

Struttura e complessità del genoma

Si ringrazia il contributo di S. Zucchelli

Struttura e dimensioni di genomi cellulari (I)

Procarioti (Bacteria e Archaea)

Tipicamente c'è **un solo cromosoma** nel citoplasma, contenente una molecola di **DNA circolare** lungo da **0,5 a 10 milioni** di paia di basi circa.

Spesso possono essere presenti anche uno o più **plasmidi**, piccoli DNA circolari accessori, costituiti da alcune migliaia di paia di basi. Sono note eccezioni: alcuni procarioti con due o più cromosomi circolari, o un cromosoma lineare.

Struttura e dimensioni di genomi cellulari (II)

Eucarioti (unicellulari, piante, funghi, animali)

Di regola c'è **più di un cromosoma nel nucleo**, mediamente poche decine, anche se sono noti casi di parecchie centinaia e, all'altro limite, di uno solo.

Ciascun cromosoma contiene una molecola di DNA lineare. Nel nucleo può essere presente un solo corredo di cromosomi diversi (la cellula si dice **aploide**), due serie (**diploide**), e sono i casi più frequenti, o più (tri-, tetra-, esa-, ...poliploide).

Le **dimensioni di un corredo aploide** variano da **una decina di milioni** di paia di basi circa (ad esempio i lieviti o qualche alga unicellulare) a un **centinaio di miliardi** (alcune felci, i pesci polmonati e gli anfibi urodela) o anche più in alcune amebe.

Struttura e dimensioni di genomi cellulari (III)

Eucarioti (unicellulari, piante, funghi, animali)

Al di fuori del nucleo:

è generalmente presente un certo numero di **mitocondri**, dotati di un proprio minigenoma, costituito da una molecola di DNA circolare di circa **15.000 – 100.000** paia di basi.

I vegetali contengono anche un certo numero di **cloroplasti**, pure dotati di un proprio minigenoma, costituito da una molecola di DNA circolare di circa **100.000** paia di basi.

Struttura e dimensioni di genomi cellulari (IV)

Casi particolari:

Alcuni unicellulari sono privi di mitocondri (es.: *Giardia lamblia*), alcune alghe unicellulari (es. *Guillardia theta*) presentano due nuclei distinti ciascuno con più cromosomi (il minore è detto **nucleomorfo**).

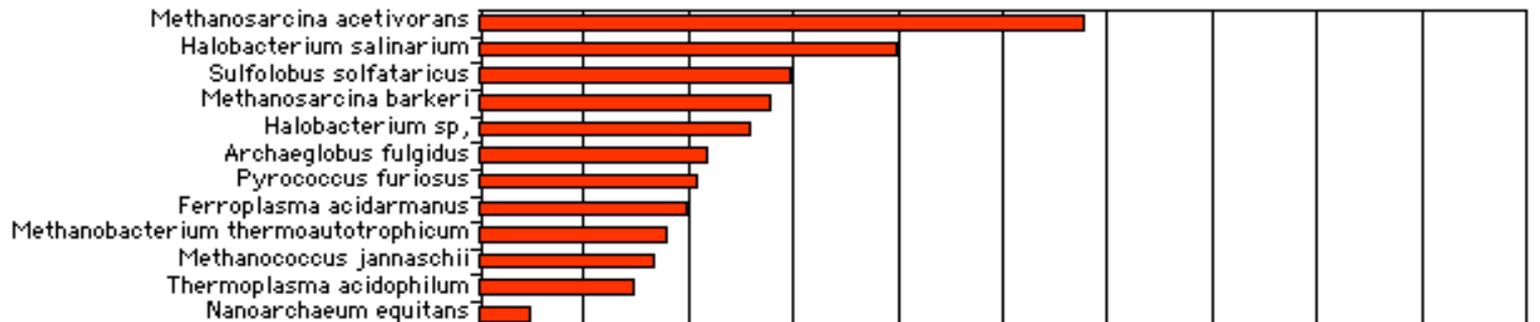
I Ciliati (es. *Tetrahymena*), accanto al nucleo normale, detto germinale o **micronucleo**, sviluppano secondariamente un secondo nucleo più grande, detto somatico o **macronucleo**, contenente molte migliaia di minicromosomi a DNA lineare, copie numerosissime di segmenti del DNA dei cromosomi del micronucleo.

Le cellule salivari di *Drosophila* replicano il proprio DNA a cascata una decina di volte senza disgiungerlo dando origine a **cromosomi giganti** di oltre 1.000 copie, detti **politenici**.

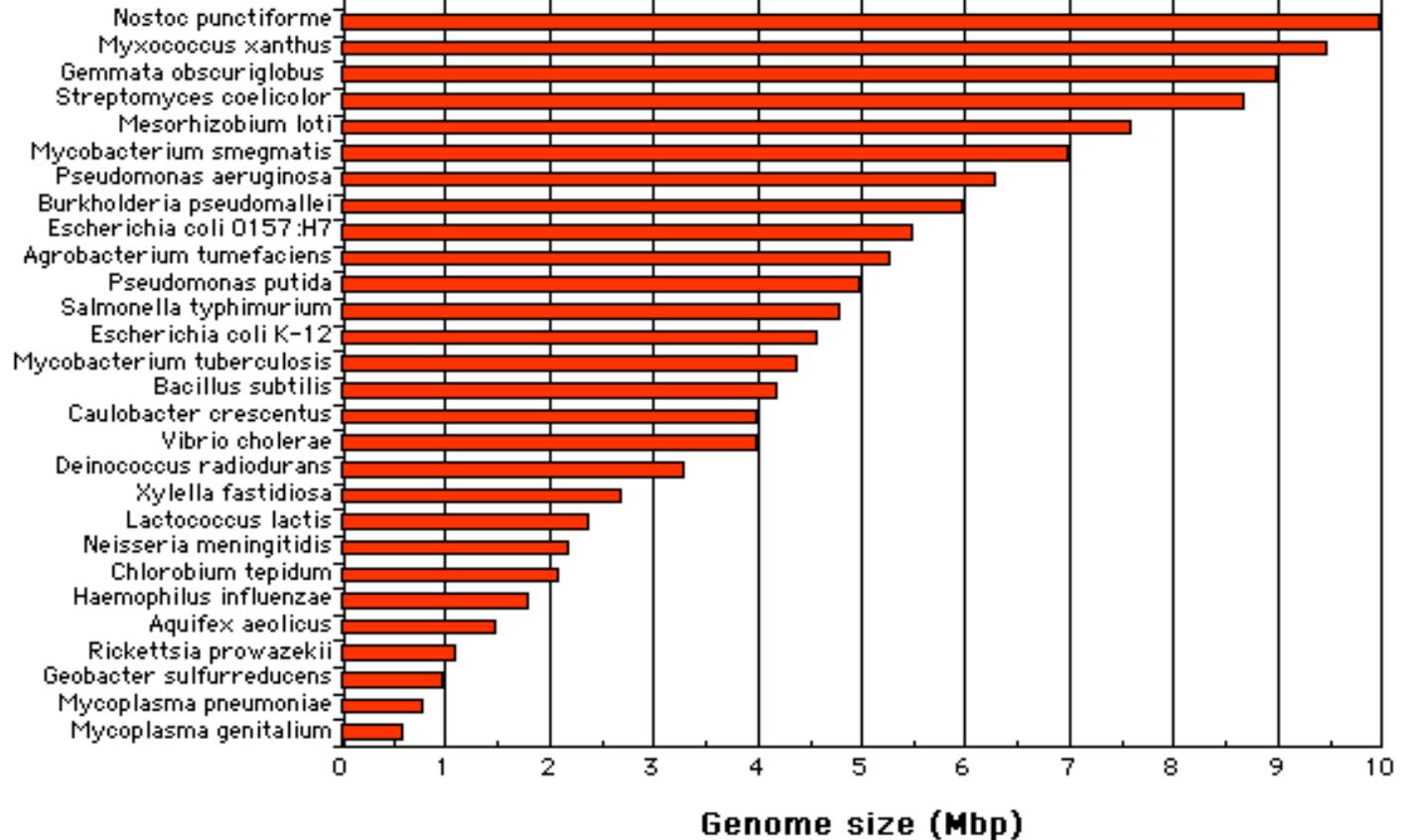
Il batterio gigante *Epulopiscium* contiene un numero molto elevato di copie del proprio genoma.

Correlazione fra dimensione del genoma e numero di geni

Archaea:



Bacteria:



Il paradosso C dei genomi eucariotici

Nei **procarioti** numero di geni e dimensioni genomiche delle varie specie sono approssimativamente proporzionali, in ragione di circa **1.000-1.200 pb/gene**. Appare ragionevole.

