoplasma genitalium*

| Category | Number of genes in *E. coli* | *H. influenzae* | *M. genitalium* |
|---|---|---|---|
| Total protein-coding genes | 4288 | 1727 | 470 |
| Biosynthesis of amino acids | 131 | 68 | 1 |
| Biosynthesis of cofactors | 103 | 54 | 5 |
| Biosynthesis of nucleotides | 58 | 53 | 19 |
| Cell envelope proteins | 237 | 84 | 17 |
| Energy metabolism | 243 | 112 | 31 |
| Intermediary metabolism | 188 | 30 | 6 |
| Lipid metabolism | 48 | 25 | 6 |
| DNA replication, recombination, repair | 115 | 87 | 32 |
| Protein folding | 9 | 6 | 7 |
| Regulatory proteins | 178 | 64 | 7 |
| Transcription | 55 | 27 | 12 |
| Translation | 182 | 141 | 101 |
| Uptake of molecules from the environment | 427 | 123 | 34 |

Taken from Fraser *et al.* (1995) and Blattner *et al.* (1997). The numbers refer only to genes whose functions are known and are therefore approximate, because each genome also contains many genes whose functions have not yet been identified.

## Box 6.4: The organization of the human genome




Based on Strachan and Read (1996).

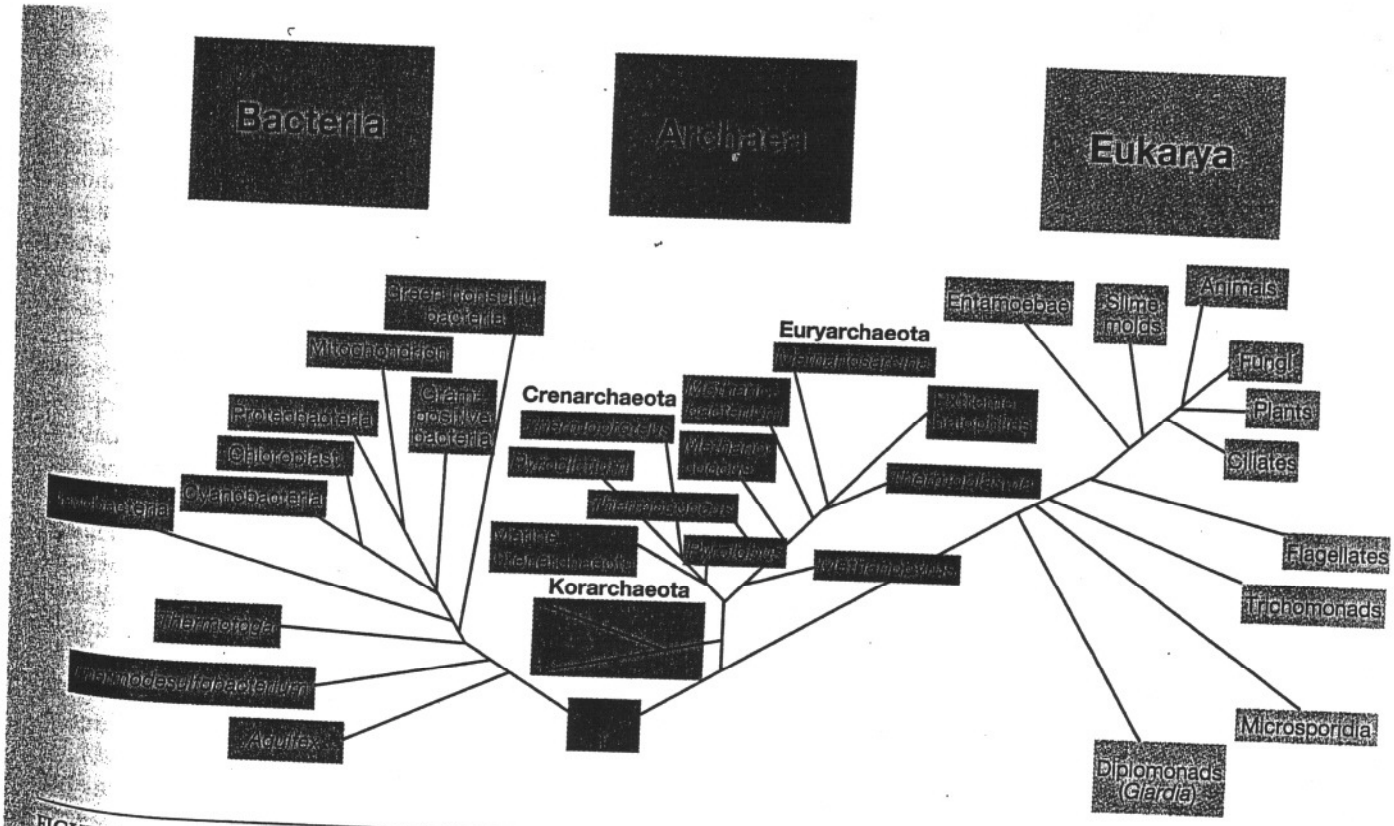



**FIGURE 12.13** Universal phylogenetic tree as determined from comparative ribosomal RNA sequencing. The data support the separation of three domains, two of which (Bacteria and Archaea) contain only prokaryotic representatives. The location highlighted in red is the hypothetical root of the tree, which represents the position of the universal ancestor of all cells. See text for discussion of the Korarchaeota.

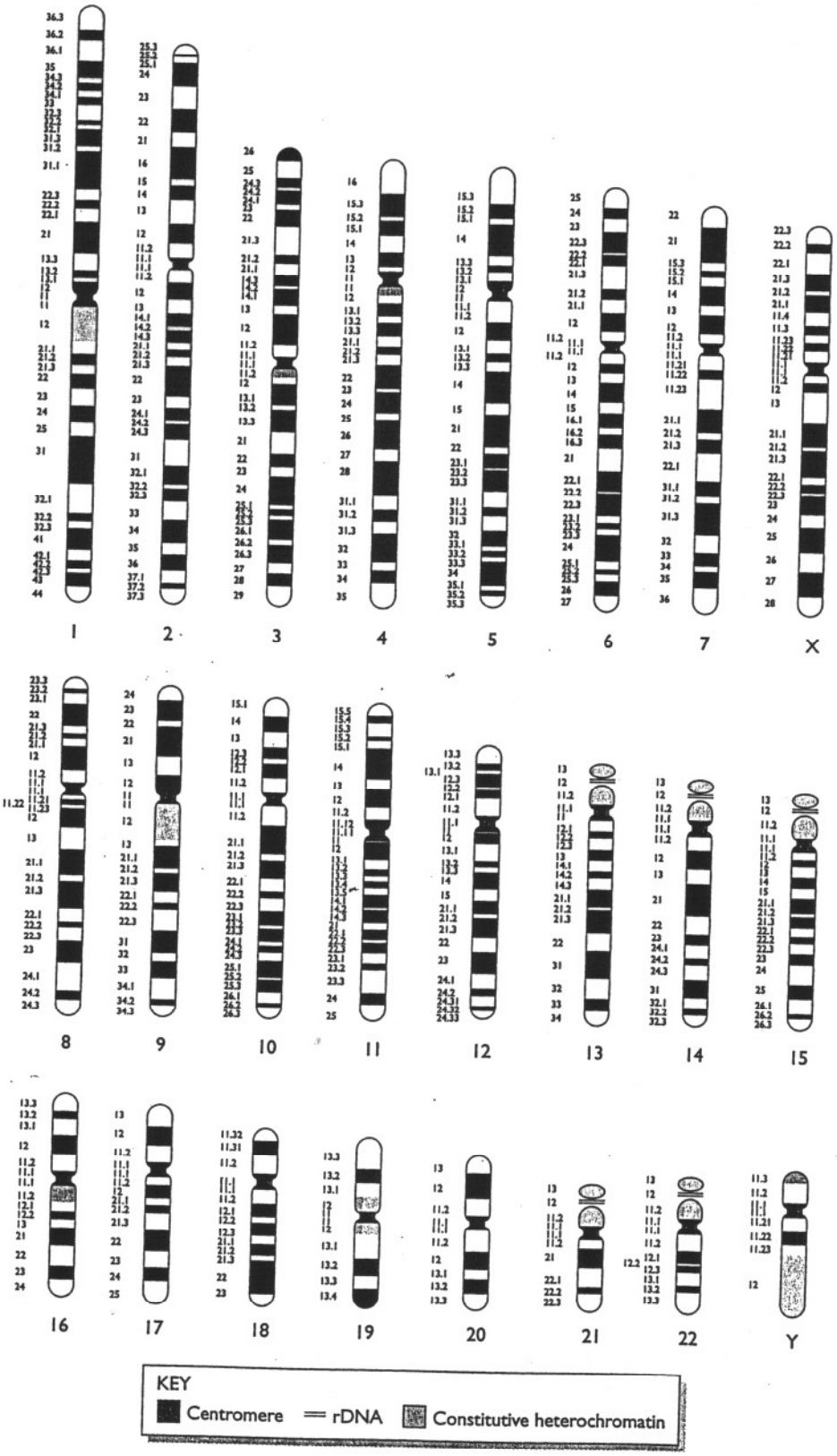



**Figure 6.4** The human karyogram.

The chromosomes are shown with the G-banding pattern obtained after Giemsa staining. Chromosome numbers are given below each structure and the band numbers to the left. 'rDNA' is a region containing a cluster of repeat units for the 28S, 5.8S and 18S rRNA genes (see p. 122). 'Constitutive heterochromatin' is very compact chromatin that lacks genes (Section 8.1.1). Redrawn from Strachan and Read (1996).

