ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΥΓΕΙΑΣ
ΤΜΗΜΑ ΜΟΡΙΑΚΗΣ ΒΙΟΛΟΓΙΑΣ ΚΑΙ ΓΕΝΕΤΙΚΗΣ

ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ | DEMOCRITUS UNIVERSITY OF THRACE

ΕΡΓΑΣΤΗΡΙΟ ΓΟΝΙΔΙΑΚΗΣ ΕΚΦΡΑΣΗΣ,
ΜΟΡΙΑΚΗΣ ΔΙΑΓΝΩΣΗΣ
& ΣΥΓΧΡΟΝΩΝ ΘΕΡΑΠΕΥΤΙΚΩΝ ΜΕΣΩΝ

DarkMatters Group

Decoding Biology's Dark Matter

The Evolution of Gene Expression

Antonis Giannakakis, PhD

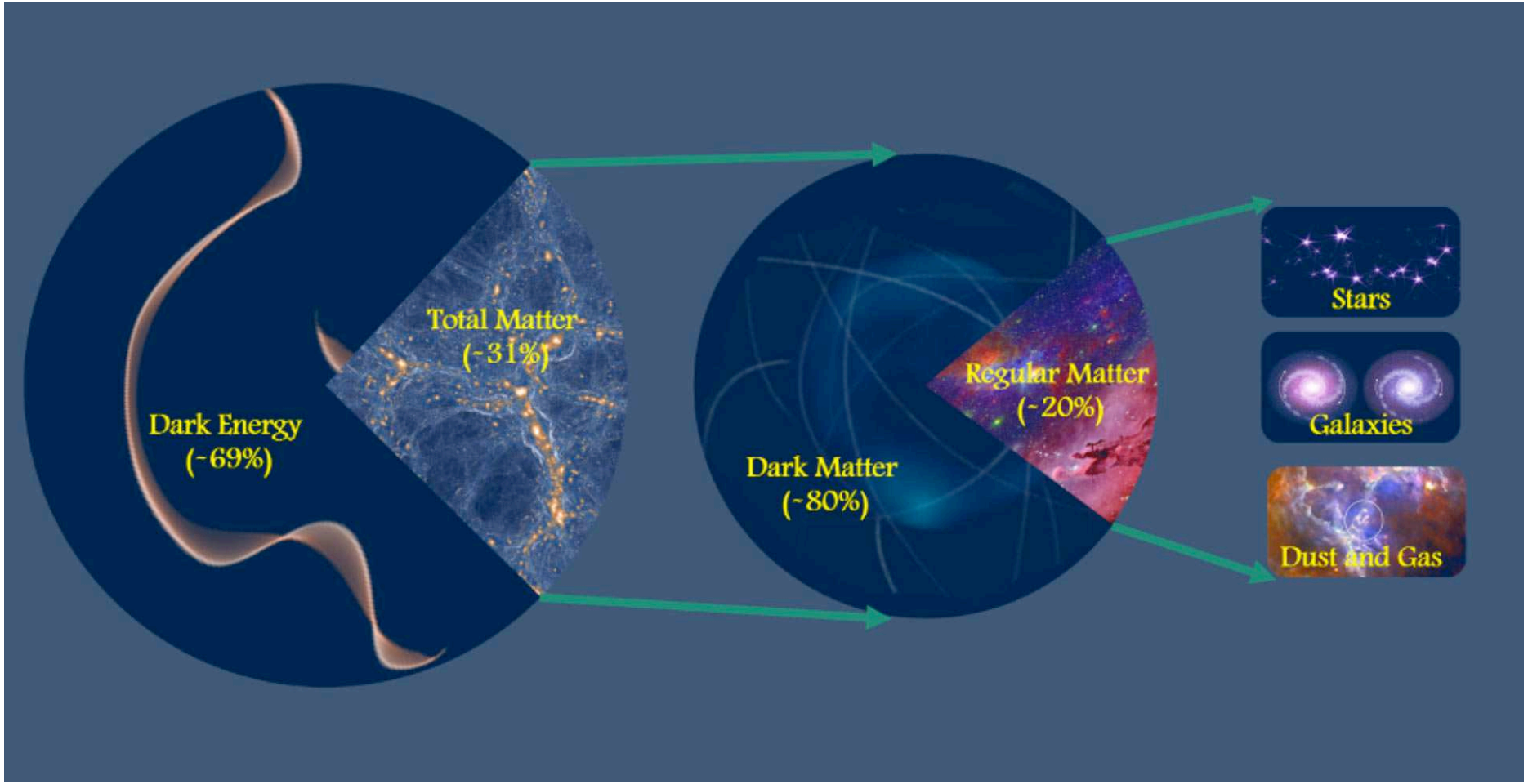
Computational Molecular Biology

For MSc in Translational Research in Biomedical Sciences

16 January 2022

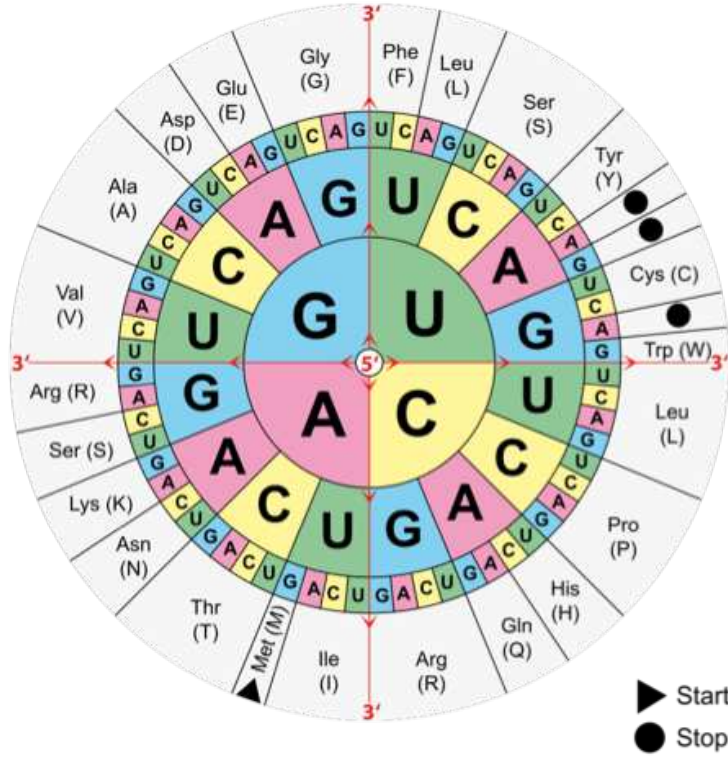
Energy and Matter

- ~70% **Dark Energy**, ~ 30% **Total Matter** of which the majority (80%) is **Dark Matter** and 20% is **Regular Matter**.
- In this grand scheme of things, only ~5% is Regular Matter and ~25% Dark Matter
- Regular Matter is divided into **Inorganic**, **Organic** and **Living**.



Evolution as the main force of Living

- Living energy requires **genetic information** and **chemistry** (i.e. metabolism)
- Life consists of **Living** (biotic) and **Non-living** (abiotic) factors
- The **Cell** represents the Basic Structural unit of ANY **Autonomous** Living entity (aka the Cellular Empire) along side **Non-autonomous** or parasitic living forms, such as viruses (aka The Viral Empire).
- The **Genetic code** of all Living entities (all the way to RNA viruses) is Universal! Is composed of both **Coding** and **Non-coding** units.
- Out of «randomness», **interconnected patterns/pathways of genetic information and metabolism emerge before they disappear (e.g. negative selection) or, become moderately or highly reproducible (i.e. dominate) (e.g. positive selection) <u>in a multi-level selection process (MLS)</u>**.
- Molecular evolution represents **organized molecular patterns** (e.g. genes, genomes, chromosomes, etc.) that are being (i) accurately reproduced, (e.g. Replication) (ii) in homeostasis (iii) **despite or in parallel of molecular entropy** a.k.a. dispersed energy (e.g. environmental insults, mutations/errors during DNA synthesis and repair).
- There is more to evolution than Natural selection of single mutations.



Evolution of cellular life

- The Origin of the Genetic code at the proto-cell was most likely based on a **pre-RNA** core able to bind by affinity peptides or lipids in order to acquire fast abilities to scavenge for genetic information and metabolic energy, self-organize a homeostasis and safeguard its accurate self-replicate.

Proto-cell

- RNA was the first “living” molecule with minimum coding capacity and ability to form by itself single and double strands.

RNA

- DNA helix represents the first evolutionary step towards reproducibility (two copies of genetic informations) and increased storage/stability of genetic information.

DNA

- Chromosomes resulted as an evolutionary step for more organized patterns of genetic information and served as Mediators of Conflict and Gatekeepers of Survival/Development and Reproduction/Hereditability (Alleles, Mitosis/Meiosis).

Chromosomes

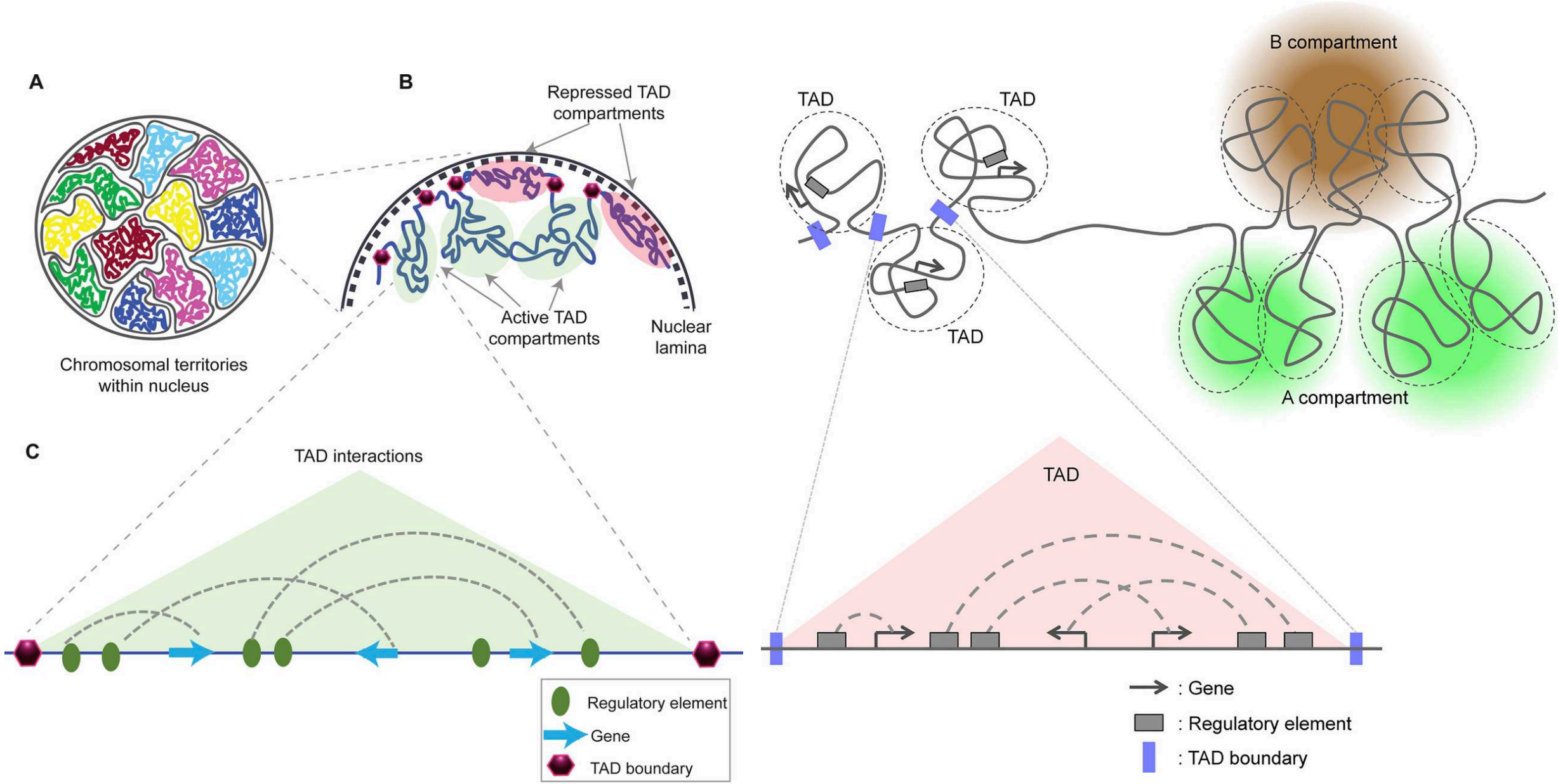
- Interplay of cell types/tissues/organs to sustain complex life forms

