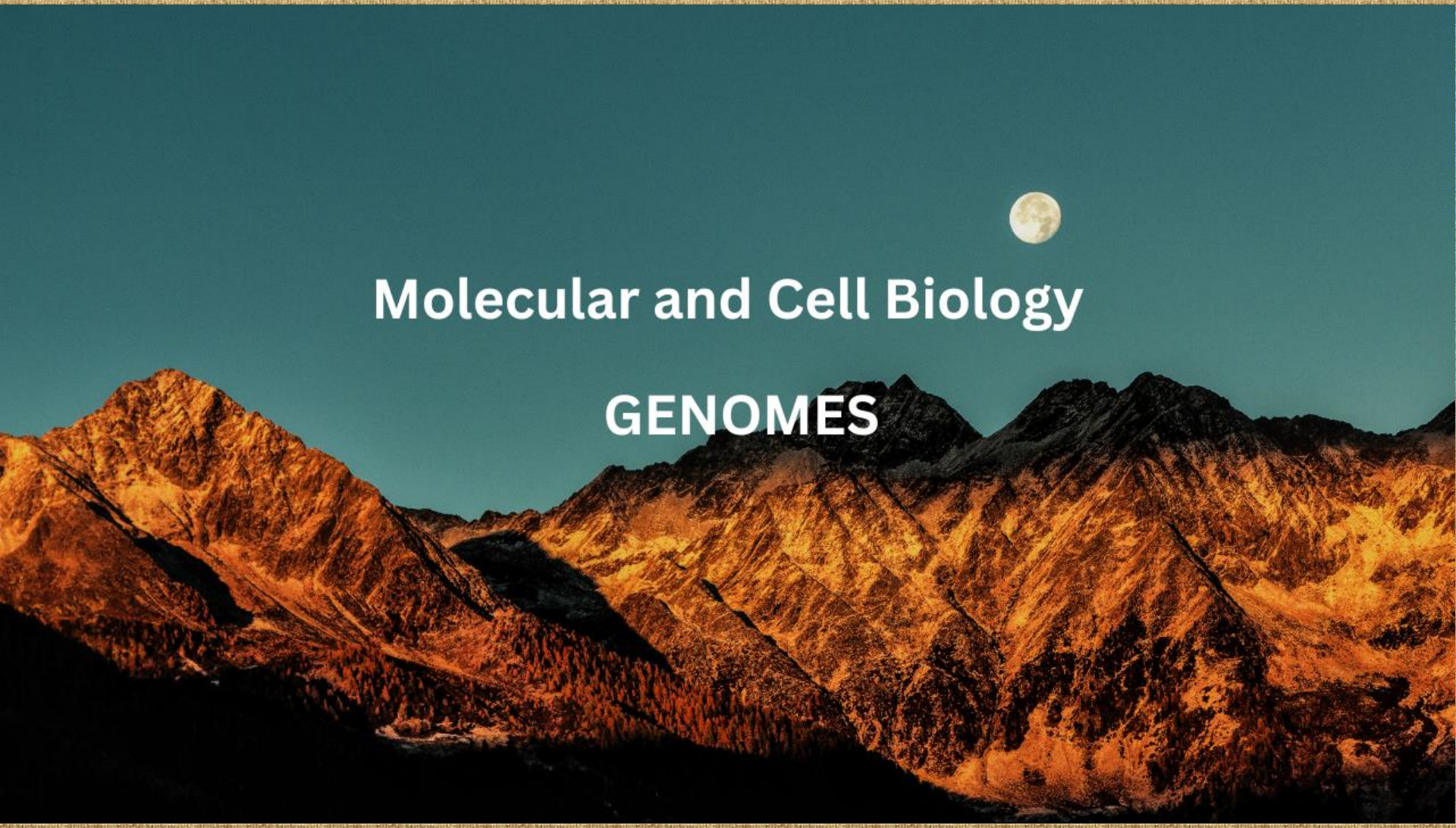

Molecular and Cell Biology
GENOMES

This is a Tutorial/Silent Lecture

What is a Tutorial/Silent Lecture?

a sequence of "slides" formatted to guide you through the exploration/study of the topic

you are the main actor in this active learning experience
think of it as working with a tutor without having to pay for it

as the slide sequence unfolds, you will get opportunities to engage with the material

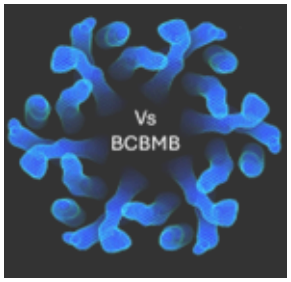
- **by thinking about/answering questions**,
(my answer is always provided on the next slide).
- **by completing a "short assignment"**
(it never will take more than a few minutes, if at all that long),
- by **watching a short video/clip**
(the embedded links will take you to my YouTube@VsBCBMB channel;
key moments are captured as still and are shown in the slide-deck, in case you don't want to watch the videos)

of course, you can skip the active learning aspect and look at the answers right away.

Why Give This a Go?

- **benefits: you set the pace** taking as much or as little time as you need.
- you **can turn tutorials/silent lectures into fully immersive experiences** (eg playing your favourite music while working through the content),
- **or invite friends to over the Q&A structured/guided materials together,** discussing the questions before looking at answers.

each of these features help you to hold on to the material.



INTRODUCING
MOLECULAR BIOLOGY and CELL BIOLOGY COLLECTION

ihis collection builds on the BIOCHEMISTRY FUNDAMENTALS chapters and assumes that your knowledge and understanding of biochemistry is at least as deep as what is presented in the FUNDAMENTALS collection.

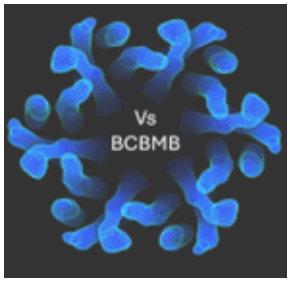
If you had no prior exposure to biochemistry, or if your introductory Molecular/Cell Biology course limits its biochemistry primer to less than four weeks worth of lectures, then I (strongly) recommend that you work through the BIOCHEMISTRY FUNDAMENTALS first.

it really will make a difference

browsing the chapters of this collection, the objective may differ from typical Molecular and Cell Biology courses in that it **prioritizes understanding over simple transmission of facts**. Moreover, the list of covered topics is not "complete" (by any means) and - at times – chapters go into significantly more depth than a typical introductory survey course would do.

ALL these aspects are intentional and reflect my preference for giving you an opportunity to understand the materials (which requires a certain depth), instead of ticking off a "mandatory" bucket list of topics.

in other words: studying the chapters in this Collection **will not** replace your own class but may serve as a useful supplement to things you cover in the class you take.

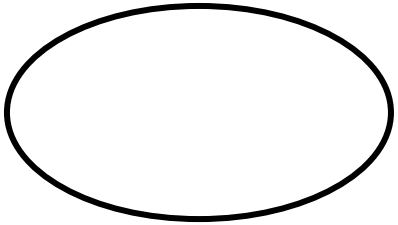


The Narrative

As for the Fundamentals and Advanced Biochemistry Collections, the chapters explore cells and their function from an "engineering" point of view = the basic questions to drive a narrative like this are: "**what** needs to be accomplished?", "**how** can you solve this problem?", and "**why** is this solution working?"