**TABLE 9.5 Prokaryotic chromosomes[a]**

| Organism | Size (base pairs)[b] | *ORFs*[c] | Comments |
|---|---|---|---|
| *Mycoplasma genitalium* | 580,070 | 470 | Bacterium, smallest known cellular genome (Section 13.20) |
| *Mycoplasma pneumoniae* | 816,394 | 677 | Bacterium, causes pneumonia (Section 13.20) |
| *Borrelia burgdorferi* | 910,725 | 853 | Bacterium, spirochete, has linear chromosome,[d] causes Lyme disease (Sections 13.30 and 24.4) |
| *Chlamydia trachomatis* | 1,042,519 | 894 | Bacterium, obligate intracellular parasite, common human pathogen (Sections 13.26 and 23.6) |
| *Rickettsia prowazekii* | 1,111,523 | 834 | Bacterium, obligate intracellular parasite, causes epidemic typhus (Sections 13.12 and 24.3) |
| *Treponema pallidum* | 1,138,006 | 1,041 | Bacterium, spirochete, causes syphilis (Sections 13.30 and 23.6) |
| *Aquifex aeolicus* | 1,551,335 | 1,512 | Bacterium, hyperthermophile (Section 13.34) |
| *Methanococcus jannaschii* | 1,664,976 | 1,738 | Archaeon, methanogen (Section 14.3) |
| *Helicobacter pylori* | 1,667,867 | 1,590 | Bacterium, causes peptic ulcers (Section 24.11) |
| *Pyrococcus horikoshii* | 1,738,505 | 2,061 | Archaeon, hyperthermophile (Section 14.9) |
| *Methanobacterium thermoautotrophicum* | 1,751,377 | 1,855 | Archaeon, methanogen (Section 14.3) |
| *Haemophilus influenzae* | 1,830,137 | 1,743 | Bacterium, can cause disease (Section 23.1) |
| *Chlorobium tepidum* | 2,100,000 | 1,900 | Bacterium, anoxygenic phototrophic green bacterium (Section 13.29) |
| *Archaeoglobus fulgidus* | 2,178,400 | 2,436 | Archaeon, hyperthermophile (Section 14.6) |
| *Synechocystis* sp. | 3,573,470 | 3,168 | Bacterium, cyanobacterium (Section 13.24) |
| *Bacillus subtilis* | 4,214,810 | 4,100 | Bacterium, gram-positive genetic model (Section 13.19) |
| *Mycobacterium tuberculosis* | 4,411,529 | 3,924 | Bacterium, causes tuberculosis (Sections 13.22 and 23.3) |
| *Escherichia coli* | 4,639,221 | 4,288 | Bacterium, gram-negative genetic model (Section 13.10) |

*a* One sometimes finds the word "genome" used to refer to a prokaryotic chromosome. While in some cases this may actually be correct, remember that the genome actually includes *all* the genes found in an organism, even those of resident plasmids and viruses. In some cases these elements may not be found, and in some cases they may be very small or the genes carried by them may be of no importance to the organism's overall metabolism. However, this need not always be the case. For example, *Borrelia burgdorferi* contains 17 different plasmids consisting of 533 kilobase pairs (Table 6.2).

*b* Information on these and other genomes can be found in the TIGR Microbial Database (www.tigr.org/tdb/mdb/mdb.html), a Web Site maintained by The Institute for Genomic Research (TIGR), Rockville, MD, a not-for-profit research institute. The data for *Chlorobium tepidum* are estimates.

*c* Open Reading Frames (Section 6.10). The purpose of reporting ORFs is to predict the total number of proteins that an organism might encode. Therefore, constraints are placed on what is reported as an ORF. Of course, genes encoding known proteins are included, as are all ORF's that could encode a protein greater than 100 amino acid residues. Smaller ORFs are typically not included unless they show similarity to a gene from another organism or unless the codon usage is typical of the organism being studied.

*d* All other chromosomes in this list are circular.

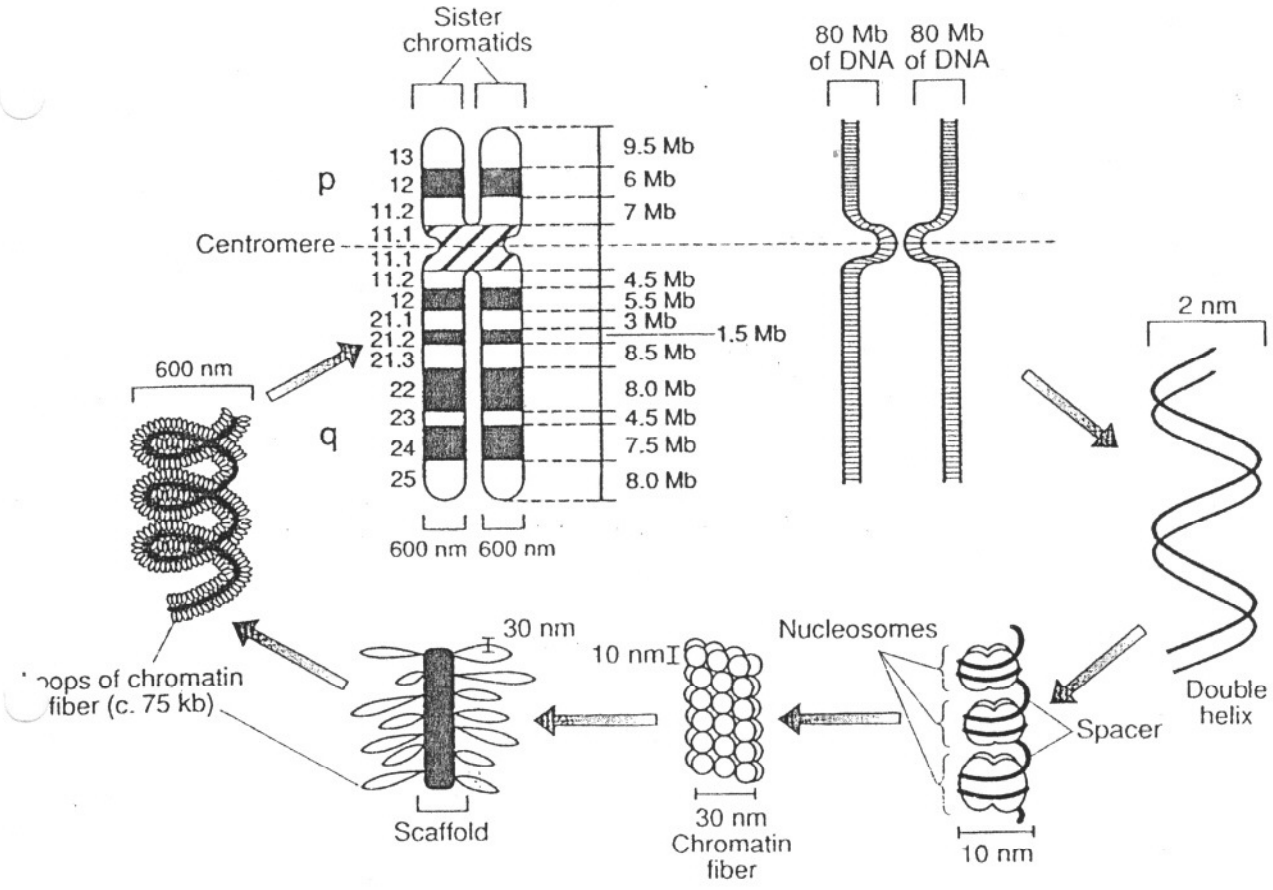

Sister chromatids
80 Mb of DNA
80 Mb of DNA
p
q
13
12
11.2
11.1
Centromere
11.1
11.2
12
21.1
21.2
21.3
22
23
24
25
9.5 Mb
6 Mb
7 Mb
4.5 Mb
5.5 Mb
3 Mb
1.5 Mb
8.5 Mb
8.0 Mb
4.5 Mb
7.5 Mb
8.0 Mb
600 nm
600 nm
600 nm
2 nm
Loops of chromatin fiber (c. 75 kb)
30 nm
Scaffold
10 nm
30 nm Chromatin fiber
Nucleosomes
Spacer
10 nm
Double helix

Table 1.1 Sizes of genomes

| Organism | Genome size (Mb) |
|---|---|
| **Prokaryotes** | |
| *Mycoplasma genitalium* | 0.58 |
| *Escherichia coli* | 4.64 |
| *Bacillus megaterium* | 30 |
| **Eukaryotes** | |
| Fungi | |
| *Saccharomyces cerevisiae* (yeast) | 12.1 |
| *Aspergillus nidulans* | 25.4 |
| Protozoa | |
| *Tetrahymena pyriformis* | 190 |
| Invertebrates | |
| *Caenorhabditis elegans* (nematode worm) | 100 |
| *Drosophila melanogaster* (fruit fly) | 140 |
| *Bombyx mori* (silkworm) | 490 |
| *Strongylocentrotus purpuratus* (sea urchin) | 845 |
| *Locusta migratoria* (locust) | 5000 |
| Vertebrates | |
| *Fugu rubripes* (pufferfish) | 400 |
| *Homo sapiens* (humans) | 3000 |
| *Mus musculus* (mouse) | 3300 |
| Plants | |
| *Arabidopsis thaliana* (vetch) | 100 |
| *Oryza sativa* (rice) | 565 |
| *Pisum sativum* (pea) | 4800 |
| *Zea mays* (maize) | 5000 |
| *Triticum aestivum* (wheat) | 17 000 |
| *Fritillaria assyriaca* (fritillary) | 120 000 |

Data taken from Brown (1998).