Multicellular complexity

- holobiome with microbes/parasites/viruses

Holobiome

Nuclear architecture



Genetic Evolution

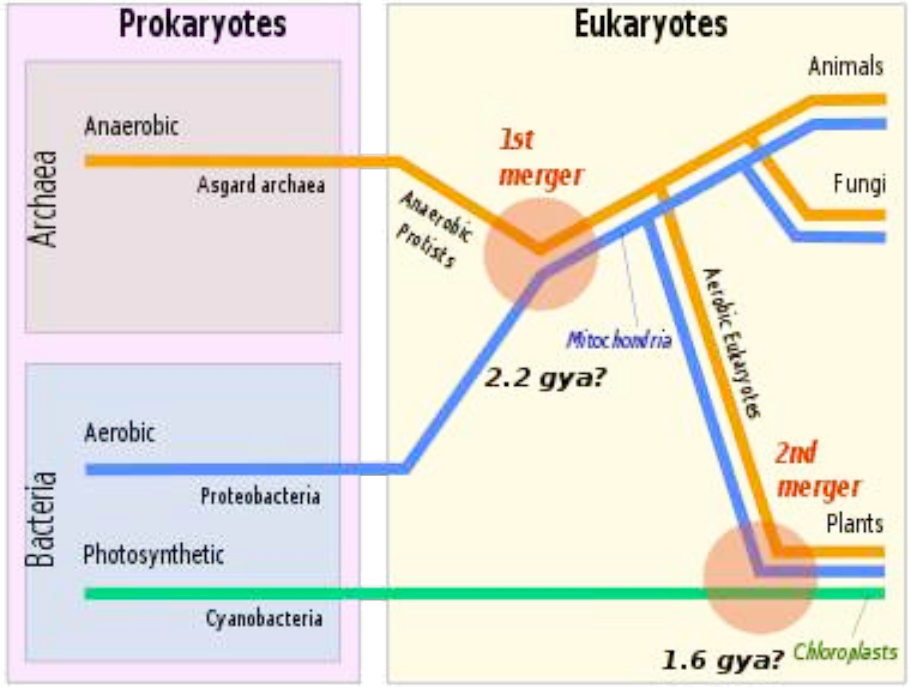
Driven by

- (i) **Natural Selection** / Survival of the fittest, (ii) **neutral evolution** (micro-evolution, long-term, ordered), (iii) **complexity,** (iv) **genetic drifts** (macro-evolution, sudden, random)
  - Fixation of a trait variation depends on population fitness MLS – survival/Adaptability and Reproduction/Heritability within a sub-population by directional (Positive) or purifying selection (Negative)
  - *The bulk amount of molecular patterns/traits (especially in diploids/eukaryotes) is not selected by strong natural selection forces but are Neutral or Balanced (and hence not fixed) within the whole population.*
  - Complexity is driven by a Multilevel Selection (MLS) process
  - Stochastic patterns or Random to the naked eye events emerging from a MLS
- **Gene expression** (DNA to RNA to protein MLS process) (genetic manifestation)
- **Epigenetics** and epigenetic-mediated variation (the MLS process for the when and where of genetic expression)
- **Organism Complexity** (tissue-specific MLS)
- Cancer evolution (evolution by genetic heterogeneity and clonal expansion MLS)

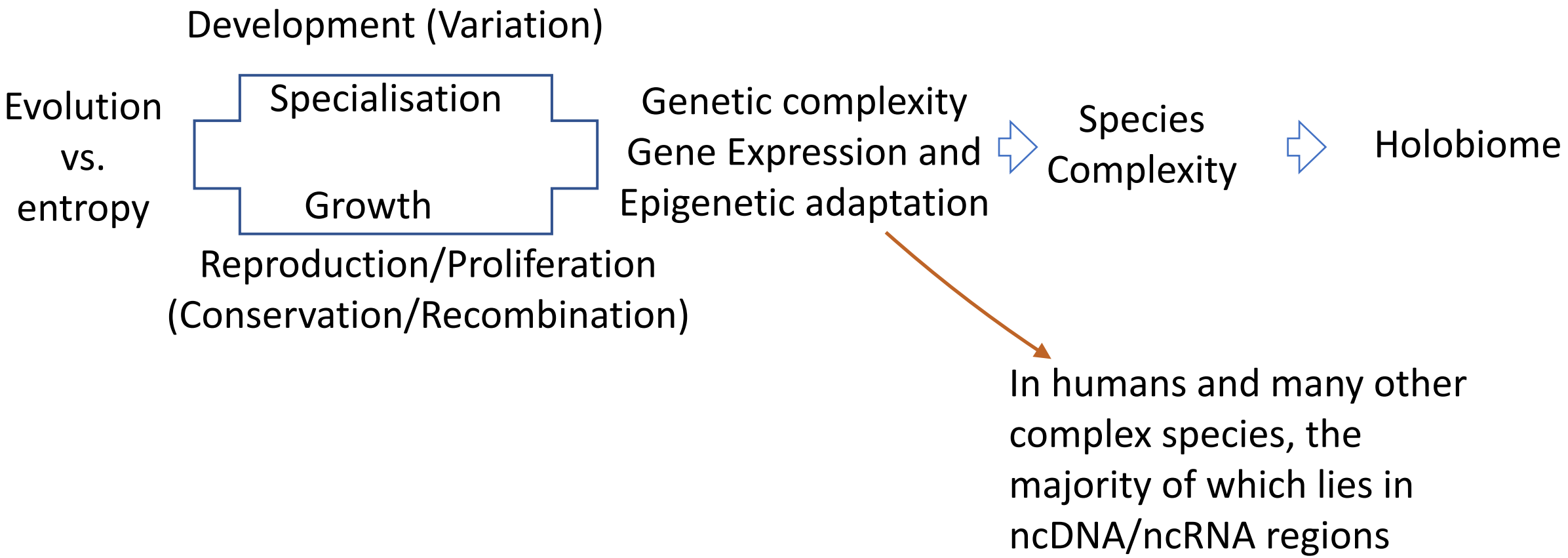
Genetic Vocabulary Extension

Extension of the needed for evolution/survival genetic vocabulary in cells is taking place by

- **In Eukaryotes**
  - Genomic fussion due to organelle endosymbiosis,
  - Mitochondrial Fusion-fission,
  - Gene duplications (Gene copy number)
  - Recombination/Replication errors/mutations (SNPs, Indels, Structural Mutations),
  - Chromosomal segregation errors (Polyploeidy/Aneyploidy)
  - Meiotic/Mitotic Drive.
- **In Prokaryotes/Archaea**
  - Cell division or Asexual reproduction via binary fission
  - Accumulation of plasmid copy numbers.
  - Genomic accumulation/segregation.
  - Horizontal Gene Transfer of bacterial/archaeal genes, plasmids, viruses/virions/bacteriophages (mobile genetic elements)

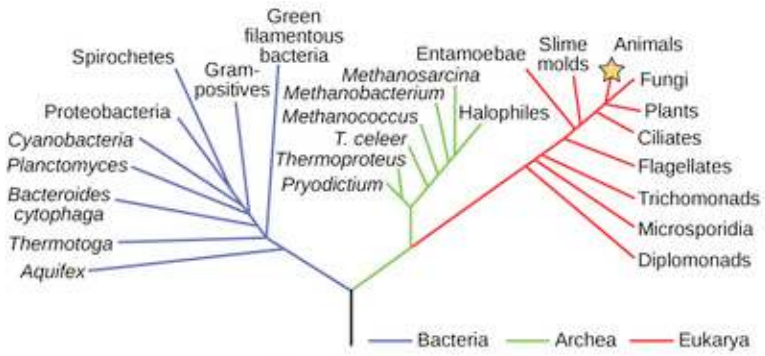


In brief so far…

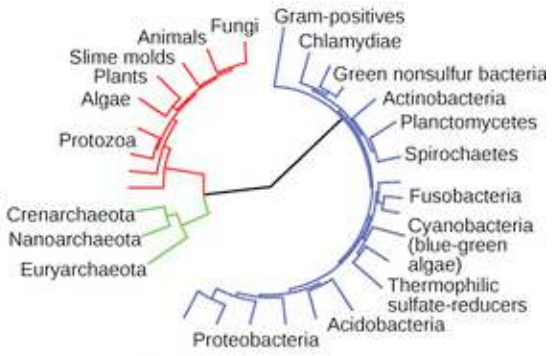


Molecular Phylogenetics

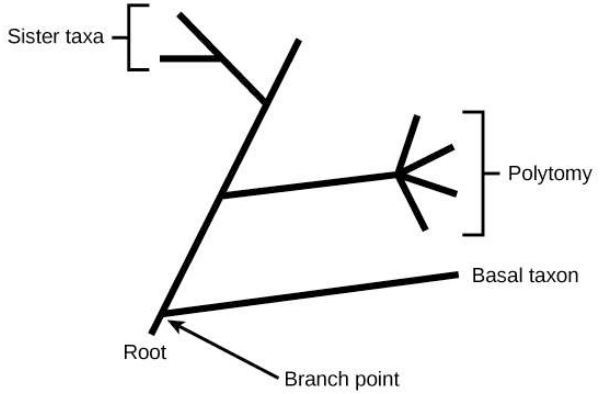
- By comparing the nucleotide sequences of the **16S rRNA** a global phylogeny of cellular organisms for the first time was made (Woese *et al.* 1990).
- This phylogeny overturned the eukaryote-prokaryote dichotomy by showing that the 16S rRNA tree neatly divided into three major branches, which became known as the three domains of (cellular) life: Bacteria, Archaea (both prokaryotes) and Eukarya.
- This discovery was enormously surprising, given that superficially the members of the new Archaea domain did not appear particularly different from bacteria.
- Since archaea and bacteria looked alike, how different could other organisms (that look alike or are different) be? **Molecular Phylogenetics**
- However, recent advances in Evolutionary and Comparative genomics and a better understanding of evolution, showed that the previous conception of the Tree of Life should be replaced by a complex network of treelike and netlike routes of evolution to depict the history of life.



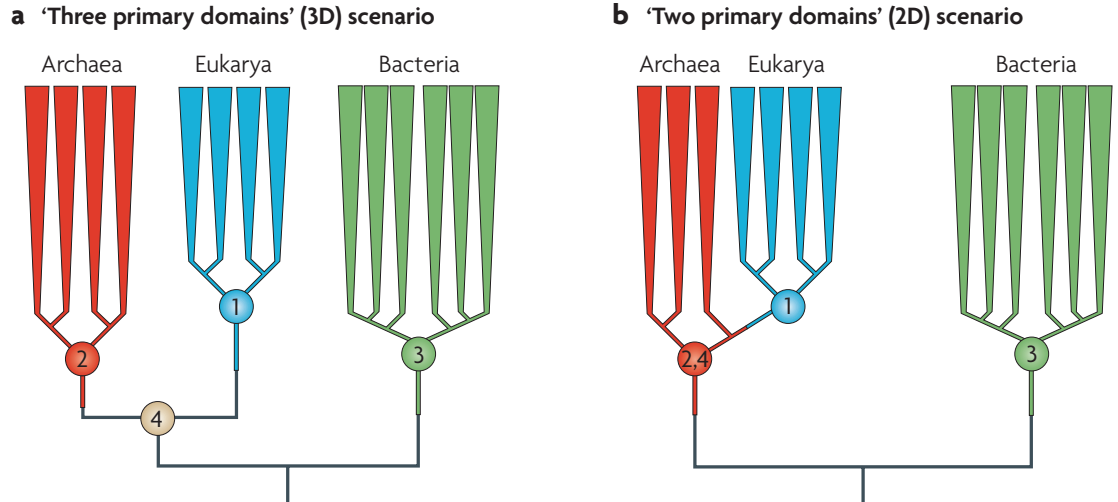
(a) Rooted phylogenetic tree



(b) Unrooted phylogenetic tree



Phylogenetic trees: Both of these phylogenetic trees shows the relationship of the three domains of life (Bacteria, Archaea, and Eukarya), but the (a) rooted tree attempts to identify when various species diverged from a common ancestor, while the (b) unrooted tree does not.