Approaching the exploration of cells and their functions this way, you have two principal options:
building the cell from the outside in or – the opposite – from the inside out.

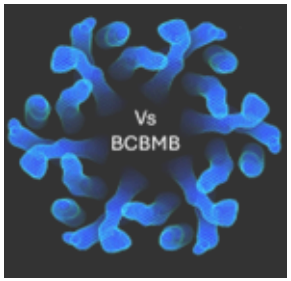
Starting with the following picture, both trajectories can produce logical storylines.



Yep – that's it! Plain and simple – this is your starting point…. the "Fundamentals Chapter" on Lipids explored how to build that boundary (in principle). Now – how do you turn this into a functional cell?

I would like to invite you to pause here and think of a sequence of chapters that could drive the storyline before continuing.

Start either on the cell surface and go down, or at the "core of the cell" and work your way up.
To make it even more fun – you can sit down with a group of friends and do this as a team, or as a friendly competition? Putting yourself to work here, in my experience, makes a big difference and gives you a different lease on studying the material (even in the actual class you take)

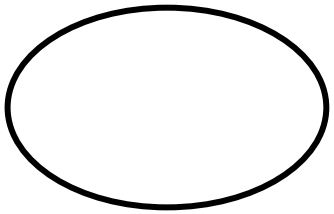


The Narrative

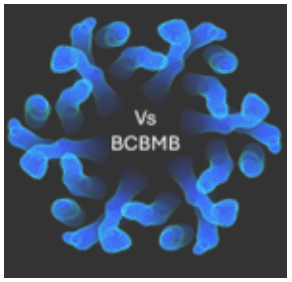
As a peace offer …. here is one way how to put a framework together

(note: these frameworks assume that certain processes like "cellular uptake of nutrients" etc just work [the detailed coverage of those is part of "Advanced Biochemistry"])

Inside → Out



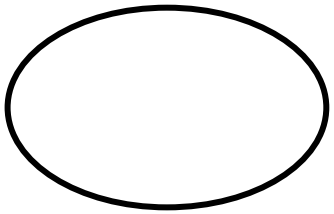
1/ What serves as "command center"?
(genome)
2/ How the "command center organized?"
(chromatin)
3/ How do you read out information?
(transcription)
4/ How do you turn 1D code into multidimensional output?
(translation)
5/ How do you regulate that step?
(regulation of gene expression)
6/ How do you duplicate the command center?
(replication)
7/ How do you separate and divide the genome/cell
(cytoskeleton, molecular motors, mitosis)
8/ How do you regulate cell division
(cell cycle, signaling, cell cycle regulation)
9/ How do you organize/coordinate different chemistries
(organelles)
10/ How do you integrate function of organelles
(vesicular transport, organelle contact sites)



The Narrative

as a peace offer …. here is one way how to put a framework together

(note: these frameworks assume that certain processes like "cellular uptake of nutrients" etc just work [the detailed coverage of those is part of "Advanced Biochemistry"]



Inside → Out

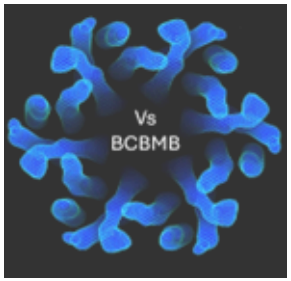
1/ What serves as "command center"?
(genome)
2/ How the "command center organized?"
(chromatin)
3/ How do you read out information?
(transcription)
4/ How do you turn 1D code into multidimensional output?
(translation)
5/ How do you regulate that step?
(regulation of gene expression)
6/ How do you duplicate the command center?
(replication)
7/ How do you separate and divide the genome/cell
(cytoskeleton, molecular motors, mitosis)
8/ How do you regulate cell division
(cell cycle, signaling, cell cycle regulation)
9/ How do you organize/coordinate different chemistries
(organelles)
10/ How do you integrate function of organelles
(vesicular transport, organelle contact sites)

Outside → In

Prokaryote	Eukaryote
1	9
2	10
3	---
4	4
5	1
6	2
7	3
8	5
	6
	7
	8

looking at these "solutions", other permutations are possible, but not all have (equally) smooth ways to "transition" between chapters.

note: in the "outside → in" narrative, you would want to distinguish between eukaryotes and prokaryotes. Possible – but not terribly "smooth". Specifically, much of that narrative relies on the question "where does the thing come from that we just used?" (eg translation needs mRNA → where does that come from, how made?). For prokaryotes – well those are single cells = you could use the "inside-out" sequence because you don't need to think about the complications that arise from having organelles.



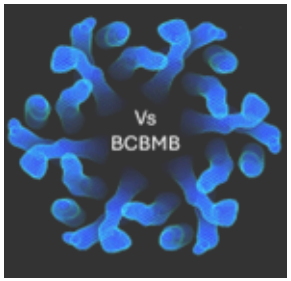
The Narrative and Goal of This First Chapter

for this collection of Chapters, I chose the "inside out" version. To me, this narrative is easier to build, starting at the very command center of a cell (its genetic material) and go from there.

Starting to design a cell from the inside out, the first pit stop is to look at how genetic material is organized

By the end of this Chapter you should know and be able to explain

What a genome is
Why genome size does not scale linearly with complexity of organism
Why “non-coding” is not a synonym for “non-functional”
Why large genomes require subcellular confinement and fragmentation



Setting the Stage

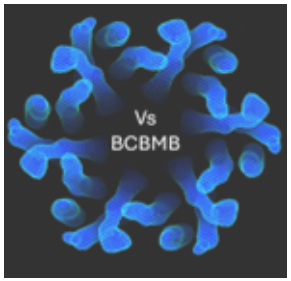
from exploration of biological polymers in the Fundamentals Collection: double stranded DNA is a conceptually simple molecular solution to a cell's need to store large amounts of information.

That said: two simple questions come to mind:

(1) just how much information is needed to encode a small cell like a bacterium?

(2) how much more information is needed for a complex organism like ourselves?

…try to think of answers…



Setting the Stage

From exploration of biological polymers in the Fundamentals Collection: double stranded DNA is a conceptually simple molecular solution to a cell's need to store large amounts of information.

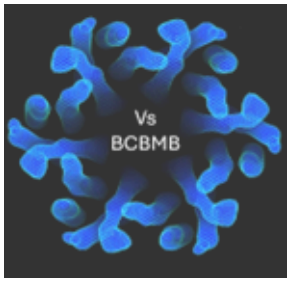
That said: two simple questions come to mind:

(1) **just how much information is needed to encode a small cell like a bacterium?**

(2) **how much more information is needed for a complex organism like ourselves?**

Answers:

(1) synthetic biologists have "streamlined" bacterial genomes to identify the minimum needed for a viable organism → **473 genes needed**, ~31% of those had unknown functions at the time of design [reference]



Setting the Stage

From exploration of biological polymers in the Fundamentals Collection: double stranded DNA is a conceptually simple molecular solution to a cell's need to store large amounts of information.

That said: two simple questions come to mind:

(1) just how much information is needed to encode a small cell like a bacterium?

(2) how much more information is needed for a complex organism like ourselves?