Table 5.1 Examples of codon bias in the human genome

| Amino acid | Codons | Codon frequency |
|---|---|---|
| Alanine | GCA | 22% |
| | GCC | 41% |
| | GCG | 11% |
| | GCT | 26% |
| Threonine | ACA | 27% |
| | ACC | 38% |
| | ACG | 12% |
| | ACT | 23% |
| Valine | GTA | 11% |
| | GTC | 25% |
| | GTG | 48% |
| | GTT | 17% |

See Section 10.1.2 for details concerning the genetic code. The data are taken from the Internet site maintained by Yasukazu Nakamura, Laboratory of Gene Structure 2, Kazusa DNA Research Institute, Japan: <http://www.dna.affrc.go.jp/~nakamura/codon.html>

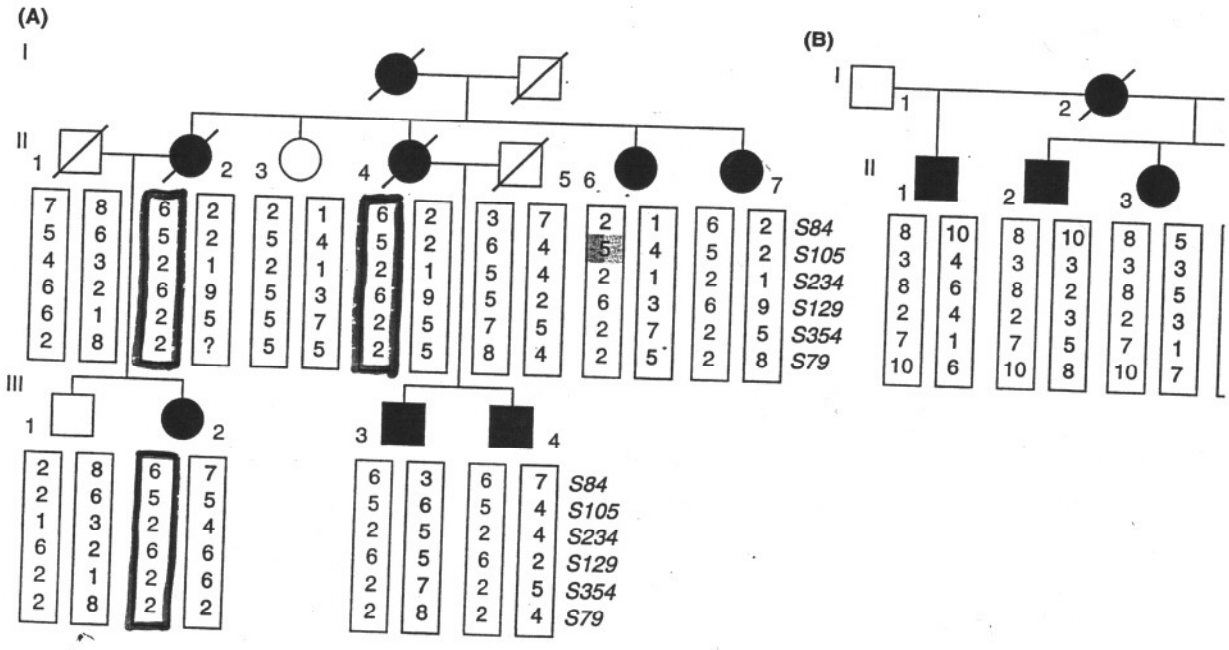
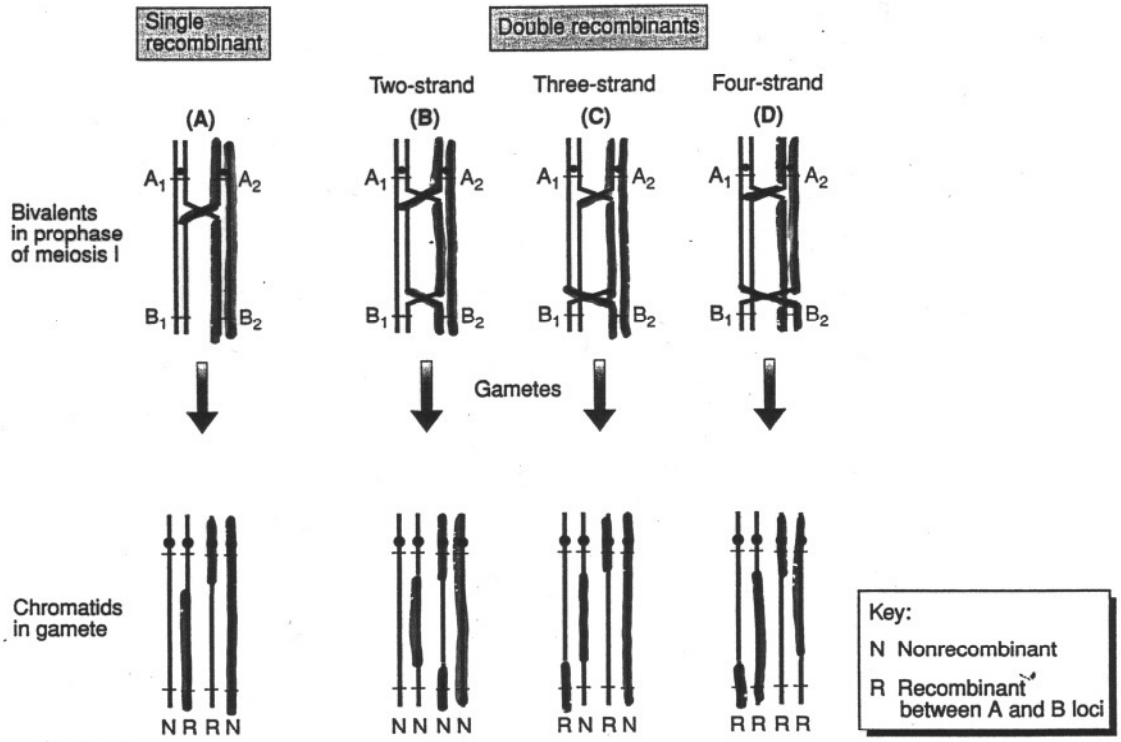



**Figure 14.5:** Crossover analysis seeks to map a gene by defining flanking proximal and distal recombinants.

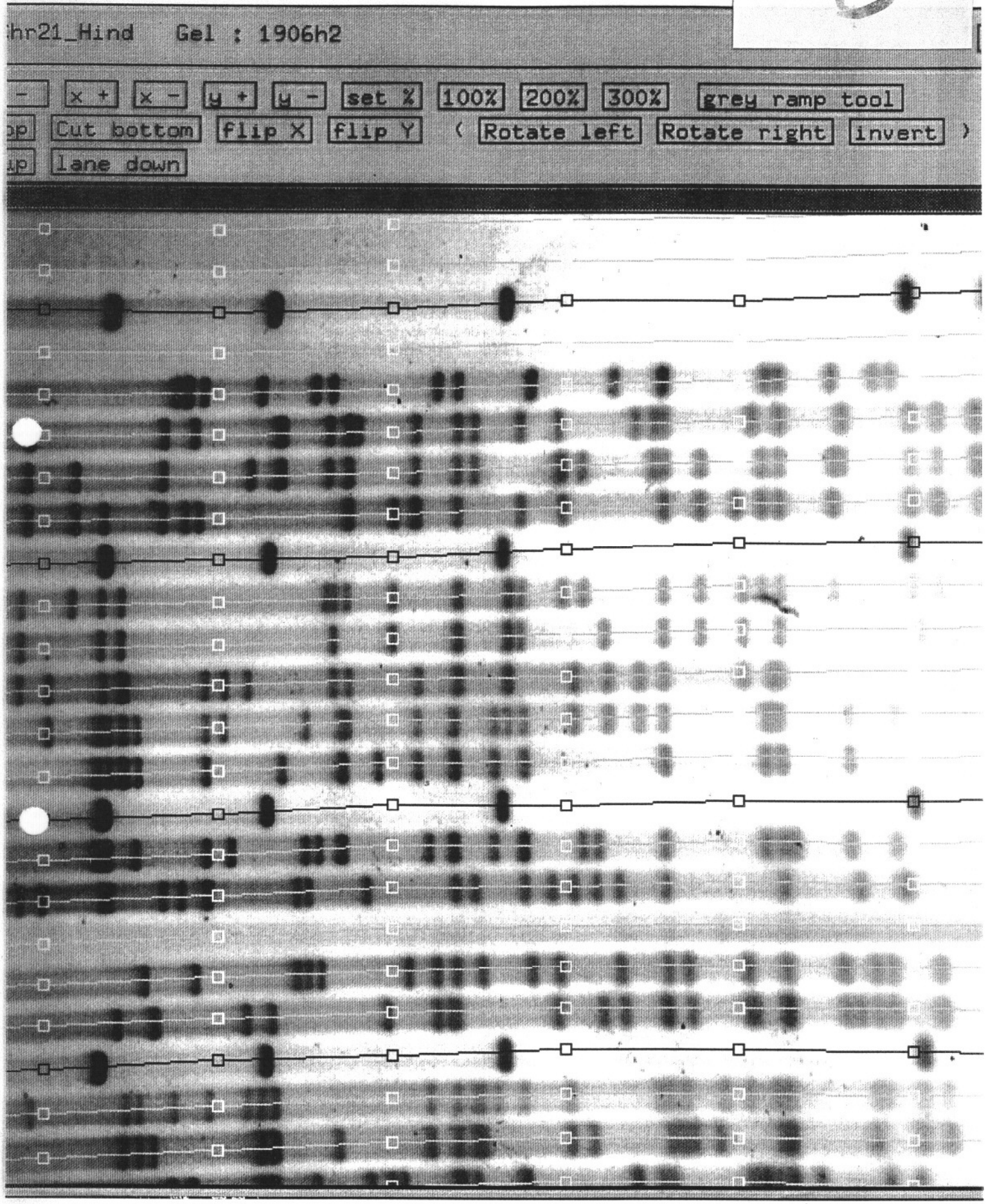
The figure shows haplotype analysis in two pedigrees with a dominantly inherited skin disorder, Darier's disease, which hac been shown to be linked to markers on 12q. Genotypes in $II_1$, $II_2$, $II_4$ and $II_5$ in pedigree **(A)** are inferred.
**(A)** In this family the disease gene segregates with the marker haplotype 6-5-2-6-2-2 between *D12S84* (proximal) and *D12*
A crossover in $II_6$ indicates that the disease gene must map distal to *D12S84* (the positioning of *D12S105* is ambiguous bec presumed homozygosity for allele 5 in $I_1$ – compare the genotypes for $II_3$ and $II_6$).
**(B)** In this family the disease gene can be deduced to be proximal to *D12S129*. The combined data indicate that the Darier's gene must map between the proximal marker *D12S84* and the distal marker *D12S129*. Reproduced from Carter *et al.* (1994 permission from Academic Press.