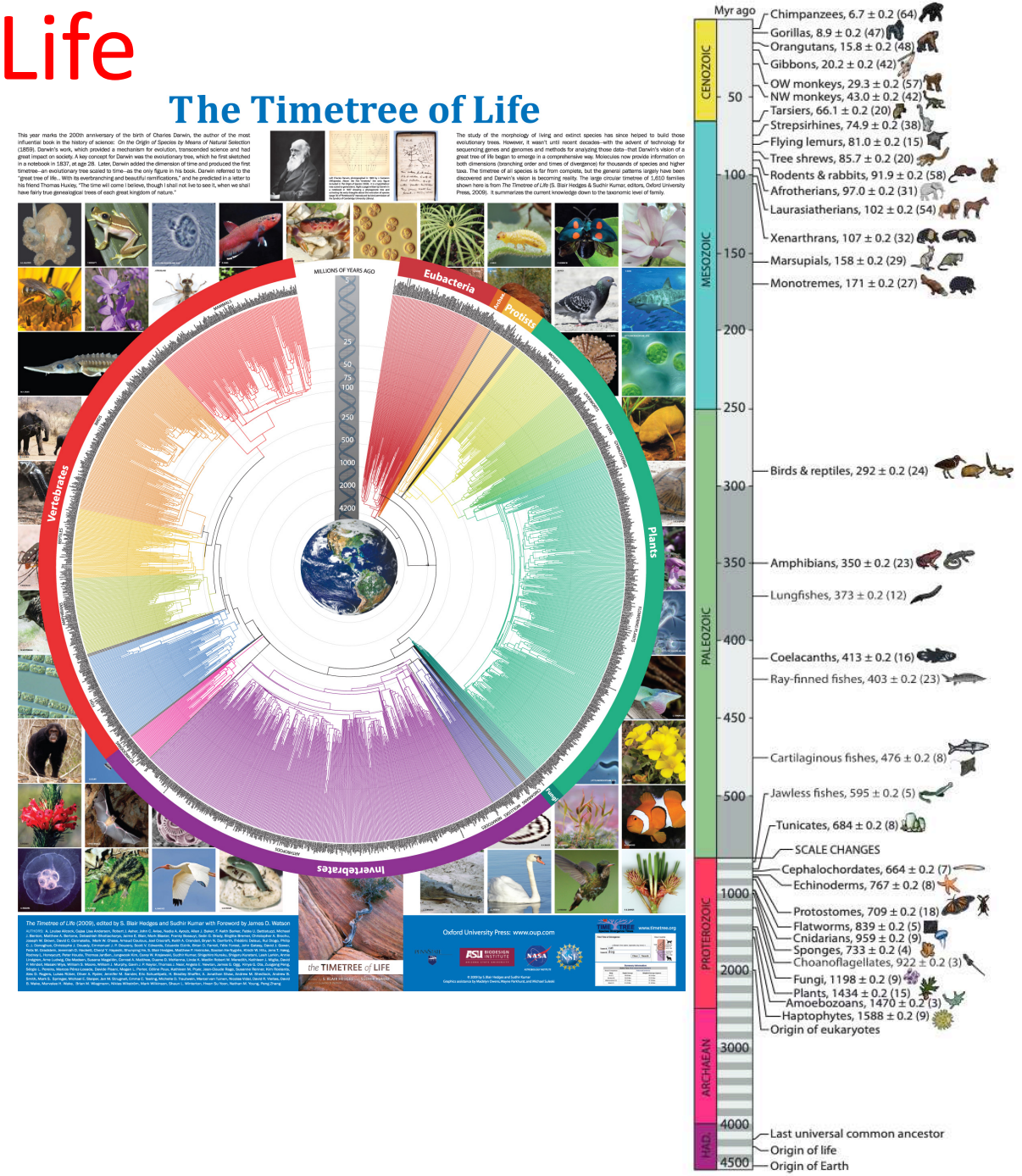


The history of Living Life

1. Unicellular prokaryotes appeared first on the Earth, followed by Archean, and Eukaryotes and subsequently the metazoan multicellularity ~600 Myr ago.

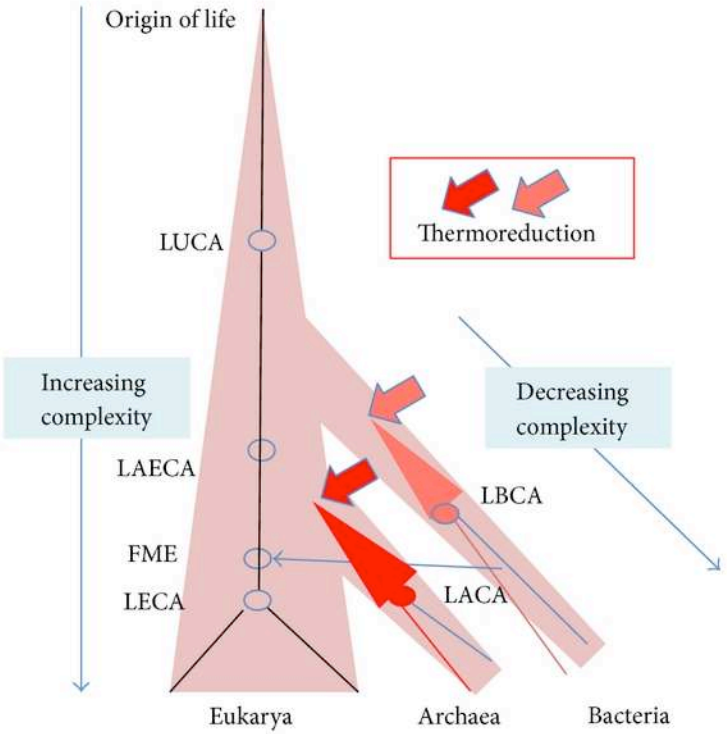
2. Eukaryotes are archaebacterial chimeras, which evolved as a result of, or at least under a strong influence of, an **endosymbiotic event that gave rise to mitochondria**:

3. **Early at in the animal lineage whole genome duplications occurred many times** to extend the vocabulary (500 MYA).

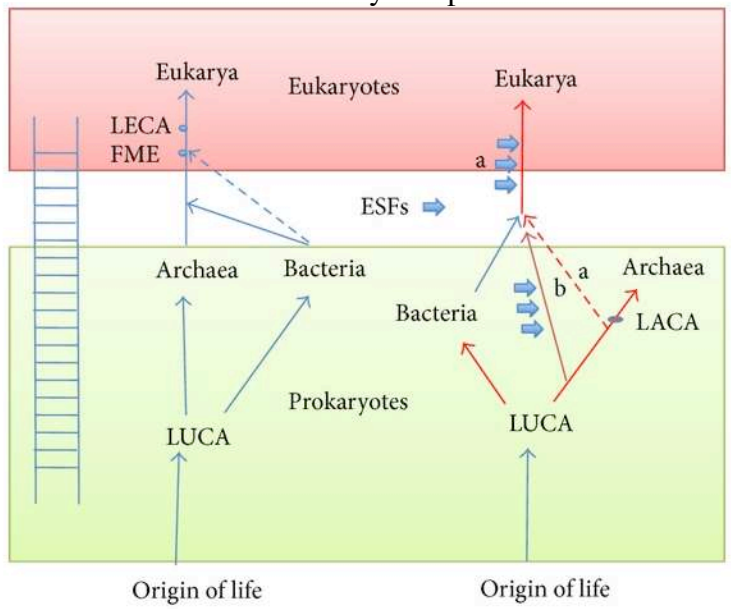


The Tree of Cellular Life

- By comparing the nucleotide sequences of the **16S rRNA** a global phylogeny of cellular organisms for the first time was made (Woese *et al*. 1990).
- This phylogeny overturned the eukaryote-prokaryote dichotomy by showing that the 16S rRNA tree neatly divided into three major branches, which became known as the three domains of (cellular) life: Bacteria, Archaea (both prokaryotes) and Eukarya.
- This discovery was enormously surprising, given that superficially the members of the new Archaea domain did not appear particularly different from bacteria.
- Since archaea and bacteria looked alike, how different could other organisms (that look alike or are different) be? Molecular Phylogenetics
- However, recent advances in Evolutionary and Comparative genomics showed that eukaryotic branching point lies WITHIN the archaea lineage



ESF: eukaryote-specific features



Genetic Variation in Eukaryotes

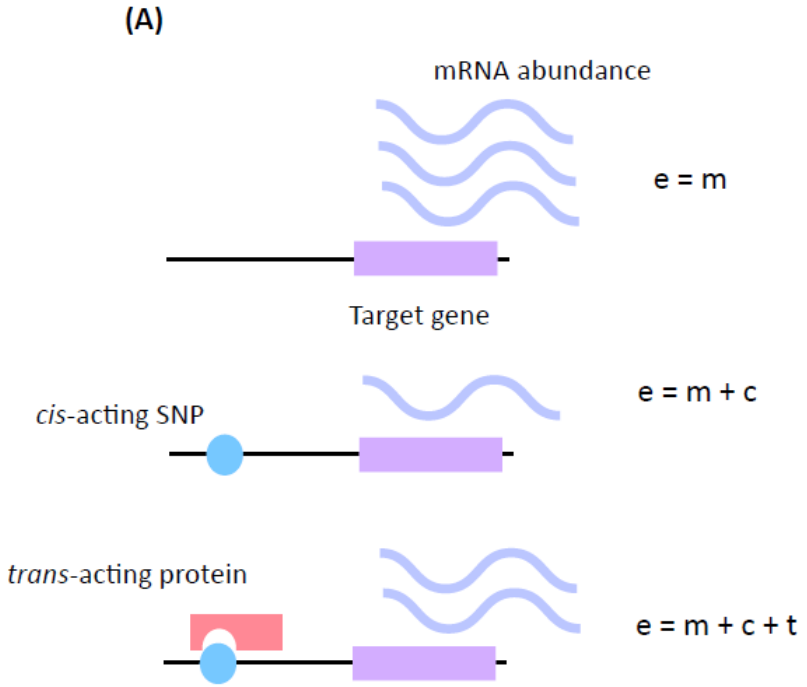
1. There is abundant gene expression variation between individuals, populations, and species.
2. Genetic variation within a eukaryotic population/species is caused by meiosis and copy number alterations (structural variants), genomic recombination/aneuploidy, DNA sequence errors (single, indels), transposition
3. Genetic Richness is mostly fueled by meiosis and structural mutations in eukaryotes.
4. Most of the DNA sequence errors and CNVs are under neutral or balanced selection and happen usually at non-coding regions.
5. SNPs (*in cis* or *in trans* factors) variation is stabilizing/balancing the expression of acquired genetic information (survival adaptation)
6. (Epi)genetic and (epi)transcriptomic variation (transposition, ncRNAs)

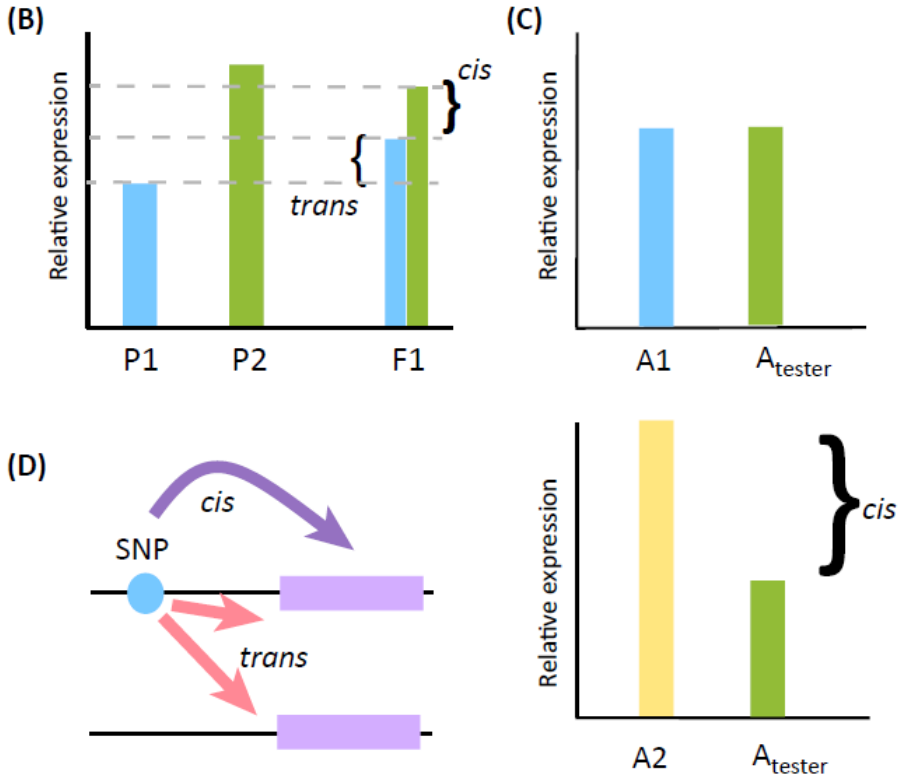
in cis and *in trans* regulation

- Trans: e.g. transcription factor
- Cis: e.g. promoter, enhancer

The Difference between the expression of each allele of F1 with the difference of expression between P1 and P2 strains correspond to the in trans effect

