6 Genes, Genomics, and Chromosomes

Review the Concepts

1. A gene is commonly defined as the entire nucleic acid sequence that is necessary for the synthesis of a functional gene product (RNA or polypeptide). This definition includes introns and the regulatory regions, e.g., promoters and enhancers of the gene. A complex transcription unit consists of multiple exons. These multiple exons can be spliced together in a number of different ways to generate different mRNAs. For example, assume that a hypothetical gene contains four exons, 1, 2, 3, and 4. Assuming that all mRNA from this gene must start with exon 1 and end with exon 4, which is not true for all genes, the possible combinations of alternatively spliced exons are exons 1-2-3-4, exons 1-2-4, exons 1-3-4, and exons 1-4. Because the sequence of each of these alternatively spliced mRNAs differ, the amino acid sequence of the encoded protein would differ.

2. Single or solitary genes are present once in the haploid genome. In multicellular organisms roughly 25–50% of the protein-coding genes are solitary genes. Gene families are sets of duplicated genes that encode proteins with similar but nonidentical amino acid sequences. An example of a gene family is the β-like globin genes. Pseudogenes are copies of genes that are nonfunctional even though they seem to have the same exon-intron structure as a functional gene. Pseudogenes probably arise from a gene duplication followed by the accumulation of mutations that render the gene nonfunctional. Tandemly repeated genes are present in a head-to-tail array of exact or almost exact copies of genes. Examples of tandemly repeated genes include ribosomal RNAs and RNAs involved in RNA splicing.

3. Satellite, or simple sequence, DNA can be categorized as microsatellite or minisatellite DNA depending upon the size of the repeated DNA sequence. Microsatellites contain repeats that contain 1–13 base pairs. Minisatellites consist of repeating units of 14–100 base pairs, present in relatively short regions of 1 to 5 kb made up of 20–50 repeat units. The number of copies of the tandemly repeated DNA sequences varies widely between individuals. DNA fingerprinting is a technique that examines the number of repetitive units at a specific genetic locus for several separate loci. This technique can distinguish and identify individuals on the basis of differences in the number of repeats in their micro- and minisatellite simple sequence DNA.

4. A bacterial insertion sequence, or IS element, is a member of the class of mobile DNA elements. The IS element usually contains inverted repeats at the end of the insertion sequence. Between the inverted repeats is a region that encodes the enzyme transposase. Transposition of the IS element is a three-step process. First, transposase excises the IS element in the donor DNA; second, it makes staggered cuts in a short sequence in the target DNA; and, third, it ligates the 3′ termini of the IS element to the 5′ ends of the cut donor DNA. The final step involves a host-cell DNA polymerase, which fills in the single-stranded gaps, generating 5–11 base pair short direct repeats that flank the IS element before DNA ligase joins the free ends.

5. Retrotransposons transpose through an RNA intermediate. One class of retrotransposons, the LTR retrotransposons, contain long terminal repeats (LTRs) at their ends and a central protein-coding region that encodes the enzymes reverse transcriptase and integrase. The retrotransposon is first transcribed into RNA by host RNA polymerase. This RNA intermediate is then converted into DNA by the action of reverse transcriptase primed with a cellular tRNA in the cytoplasm. The double-stranded DNA copy generated is then imported into the nucleus and inserted into chromosomal DNA by the action of

an integrase that is similar to the transposases of DNA transposons. Retrotransposons that lack LTRs transpose by a different mechanism. Non-LTR retrotransposons, of which LINES are an example, consist of direct repeats that flank a region that encodes two proteins: ORF1, an RNA-binding protein, and ORF2, which is similar to reverse transcriptase. The LINE element is first transcribed by host RNA polymerase and exported to the cytoplasm, resulting in the translation of ORF1 and ORF2. These proteins bind the LINE RNA and import it into the nucleus, where ORF2 makes staggered nicks in A/T-rich target DNA. The resulting T-rich strand of chromosomal DNA then hybridizes to the poly(A) tail at the 3′-end of the LINE RNA and primes reverse transcription of the RNA by ORF2 protein. The RNA strand of the resulting RNA/DNA hybrid is replaced by DNA and both resulting DNA strands are ligated to the chromosome ends generated by the original ORF2 cut through the action of host-cell enzymes that normally replace the RNA primers of Okazaki fragments and ligate them together during cellular DNA synthesis.

6. Insertion of transposons can generate spontaneous mutations that may influence evolution. In addition, unequal crossing over between homologous mobile elements at different chromosomal locations leads to exon duplications, gene duplications, and chromosomal rearrangements that can generate new combinations of exons. Subsequent divergence of duplicated genes leads to members of gene families with distinct functions. The inclusion of flanking DNA during transposition also results in the movement of genomic DNA to another region of the genome. This can result in new combinations of exons, an evolutionary process known as exon shuffling, as well as new combinations of transcriptional control regions.