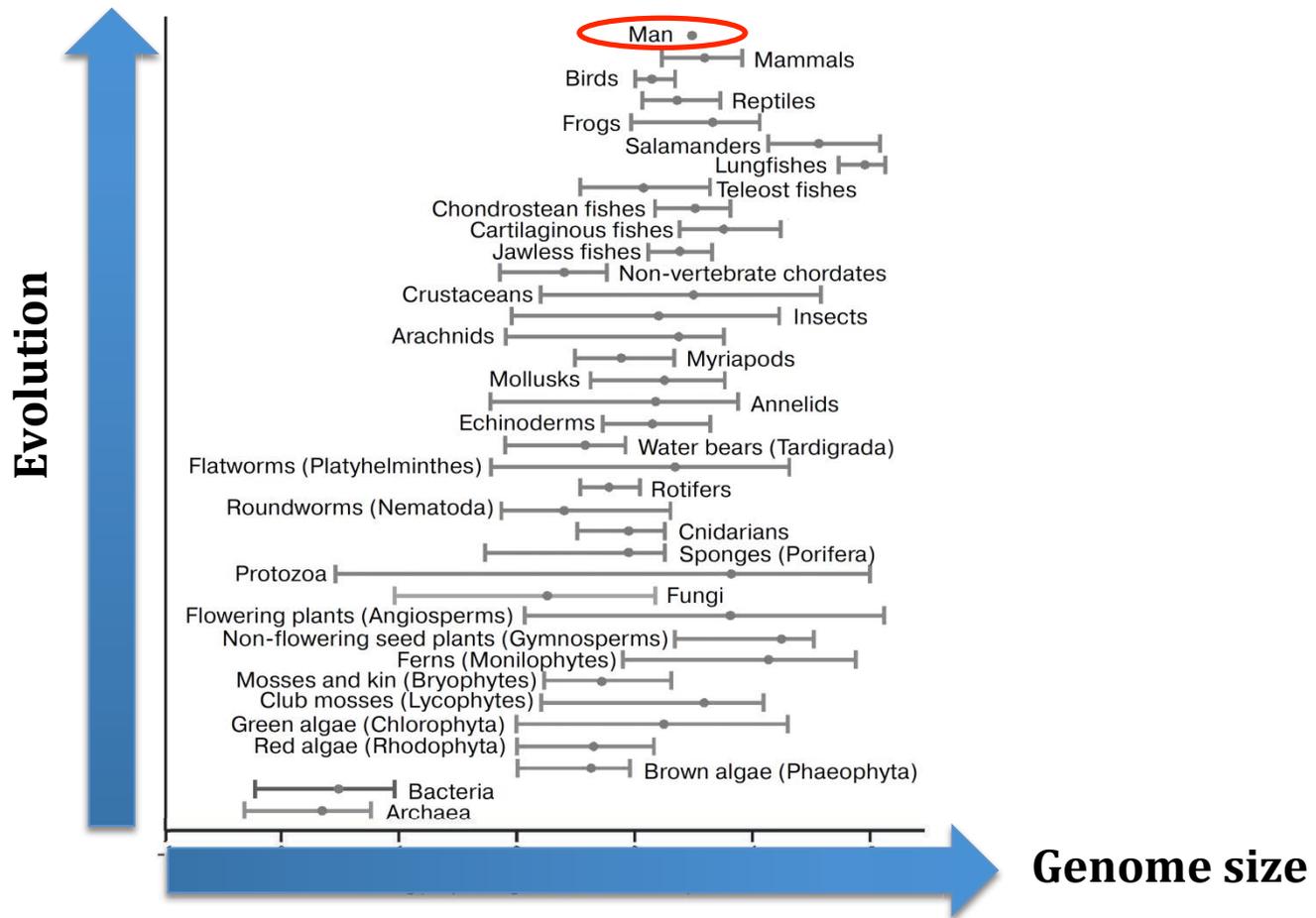
Le **dimensioni genomiche** (corredo aploide) nelle **specie eucariotiche** per contro variano enormemente (da circa 10^7 a più di 10^{11}) senza alcuna relazione con la **complessità dell'organismo**. Ad esempio, negli unicellulari alcune amebe hanno le massime dimensioni genomiche mentre alcune alghe unicellulari e alcuni lieviti le minime. Nelle piante superiori si va da circa 10^8 a circa 10^{11} e lo stesso succede con gli animali.

Questa osservazione è stata chiamata il **Paradosso C** (Complessità).

Genome Size and Developmental Complexity

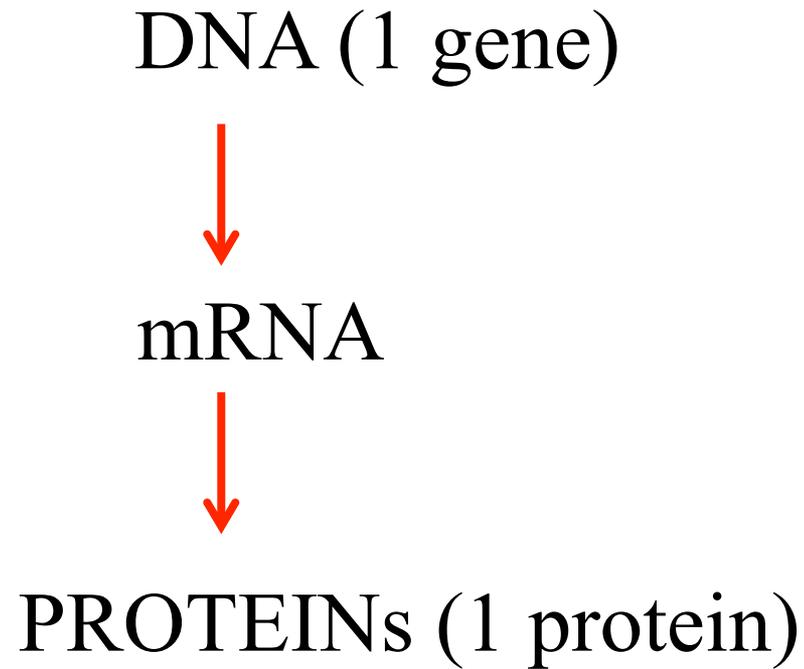
C- value or genome size paradox

The amount of genomic DNA does not correlate with organismal complexity



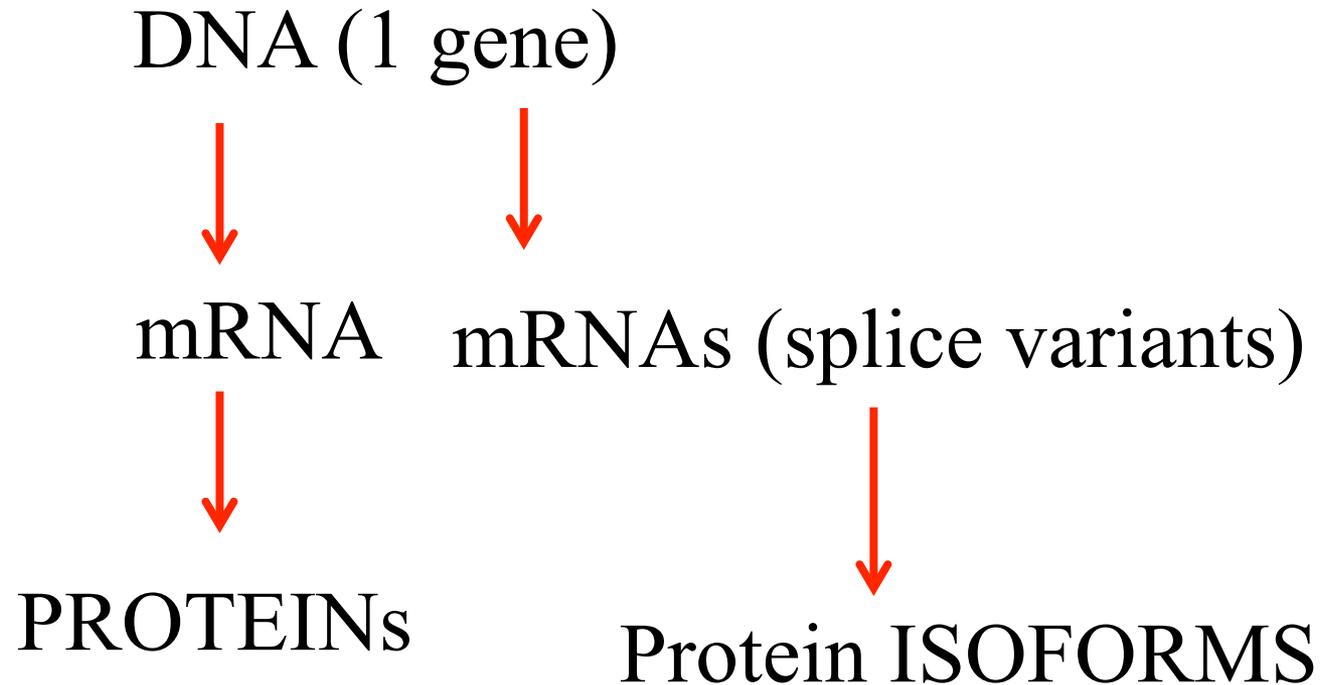
The central DOGMA

1958 (Crick)



