RAMSADAY COLLEGE

Content Type: PDF

Department: BOTANY

Teacher: Dr. Arnab Kumar De

Caption: e-Material of BOT-A-CC-4-10-TH

GENETICS

7. Structural organisation of Gene:

7.4. Repititive DNA.

Repititive DNA

Repetitive DNA are the DNA whose base sequence is repeated many times throughout the genome of an organism.

- It is common in eukaryotes, accounting for about half of the total DNA in mammals, for example, and can be divided into various types. Some serves a useful purpose, but a significant proportion is of uncertain function, and may be 'junk', or selfish DNA.
- One important type consists of multiple copies of particular genes or gene sequences; these may represent the members of gene families or be duplicates of genes encoding histones or ribosomal RNAs, which often form tandem arrays.
- Repeats of short DNA sequences, typically less than 10 bp, flank the centromeres of each chromosome, stretching for hundreds of kilobases along either arm of the chromosome and forming centromeric heterochromatin.
- On centrifugation of the total DNA, this separates out as a distinct band, called satellite DNA. Tandemly repeated short sequences also occur at each chromosome tip (telomeric DNA). Both types are important for maintaining chromosome structure.
- Other distinct types of repetitive DNA lie dispersed throughout the genome, both in noncoding introns within genes and between genes, where they may act as 'spacer' DNA. Among these are variable number tandem repeats (VNTRs), sequences of 15–100 nucleotides repeated hundreds or thousands of times at numerous sites within the genome, and represented by minisatellite DNA. Repeats of shorter sequences (2–10 bp) form so-called microsatellite DNA.
- Many transposons also occur as numerous copies throughout the genome, and so contribute to repetitive DNA.

Eukaryote and also human DNA contains large portion of noncoding sequences. As for the coding DNA, the noncoding DNA may be unique or in more identical or similar copies. DNA sequences with high copy numbers are then called repetitive sequences.

Prokaryotes:

1) Most DNA is in the form of "unique" sequences. Exceptions are the genes encoding ribosomal RNA (rDNA, 10-20 copies) and various "recognition" sequences (e.g., promoters).
2) rRNA, mRNA, tRNA, and other RNA types (eukaryotes only)
   a) snRNA (Snerps)
   b) snoRNA (Snors)
   c) RNAi (interference RNAs) (non-coding) DNA → Recognition sites → Spacer bases (few)
3) DNA sequence organization is represented by a linear array of unique sequences (genes or recognition sites), with one or a few bases (spacer bases) separating the functional sequences.
4) Cesium-chloride (CsCl) density-gradient analysis of semi-sheared DNA yielded only a single "main band" of DNA, indicating that most "pieces" of DNA had similar

buoyant densities (meaning base composition). In E. coli, for example, the buoyant-density profile indicates the base composition to be about 50% AT base pairs (and 50% GC base pairs).

Eukaryotes:

Much more complex A. CsCl analysis (initial discovery) of eukaryotic DNA revealed one large "main band" along with several, minor satellite bands (that presumably differed in AT/GC composition). So-called "heavy satellites had a higher percentage (proportion) of GC base pairs composition; whereas socalled "light" satellites had a higher percentage of AT base pairs.

Eukaryotic genomes can be divided coarsely into three categories (largely for convenience):

- **Highly repeated sequences (>105 copies/genome) -- satellite DNA**
- **Moderately repeated DNA sequences (100 - 105 copies/genome)**
- **Unique or single-copy DNA sequences (1 - 10 copies/genome)**

I. **Highly repeated sequences (>105 copies/genome) -- satellite DNA**: The satellite bands were referred to as "satellite DNA" and turned out to be tandem (104 - 107) copies of short DNA sequences (from 10-300 bases pairs per monomer) that were rich in GC base pairs ("heavy" satellites), rich in AT base pairs ("light" satellites), or neither ("cryptic" satellites). Satellite DNA include highly repeated (repetitive) DNA and simple sequence DNA.

II. **Unique or single-copy DNA sequences (1 - 10 copies/genome):** A significant fraction of unique sequence DNA is "coding" for proteins and various types of RNA sequences.
(i) included are transcriptionally active units such as enzymes, structural proteins, and proteins involved in normal cellular function (structural genes, e.g., enzymes in metabolic pathways), gene regulation (regulatory genes, e.g., transcription factors), and (recently discovered) retrogenes
(ii) also included are potentially functional retroviruses and intact transposons and retroposons (transposable elements)

III. **Moderately repeated DNAs (100 – 105 copies):** An extremely heterogeneous class in terms of sequence complexity, with relatively little known about possible functional roles. Can be subdivided into:
(i) coding DNAs that contain structural gene sequences, SINEs, and LINEs
(ii) non-coding DNAs that contain common regulatory sequences, fossil repeats, and introns of moderately-repeated structural gene sequences

SINEs [Short Interspersed Nuclear Element(s)]

(i) 150-300 base pair (bp) repeated elements that are found in the "short interspersion pattern" – typically possess an 8-20 bp inverted repeat (characteristic of "insertion" sequences) called 'target-site duplications'
(ii) exhibit a highly variable pattern among organisms.

LINEs [Long Interspersed Nuclear Elements]

i. An interesting and heterogeneous class of sequences comprised in part of transposons and retrotransposons
   a) Transposons: mobile genetic elements, jumping genes, nomadic sequences, etc.
   b) Retrotransposons: mobile elements depending on reverse transcription
ii. Elements that are 3,000 - 5,000 bp in length that are dispersed (interspersed) throughout genomes (hence LINE)
iii. Clearly mobile (able to "move" from location to location within a genome) and inducible. The latter accounts for the phenomenon of "hybrid dysgenesis."

Definite involvement of transposable elements in mutation and chromosomal rearrangement.