Answers:

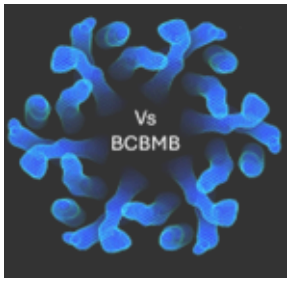
(1) synthetic biologists have "streamlined" bacterial genomes to identify the minimum needed for a viable organism → **473 genes needed**, ~31% of those had unknown functions at the time of design [reference]

outside the synthetic biology realm: typical **bacterial genome is ~4.6x10^6 bp.**

- for simplicity assume: all of this codes for proteins (main engines that drive the chemistry) = repertoire of 20 amino acids; to code: need 3 bp/amino acid (see chapter "TRANSLATION")
  - Assume that typical protein is **300-350 amino acids long ≈ 1000 bp/protein**

→ **predict: 4,600 proteins** encoded, **actual: 4255 proteins in E. coli**
= good estimate = very "lean"....which is what one would expect (why? → intuitively makes no sense to carry excess material that does not code for anything = expensive to maintain, slower to copy, harder to manage)

➔ an actual and typical bacterium has a huge "excess" of genes over the bare minimum ....however....that additional information increases fitness and adaptability. This is what the synthetic biologists say about it: >>*The work described here has been conducted in medium that supplies virtually all the small molecules required for life. A minimal genome determined under such permissive conditions should reveal a core set of environment independent functions that are necessary and sufficient for life.* ***Under less permissive conditions, we expect that additional genes will be required.***<< *(wow….)*



Setting the Stage - Continued

(1) just how much information is needed to encode a small cell like a bacterium?

(2) how much more information is needed for a complex organism like ourselves?

Answer:

complement of **protein-dependent functions** required to construct all cell types in **humans is ~20,000** with ~10,000 (or less) expressed in any one cell type

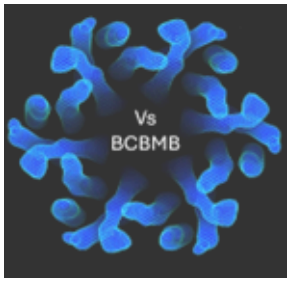
- Assume same ≈ 1000 bp/protein

→ **predict:** genome size of **21×10^6**

→ **actual:** **~3.3×10^9 (haploid n=1) → ~100x larger than expected!** → not intuitive

➔ **what causes the huge discrepancy between expected and actual size of mammalian genomes?**

...try to come up with hypotheses....



Setting the Stage - Continued

(1) Just how much information is needed to encode a small cell like a bacterium?

(2) How much more information is needed for a complex organism like ourselves?

Answer:

complement of **protein-dependent functions** required to construct all cell types in **humans is ~20,000** with ~10,000 (or less) expressed in any one cell type

- Assume same ≈ 1000 bp/protein

→ **predict:** genome size of **21×10^6**

→ **actual: ~3.3×10^9 (haploid n=1) → ~100x larger than expected!** → not intuitive

➔ **what causes the huge discrepancy between expected and actual size of mammalian genomes?**

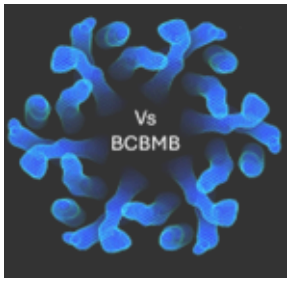
possible Answers:

- several copies for each protein, **or**
- a lot of DNA that does not code for proteins, **or**
- both

how can you test that?

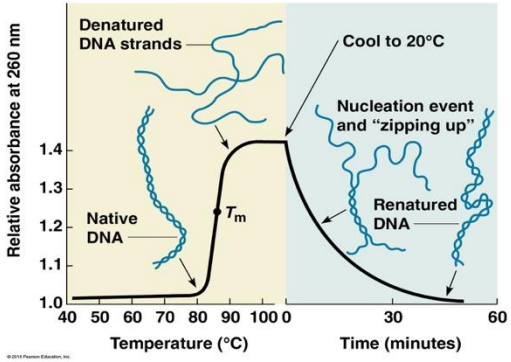
thought experiment

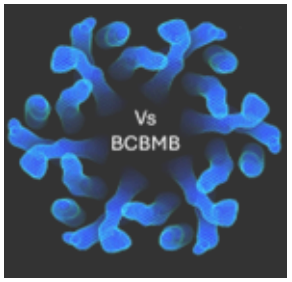
take the genomes of E coli and mammalian cells → mechanically break into small random fragments → heat to allow "melting" of the double strand → <u>slowly</u> cool down and observe how quickly strands anneal as complementary sequences come back together.



Molecular Cause for Genome Size Paradox

background: DNA strand separation and annealing can be followed by measuring the absorbance at 260nm (T_m = melting temperature)

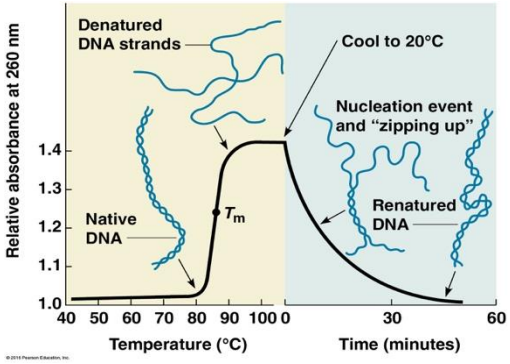




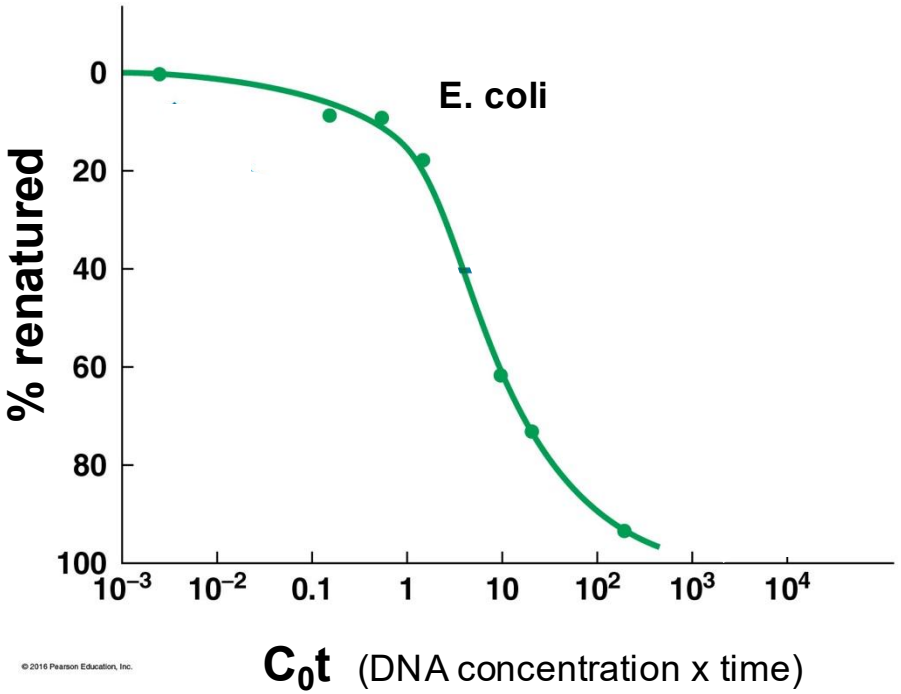
Molecular Cause for Genome Size Paradox

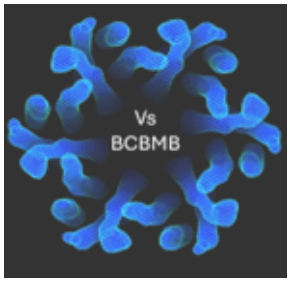
what do you see?
...try to answer ....

background: DNA strand separation and annealing can be followed by measuring the absorbance at 260nm (T_m = melting temperature)



interpretation:
...try to interpret....