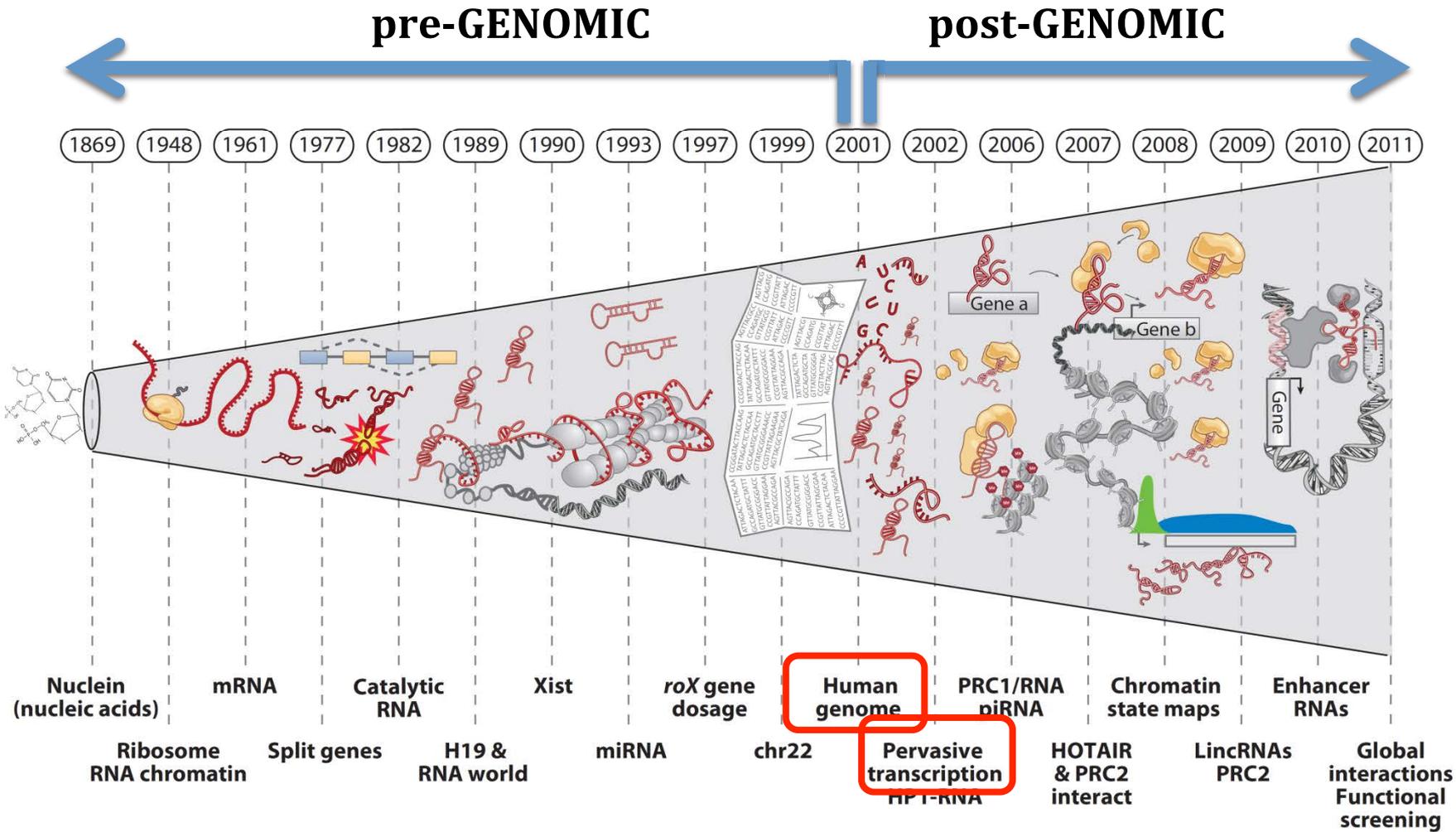
The central DOGMA

1970s-80s

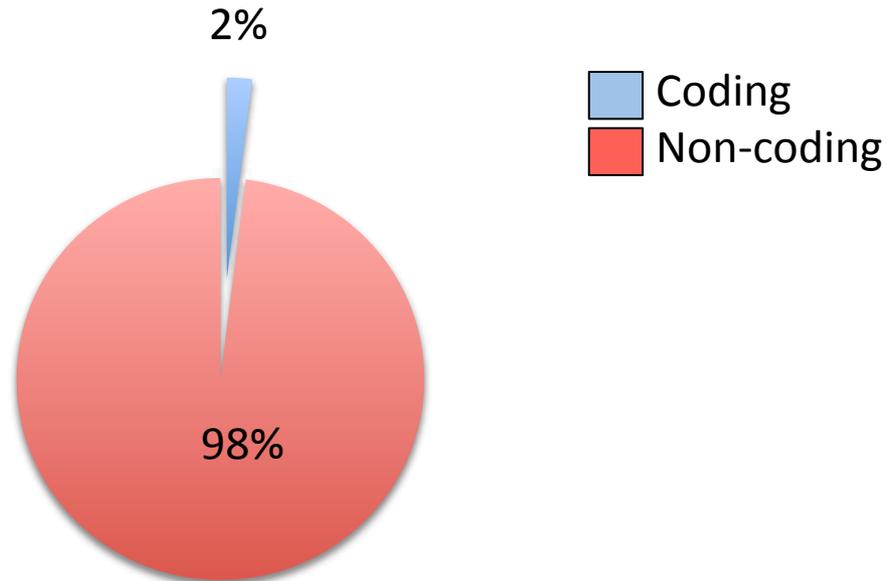


Era pre-GENOMICA e post-GENOMICA

Rinn & Chang Annu Rev Biochem 2012



The Whole Genome Sequence

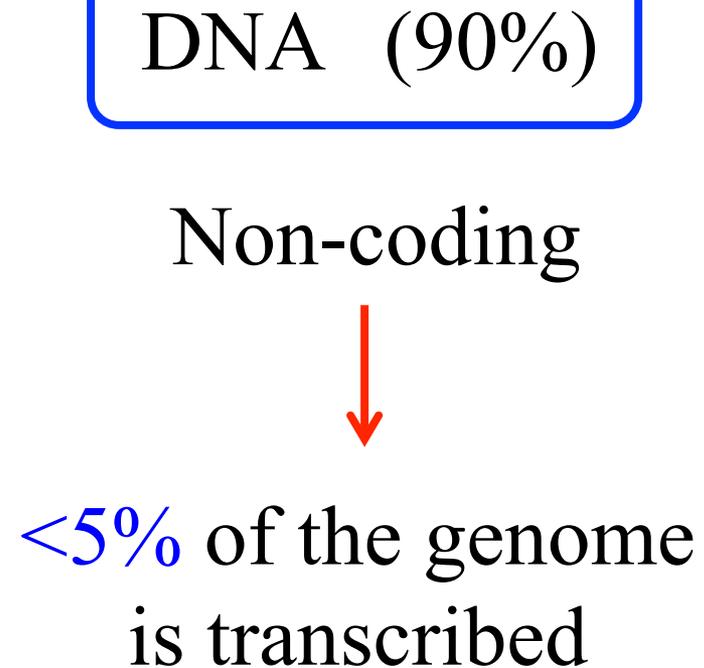
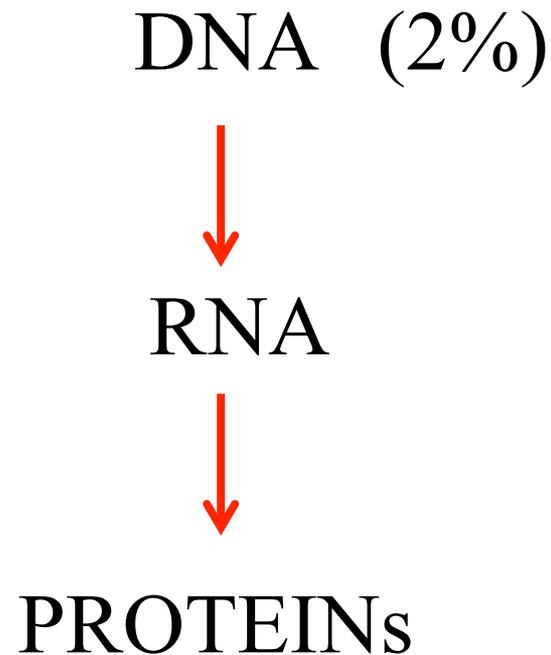


Issues:

#1 the vast majority of the genome is “Junk”

#2 the number of protein-coding genes does not correlate with genome evolution

The central dogma DERAILED



La grandezza del genoma è correlata alla complessità dell'organismo (?)