7. Mitochondrial genomes encode their own mitochondrial rRNAs, tRNAs, and some essential mitochondrial proteins. Plant mitochondrial genomes are generally larger and more variable in size than the mitochondrial genomes of other organisms. For example, watermelon mitochondrial DNA is about 330 kb, while human mitochondrial DNA is about 16.6 kb, and that of yeast is about 78 kb. Furthermore, plant mitochondrial genomes contain multiple mitochondrial DNA molecules that recombine with one another, whereas mammals contain a single circular mitochondrial DNA molecule. In addition, plant mitochondrial genomes encode a unique plant 5S mitochondrial rRNA and the α subunit of the F_1 ATPase. Plant mitochondria use the standard genetic code for protein translation, whereas the mitochondria from mammals and fungi (yeast) use a modified genetic code.

8. Similarities between bacteria, mitochondria, and chloroplasts reflect the proposed endosymbiotic origin of mitochondria and chloroplasts. Mitochondrial and bacterial ribosomes resemble each other and differ from eukaryotic cytosolic ribosomes in their RNA and protein compositions, their size, and their sensitivity to antibiotics. Bacterial and mitochondrial ribosomes are sensitive to chloramphenicol but resistant to cycloheximide. Eukaryotic cytosolic ribosomes are sensitive to cycloheximide and resistant to chloramphenicol. Also, comparing the mitochondrial DNA of multiple classes of eukaryotes, both unicellular and multicellular, all can be seen to derive from a common ancestor with a genome similar to contemporary symbiotic bacteria that invade host eukaryotic cells. Mitochondrial DNA in different contemporary eukaryotes can be derived from this common ancestor by deletion of different sets of genes in the mitochondria of different eukaryotes and the transfer of genes essential for mitochondrial function to the nucleus.

9. Genomics is defined as the genome-wide analysis of the organization, structure and expression of genes. Proteomics is the global analysis of the function and expression of proteins. Computer searches for open reading frames are more useful for bacterial genomes because bacterial genes are present as an uninterrupted stretch of nucleotides. In contrast, in eukaryotes many of the genes are divided into exons and introns, which makes the search for genes more complex. Paralogous genes are genes that have diverged as a result of a gene duplication, i.e., two genes in an organism that have different functions but very similar

nucleotide sequences. Orthologous genes are genes that arose because of speciation, i.e., genes found in different species that have very similar nucleotide sequences and functions. Because of alternative splicing, a gene can give rise to numerous protein products. Thus a small increase in gene number could result in a very large increase in protein number. Thus the number of proteins and protein-protein interactions could be much greater in the organism with the larger genome. Also, complexity among multicellular organisms arises largely from organizing larger number of cells into more complex groups of interacting cells. This requires evolution of control regions that regulate transcription and cell division as well as the evolution of new proteins.

10. A nucleosome consists of a protein core of histones with DNA wound around its surface. The protein core consists of an octamer, containing two copies of histones H2A, H2B, H3, and H4. Approximately 150 base pairs of DNA are wrapped less than two complete turns around the octameric histone core. The histone H1 binds to the linker region, which varies in length from 10 to 90 base pairs and is located between nucleosomes. The nucleosomes are folded into a two-start helix (see Figure 6-30) to form a 30-nm fiber.

11. Actively transcribed genes (euchromatin) are in regions where the histone tail lysines are hyperacetylated compared to transcriptionally repressed genes (heterochromatin) that are associated with histones with hypoacetylated tails. Lysine 4 of histone H3 is trimethylated in euchromatin, whereas histone H3 lysine 9 is trimethylated in heterochromatin. Heterochromatin protein 1 (HP1) is a major component of heterochromatin that binds to the histone H3 tail when it is trimethylated on lysine 9. HP1 also binds the histone methyl transferase that methylates H3 lysine 9, causing the spreading of heterochromatin along regions of the genome.

12. A eukaryotic chromosome consists of a long, linear DNA molecule. The DNA is wrapped around octameric histone cores to form chromatin, which is then condensed into a 30-nm fiber. Long loops of the 30-nm fiber are thought to be tethered at the base by SMC proteins that encircle chromatin fibers (see Figure 6-38). Scaffold associated regions (SARs) or matrix attachment regions (MARs) are regions in the DNA thought to be associated with the bases of these loops. Genes are primarily located within chromatin loops, i.e., between SARs or MARs.