Differences between the two alleles in F1 progeny are due to in cis regulation





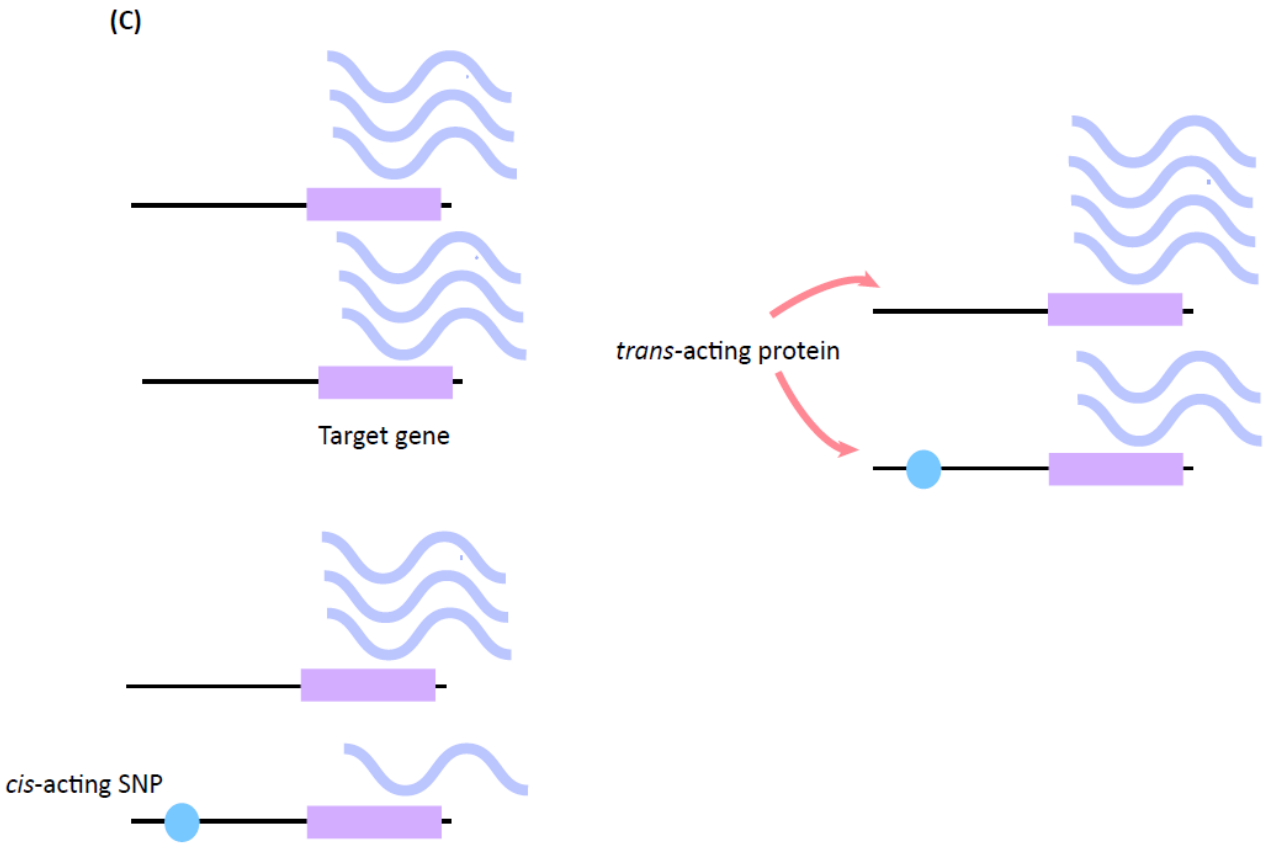
Allele of a individual from a crossed test population, Atest

Alleles from an individual from populations A1 and A2

Signor, S. A. & Nuzhdin, S. V. The Evolution of Gene Expression in cis and trans. *Trends in Genetics* (2018) doi:10.1016/j.tig.2018.03.007.

in cis and *in trans* SNPs are balancing each other

- cis differences are local to the affected gene and allele specific
- large cis effects are more commonly negative, and large trans effects are more commonly positive.
- trans changes can be linked or unlinked but affect both alleles and many more
- cis–trans effects most frequently compensate one another



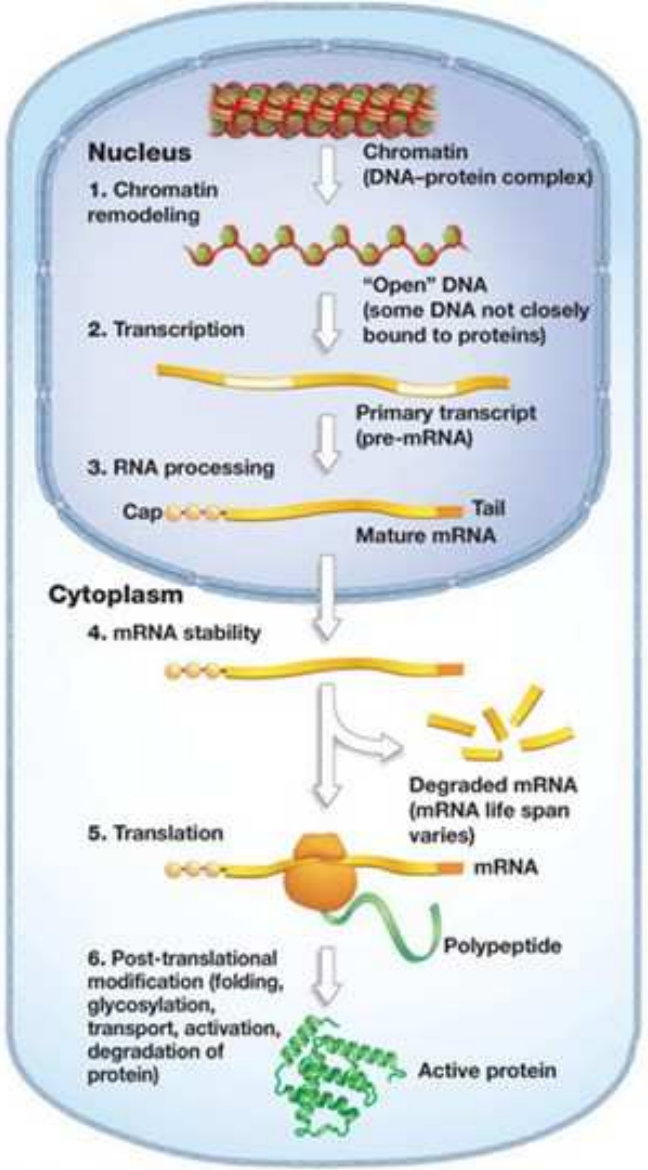
Signor, S. A. & Nuzhdin, S. V. The Evolution of Gene Expression in cis and trans. *Trends in Genetics* (2018) doi:10.1016/j.tig.2018.03.007.

Evolution of Gene expression

- The central dogma of Gene expression is the spatial and temporal pattern of a gene product, such as mRNA or protein
- In prokaryotes is regulated mostly at transcriptional level.
- However, in eukaryotes not only it has a complex regulation but also this regulation takes place mostly in non-coding regions (e.g. promoters, enhancers, splicing etc).
- Evolution and biological diversity is fueled by genetic variation (both in quantity and sequence) of gene expression, and is regulated by epigenetic control in combination with phenotypic variations.

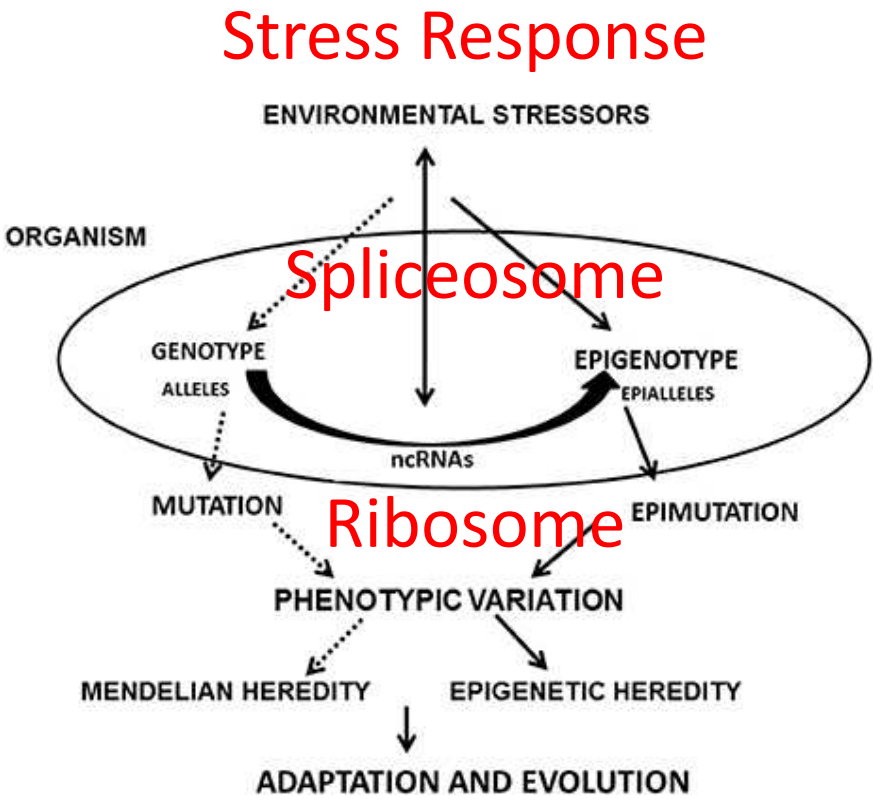
Regulatory Dimensions of GE

- **Chromatin remodeling**
- **Transcriptional**
- **Epitranscriptional**
- **Translational**
- **Post-translational**