TABELLA 7.2 Comparazione della densità genica in genomi di differenti organismi

specie	Grandezza del genoma (Mb)	Numero approssimativo di geni
PROCARIOTI (batteri)		
<i>Mycoplasma genitalium</i>	0,58	500
<i>Streptococcus pneumoniae</i>	2,2	2300
<i>Escherichia coli</i> K-12	4,6	4400
<i>Agrobacterium tumefaciens</i>	5,7	5400
<i>Sinorhizobium meliloti</i>	6,7	6200
EUCARIOTI		
Funghi		
<i>Saccharomyces cerevisiae</i>	12	5800
<i>Schizosaccharomyces pombe</i>	12	4900
Protozoi		
<i>Tetrahymena thermophila</i>	125	27 000
Invertebrati		
<i>Caenorhabditis elegans</i>	103	20 000
<i>Drosophila melanogaster</i>	180	14 700
<i>Ciona intestinalis</i>	160	16 000
<i>Locusta migratoria</i>	5000	nd
Vertebrati		
<i>Fugu rubripes</i> (pesce palla)	393	22 000
<i>Homo sapiens</i>	3200	20 000
<i>Mus musculus</i> (topo)	2600	22 000
Piante		
<i>Arabidopsis thaliana</i>	120	26 500
<i>Oryza sativa</i> (riso)	430	~45 000
<i>Zea mays</i> (granoturco)	2200	> 45 000
<i>Triticum aestivum</i> (frumento)	16 000	nd
<i>Fritillaria assyriaca</i> (tulipano)	~120 000	nd

Il numero dei geni piuttosto che le dimensioni del genoma correlano con la complessità dell'organismo (?)

Il paradosso C dei genomi eucariotici

Il paradosso (in parte) rientra quando si scopre che la variazione nel **numero di geni** che codificano per proteine (che vanno da circa **5.000** a circa **30.000**) risulta molto più contenuta e approssimativamente connessa alla complessità dell'organismo (**paradosso G**).

Approssimativamente prima che si sequenziasse il genoma umano ci si aspettava di trovare un numero di geni codificanti per proteine fra 50.000 e 140.000!!!

Il genoma di *E. coli* è formato quasi interamente da geni

La maggior parte del cromosoma di *E. coli* codifica per proteine o RNA strutturali.

La maggior parte delle sequenze non codificanti è adibita a regolazione trascrizionale dei geni.

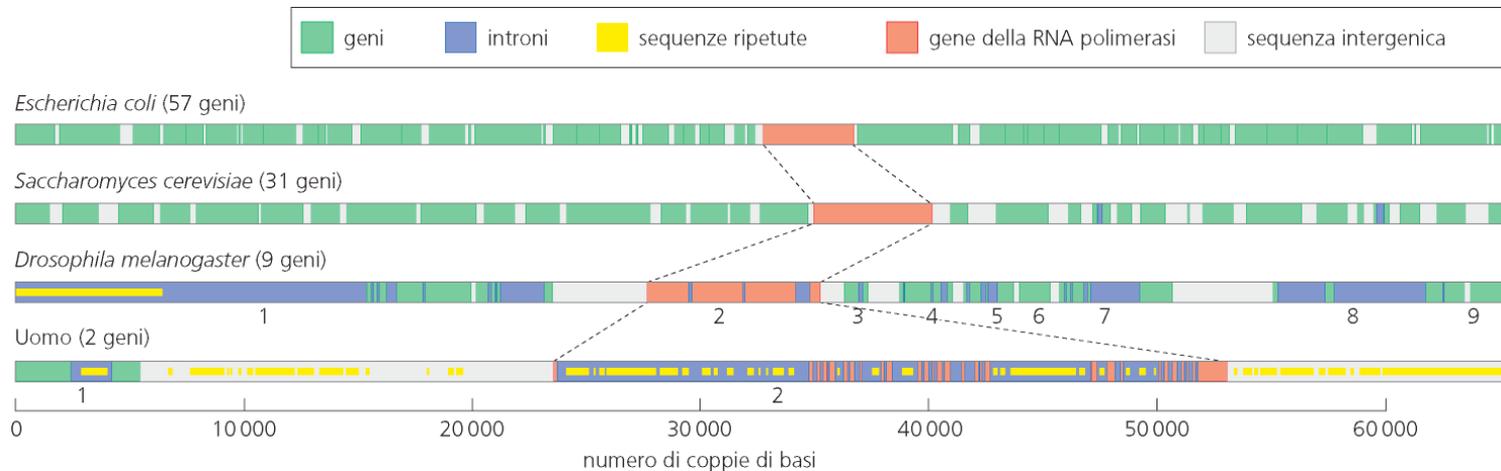
Spesso un solo sito di inizio trascrizione è usato per il controllo dell'espressione di più geni quindi queste **regioni regolative** sono **ridotte al minimo**.

Anche una origine di replicazione ma di poche centinaia di pb.

Gli organismi più complessi hanno una minor densità genica

Densità genica = n° geni /Mb.

Esiste una **approssimativa correlazione inversa** fra **densità genica** e la **complessità** di un **organismo**. Le densità geniche più elevate si trovano nei virus dove entrambi i filamenti di un tratto di DNA possono codificare per geni sovrapposti.



Regione genica codificante la subunità maggiore della RNA pol

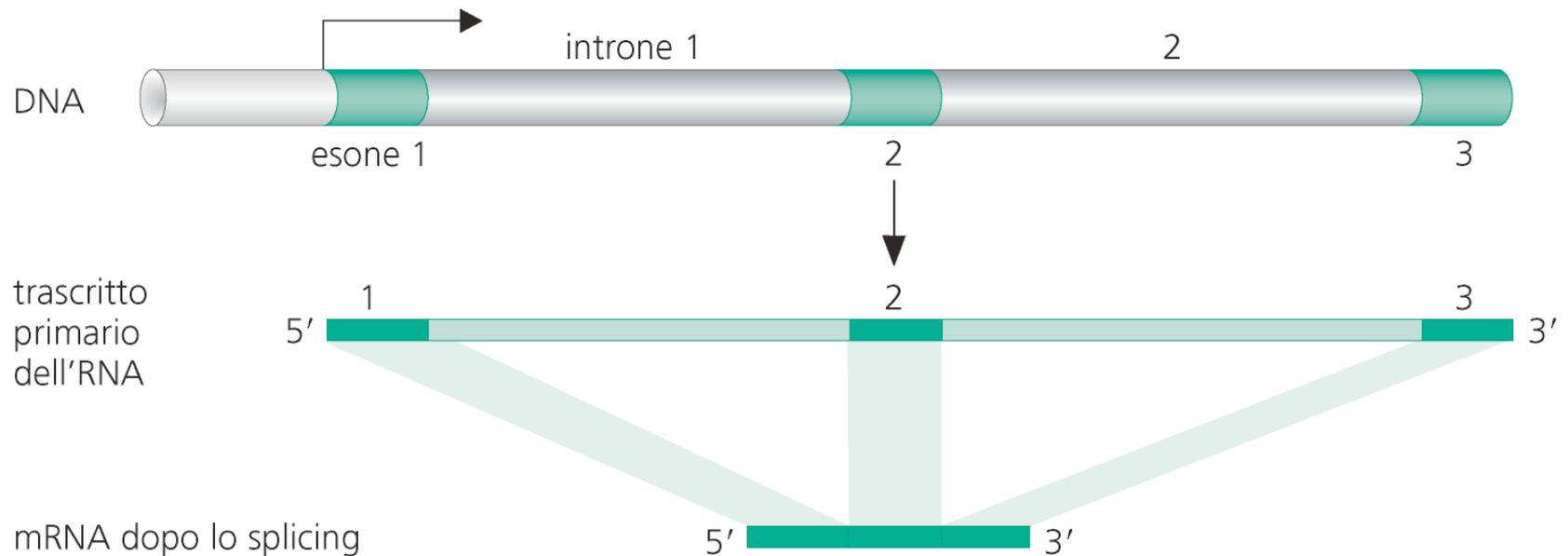
Gli organismi più complessi hanno una minor densità genica