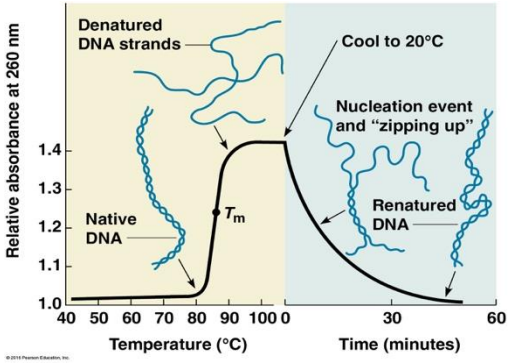
Molecular Cause for Genome Size Paradox

what do you see?

E coli curve:

- x-axis is a log scale (!), weird unit ; y-axis is inverted
- Shape: Initial plateau → steep drop off →"asymptotic" behavior at large "x"

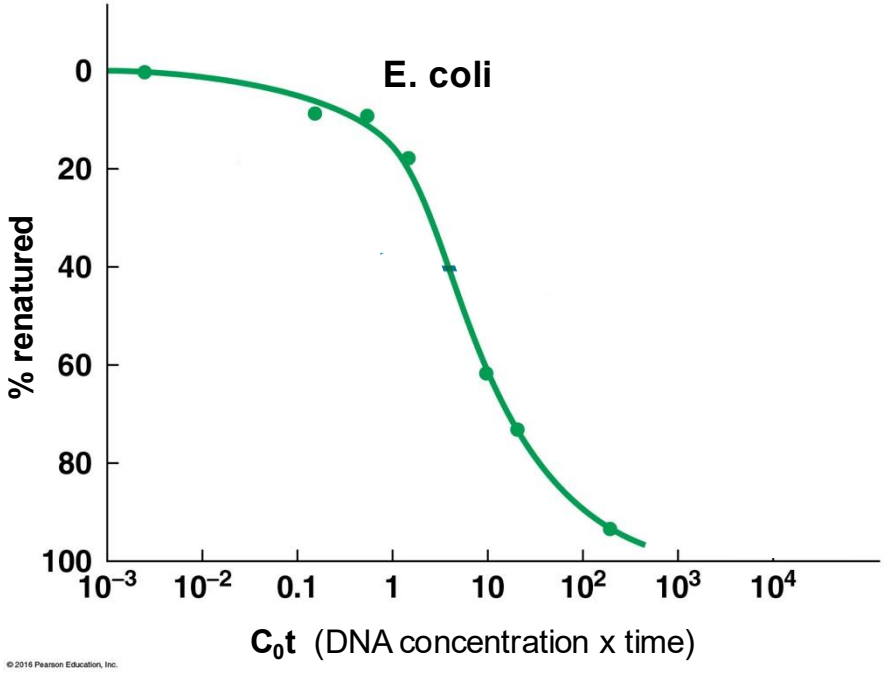
background: DNA strand separation and annealing can be followed by measuring the absorbance at 260nm (T_m = melting temperature)

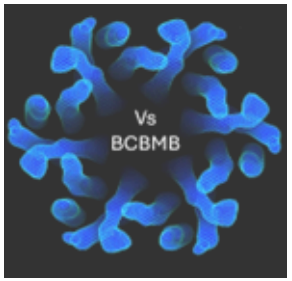


interpretation: x-Axis C_0t = normalize for amount (C_0) of DNA used so you can compare results between expriments.

- initial plateau (lag phase) = unique sequences (like coding regions for proteins) need to find each other (this is concentration dependent → unique means effective [ ] is quite low) ...but once they do, annealing is very fast (= steep drop off) ...until it slows again as it approaches 100% renaturation;

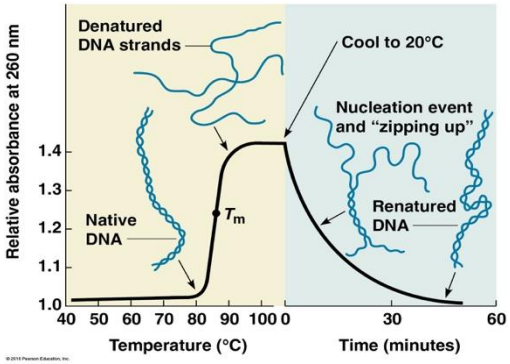
- plateau also says that the total amount of repetitive sequences (those that exist severall/manny times is low since these would find each other quicky and lead ot a significant increase in renatured material right away)





Molecular Cause for Genome Size Paradox

background: DNA strand separation and annealing can be followed by measuring the absorbance at 260nm (T_m = melting temperature)



Calf DNA – What do you see?

…try…..

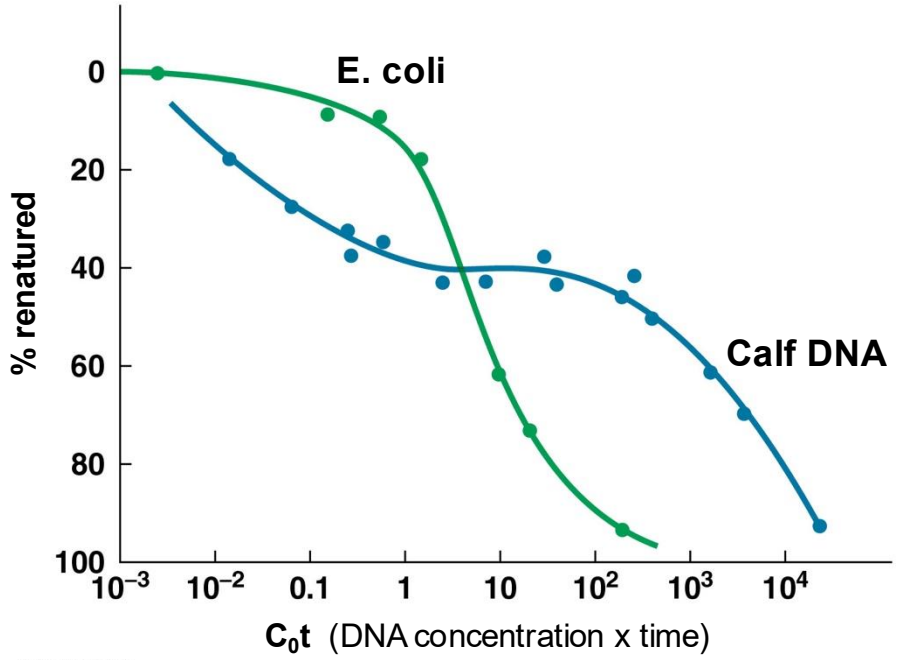
what do you see?
E coli curve:

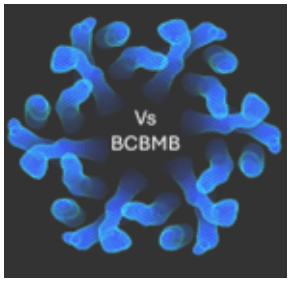
- x-axis is a log scale (!); weird unit; y-axis is inverted
- Shape: Initial plateau → steep drop off →"asymptotic" behavior at large "x"

interpretation: x-Axis C_0t = normalize for amount (C_0) of DNA used so you can compare results between expriments

- initial plateau (lag phase) = unique sequences (like coding regions for proteins) need to find each other (this is concentration dependent → unique means effective [ ] is quite low) ...but once they do, annealing is very fast (= steep drop off) ...until it slows again as it approaches 100% renaturation;

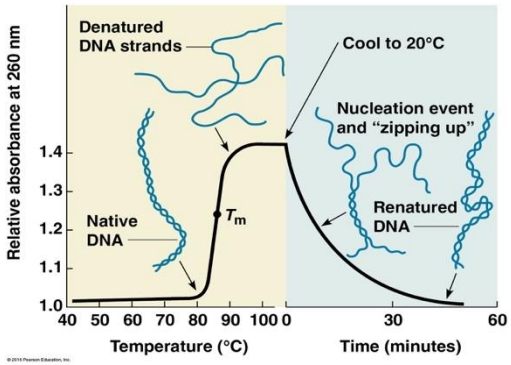
- plateau also says that the total amount of repetitive sequences (those that exist severall/manny times is low since these would find each other quicky and lead ot a significant increase in renatured material right away)





Molecular Cause for Genome Size Paradox

background: DNA strand separation and annealing can be followed by measuring the absorbance at 260nm (T_m = melting temperature)



Calf DNA: biphasic
interpretation:

- must have two populations of sequences ...some renature very quickly (→ effective [conc] = high → **repetitive sequences !**)
- and some that are unique; but
- unique renaturing **more slowly than in E. coli = there are more** unique sequences in calf genome → their effective [conc] is lower than E coli → slower renaturation.

what do you see?

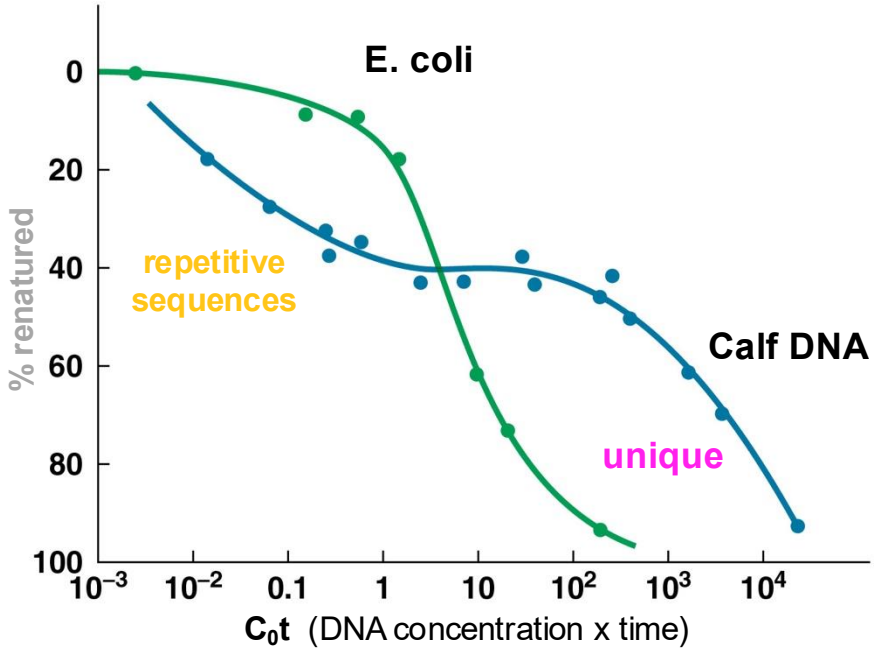
E coli curve:

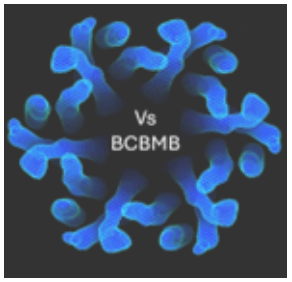
- x-axis is a log scale (!); weird unit; y-axis is inverted
- Shape: Initial plateau → steep drop off →"asymptotic" behavior at large "x"

interpretation: x-Axis C_0t = normalize for amount (C_0) of DNA used so you can compare results between expriments

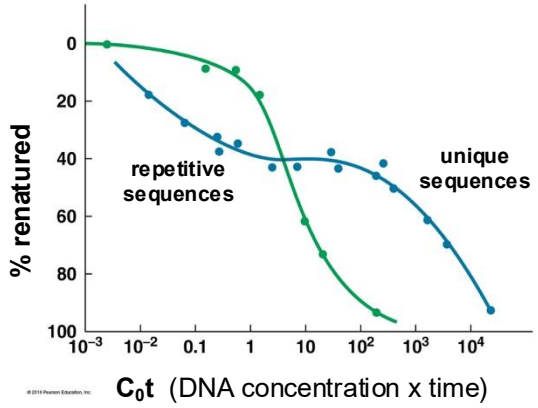
- initial plateau (lag phase) = unique sequences (like coding regions for proteins) need to find each other (this is concentration dependent → unique means effective [ ] is quite low) ...but once they do, annealing is very fast (= steep drop off) ...until it slows again as it approaches 100% renaturation;

- plateau also says that the total amount of repetitive sequences (those that exist severall/manny times is low since these would find each other quicky and lead ot a significant increase in renatured material right away)





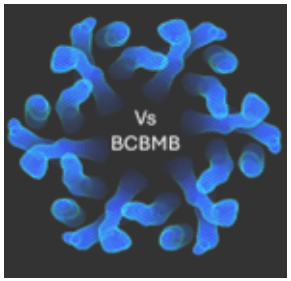
Repetitive Sequences Dominate Mammalian Genomes



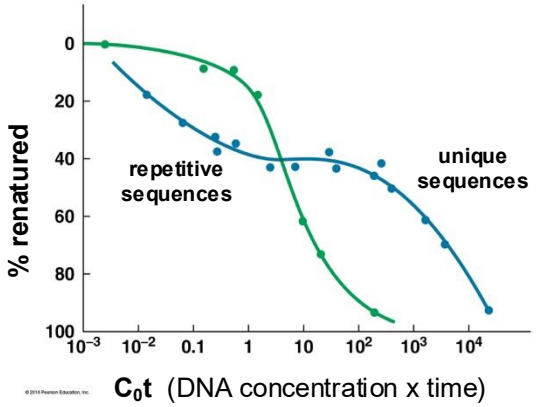
the discovery that mammalian DNA contains large amounts of repetitive sequences invites several questions:

- exactly what fraction is repetitive, and is that
- fraction the same for all mammals?
- is it all randomly distributed or not?
- does it have any function(s)?
- what are potential advantages?