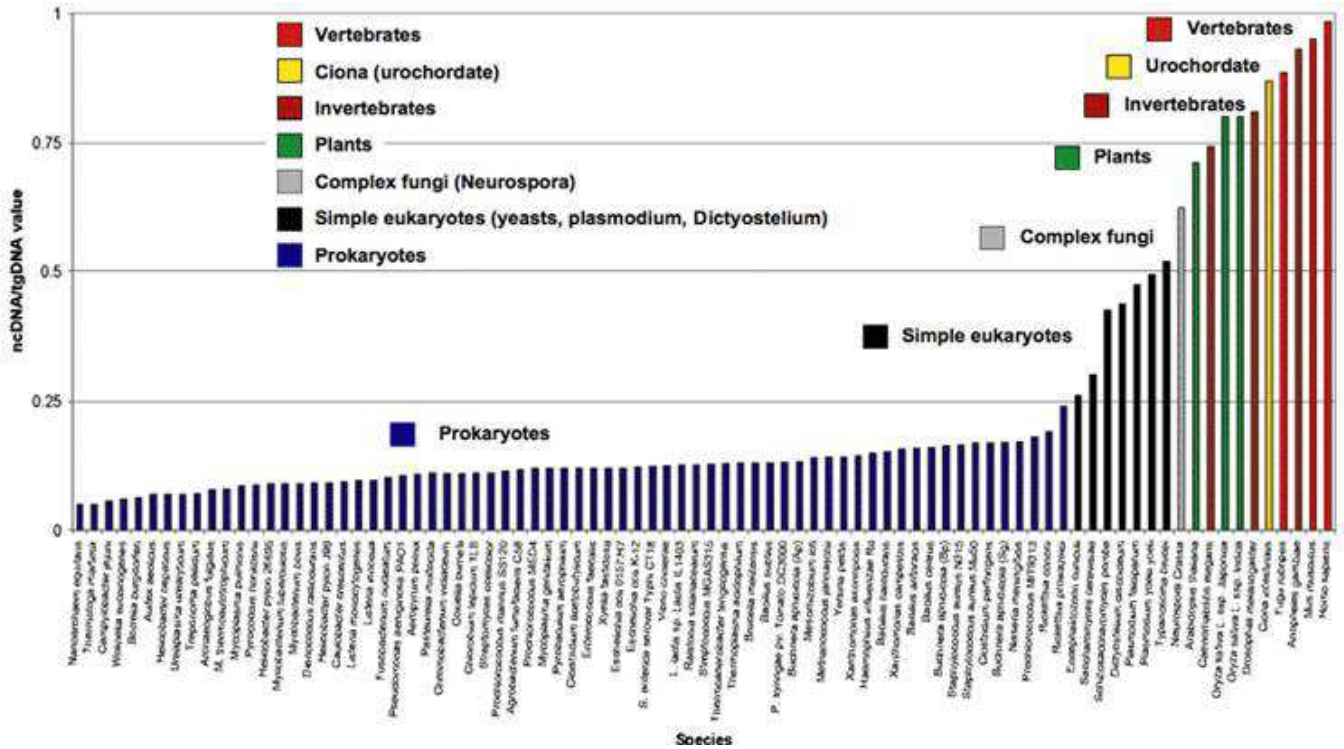
Evolution of gene expression in eukaryotes is not gene-centric anymore

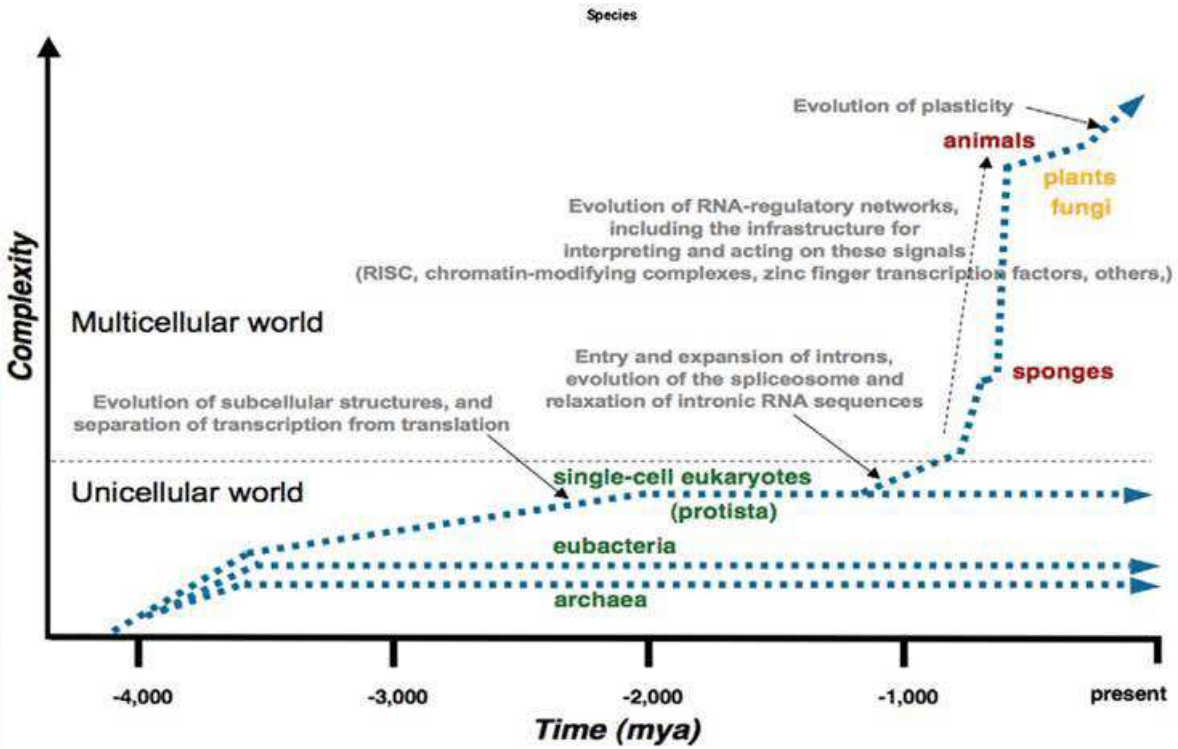
- For over 50 years the term gene was synonymous with coding a protein
- However the last 20 years we have discovered that human genome not only shows extensive ncDNA-mediated regulation **it also pervasively transcribes thousands of different types of ncRNAs molecules**



Frías-Lasserre, D. & Villagra, C. A. The importance of ncRNAs as epigenetic mechanisms in phenotypic variation and organic evolution.

Non-coding RNAs hold the key to development and evolution of complexity

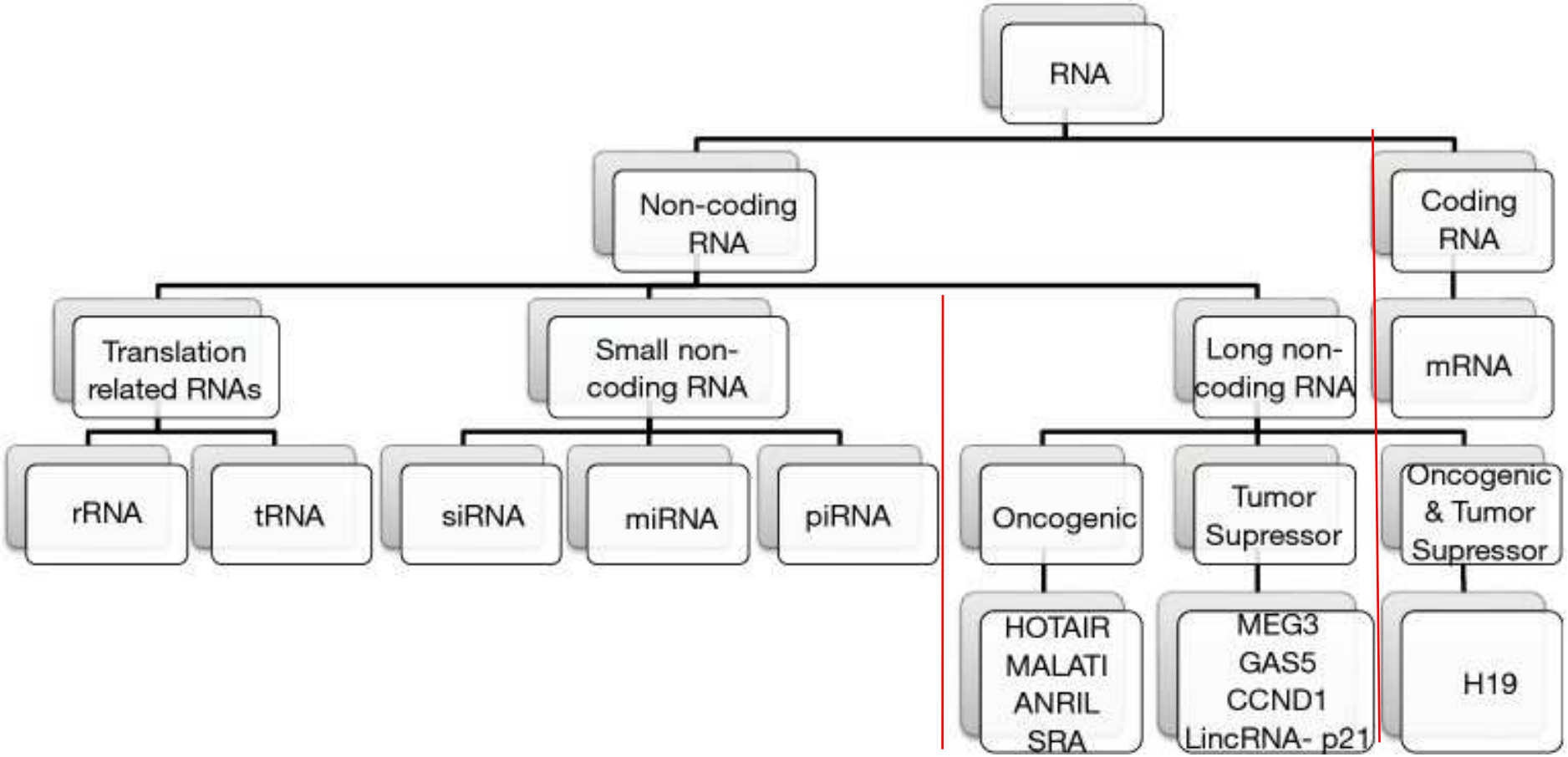




Non-Coding RNA and Evolution of Complexity

- **The ratio of the non-coding size to the overall size** of a given organism's genome (not the genome size, nor the number of mRNA genes) could be associated to evolutionary complexity

The brave new world of RNAs



- **Genes are not so important any more**-we look for transcripts.
- Long non-coding RNAs (>200nts in size) act *in cis* and *in trans* transcribed from noncoding regions that many are enhancers and promoters
- These regions evolve quickly through gain and loss of binding sites while retaining a conserved function (the role of expressed ncRNA structures).
- Against this background of extensive turnover, some mutations generate novel regulatory function, which contribute immensely to biological diversity.

