By: Dr. Shubhi Avasthi (Botany Department)

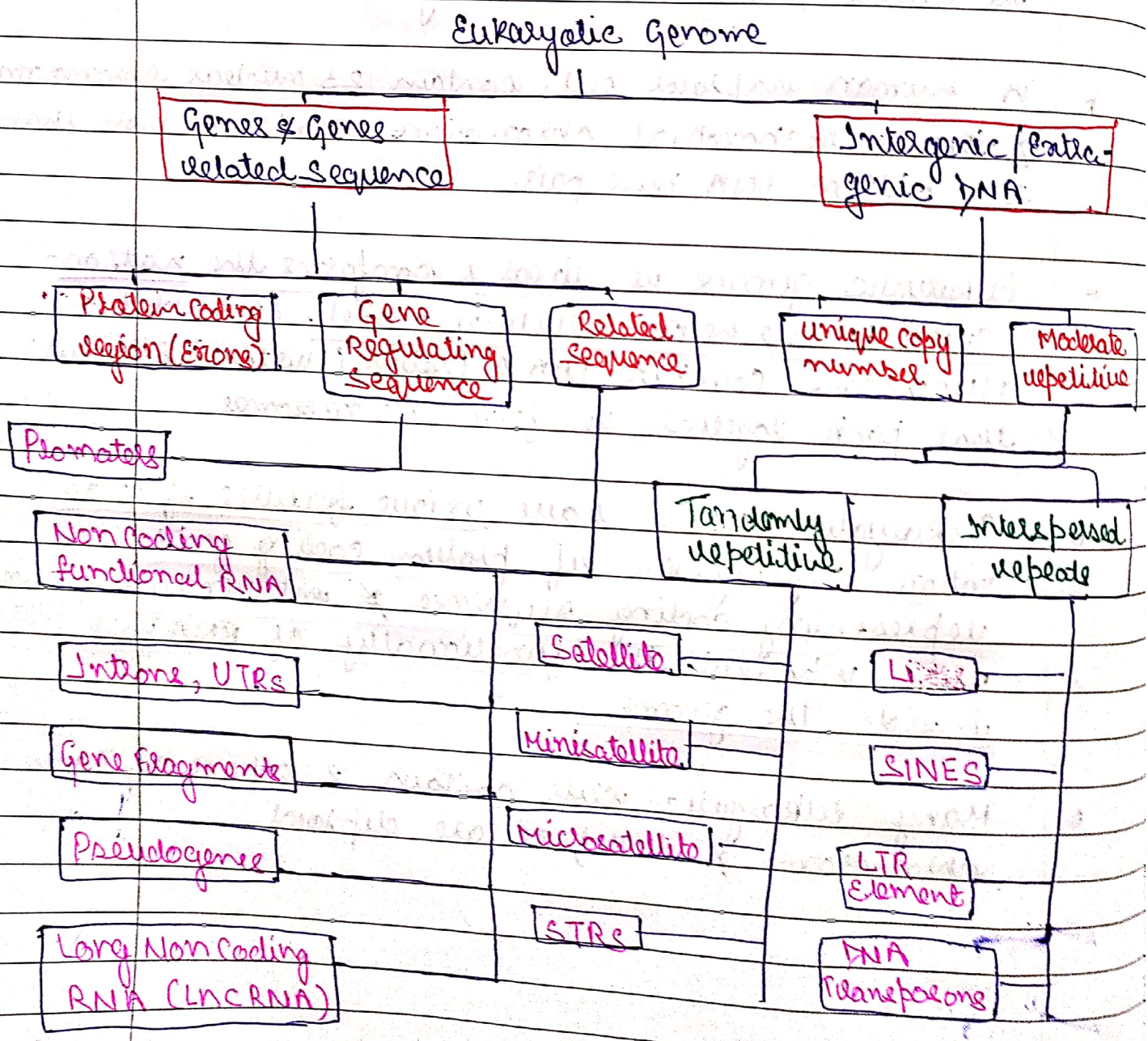
DATE

Eukaryotic Genome Organization (Unit-II)

* A genome is an organism's complete set of DNA, comprising of nuclear & mitochondrial DNA
* Each genome contains all of the information needed to build & maintain the organism
* A human haploid cell, contain 23 nuclear chromosome & one mitochondrial chromosome, contains more than 3.2 billion DNA base pair.
* Eukaryotic genome is linear & conforms the watson-crick double helix structural model. Embedded in Nucleosome-complex DNA & Protein (Histone) structure that pack together to form chromosomes.
* Eukaryotic genome have unique features of exon-intron organization of protein coding genes, representing coding sequence & intervening sequence that represent the functionally of RNA part inside the genome.
* Many eukaryotic cells contain 2 copies of each chromosome & therefore are diploid.

DATE

+ The configuration of eukaryotic genome includes protein region, gene regulating region, gene related sequence & intergenic DNA or extra genic DNA which include low copy number & moderate or high copy number repetitive sequence



DATE

1. Genes & Gene related sequence

A) Protein coding regions (Exons)

A Protein coding sequence are the DNA sequence that are transcribed into mRNA later translated to protein.

A The complete protein coding genes capacity of the genome is contained within the exome (the part of the genome formed by exons, the sequence which when transcribed remain with in the mature RNA sequence after introns & are removed using RNA splicing) and consist of DNA sequence encoded by exons that can be later translated into protein.