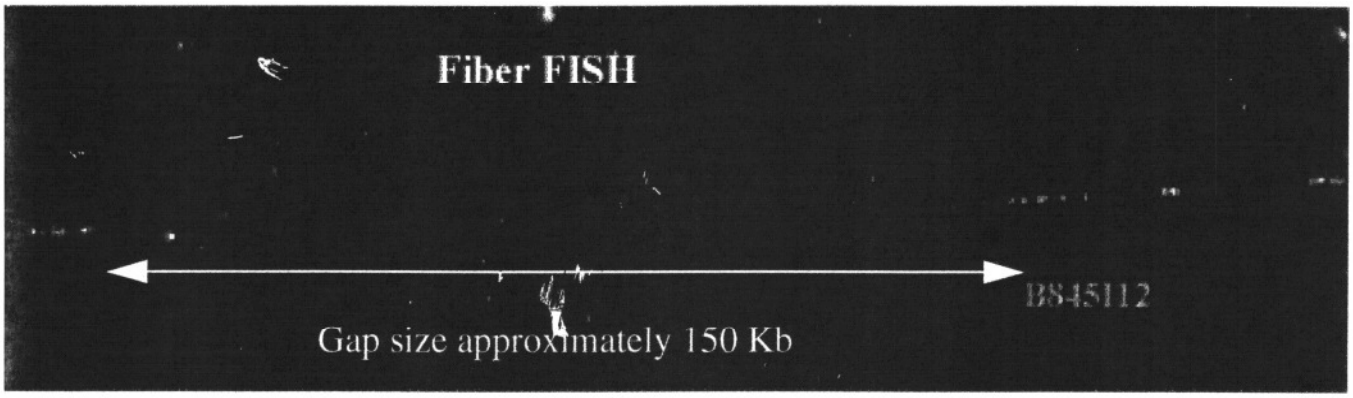

**ure 12.2:** Single and double recombinants.

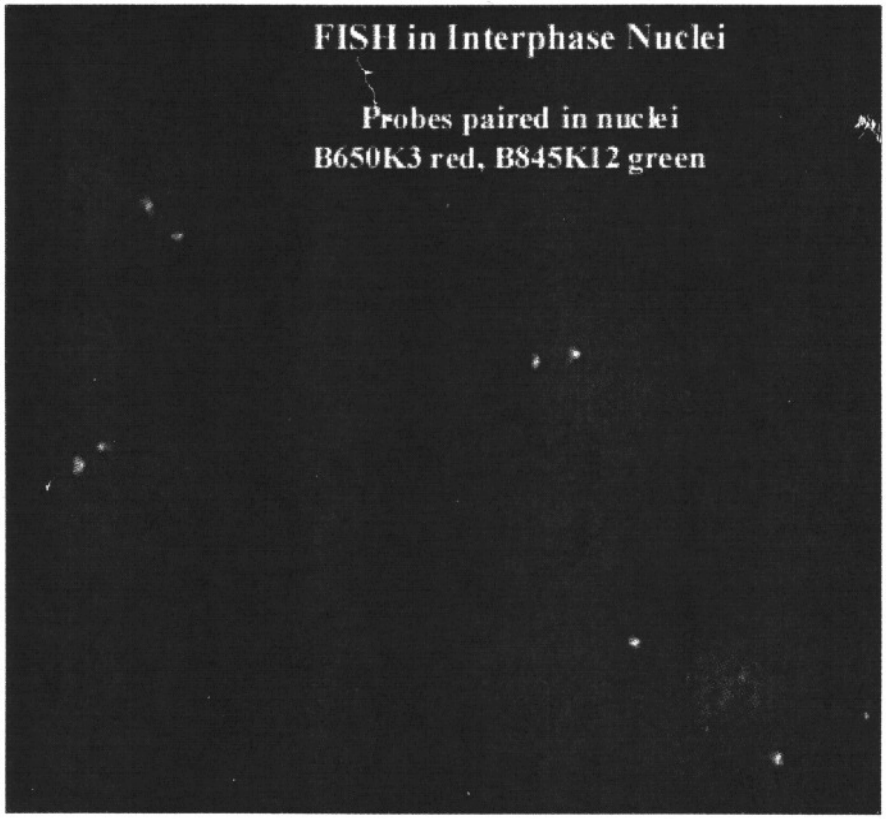
h crossover involves two of the four chromatids of the two synapsed homologous chromosomes. The black chromosome carries es $A_1$ and $B_1$ at two loci, while the red chromosome carries alleles $A_2$ and $B_2$. Gametes in which the chromatid is the same color at two loci are nonrecombinant for these loci, those where the chromatids are different colors are recombinant.
A single crossover generates two recombinant and two nonrecombinant chromatids.
A two-strand double crossover leaves flanking markers nonrecombinant on all four chromatids.
A three-strand double crossover leaves flanking markers recombinant on two of the four strands.
A four-strand double crossover generates 100% recombinants. The three types of double crossover occur in random proportions, so average effect of a double crossover is to give 50% recombinants.

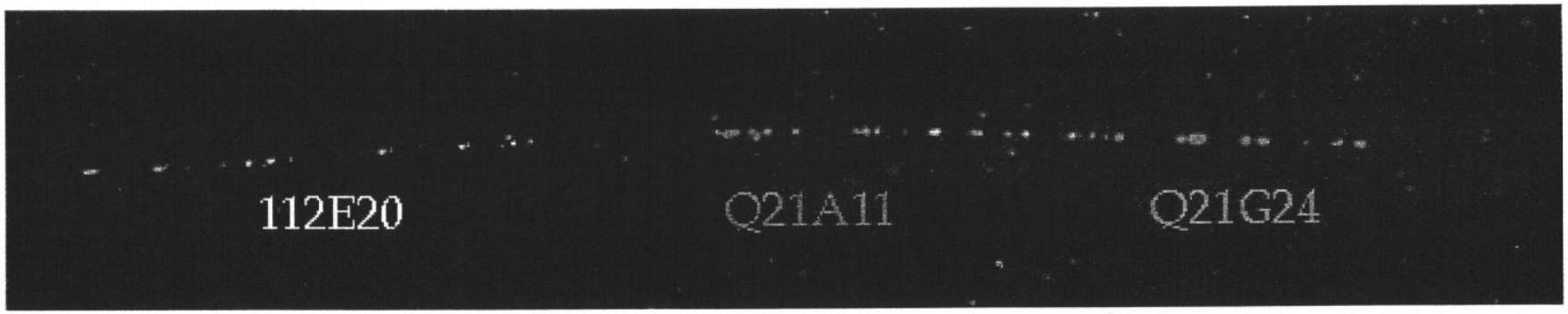
B


hr21_Hind Gel : 1906h2
x +
x -
y +
y -
set %
100%
200%
300%
grey ramp tool
Cut bottom
flip X
flip Y
Rotate left
Rotate right
invert
lane down