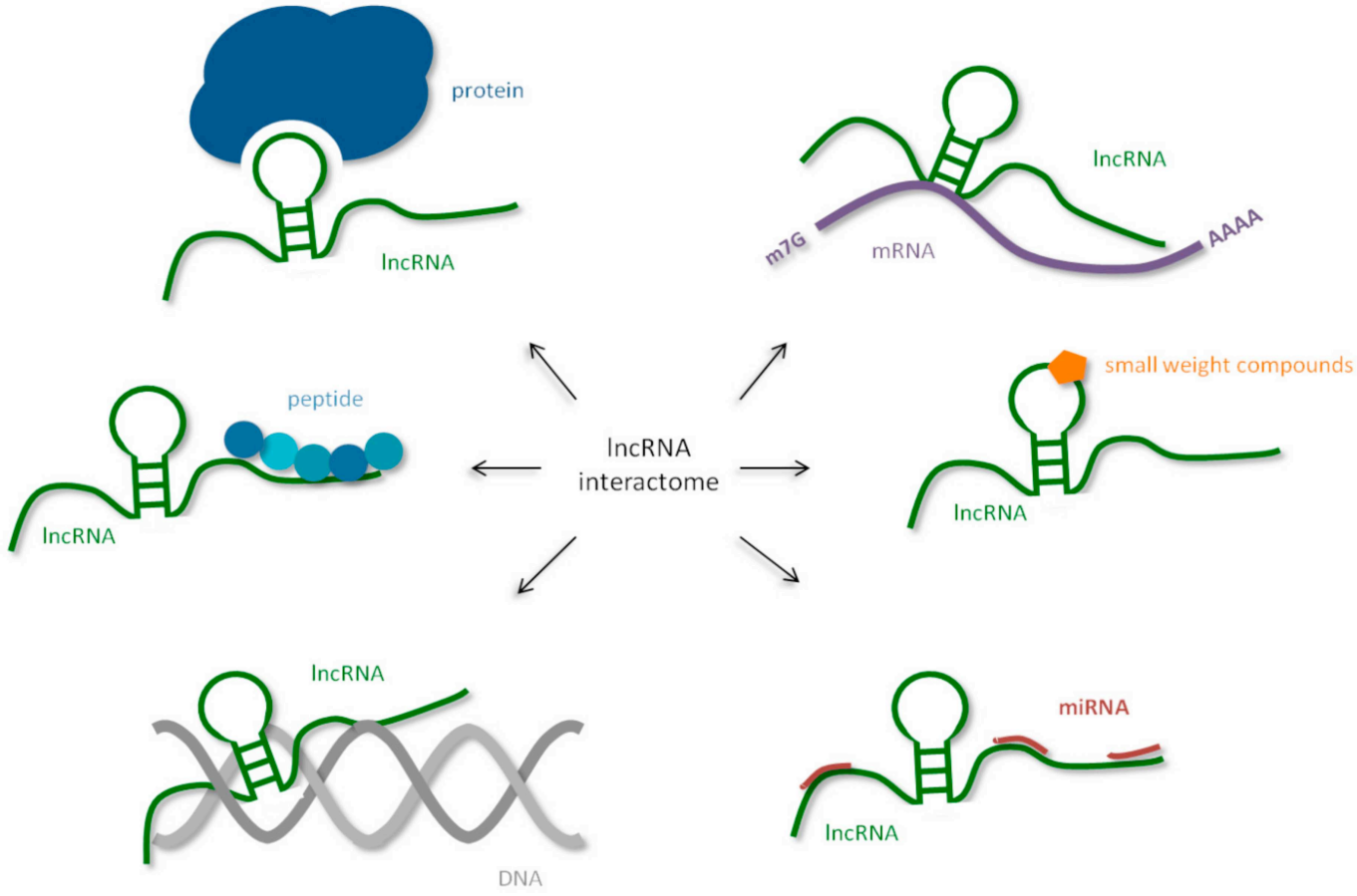
Long non-coding RNA resemble gene units

Differences

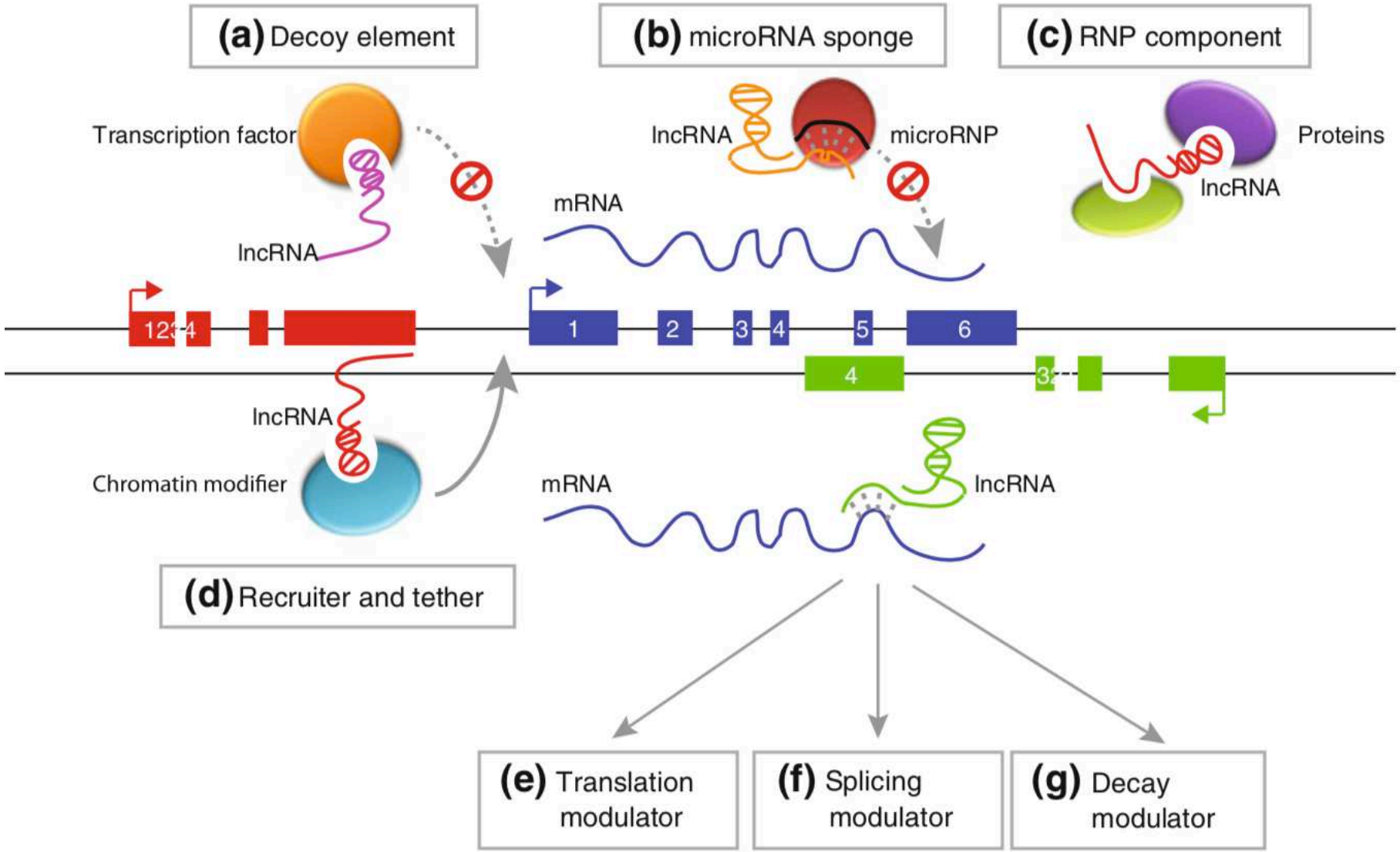
- Short (100-300nt) or absent open reading frames
- Unconserved in primary sequence (thus can purify mutations?) but otherwise conserved in secondary and tertiary structure, due to repeats forming long double strand RNAs)
- On average 10 times less expression levels compared to codons.

Similarities

- Like mRNAs, many lncRNAs are transcribed by RNA polymerase II (Pol II) from genomic loci with similar chromatin structures
- Histone 3 trimethylated lysine 4 (H3K4) at their promoters
- Histone 3 trimethylated lysine 3 (H3K36) along their gene body
- often (but not always) the structure 5’ cap, introns and tail very(A).
- In most cases they do not differ biochemically from mRNAs, except for the **absence of a translated open reading frame** and the presence of **repeats elements**.



lncRNAs Functions I

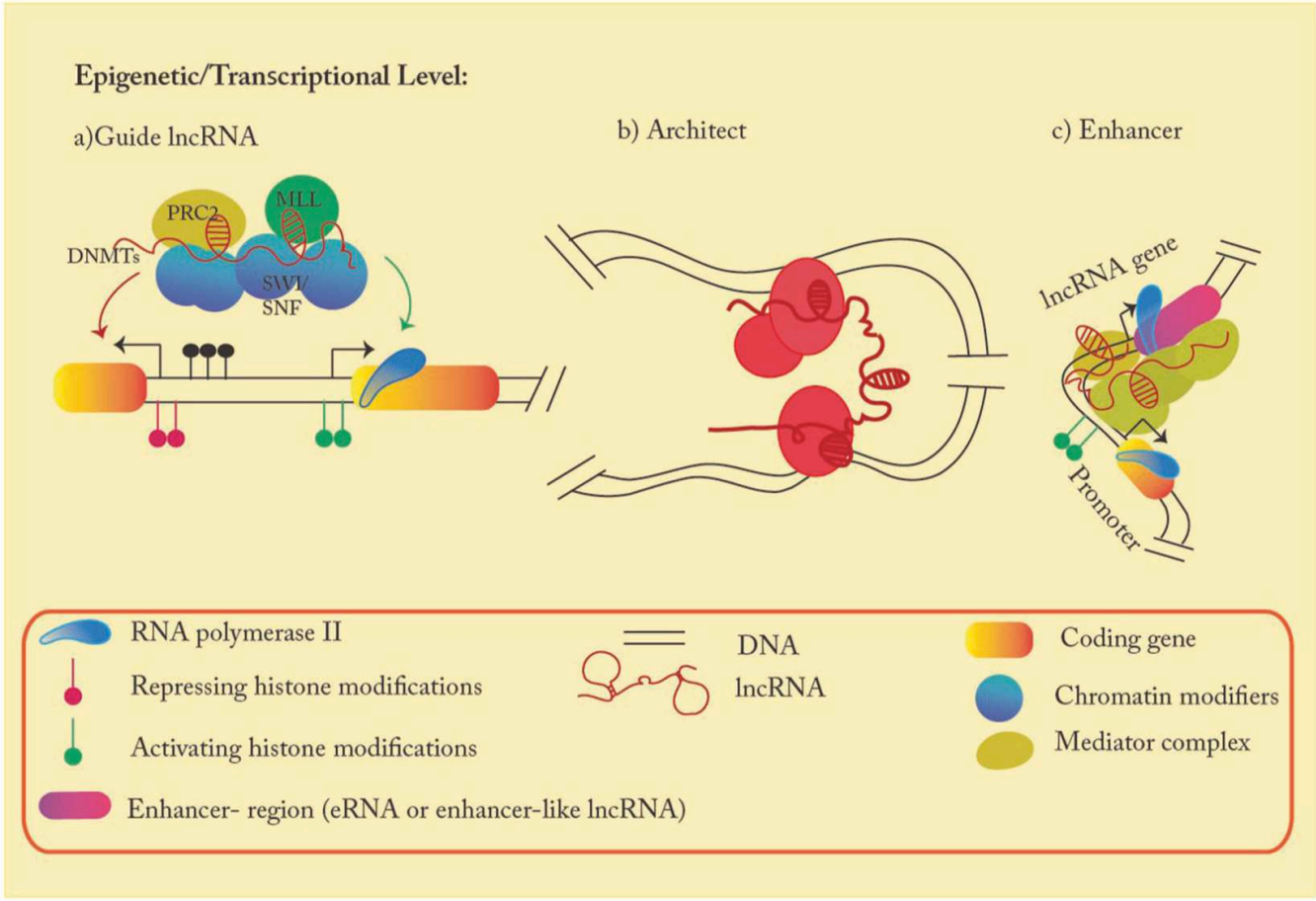


Transcriptional-mediated epigenetic regulation by LncRNAs

a. Guide lncRNAs (tethered scaffold model) – DNA methylation

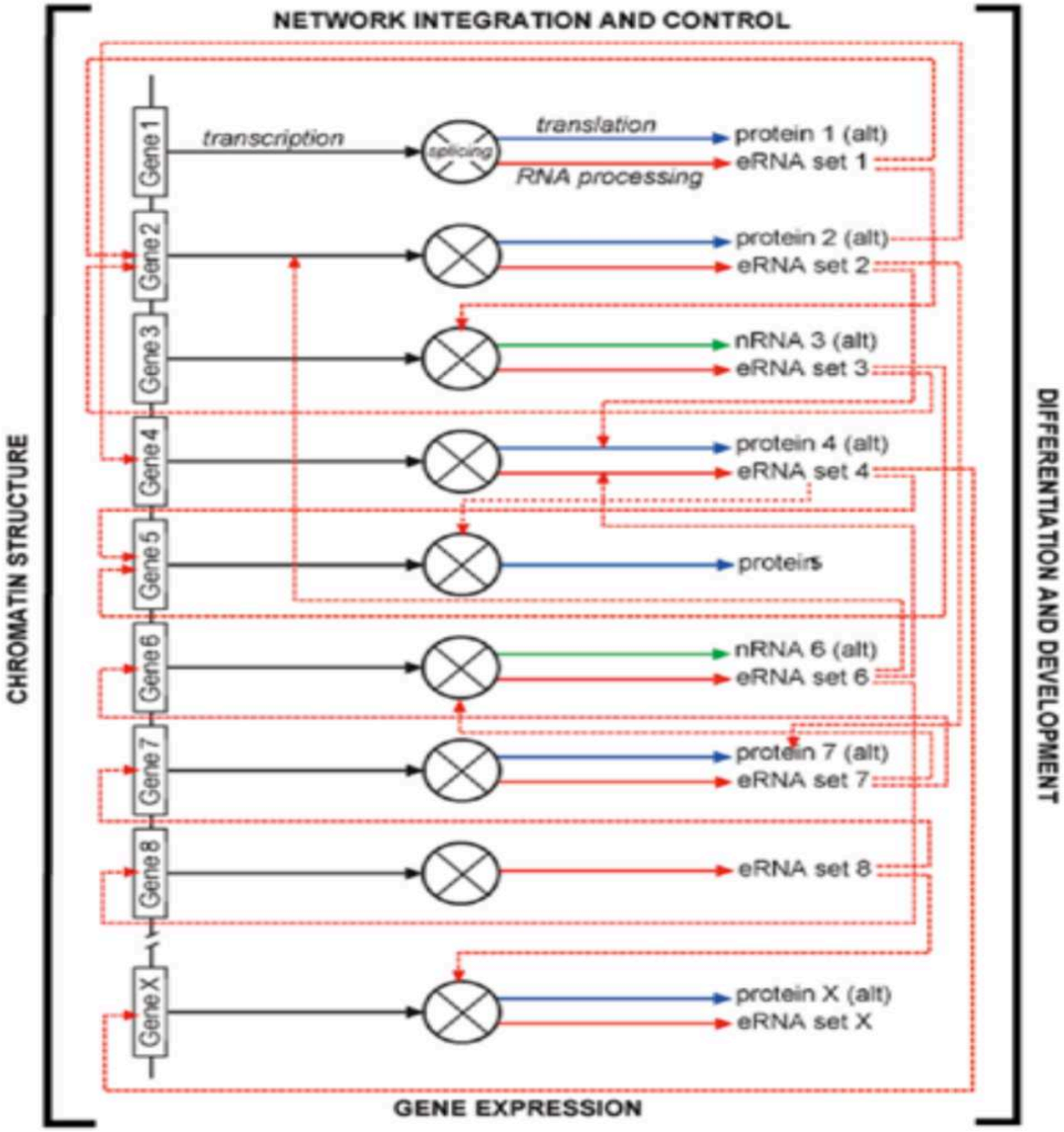
b. Architect lncRNAs (the bridging scaffold model): - Chromatin Remodeling

c. eRNAs, enhancer-like lncRNA: Coherent recruitment of TFs, Chromatin modifiers, RNA polymerase in a pathway of coding genes