B. Gene regulating sequence

Q- Promoters are combinations of short sequence elements usually located in the immediate upstream region of
A the gene often within 200bp of the transcription start site) which serve to initiate transcription. They can be sub divided into different components

i) Core promoter directs the basal transcription complex to initiate transcription of the gene. They include TATA box

DATE

ii) The proximal promoter region is the sequence located upstream of the core promoter, usually from -50 to -200 bp. They include GC boxes, the consensus sequence is GGGCGG which is often found in multiple copies within 100 bp of the transcription sites.

iii) Enhancers are positive transcriptional control elements which are particularly prevalent in the cells of complex eukaryotes. They serve to increase the basal level of transcription which is initiated through the core promoter element. Enhancer contain with in a span of only 200-300 bp.

iv) Silencers serve to reduce transcription levels. Silencer elements have been reported in various position : Close to the promoter, some distance upstream & also within introns

V) Boundary elements are regions of DNA, often spanning from 0.5 Kb to 3 Kb, which function to block the spreading of the influence of agents that have a positive effect on transcription.

Vi) Response elements modulate transcription in response to specific external stimuli. They are usually located a short distance upstream of the promoter elements.

classmate

PAGE

DATE

c) Related Sequence

i) Non Coding functional RNA

* It is an RNA molecule that is not translated into a protein. Less-frequently used synonyms are non-protein-coding RNA, non-messenger RNA & functional RNA.

i) Intons, UTRS

* The term intron refers to both the DNA sequence within a gene & the corresponding sequence in RNA transcripts.

* An An untranslated region (UTR) refers to either of two sections, one on each side of a coding sequence on a strand of mRNA. If it is found on the 5' side, it is called 5' UTR.

ii) Gene fragments

* They are pieces of genes containing only the exons. They are composed of cDNA.

iii) Pseudogenes

* They are dysfunctional relative of genes that have lost their gene expression in the cell or their ability to code protein. Depending on their DNA sequence characteristics pseudogenes are of 2 type

DATE

- Processed pseudogenes :- They have normal protein coding gene
- unprocessed pseudogenes :- They lacks intervening non-protein coding sequence.

(iv) Long Non Coding RNA

* They are a type of non coding RNAs that exceed 200 nucleotide in lenght.

* Human genome contains many thousand of Lnc-RNA & these is again clivided into following categories

- Macro Lnc+RNA
- Bidirectional Lnc-RNA
- Sense-overlapping Lnc-RNA
- Sense Intronic Lnc-RNA
- Retained Intron
- NC-RNA Host
- LINC RNA
- Antisense.
- Ambiguous ORF
- 3 Prime overlapping NCRNA
- Non coding.

DATE

2. Intergenic / Extragenic DNA

- IGR is a stretch of DNA sequence located between genes. Intergenic regions are a subset of noncoding DNA.
- In humans, intergenic regions comprise about 75% of the genome, whereas this number is much less in bacteria (15%) & yeast (30%)
- They are thought to have regulatory functions.

A) Unique / low Copy number

- LCN is a DNA profiling or DNA testing technique developed by FSS in 1999. It is a sensitive technique because it involves a greater amount of copying via PCR from a smaller amount of starting material.

B) Moderate / Highly repetitive

- Genome contain some repetitive DNA sequence, including repetitive coding DNA. However, the majority of highly repetitive DNA sequence occur outside genes.

1) Tandemly repetitive region

- Highly repeated noncoding human DNA often occur in arrays of tandem repeats of sequence which may be simple one or a moderately complex one.

classmate

PAGE

DATE

1a Satellite DNA

- It is transcriptionally inactive as is the vast majority of minisatellite DNA, but in the case of minisatellite DNA a significant % is located in coding DNA

1(b) Minisatellite DNA

- It comprises of a collection of moderately sized alleys of tandemly repeated DNA sequence which are dispersed over considerable portion of nuclear genome
- It is a tract of repetitive DNA in which certain DNA motifs are typically repeated 5-50 times.
- They are prominent in the centromeres & telomeres of chromosomes, that latter protecting the chromosomes from damage.

1(c) Microsatellite DNA

- It is simple sequence repeats (SSR) are small alleys of tandem repeats of a simple sequence.
- They are interspersed throughout the genome, accounting for over 60 Mb (2% of genome).
- It occur at thousands of locations in human genome & they are notable for their high mutation rate & high diversity in the population.

CLASSMATE

PAGE

DATE

1(d) STRS:

* Short tandem repeat is a microsatellite, consisting of a unit of 2 to 13 nucleotides repeated hundred of times in a row on the DNA strand.

2. Interspersed Repeat

* It mainly come from transposable element, but also include some protein coding gene families & pseudogenes

* Transposable elements are able to integrate into the genome at another site within the cell.

2(a) LINES

* Long interspersed nuclear elements have a comparatively long evolutionary history.

* Human LINES consist of 3 distantly related families LINE-1, LINE-2, & LINE-3 collectively comprising 20% of the genome.

* They are located in euchromatic region & are located in the dark AT rich G bands.

DATE

2(b) SINES

A Short interspersed nuclear element are retrotransposons about 100 – 400 bp in lenght.

A They do not encode protein & can't transpose independently.

A SINES & LINES share sequence at their 3' end and SINES have been shown to be mobilized by neighboring LINES.

2(c) LTR elements

A LTR transposons include autonomous & non autonomous retrovirus like element that are flanked by long terminal repeats.

A They are able to transpose independently. There are 3 major classes of human endogenous retrovirus sequences, with a cumulative copy number of about 240,000, accounting for a total of about 4.6% of the human genome

2(d) DNA Transposons

P The cut & paste transposition mechanism of class II TEs do not involve an RNA intermediate.

DATE

* DNA transposons have terminal inverted repeats & encode a transposase that regulates transposition.
* They account for close to 3% of the human genome.
* They have a short life span, within a species.