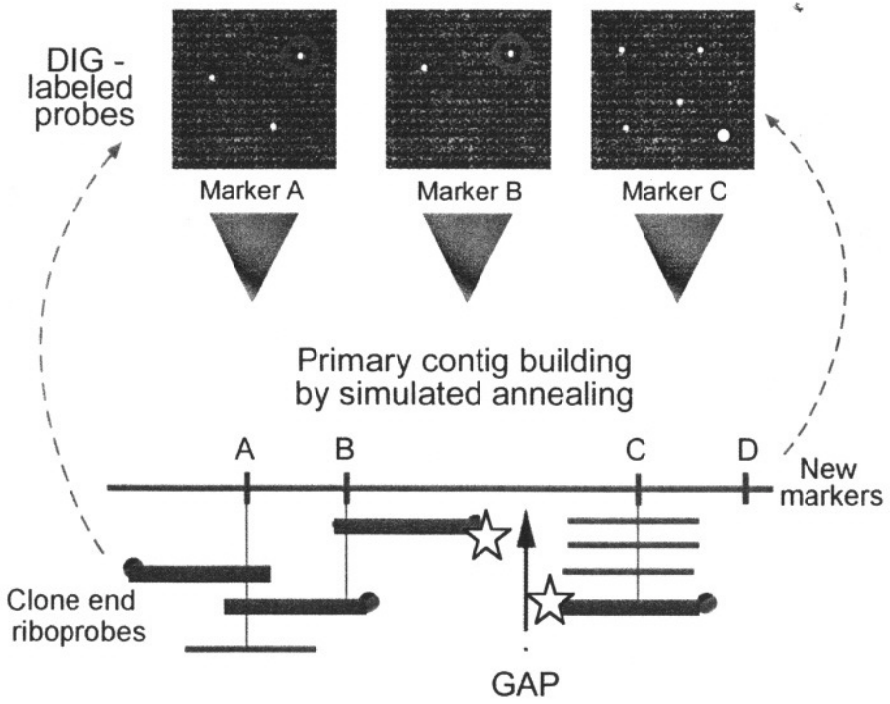
# Estimation of Gap Size by FISH

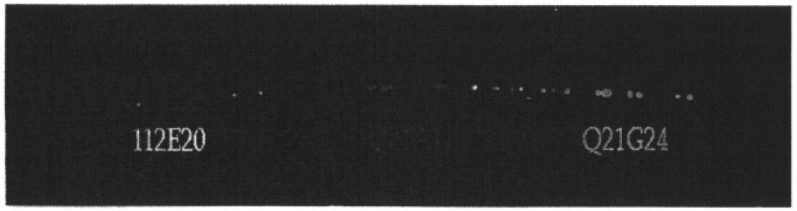



112E20
Q21A11
Q21G24

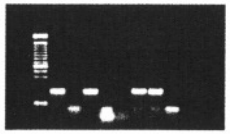
# Building "sequence-ready " maps by high-throughput hybridization

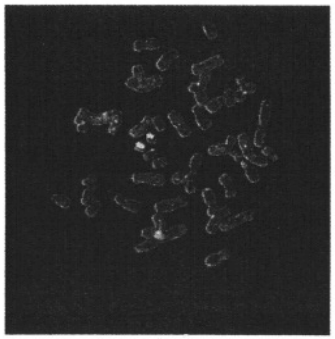



Visual mapping by fibre-FISH
resolution < 10-300 kb >

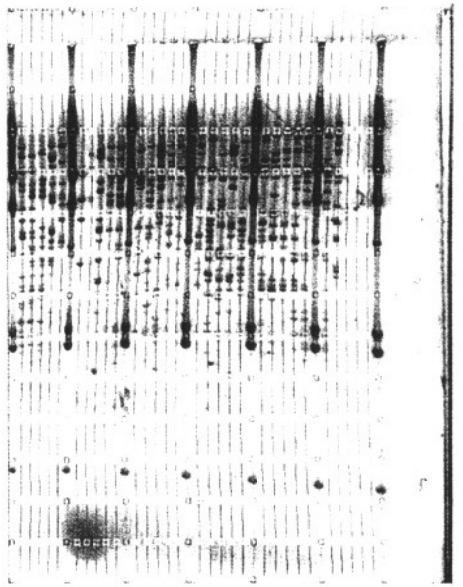
STS content validation





Metaphase FISH

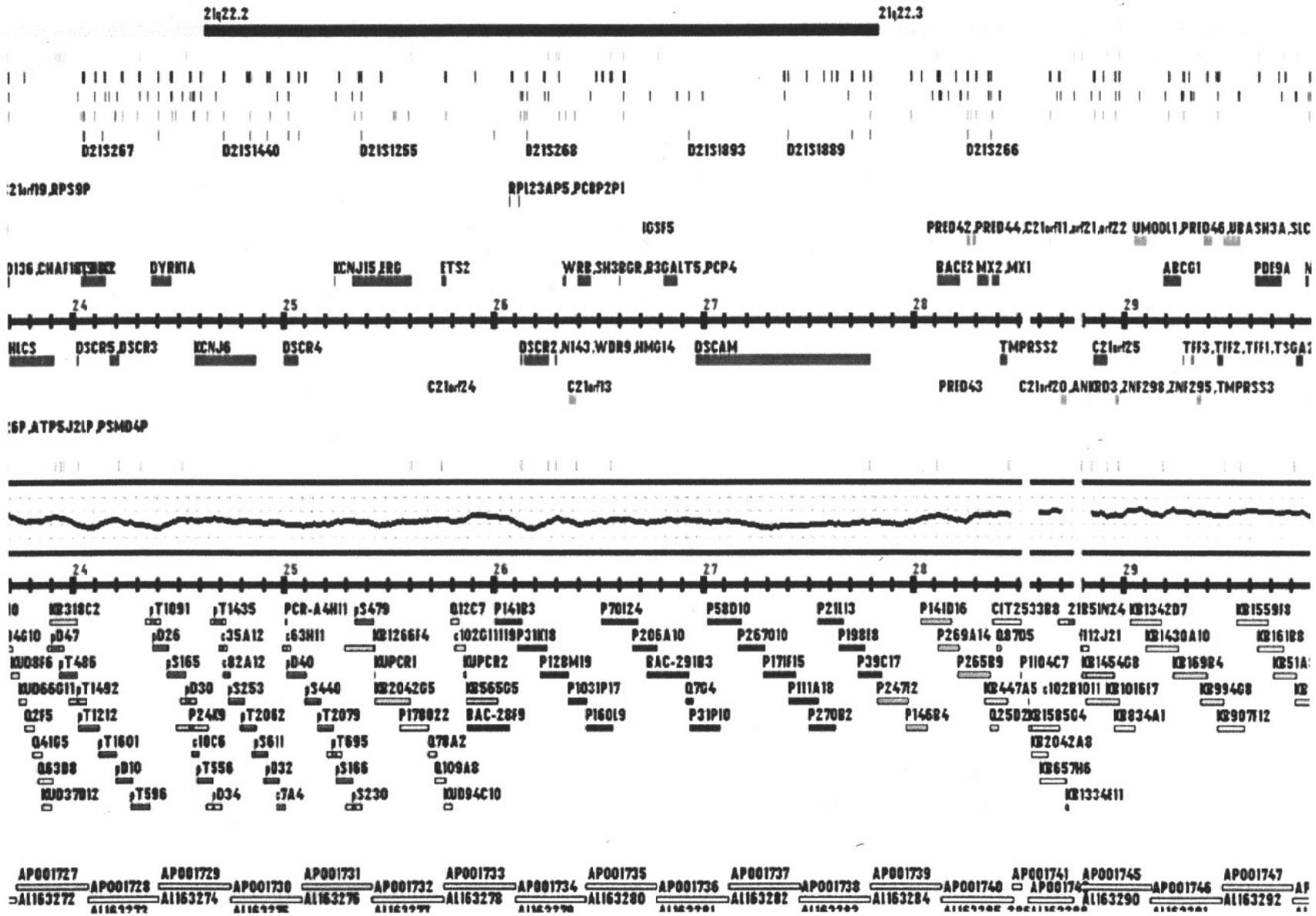
Restriction Fingerprinting



IM3/FPC programs

Minimum Tiling Path

Clone contig covering whole chr. 21


21q22.2
21q22.3
D21S267
D21S1440
D21S1255
D21S268
D21S1893
D21S1889
D21S266
RPL23AP5,PCBP2P1
DYRK1A
KCNJ15,ERG
ETS2
WRB,SH3BGR,B3GALT5,PCP4
BACE2,MX2,MX1
ABCG1
PDE9A
24
25
26
27
28
29
HLCS
DSCR5,DSCR3
KCNJ6
DSCR4
DSCR2,N143,WDR9,HMG14
DSCAM
TMPRSS2
C21orf25
TFF3,TFF2,TFF1,TSGA
C21orf24
C21orf13
PRED43
C21orf20,ANKRD3,ZNF298,ZNF295,TMPRSS3
ATP5J2L,PSMD4P
AP001727
AP001728
AP001729
AP001730
AP001731
AP001732
AP001733
AP001734
AP001735
AP001736
AP001737
AP001738
AP001739
AP001740
AP001741
AP001745
AP001746
AP001747
AL163272
AL163274
AL163276
AL163278
AL163280
AL163282
AL163284
AL163290
AL163292