TABELLA 7.2 Comparazione della densità genica in genomi di differenti organismi

specie	Grandezza del genoma (Mb)	Numero approssimativo di geni	Densità genica (geni/Mb)
PROCARIOTI (batteri)			
<i>Mycoplasma genitalium</i>	0,58	500	860
<i>Streptococcus pneumoniae</i>	2,2	2300	1060
<i>Escherichia coli</i> K-12	4,6	4400	950
<i>Agrobacterium tumefaciens</i>	5,7	5400	960
<i>Sinorhizobium meliloti</i>	6,7	6200	930
EUCARIOTI			
Funghi			
<i>Saccharomyces cerevisiae</i>	12	5800	480
<i>Schizosaccharomyces pombe</i>	12	4900	410
Protozoi			
<i>Tetrahymena thermophila</i>	125	27 000	220
Invertebrati			
<i>Caenorhabditis elegans</i>	103	20 000	190
<i>Drosophila melanogaster</i>	180	14 700	82
<i>Ciona intestinalis</i>	160	16 000	100
<i>Locusta migratoria</i>	5000	nd	nd
Vertebrati			
<i>Fugu rubripes</i> (pesce palla)	393	22 000	56
<i>Homo sapiens</i>	3200	20 000	6,25
<i>Mus musculus</i> (topo)	2600	22 000	8,5
Piante			
<i>Arabidopsis thaliana</i>	120	26 500	220
<i>Oryza sativa</i> (riso)	430	~45 000	~100
<i>Zea mays</i> (granoturco)	2200	> 45 000	> 20
<i>Triticum aestivum</i> (frumento)	16 000	nd	nd
<i>Fritillaria assyriaca</i> (tulipano)	~120 000	nd	nd

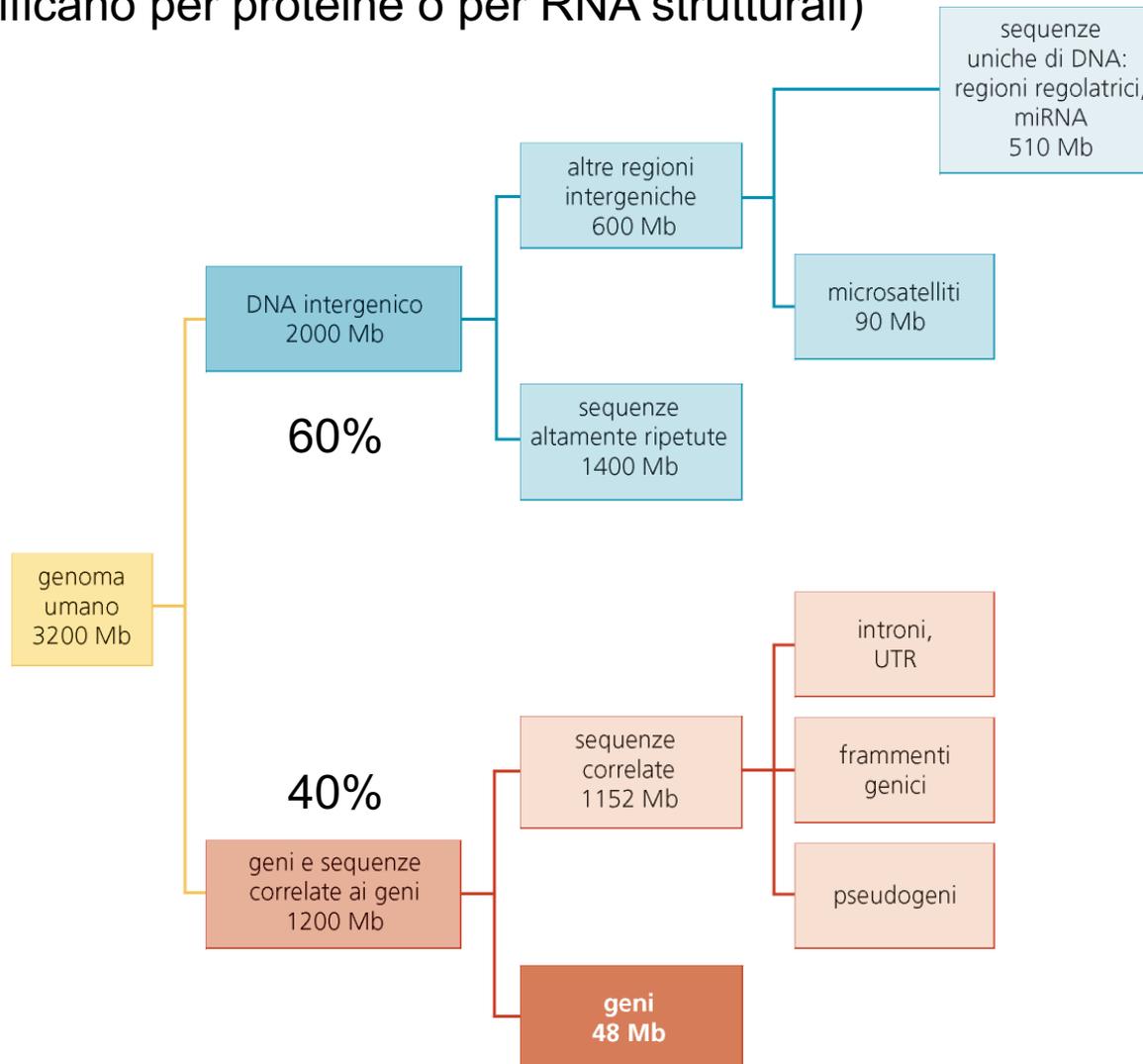
I geni rappresentano una piccola porzione del DNA cromosomico eucariotico

La diminuzione della densità genica è dovuta a:
1) Aumento delle dimensioni dei geni -> splicing

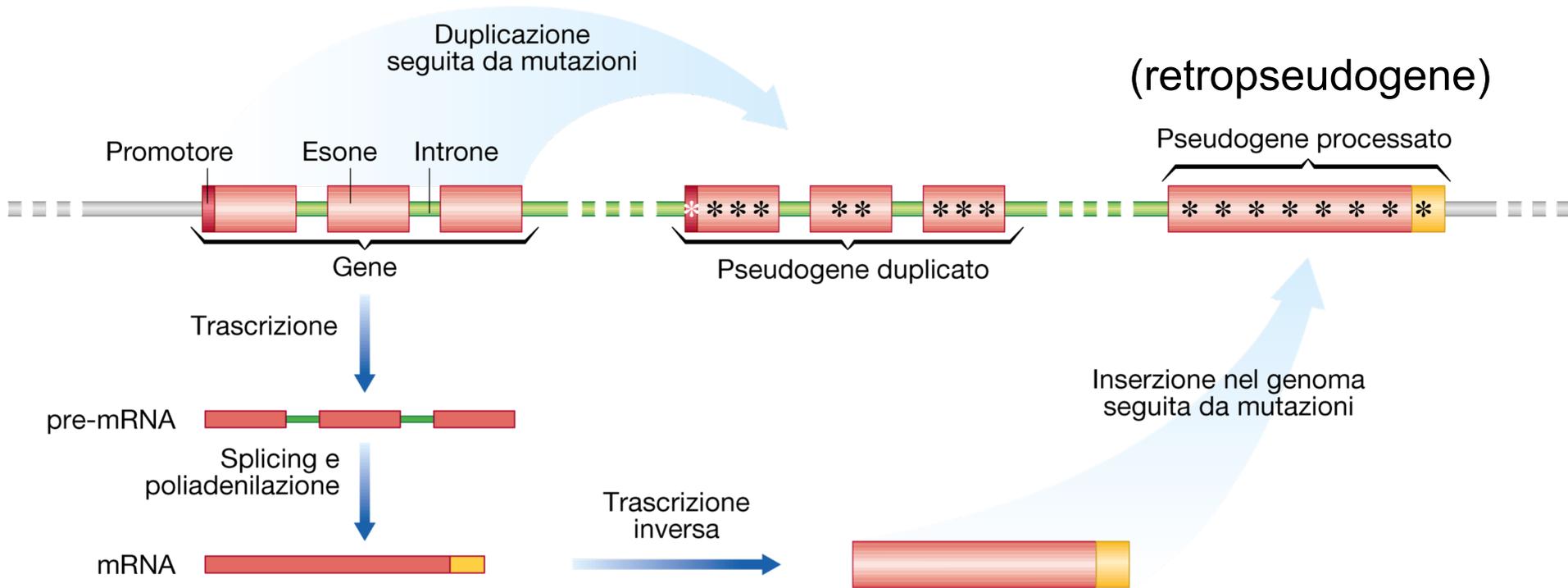


Organizzazione e contenuto del genoma umano

2) Aumento della quantità del DNA fra i geni -> **sequenze intergeniche** (non codificano per proteine o per RNA strutturali)