....collect ideas/answers before moving on….



Repetitive Sequences Dominate Mammalian Genomes



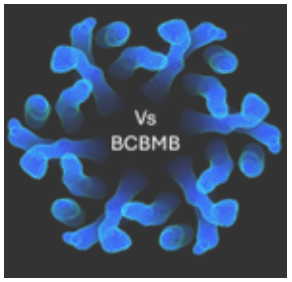
the discovery that mammalian DNA contains large amounts of repetitive sequences invites several questions:

- exactly what fraction is repetitive, and is that
- fraction the same for all mammals?
- is it all randomly distributed or not?
- does it have any function(s)?
- what are potential advantages?

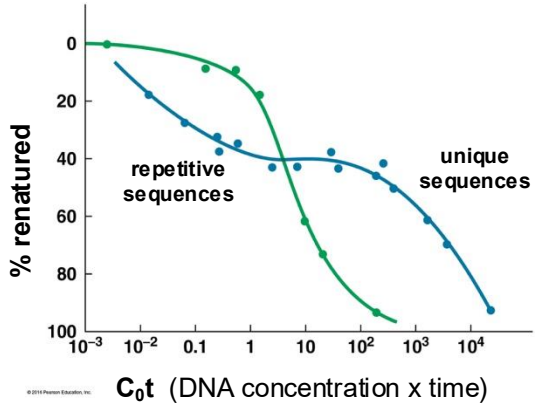
Fraction: compare **human 3.3 Gbp (~20,000 genes); mouse 2.7Gbp (~20,210 genes)**
→ not the same = repetitive DNA varies from species to species and between kingdoms (from close to 0% in bacteria to ~80% in some plants and amphibians**);** **~70% in humans**

→ **can you infer from fraction whether distribution is random or not**?

...*what do you think?....*



Repetitive Sequences Dominate Mammalian Genomes



the discovery that mammalian DNA contains large amounts of repetitive sequences invites several questions:

- exactly what fraction is repetitive, and is that
- fraction the same for all mammals?
- is it all randomly distributed or not?
- does it have any function(s)?
- what are potential advantages?

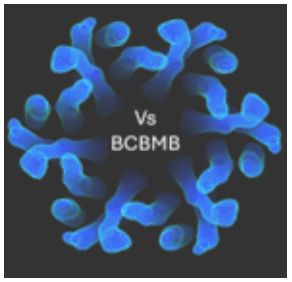
Fraction: compare **human 3.3 Gbp (~20,000 genes); mouse 2.7Gbp (~20,210 genes)**
→ not the same = repetitive DNA varies from species to species and between kingdoms (from close to 0% in bacteria to ~80% in some plants and amphibians**); ~70% in humans**

→ **can you infer from fraction whether distribution is random or not**?

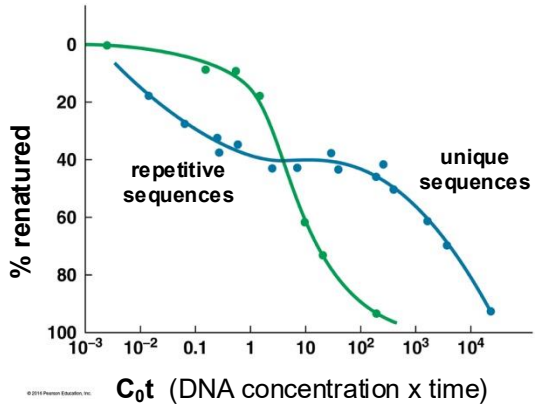
Answer: no – knowing the fraction does not tell you how these sequences are organized or where they are – **but because of the sheer amount of repetitive sequences you can say** that some fraction of repetitive DNA is not randomly dispersed.

→ **what about function?**

...your thoughts?...



Repetitive Sequences Dominate Mammalian Genomes



the discovery that mammalian DNA contains large amounts of repetitive sequences invites several questions:

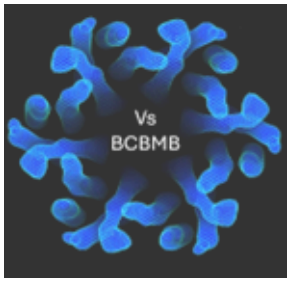
- exactly what fraction is repetitive, and is that
- fraction the same for all mammals?
- is it all randomly distributed or not?
- does it have any function(s)?
- what are potential advantages?

Fraction: compare **human 3.3 Gbp (~20,000 genes); mouse 2.7Gbp (~20,210 genes)**
→ not the same = repetitive DNA varies from species to species and between kingdoms (from close to 0% in bacteria to ~80% in some plants and amphibians**); ~70% in humans**

→ **can you infer from fraction whether distribution is random or not**?

Answer: no – knowing the fraction does not tell you how these sequences are organized or where they are – **but because of the sheer amount of repetitive sequences you can say** that some fraction of repetitive DNA is not randomly dispersed.
→ **what about function?**

Answer: if it had no function at all....it wouldn't be there → while we may not be able to explain the functions of all repetitive DNA (yet; though AI may find some (possibly true and relevant) clues in large data sets), there are some aspects of it that we can understand even now.
However, to reach that point requires looking at another set of questions first ….can you think of what they might be?



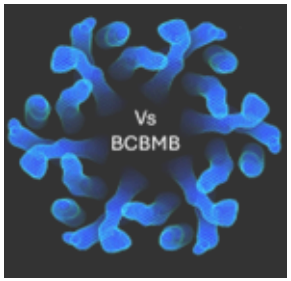
Deducing Functional Roles of Repetitive DNA

Contemplating 4.6×10^6 vs 3×10^9, questions that may help with understanding the functional roles of repetitive DNA are ....

- "how long are these molecules/sequences?",
- "how does the length of the genome relate to the size of the cell?",
- "is the genome a single piece, or is the information split into >1 piece?",
- "how is the genetic material accommodated inside the cell? Does it just sit there?"

why are these questions helpful?

...your turn! ....



Deducing Functional Roles of Repetitive DNA

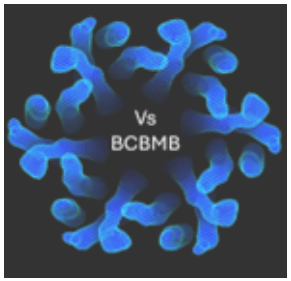
Contemplating 4.6×10^6 vs 3×10^9, questions that may help with understanding the functional roles of repetitive DNA are ....

- "how long are these molecules/sequences?",
- "how does the length of the genome relate to the size of the cell?",
- "is the genome a single piece, or is the information split into >1 piece?",
- "how is the genetic material accommodated inside the cell? Does it just sit there?"

why are these questions helpful?

Answer: questions about dimensions and number of pieces direct us towards **structure** …and ....**let's hear it ..... :) ..... (Structure Determines Function....**you may remember this paradigm from the Fundamentals Tutorials**)**

→ Start with: "**how long?"**



Deducing Functional Roles of Repetitive DNA

Contemplating $4.6x10^6$ vs $3x10^9$, questions that may help with understanding the functional roles of repetitive DNA are ....

- "how long are these molecules/sequences?",
- "how does the length of the genome relate to the size of the cell?",
- "is the genome a single piece, or is the information split into >1 piece?",
- "how is the genetic material accommodated inside the cell? Does it just sit there?"