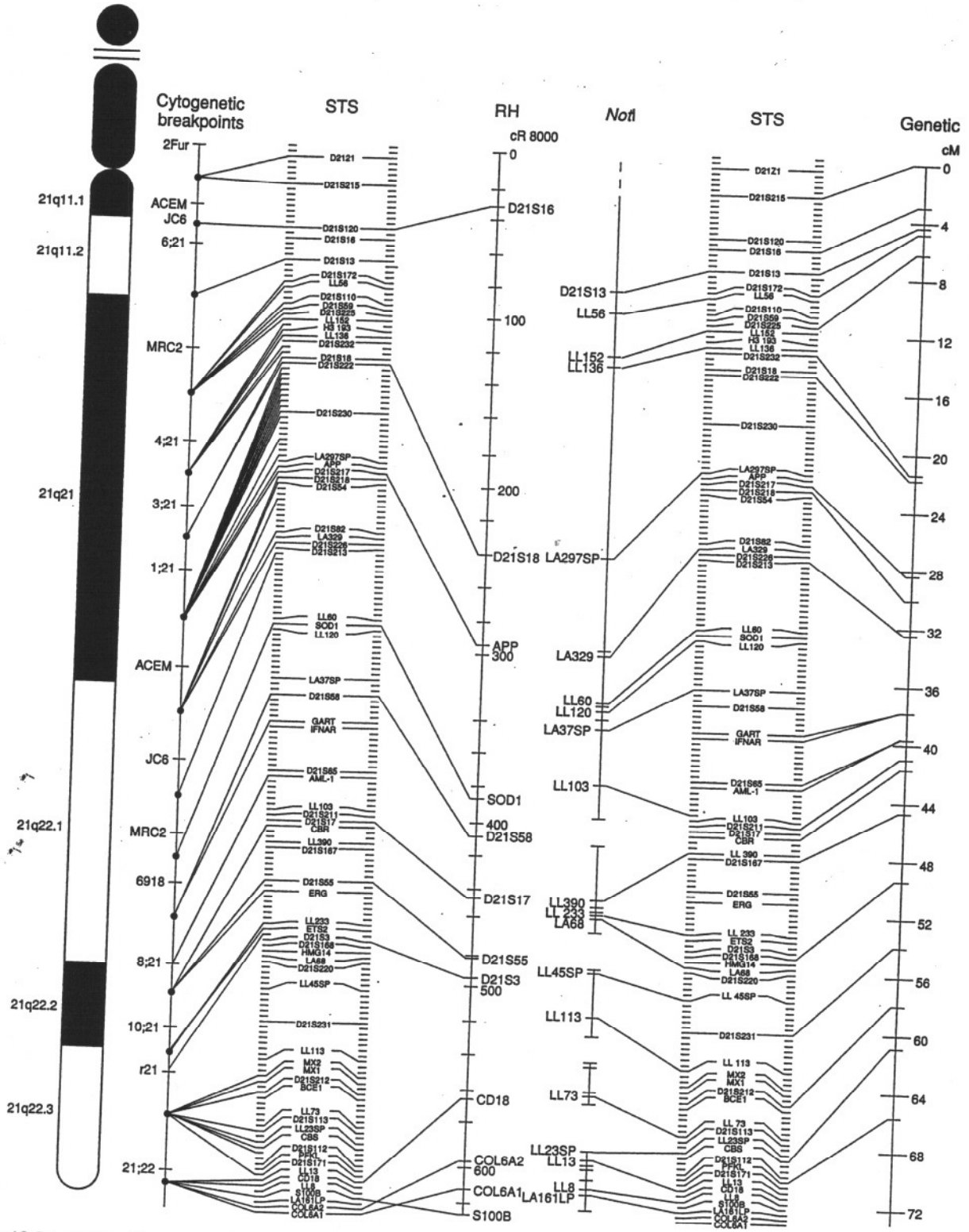



**Figure 13.2:** Several types of physical map are being constructed for human chromosomes.

The figure shows integration of several physical maps for the long arm of human chromosome 21. Next to the standard cytogenetic map on the left are the positions of chromosome 21 breakpoints observed largely from studying chromosome 21 translocations (6;21, 4;21, 3;21, 2;21, etc.). The STS map is shown twice to facilitate comparisons with the other maps, which also include the genetic linkage map of Chumakov *et al.* (1992), the PFGE-based *Not*I restriction map (Ichikawa *et al.*, 1993) and the radiation hybrid map (measured in centiRays after exposure to 8000 rad). Reproduced with permission from *Nature*, vol. 359, page 385. Copyright 1992 Macmillan Magazines Limited.

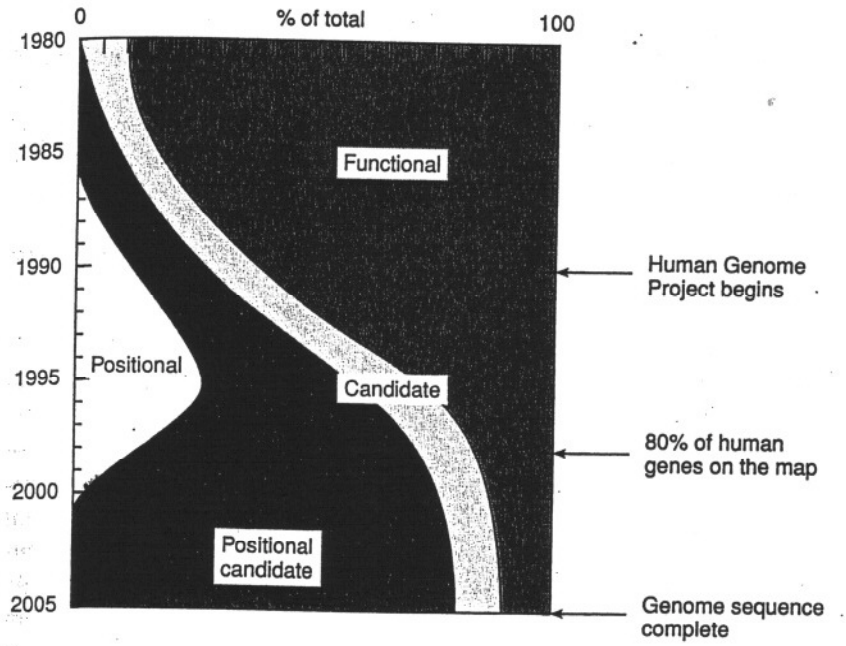



**Figure 14.3:** Positional candidate gene approaches are set to dominate other methods of identifying human disease genes.

Modified from Collins (1995) with permission from Nature America Inc.

**TABLE 14.2** Taxonomy of some extremely halophilic Archaea

| Genus | Morphology | DNA (mol % GC) | Habitat and comments |
|---|---|---|---|
| ***Halobacterium*** | Rods | | |
| *H. salinarum* | | 66–71 | Isolated from salted fish, hides, hypersaline lakes; related organisms, probably all the same species are *H. halobium* and *H. cutirubrum* |
| ***Halorubrum*** | Rods | | |
| *H. sodomense* | | 68 | Dead Sea; requires high $Mg^{2+}$ |
| *H. saccharovorum* | | 71 | Salterns; uses sugars |
| *H. lacusprofundi* | Pleomorphic rods | 65–66 | From an Antarctic lake; grows at 4°C |
| *H. trapanicum* | | 64 | From salt works, Trapani, Italy |
| *H. vacuolatum* | | 63 | Alkaliphilic; contains gas vesicles |
| ***Halobaculum*** | Rods | | |
| *H. gomorrense* | | 70 | Dead Sea; requires high $Mg^{2+}$, optimal NaCl about 1.5 M |
| ***Haloferax*** | Flattened disc or cup-shaped | | |
| *H. volcanii* | | 63–66 | Dead Sea; requires high $Mg^{2+}$ |
| *H. mediterranei* | | 60–62 | Salterns; uses starch; contains gas vesicles |
| *H. gibbonsii* | | 62 | Marine salterns in Spain |
| *H. denitrificans* | | 64 | Baja California saltern; capable of dissimilative nitrate reduction |
| ***Haloarcula*** | Irregular discs, triangles, rectangles | | |
| *H. vallismortis* | | 65 | Death Valley, CA |
| *H. hispanica* | | 63 | Marine salterns in Spain |
| ***Halococcus*** | Cocci | | |
| *H. morrhuae* | | 60–66 | Salted fish |
| *H. saccharolyticus* | | 59 | Salterns; uses sugars |
| ***Natronobacterium*** | Rods | | |
| *N. gregoryi* | | 65 | Isolated from highly saline soda lakes; optimum pH for growth, 9.5 |
| ***Natrialba*** | Rods | | |
| *N. magadii* | | 63 | Alkaliphile from Lake Magadi, Kenya (see Figure 14.2*c*) |
| ***Natronosomonas*** | Rods | | |
| *N. pharaonis* | | 64 | |
| ***Natrononococcus*** | Cocci | | |
| *N. occultus* | | 64 | Highly saline soda lakes |
| *N. amylolyticus* | | 63 | Lake Magadi; pigmented orange-red; uses starch |