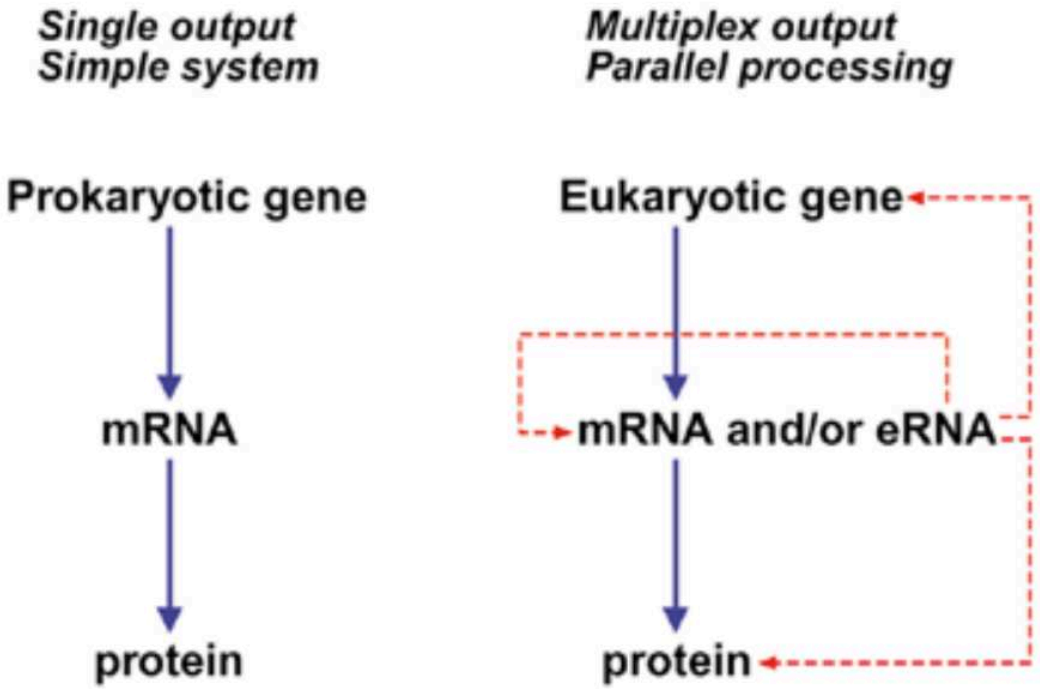
Splicing a major driver of eukaryotic evolution

- A more detailed schematic of the proposed role of eRNAs in eukaryotic system networking and control. Genes, packaged in chromatin, express primary transcripts which are then (alternatively) spliced to yield an mRNA and/or *n* introns, which may be further processed to form multiple smaller species, such as *let-7*. Some noncoding RNA genes may yield functional RNAs from both introns and exons (nRNA). These RNAs may then act as signaling or guide molecules to integrate activity at this locus with that of related parts of the network, via effects on chromatin structure, **transcription, splicing, other levels of RNA processing, mRNA translation, mRNA stability and other levels of RNA-mediated signal transduction within the cell**. The evidence indicates that many if not most of these interactions will be homology (primary sequence) dependent, and involve RNA–DNA, RNA– RNA and RNA–protein interactions, but others may involve secondary or tertiary RNA structures and RNA-mediated catalysis. This scheme is not comprehensive, but is intended to give a sense of the complexity and potential of such networks for programmed control and system integration of complex suites of gene activity in differentiation and development.



The case of long ncRNAs in gene definition

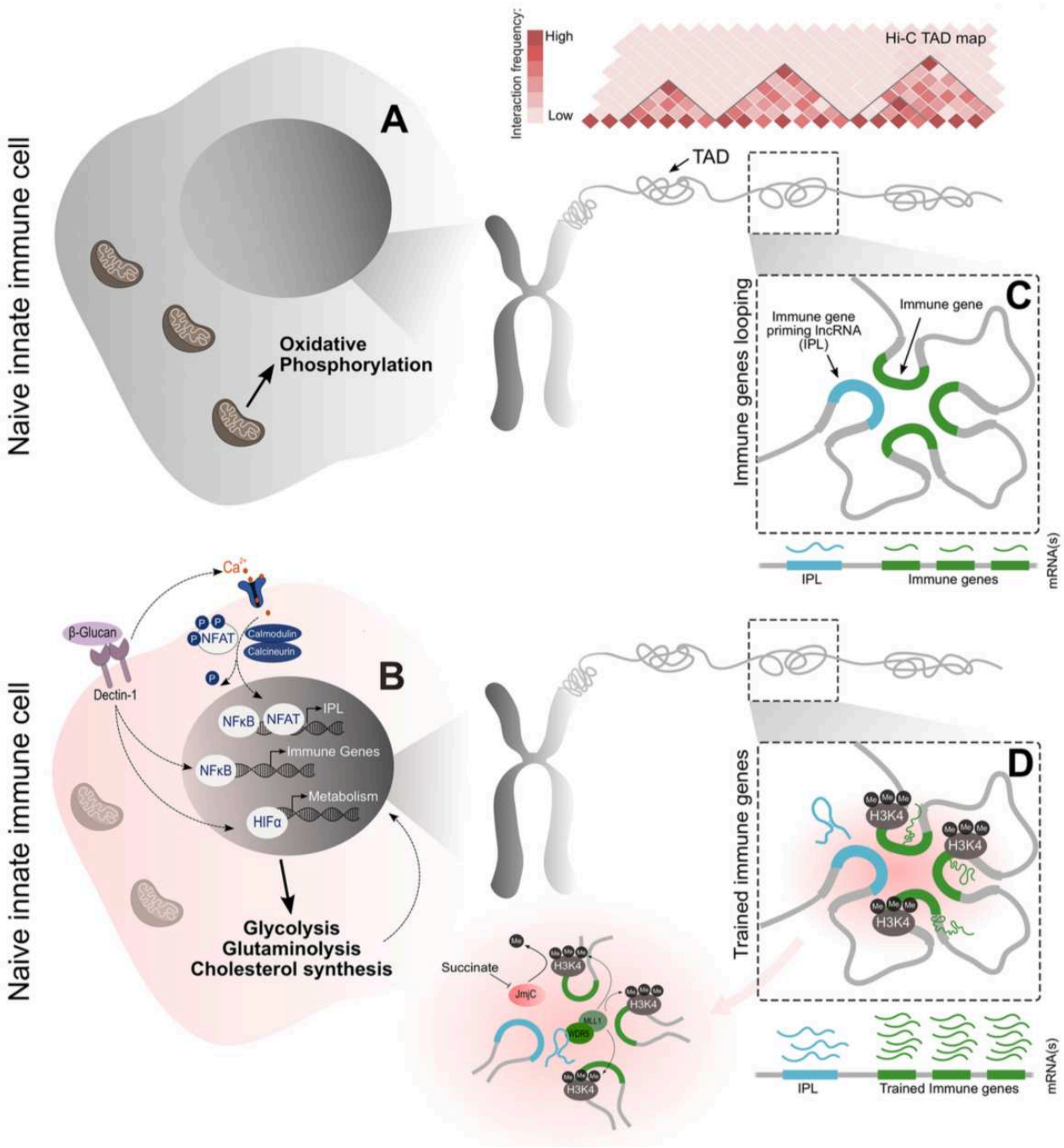
- In prokaryotes the central dogma of gene expression still holds: genes code, via mRNA, for proteins, which carry out the catalytic, structural, signal transduction and regulatory functions of the cell.
- In eukaryotes, genes may express two levels of information: mRNA for proteins, and lncRNAs (for example eRNAs) that carry out concomitant networking and other functions within the organism.
- Thus there are **three types of genes in eukaryotes**: (i) those that encode only protein (which are rare), (ii) those that encode only lncRNAs, and (iii) those that encode both (exon/intron/UTRs).



- Long intergenic non-coding RNAs transcribed from regions that many are enhancers and promoters
- These regions can evolve quickly through gain and loss of binding sites while retaining a conserved function (the role of ncRNA structures in retaining enhancer function).
- Against this background of extensive turnover, some mutations generate novel enhancer functions, which contribute immensely to biological diversity.

Epigenetic Training of Innate Immunity

- The last few years have witnessed an increasing body of evidence that challenges the traditional view that immunological memory is an exclusive trait of the adaptive immune system. Myeloid cells can show increased responsiveness upon subsequent stimulation with the same or a different stimulus, well after the initial challenge. This de facto innate immune memory has been termed “trained immunity” and is involved in infections, vaccination and inflammatory diseases. **Trained immunity is based on two main pillars: the epigenetic and metabolic reprogramming of cells.**
- **Immune-gene Priming lncRNAs (IPLs)**
- An overview of the molecular events that lead to the establishment of the epigenetic memory underlying trained immunity. (A,B) Training begins with the primary exposure of monocytes to β-glucan. This activates the Dectin-1 receptor and calcium-dependent NFAT signaling to initialize transcriptional programs related to immunity and metabolism. The metabolic signaling results in changes to glucose, glutamine, and cholesterol metabolism, which together, supply the metabolites and co-factors essential for the induction and maintenance of the epigenetic changes that are causal to the trained phenotype. NFAT signaling induces the transcription of the newly identified IPLs within immune TADs. (C,D) These lncRNAs facilitate the transcriptional priming of the trained immune genes by recruiting the WDR5/MLL histone methyltransferase complex and exploiting the spatial proximity of immune genes to discreetly deposit the H3K4me3 epigenetic mark on their promoters. These events culminate in a more powerful pro-inflammatory response through the enhanced transcription of trained immune genes upon secondary stimulation.