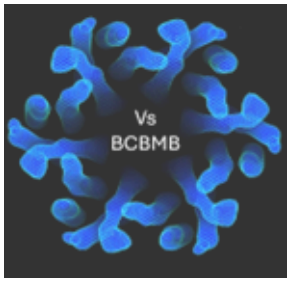
why are these questions helpful?

Answer: questions about dimensions and number of pieces direct us towards **structure** …and ....**let's hear it .....** **:)** **.....** **(Structure Determines Function....**you may remember this paradigm from the Fundamentals Tutorials**)**

→ Start with: "**how long?"**

- **Answer:**
- **E coli** $4.6x10^6$ bp ...assume all regular "B-DNA" (the most common form of DNA, covered in more detail in the "Advanced Biochemistry Collection) → ~1,600μm (1.6 mm = 0.063 in)
- **Human ~**$3x10^9$ bp (haploid germ cell) ➔ ~1 m
- that makes the **question of cell size quite interesting**.....

....*what do you think/know about it? .....*



Deducing Functional Roles of Repetitive DNA

- "how long are these molecules/sequences?",
- "how does the length of the genome relate to the size of the cell?",
- "is the genome a single piece, or is the information split into >1 piece?",
- "how is the genetic material accommodated inside the cell? Does it just sit there?"

why are these questions helpful?

Answer: questions about dimensions and number of pieces direct us towards **structure**
…and ….**let's hear it ….. :)** **…… (Structure Determines Function….**you may remember this paradigm from the Fundamentals Tutorials**)**

→ Start with: "**how long?"**

- **Answer:**
- **E coli** 4.6×10^6 bp ...assume all regular "B-DNA" (the most common form of DNA, covered in more detail in the "Advanced Biochemistry Collection) → ~1,600µm (1.6 mm = 0.063 in)

- **Human** ~3×10^9 bp (haploid germ cell) ➔ ~1 m

- that makes the **question of cell size quite interesting**…….

Answer:
- **E coli: ~**2x1µm**;**
- **Human: ~**10-100µm diameter (quite variable)

(more detail on that: Advanced Biochemistry Collection – CHEMICAL EVOLUTION, slides 70-71)

➔ one structural challenge becomes apparent immediately….how do you fit this into a cell??

Deducing Functional Roles of Repetitive DNA

being aware of how challenging it will be to fit the DNA into the cell ...

will this solution leave it all in one piece...or do you split it into several pieces (which should be easier to fit)?

...take a guess and justify (if you can)....

Deducing Functional Roles of Repetitive DNA

being aware of how challenging it will be to fit the DNA into the cell ...
will this solution leave it all in one piece...or do you split it into several pieces (which should be easier to fit)?

...take a guess and justify (if you can)....

Answer: easy to answer for human genome → there is NO WAY how this can be kept and propagated in one piece
….why?

Deducing Functional Roles of Repetitive DNA

Being aware of how challenging it will be to fit the DNA into the cell
... **will this solution leave it all in one piece...or do you split it into several pieces (which should be easier to fit)?**

Answer: easy to answer for human genome → there is NO WAY how this can be kept and propagated in one piece ….why?

Answer: intuitively – if kept as single piece, the mechanics of separating the two copies during cell division becomes impossible

- entanglement of the two copies that need to be separated
- big drag forces to overcome resistance of "gooey cytoplasm"
- insufficient mechanical stability of apparatus guiding the separation

→ if there were a machinery capable of separating this, it would have to be able to create very significant force + be able to disentangle “knots” "on the go"

also - elastic properties: like every polymer, DNA has an upper physical limit for how much it can be compacted. At this point in our discussion, it is unclear if DNA could be folded somehow to reduce its length 100,000-fold to fit 1m → ~10µm (size of smallest cells that can still divide)

Deducing Functional Roles of Repetitive DNA

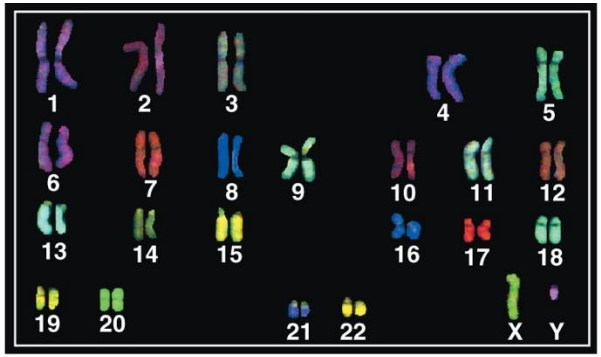
Being aware of how challenging it will be to fit the DNA into the cell
... **will this solution leave it all in one piece...or do you split it into several pieces (which should be easier to fit)?**

Answer: easy to answer for human genome → there is NO WAY how this can be kept and propagated in one piece ….why?

Answer: intuitively – if kept as single piece, the mechanics of separating the two copies during cell division becomes impossible

- entanglement of the two copies that need to be separated
- big drag forces to overcome resistance of "gooey cytoplasm"
- insufficient mechanical stability of apparatus guiding the separation

→ If there were a machinery capable of separating this, it would have to be able to create very significant force + be able to disentangle “knots” "on the go"

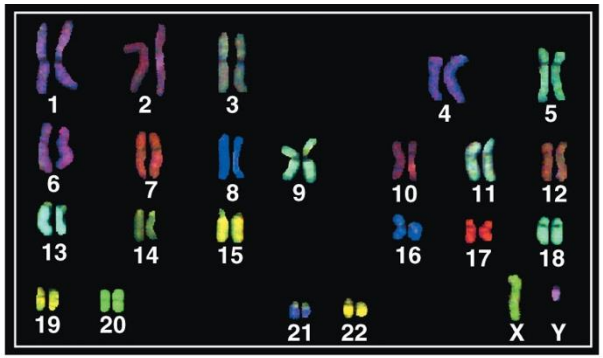


(a) Metaphase chromosomes

also - elastic properties: like every polymer, DNA has an upper physical limit for how much it can be compacted. At this point in our discussion, it is unclear if DNA could be folded somehow to reduce its length 100,000-fold to fit 1m → ~10µm (size of smallest cells that can still divide)

→ connects to what you know: 23 chromosomes (to be more exact: 2 sets of those are in each cell …making for a total of 6.6Gbp..., and a small amount of genetic material that is stored in mitochondria (= the cell’s power plant)

Deducing Functional Roles of Repetitive DNA



(a) Metaphase chromosomes

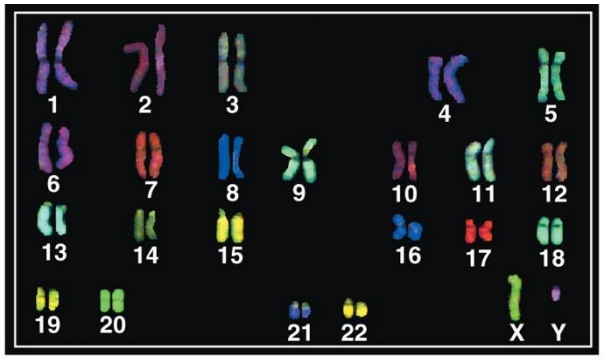
→ connects to what you know: 23 chromosomes (to be more exact: 2 sets of those are in each cell …making for a total of 6.6Gbp..., and a small amount of genetic material that is stored in mitochondria (= the cell's power plant)

What about bacteria?