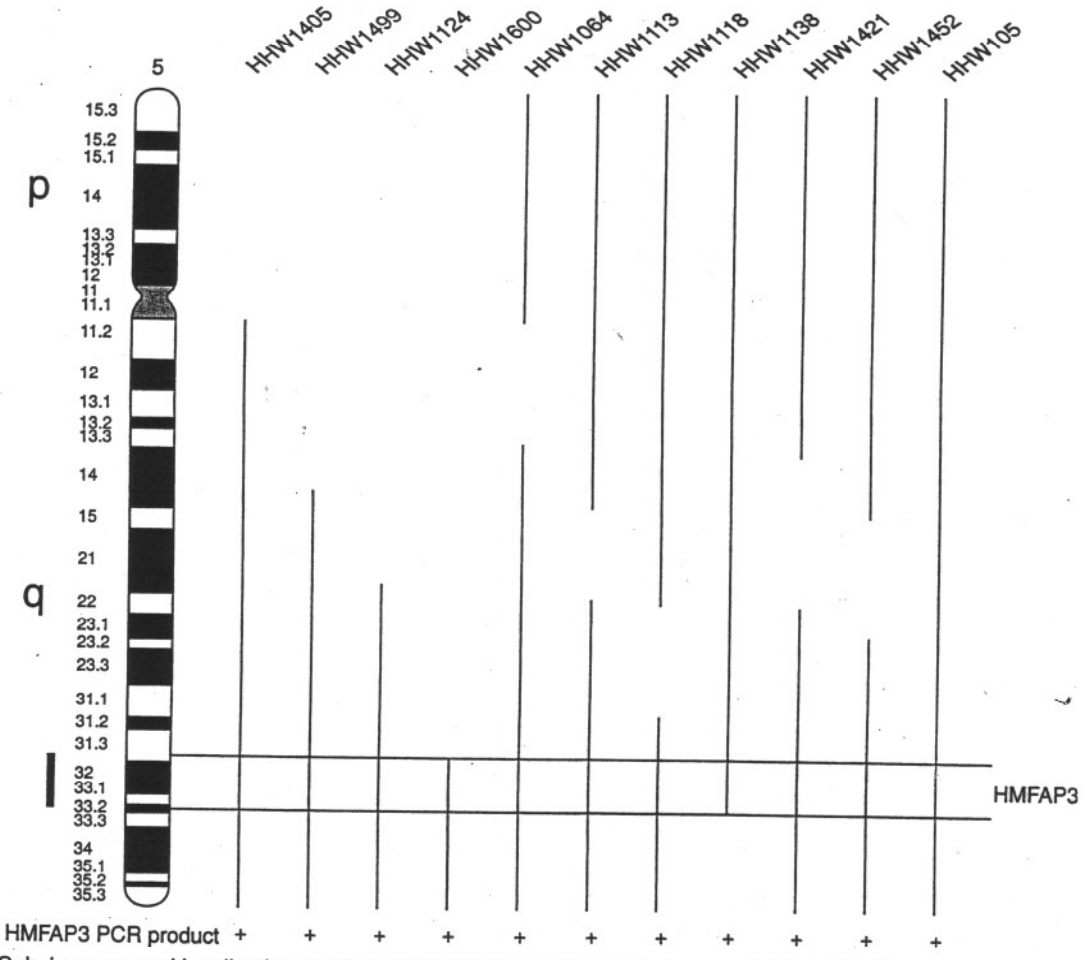
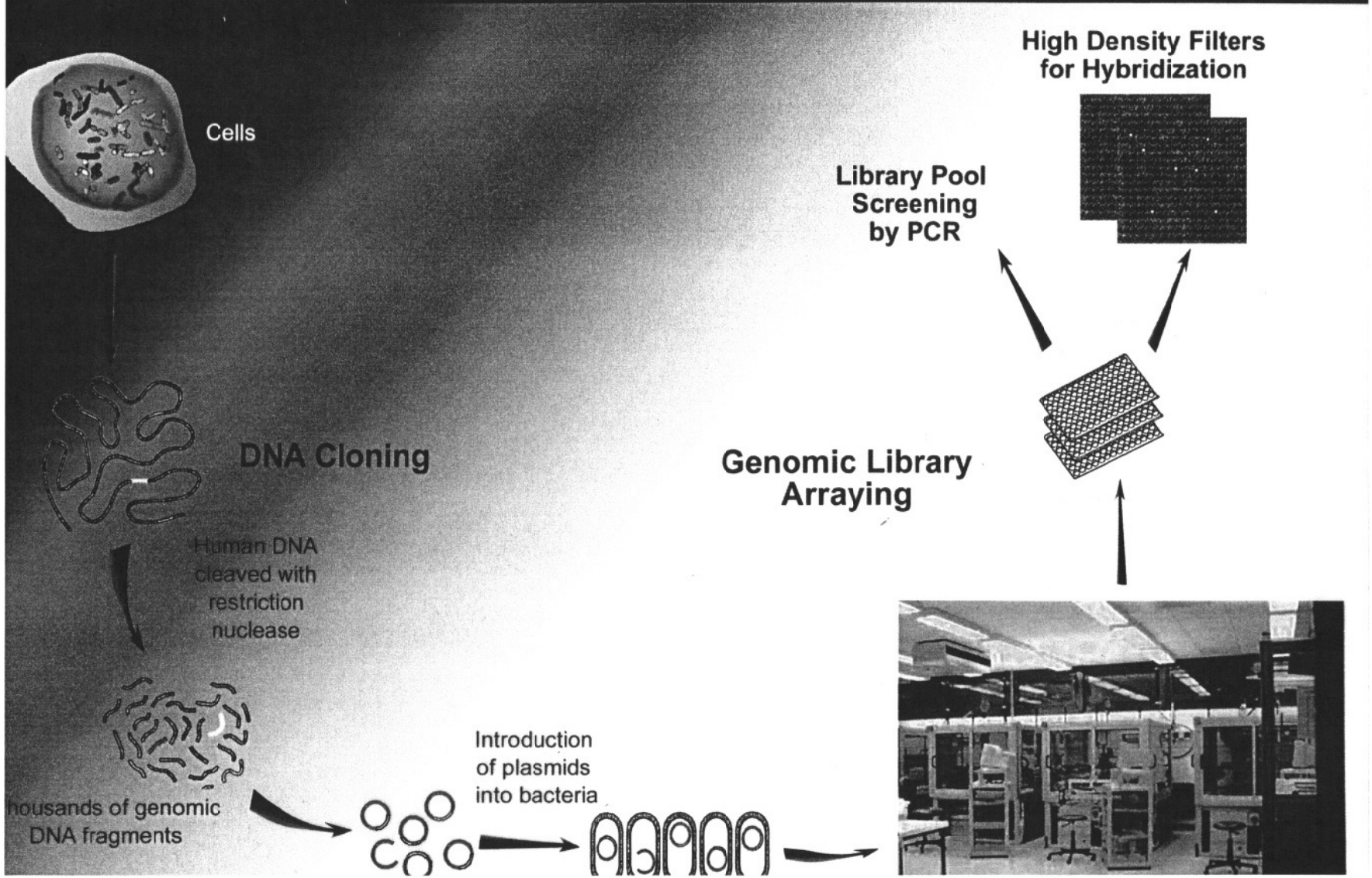



**Figure 11.2:** Subchromosomal localization can be achieved by mapping against a panel of hybrid cells containing transl deletion chromosomes.

The figure illustrates PCR-based mapping of the human microfibrillar protein MFAP3 using a panel of 5q translocation and hybrids. Vertical black bars to the right indicate the extent of human chromosome 5 sequences which are retained in the hy Hybrids HHW1405, 1499, 1124 and 1600 contain translocation chromosomes with 5q breakpoints and retention of the segi the breakpoint. By contrast, translocation hybrid HHW1138 retains material proximal to the 5q breakpoint. Hybrids HHW10 1118, 1421 and 1452 have different interstitial deletions of 5q. The solid red vertical bar to the left indicates the inferred subchromosomal location as defined by breakpoints in hybrids HHW1600 and HHW1138 (red horizontal lines near bottom) from Abrams *et al.* (1995) with permission from Academic Press Inc.