Origine degli pseudogeni



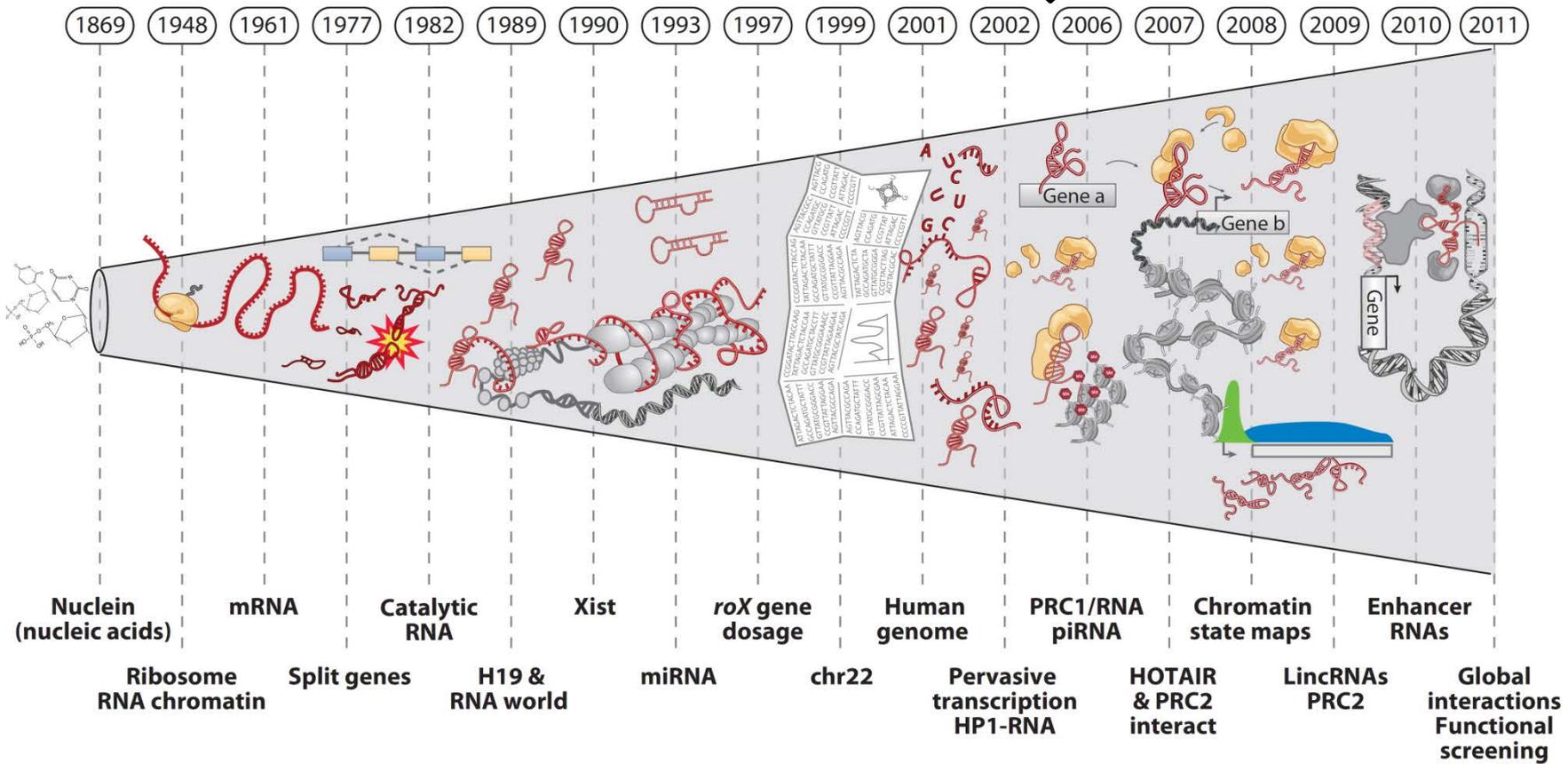
La maggior parte delle sequenze intergeniche umane è formata da DNA ripetuto

Studi sulle cinetiche di riassociazione di DNA genomico :

- 1) **Sequenze “uniche”**: quasi uniche non ripetute. Contiene la quasi totalità dei geni codificanti per proteine.
- 2) **Sequenze mediamente ripetute**:
 - a) **sequenze ripetute intersperse**, trasposoni (LINE e SINE) e retrotrasposoni attivi, alcuni geni presenti in elevato numero di copie (RNAr).
 - b) **microsatelliti**, monomeri intorno a 10 pb ripetuti in tandem 10-30 volte. Elementi SSR (Simple Sequence Repeats o STR (Short Tandem Repeats).
 - c) **minisatelliti**, monomeri 11-100 pb ripetuti fino a 20 kpb. Mini e micro polimorfici nella popolazione umana (scivolamento nella replicazione) usati come marcatori.
- 3) **Sequenze altamente ripetute: DNA satellite**, costituito da numerose (da decine di migliaia a milioni) ripetizioni in tandem di una breve sequenza (alcuni nucleotidi). Nell'uomo vari satelliti (ad es. satelliti alfa). Originano dalla difficoltà del duplicare DNA con accuratezza (usati anche come marcatori).

RNA Biology and Whole Transcriptome Analysis

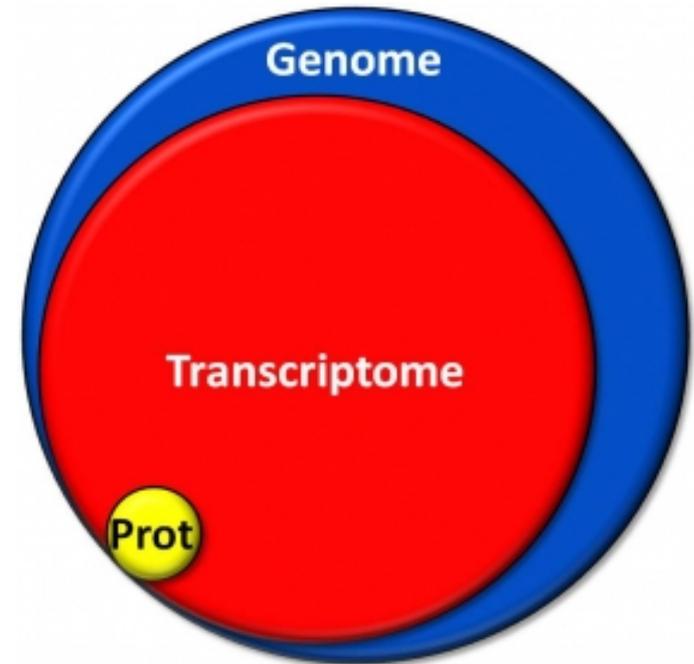
The FANTOM3 Consortium et al., 2005



PERVASIVE Transcription

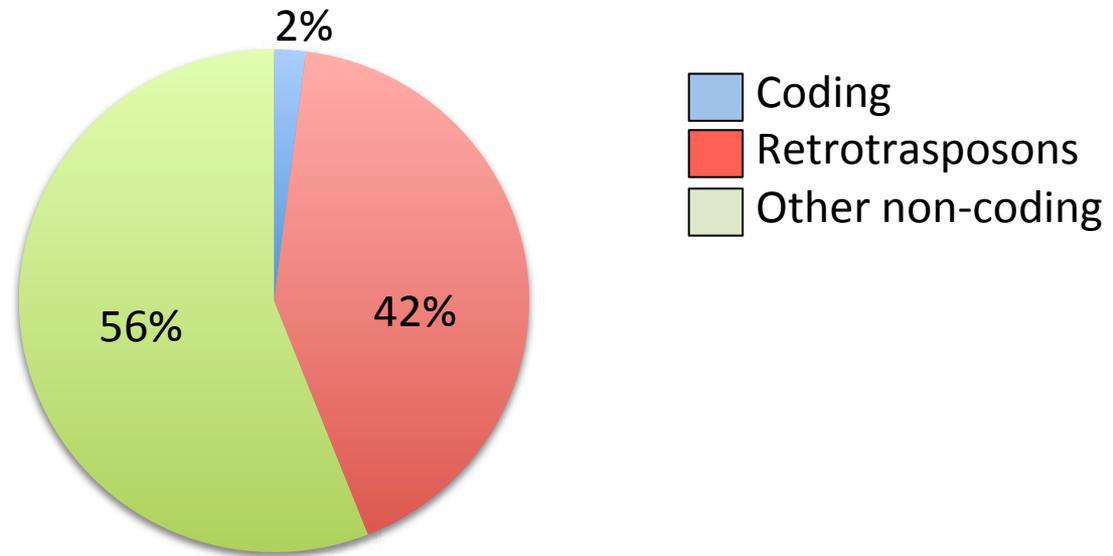
2003

Concept of “Pervasive Transcription”