13. FISH, or fluorescent in situ hybridization, is one of many related techniques used to detect DNA (or RNA) sequences in cells or tissues. In the case of DNA, a fluorescent probe to a specific sequence is made and then hybridized to its complementary sequence directly on the chromosome(s). The signal is detected using fluorescence microscopy. Multicolor FISH is used to detect chromosomal translocations. For example, patients with chronic myelogenous leukemia possess leukemic cells with the Philadelphia chromosome, a shortened chromosome 22 containing genetic material translocated from chromosome 9. These patients also have an abnormally long chromosome 9, resulting from a considerable amount of DNA that translocated from the long arm of chromosome 22. By labeling leukemic cells with chromosome 9-specific and chromosome 22-specific probes, each labeled with a different fluorescent marker, one is able to identify and compare the size of the normal chromosomes to those having undergone a translocation.

14. Metaphase chromosomes can be identified by size, shape, banding patterns, or hybridization to fluorescent probes (chromosome painting). G banding involves treating metaphase chromosomes with mild heat or proteolysis and staining them with Giemsa reagent. G band staining corresponds to regions that have low G + C content. Treating chromosome spreads with a hot alkali treatment prior to Giemsa staining results in a banding pattern almost the inverse of G bands, known as R bands. Chromosome painting is a technique for visualizing chromosomes using fluoresecent probes. In this method, DNA probes labeled with specific fluorescent tags are hybridized to metaphase chromosomes. After unbound probes are washed off, the chromosomes are visualized with fluorescence microscopy. Each chromosome then

fluoresces with a different combination of fluorescent tags. Then a computer analyzes the tags and assigns a false color image. This way each chromosome can be identified by its false-color image and size.

15. Polytene chromosomes are present in larval salivary glands of the fruit fly *Drosophila melanogaster*, and are also present in cells in other dipteran insects and in plants. These enlarged interphase chromosomes, which can be observed with a light microscope, form as a result of multiple rounds of DNA replication (polytenization) without chromosome separation or cell division. Polytene chromosomes consist of multiple gene copies, which when transcribed provide the cells with an abundance of mRNA encoding proteins required for larval growth and development.

16. Replication origins are the points at which DNA synthesis is initiated. The centromere is the region to which the mitotic spindle attaches. The telomeres are specialized structures located at the ends of linear chromosomes. Telomerase is essential for maintaining the ends of chromosomes. Because of the mechanism of DNA replication, the lagging strand at the ends of a linear DNA molecule would shorten after each round of replication. Telomerase is a RNA-dependent DNA polymerase that regenerates the ends of linear chromosomes.

Analyze the Data

a. If the kanamycin-resistance gene were localized solely in the chloroplast DNA of leaf cells of the spectinomycin-resistant plants, no kanamycin-resistant plants could be generated from the leaves since the kanamycin-resistance gene is under the control of a nuclear promoter. However, kanamycin-resistant offspring were obtained from leaves of the original spectinomycin-resistant transgenic plants. Thus these data suggest that the kanamycin-resistance gene was transferred to the nucleus of the leaf cells that grew into kanamycin-resistant plants. A 3:1 ratio of kanamycin-resistant offspring to sensitive offspring among self-pollinating kanamycin-resistant plants indicates a Mendelian (i.e., nuclear) pattern of inheritance. A cross between two resistant plants in which a single copy of the kanamycin-resistance gene had been integrated into one autosomal chromosome would give a 3:1 ratio of kanamycin-resistant to sensitive offspring.

b. Because no chloroplasts are inherited from the paternal (pollen producing, transgenic plant), the only source of a kanamycin-resistance gene in the progeny seedlings was the haploid pollen nucleus from the transgenic, kanamycin- and spectinomycin-resistant parent. Each kanamycin-resistant seedling contained the kanamycin-resistance gene, as expected. But every one of the kanamycin-resistant progeny seedlings also contained the spectinomycin-resistance gene (seedlings 2, 3, 4, 7, 9). These findings suggest that the two antibiotic resistance genes were transferred to the nucleus together in each case. Because only the spectinomycin-resistance gene was traanscribed in chloroplasts and not the kanamycin-resistance gene, it is unlikely that the original mode of transfer of the kanamycin-resistance gene from a chloroplast to the nucleus was via an RNA intermediate. It is much more likely that a fragment of chloroplast DNA containing both the kanamycin- and spectinomycin-resistance genes was transferred from chloroplast DNA into the nuclear DNA of some of the leaf cells—i.e., the leaf cells that generated kanamycin-resistant plants.

c. Most of the original, spectinomycin-resistant plants had incorporated the spectinomycin- and kanamycin-resistance genes into their chloroplast DNA molecules, making them spectinomycin resistant. However, most of these plants did not integrate the kanamycin-resistance gene into one of their nuclear chromosomes. Only rare leaf cells selected for resistance to kanamycin had transferred the kanamycin-resistance gene into a nuclear chromosome. In the plants generated from these rare kanamycin-resistant leaf cells, kanamycin-resistance was inherited like any other nuclear autosomal gene, exhibiting Mendelian segregation patterns.