# apping strategy using arrayed genomic libraries (BACs, PACs, ....)



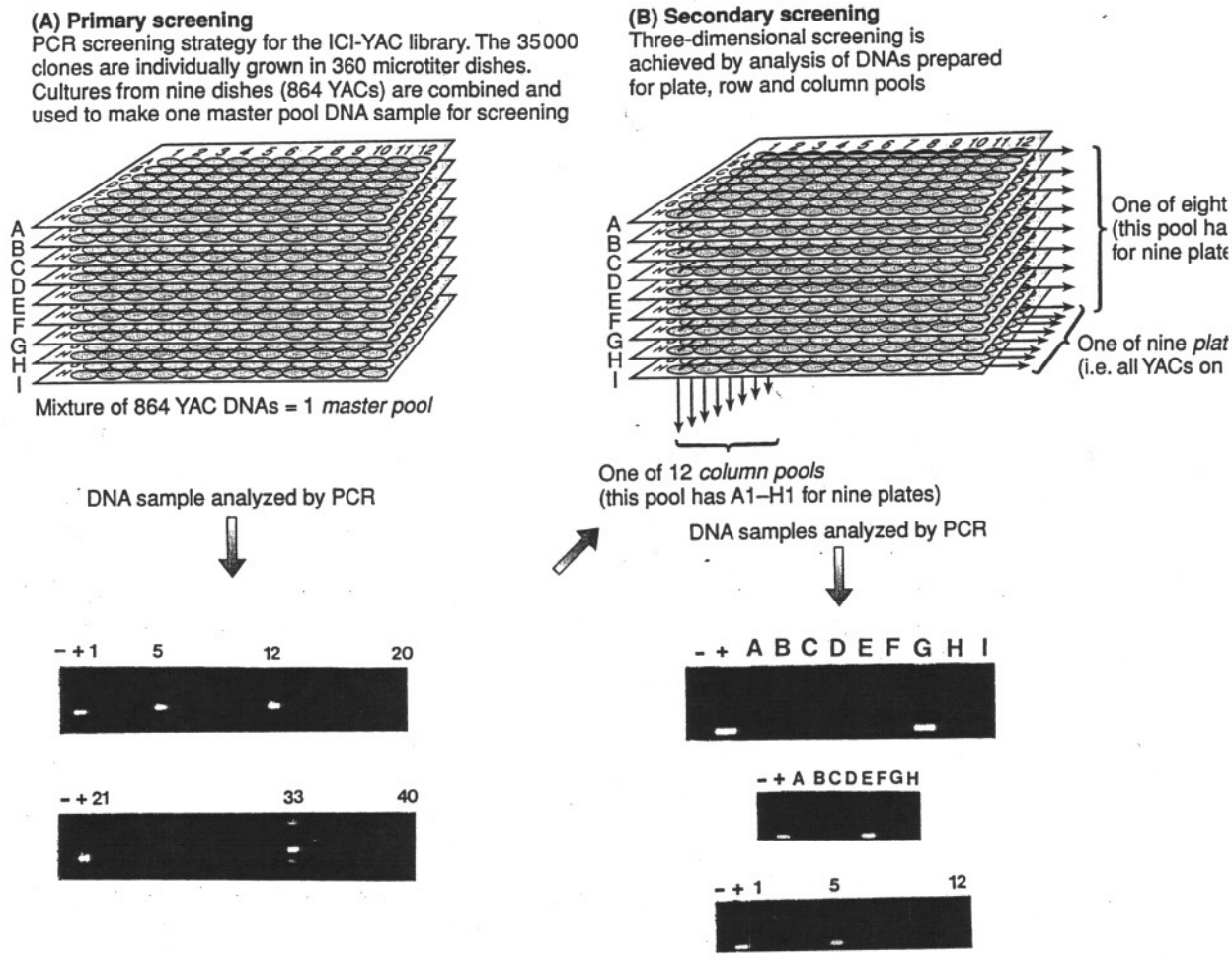



**Figure 11.14:** Screening of YAC libraries by PCR.

A YAC library can be screened for the presence of clones containing a specific sequence, provided there is a specific PCR that sequence. Amongst other applications, this provides a convenient method for chromosome walking using YACs: a PC sequence at the end of one YAC can be used to identify other YACs with overlapping sequences. The example illustrates the human ICI YAC library generated by Anand *et al.* (1990). Approximately 35 000 individual clones were individually dep 96 wells of 360 microtiter dishes. To facilitate screening, a total of 40 *master pools* were generated by combining all 864 cl 9 microtiter dishes (plates A–I). Modified from Jones *et al.* (1994) with permission from Academic Press Inc.

**(A)** *Primary screening* involves PCR assay of the 40 master pools. In this example, three master pools were positive when against positive (+) and negative (–) controls: pools 5, 12, 23.

**(B)** *Secondary screening* identifies single YACs by assaying different subsets of the 864 YACs in a positive master pool, ir master pool 12. Three-dimensional screening of each of nine *plate pools* (of 96 YACs each), eight *row pools* (of 106 YACs *column pools* (of 72 YACs each) and identified a positive YAC in plate 12G (top panel), row E (middle panel), column 5 (bc The example here involved screening for YACs containing an anonymous X chromosome sequence. Photos were kindly p Dr Sandie Herrell, University of Newcastle upon Tyne.

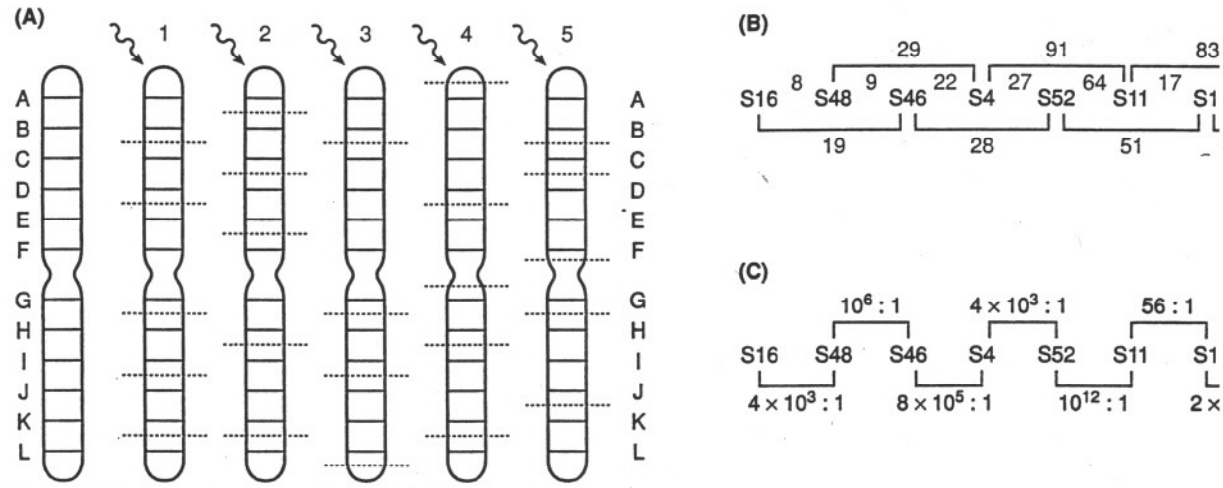



**Figure 11.4:** Constructing radiation hybrid maps.

**(A)** Breakpoints occur randomly. Five possible examples of breakpoints (dashed red lines) on the same type of chromosc Markers close together will tend to occur on the same fragment, e.g. A and B in all cases other than example 2. Thus, if a hybrid contains marker A it will frequently also contain marker B, but rarely a distant marker such as L.

**(B)** Ordering of markers on human 21q. The order of markers *D21S16–D21S8* as inferred by Cox *et al.* (1990) from radi mapping is shown. Figures on the top panel refer to distances between markers in centiRays$_{8000}$. For example, the S16– $cR_{8000}$: at a radiation dose of 8000 rad, there is 8% frequency of breakage between them, and so a 92% chance they will on one fragment.

**(C)** Odds ratios refer to the likelihood of the indicated order for pairs of markers compared with that with the markers inve example, the calculated likelihood for the order S16–S48–S46–S4 is $10^6$ times greater than for the order S16–S46–S48–

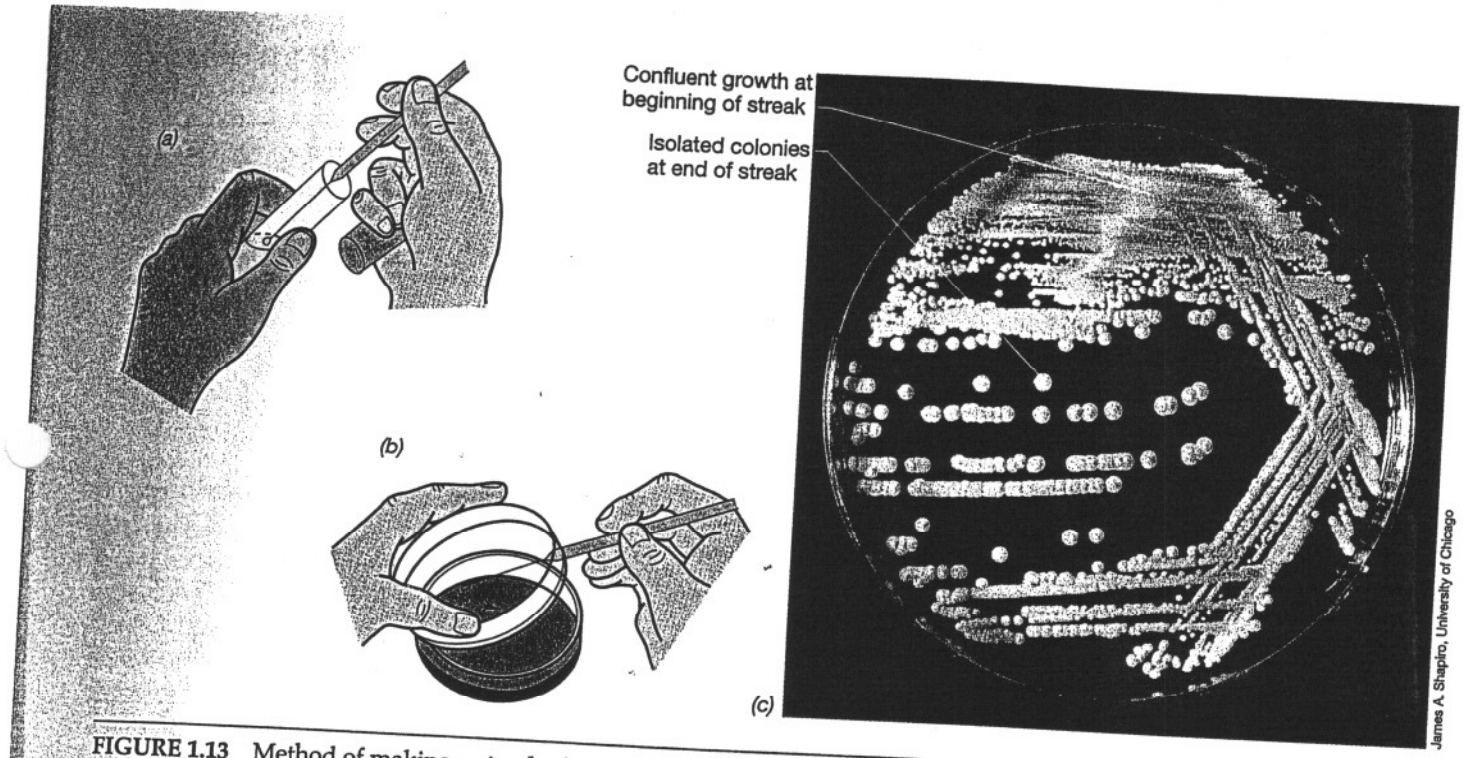



**FIGURE 1.13** Method of making a streak plate to obtain pure cultures. (a) Loop is sterilized, and then a loopful of inoculum is removed from tube. (b) Streak is made over a sterile agar plate, spreading out the organisms. Following the initial streak, subsequent streaks are made at angles to it, the loop being resterilized between streaks. (c) Appearance of the streaked plate after incubation. Colonies of the bacterium *Micrococcus luteus* grown on blood agar plates. It is from such well-isolated colonies that pure cultures can usually be obtained.

**Figure 21.1** DNA content of the haploid genome is related to the morphological complexity of lower eukaryotes, but varies extensively among the higher eukaryotes. The range of DNA values within a phylum is indicated by the shaded area.