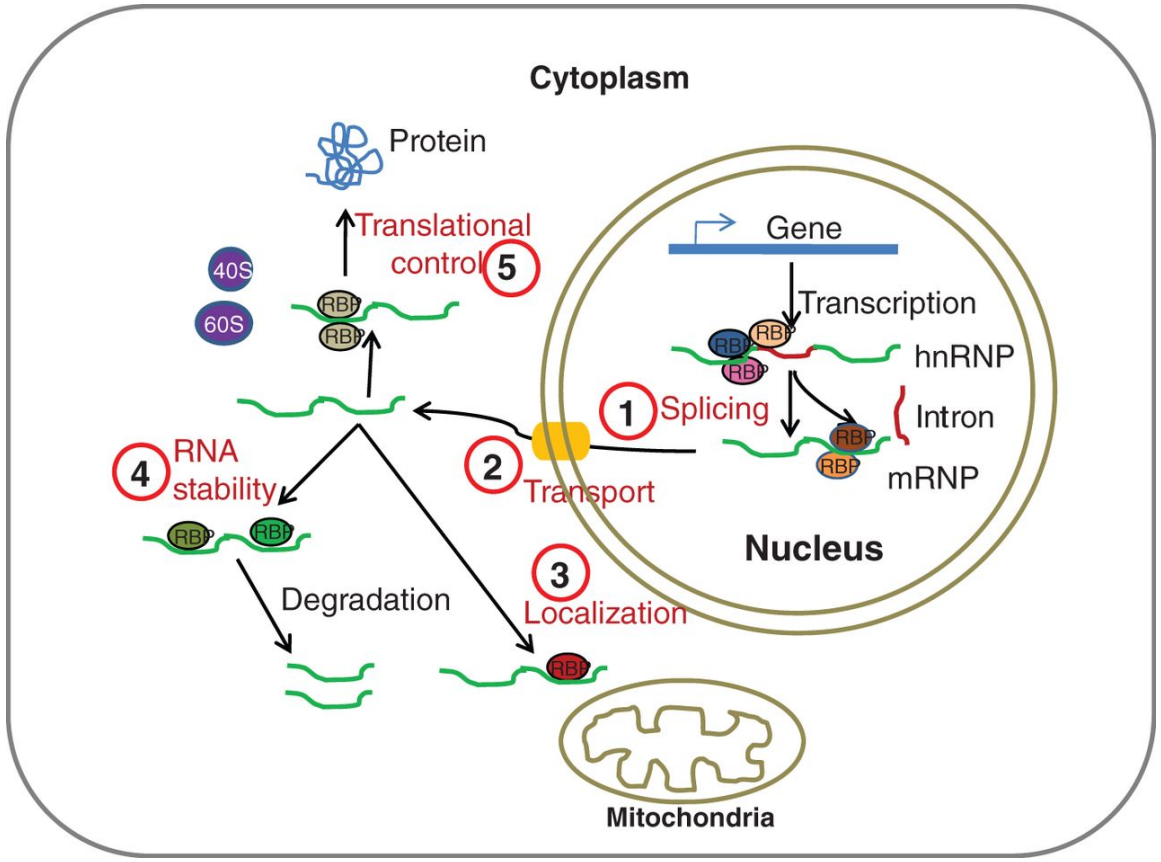


The regulatory non-coding RNAs

Newly evolved low-copy mRNA-like transcripts (>200nts in size) that cover as a class a very large fraction of the human genome.

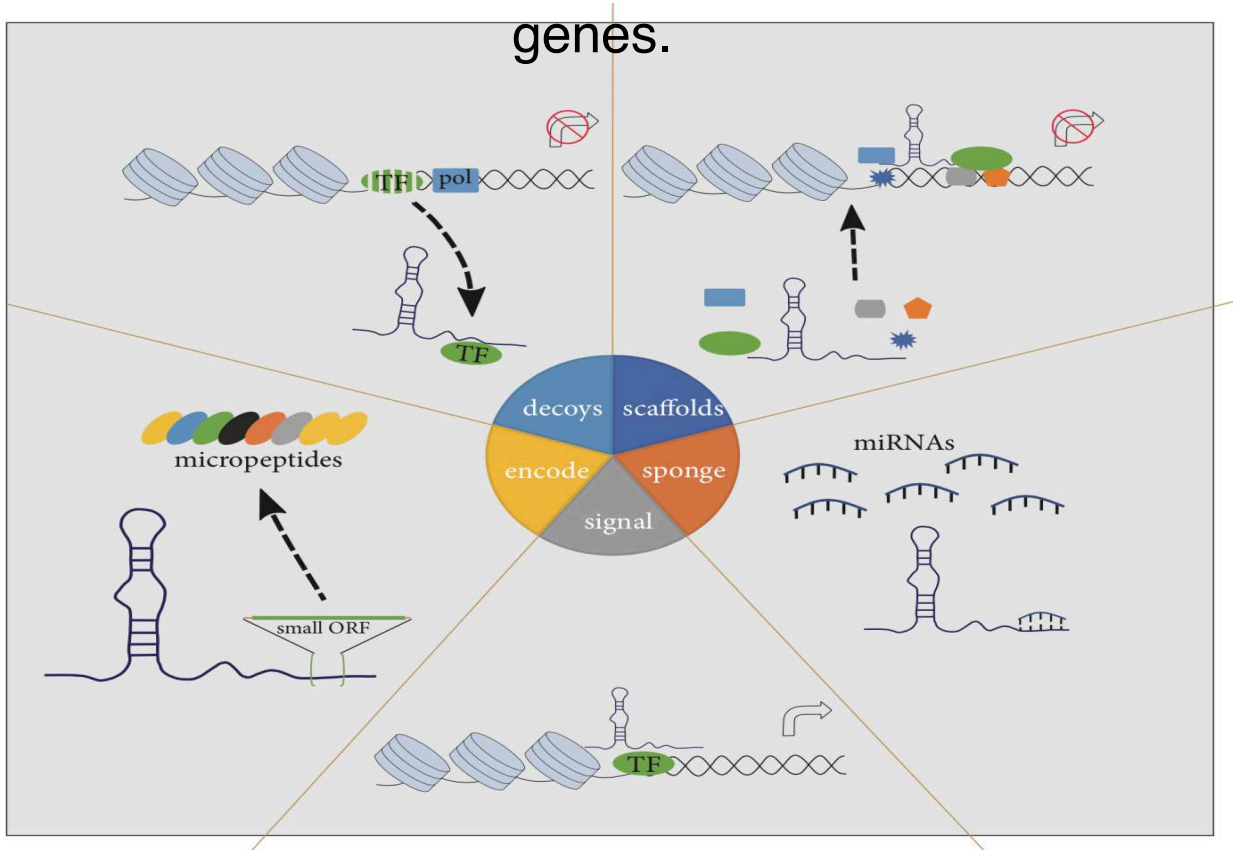
Multi-functional (like miRs)

Fine-tuners of all steps of gene expression



Multi-dimensional

Regulators, modifiers, sensors, scaffolds of all major biopolymers and source of de-novo genes.



The Evolutionary Dimensions of Gene Expression

> **Evolution of Genetic Information**

- •Acquisition, storage and replication of phenotypically advantageous genetic information (coding DNA)
- •Acquisition of Genetic and Epigenetic Variation (non-coding DNA and copy number alterations)
- •Fixation or not of a genetic variation based on population genetics and evolutionary dynamics

> **Evolution of Regulation (non-coding RNAs)**

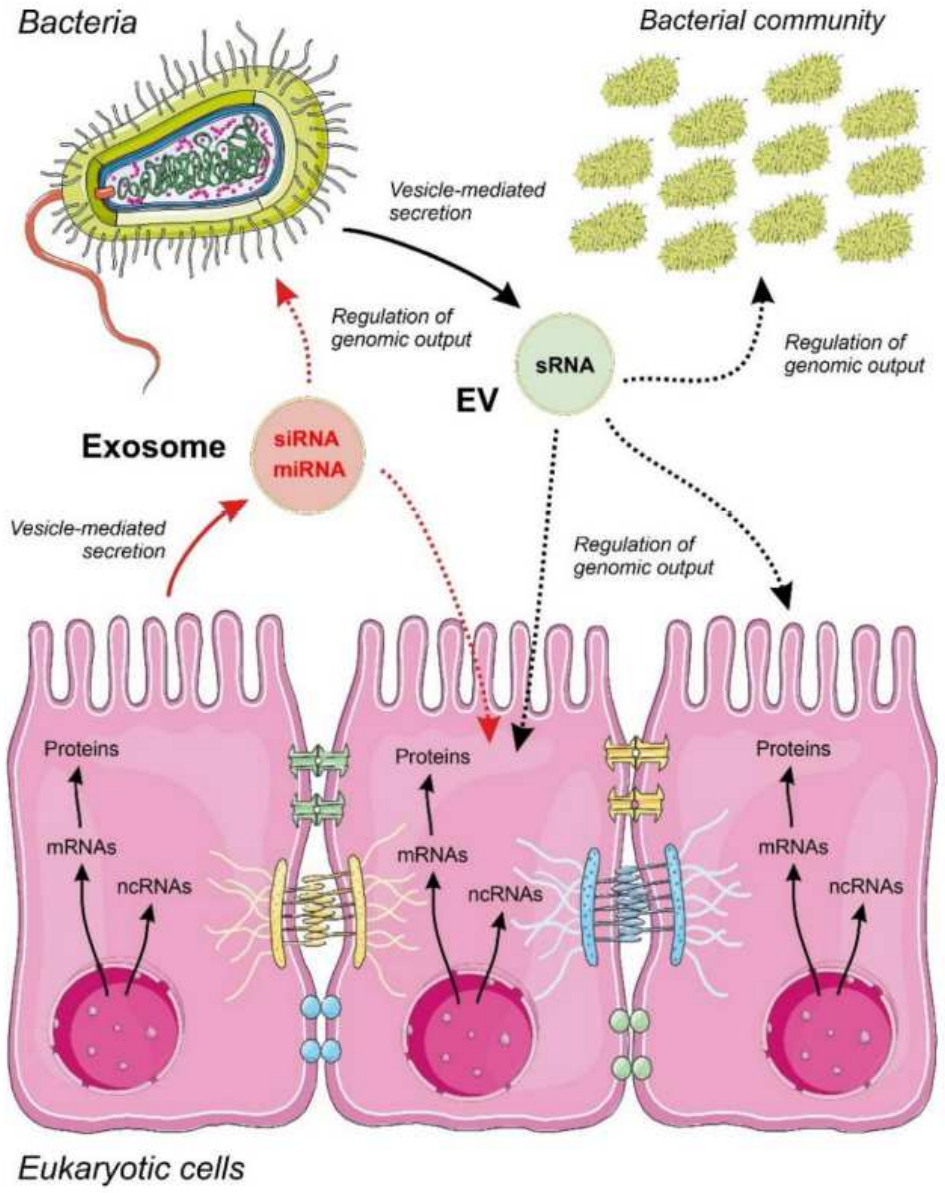
- • Transcription factors binding sites (expression of lncRNA transcripts)
- • Spliceosome/Ribosome (Intron sequences, in cis/in trans RNA regulatory elements)
- • RNA-mediated Chromatin remodeling (Histone codes and expression at promoters and enhancers)
- • An Armory of Regulatory Noncoding RNAs (miRs, ceRNAs, circRNAs, lncRNAs)

> **Evolution of Complexity**

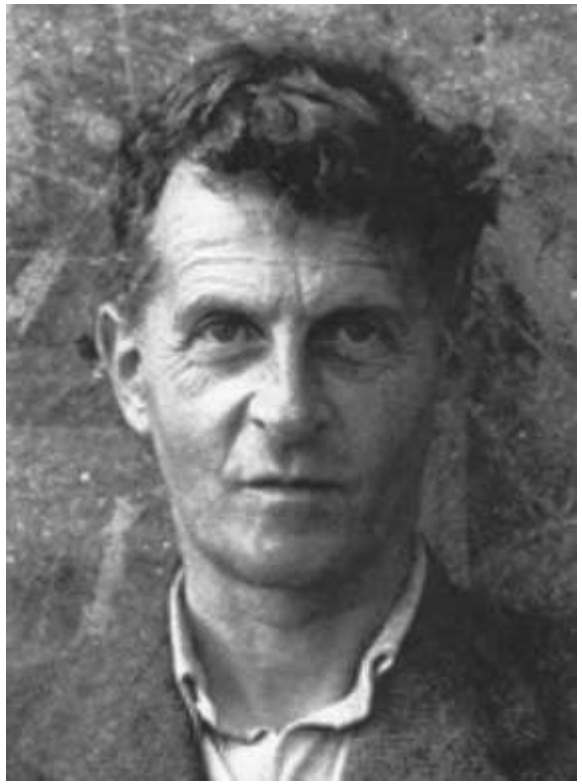
- • Response to Stress
- • Cell-to-cell communication,
- • Multicellular organisms, Development
- • Epigenetic inheritance systems with high hereditary potential
- • Cognition-based evolution
- • Formation of super-organisms, societies.

Cognition based evolution

- holobiome
- hologenome (hosting organism + all the associated microbiota)
- cell-to-cell communication (exosomes)
- acquisition, distribution, and management of information
- base pair mutations
- horizontal gene transfer
- recombination
- gene loss and duplication
- microbial loss/amplification
- ncRNA-mediated crosstalk between species within holobionts



Interspecies Communication in Holobionts by Non-Coding RNA Exchange



“I have always thought that Darwin was wrong: his theory doesn’t account for all this variety of species. It hasn’t the necessary multiplicity. Nowadays some people are fond of saying that at last evolution has produced a species that is able to understand the whole process which gave it birth. Now that you can’t say!’’

Ludwig Wittgenstein, 1951

Ευχαριστώ πολύ!