- **The vast majority of the genome is TRANSCRIBED**
- **The vast majority of the transcripts DO NOT ENCODE for PROTEINS**

Pervasive transcription is REAL



#1 Identification of more than 34,000 non protein-coding transcripts

#2 About 50% of transcripts derive from repetitive elements



Piero Carninci

**Head of Genomics Research Centre - Functional Genomics, Genomica
Research Group Leader, Carninci Group**

Piero Carninci è un genetista, attualmente Team Leader del Laboratory for Transcriptome Technology e vice direttore del centro RIKEN per le scienze mediche integrative di Yokohama (Giappone). Piero è responsabile della creazione e dello sviluppo di svariate nuove tecnologie per il sequenziamento e l'analisi del DNA e del RNA. Ha partecipato e guidato numerose iniziative nazionali ed internazionali su larga scala, quali FANTOM, ENCODE e lo Human Cell Atlas. Da oltre 20 anni in Giappone, metterà la sua pluriennale esperienza al servizio del Centro di Genomica permettendo lo sviluppo e l'avviamento del Centro.

[Maggiori info](#)

<https://humantechnopole.it/it/people/piero-carninci/>

Marino Zerial nominato nuovo Direttore Human Technopole

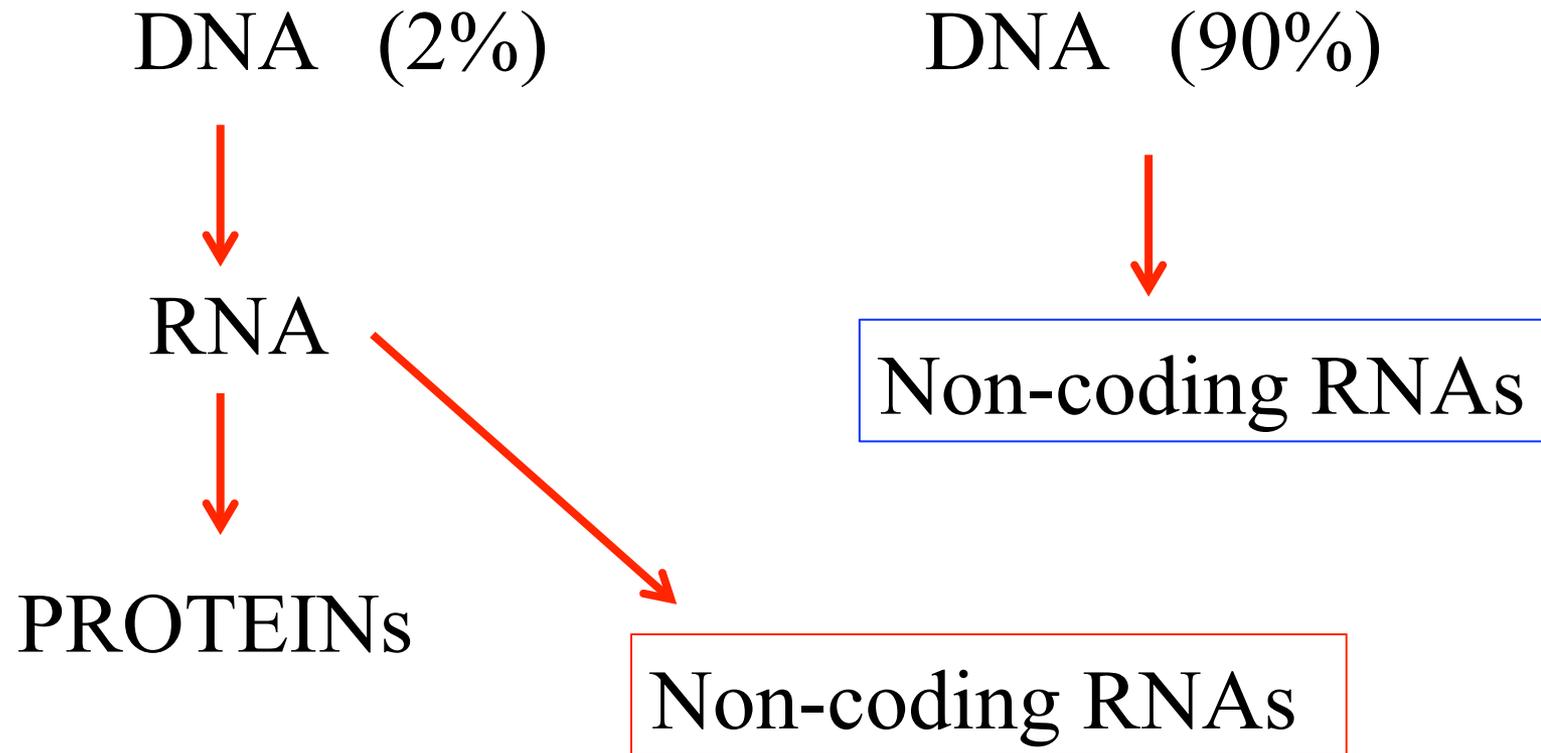


Il Consiglio di Sorveglianza della Fondazione ha nominato **Marino Zerial** quale nuovo Direttore di Human Technopole.

Sino ad oggi Direttore dell'Istituto Max Planck di Biologia Molecolare Cellulare e Genetica di Dresda in Germania, ente di ricerca che ha contribuito a fondare oltre 25 anni fa, Marino Zerial è inoltre professore onorario presso la Facoltà di Medicina della Technische Universität Dresden. Il suo percorso professionale, iniziato con una Laurea in Biologia presso l'Università di Trieste (1982), si è svolto interamente all'estero con esperienze di post-dottorato presso l'Institut Jacques Monod di Parigi e lo European Molecular Biology Laboratory di Heidelberg dove ha avviato il proprio gruppo di ricerca nel 1989.

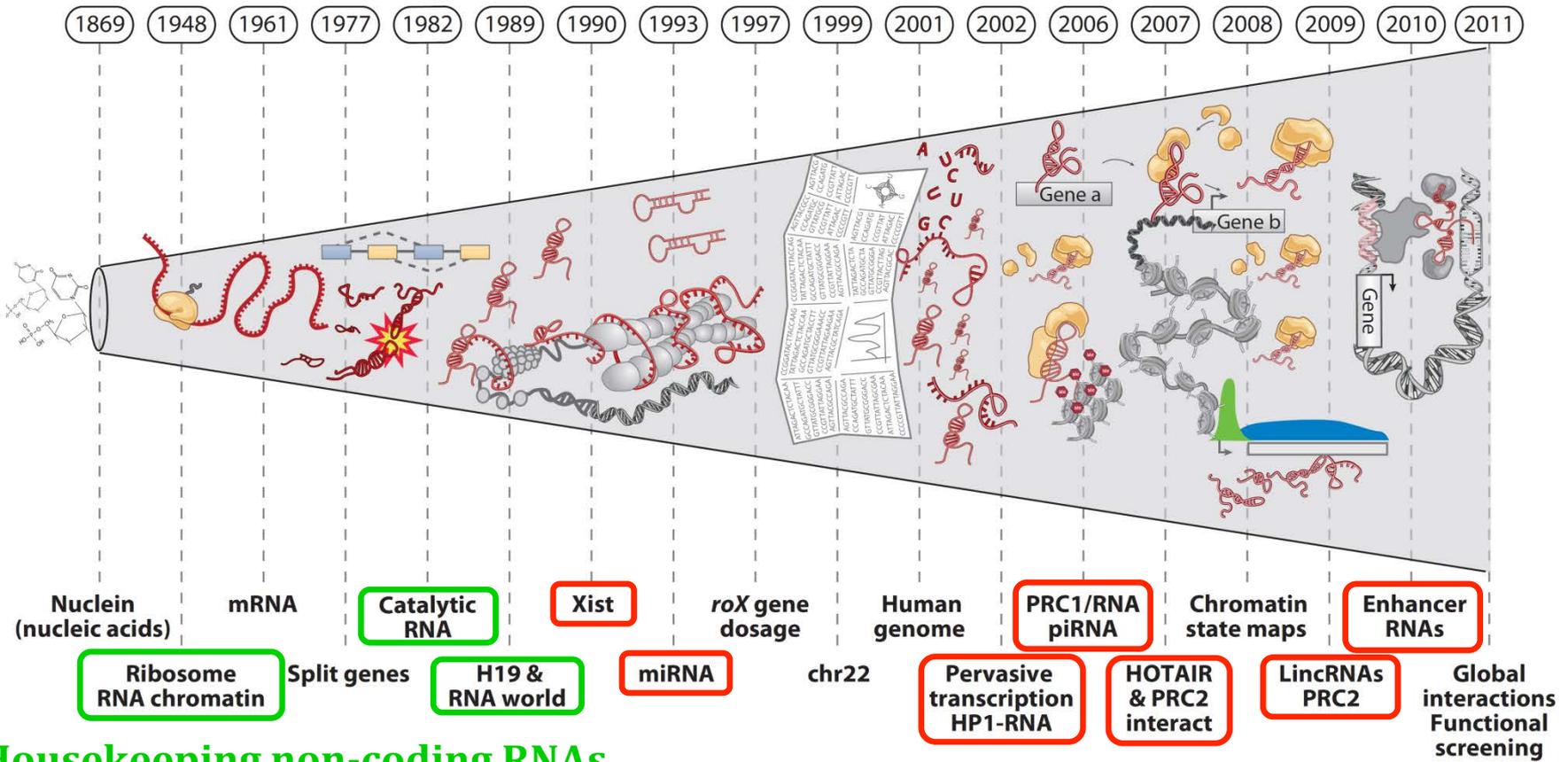
<https://humantechnopole.it/it/news/marino-zerial-nominato-nuovo-direttore-human-technopole>

The central dogma REVISITED



Next Generation Sequencing

Rinn & Chang Annu Rev Biochem 2012

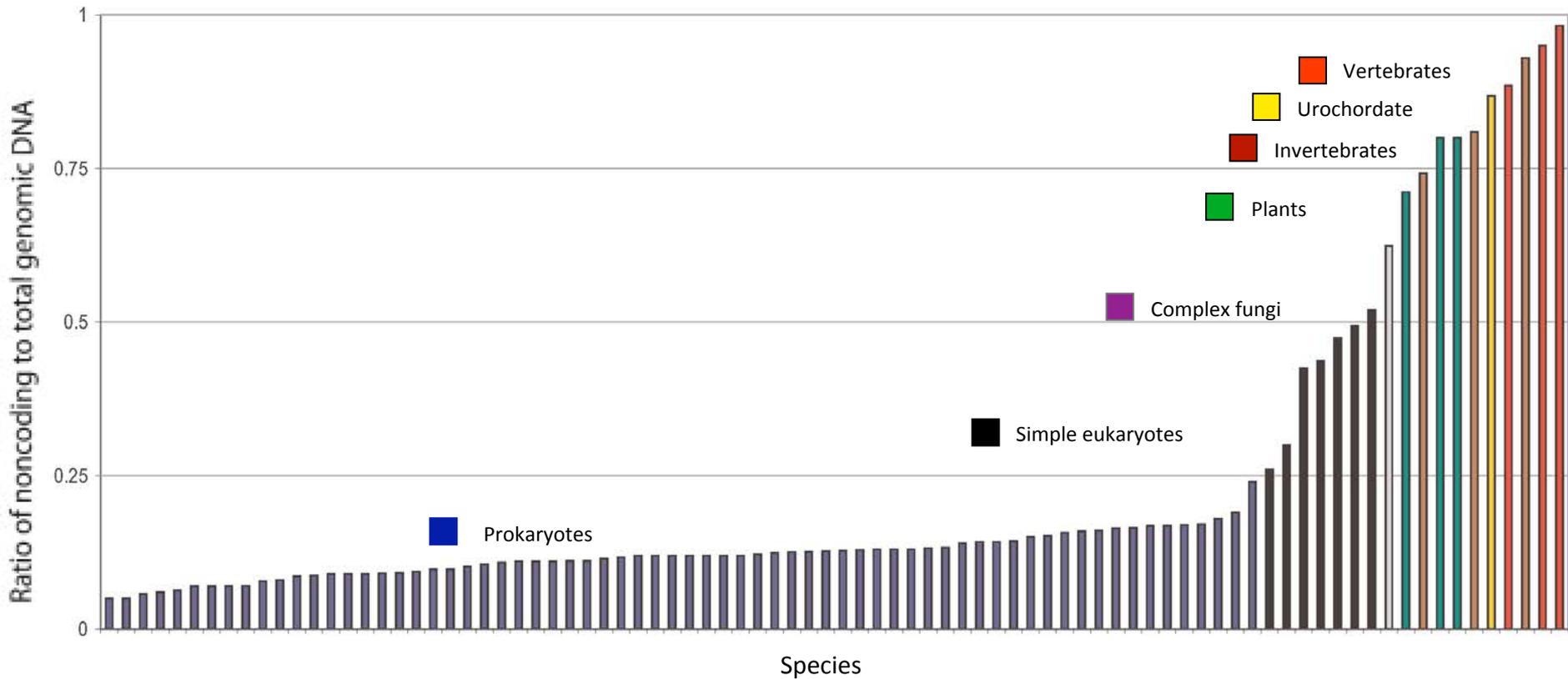


Housekeeping non-coding RNAs

Over 40,000 regulatory non-coding RNAs

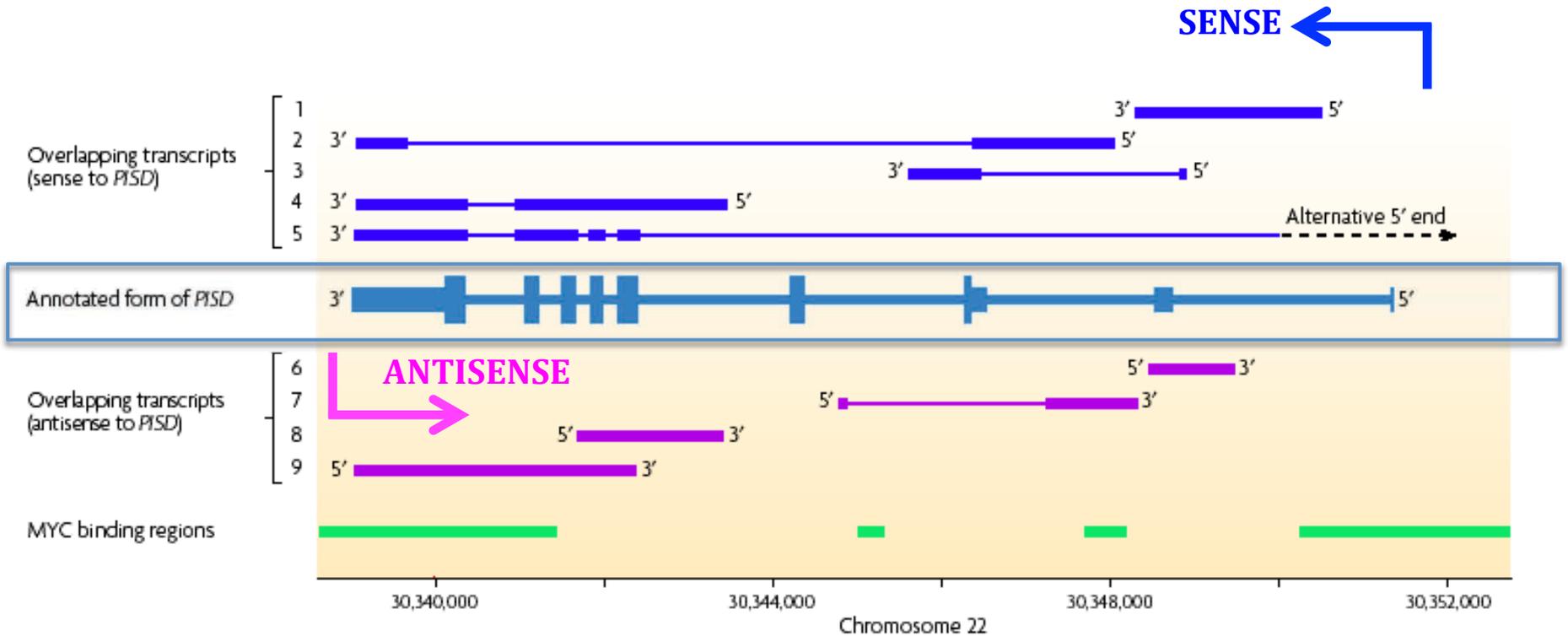
The Proportion of ncDNA Increases with Developmental Complexity

C- value paradox SOLVED



The concept of GENE is REVISITED

ONE GENE ->> SEVERAL TRANSCRIPTS
->> SEVERAL REGULATORY regions



Kapranov et al. Nature Review, 2007