Answer: less clear cut.
at 4.6Mb, can be kept as single piece, and **most bacteria** do exactly that = have **one large circular chromosome** (circular helps with partitioning between cells, and chromosome integrity).

besides, bacteria can have additional small circular DNAs (called plasmids) with various functions. These plasmids replicate independently and get randomly distributed between cells.

Deducing Functional Roles of Repetitive DNA



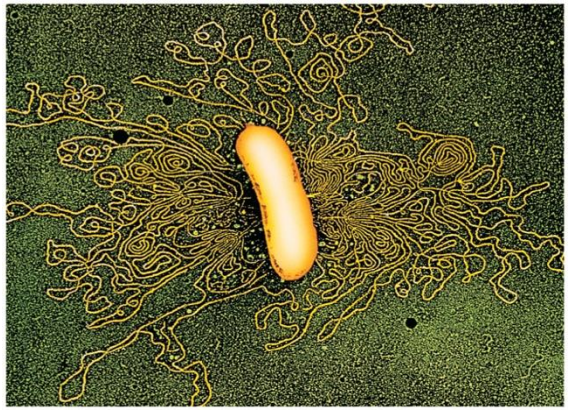
(a) Metaphase chromosomes

→ connects to what you know: 23 chromosomes (to be more exact: 2 sets of those are in each cell …making for a total of 6.6Gbp..., and a small amount of genetic material that is stored in mitochondria (= the cell's power plant)

What about bacteria?

Answer: less clear cut.
at 4.6Mb, can be kept as single piece, and **most bacteria** do exactly that = have **one large circular chromosome** (circular helps with partitioning between cells, and chromosome integrity).

besides, bacteria can have additional small circular DNAs (called plasmids) with various functions. These plasmids replicate independently and get randomly distributed between cells.



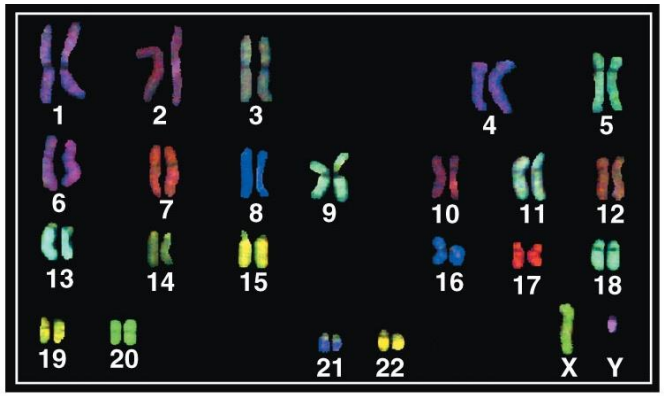
that said: gentle disruption of the bacterial cell wall allows the genome to "escape" ….
….just looking at this picture gives you an idea of the challenge cells face to accommodate their genome within the space that is available …

the picture also tells you that there must be some mechanism to "pack up" the genome so it will take up less space….and indeed, we will deal with that in the next chapter (CHROMATIN)

Deducing Functional Roles of Repetitive DNA

how does the "low resolution" (1 vs several pieces) **knowledge about genome structure help with understanding the potential usefulness of repetitive DNA sequences?**

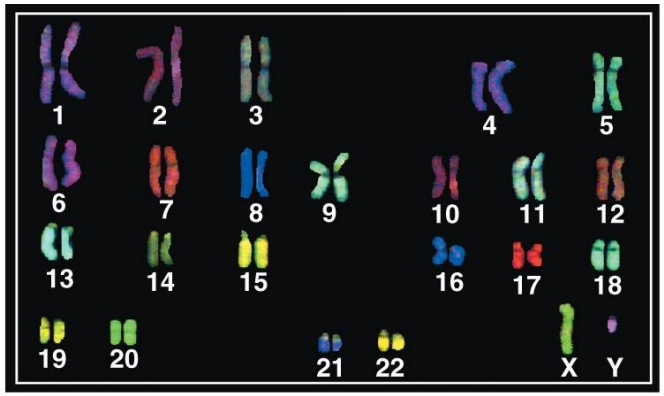
Answer: *…try….*



(a) Metaphase chromosomes

Deducing Functional Roles of Repetitive DNA

how does the "low resolution" (1 vs several pieces) **knowledge about genome structure help with understanding the potential usefulness of repetitive DNA sequences?**



(a) Metaphase chromosomes

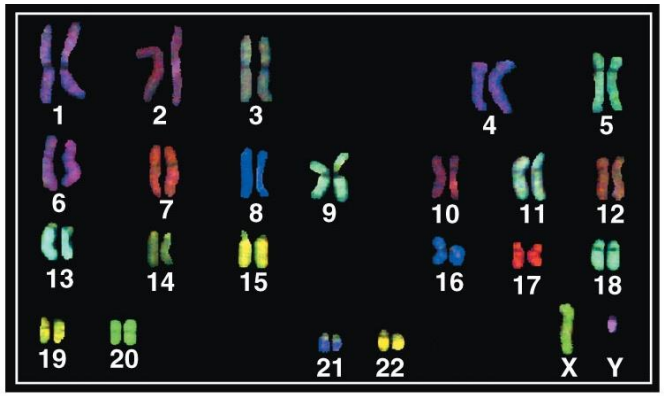
Answer: simple minded, but nevertheless true …going from 1 chromosome to 23 means....

- instead of having no end (circular genome in bacteria) you have 46 ends that require protection (Chapter: REPLICATION)
- instead of one site that helps to organize physical separation of genomes after duplication you need 23.... (Chapter: Cell Cycle 1 – MITOSIS)

so…by definition: 1→ >23 makes it repetitive, but beyond that "trivial" statement, the significance of "functional requirements" (chromosome separation and end protection) for "repetitiveness" becomes clear once you think about how to construct such sites

Deducing Functional Roles of Repetitive DNA

how does the "low resolution" (1 vs several pieces) **knowledge about genome structure help with understanding the potential usefulness of repetitive DNA sequences?**



(a) Metaphase chromosomes

Answer: simple minded, but nevertheless true …going from 1 chromosome to 23 means....

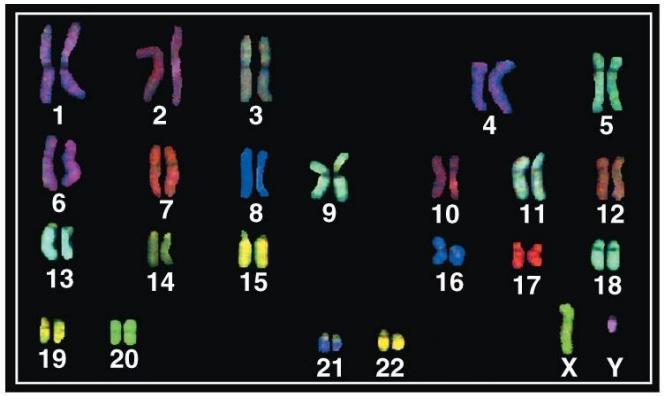
- instead of having no end (circular genome in bacteria) you have 46 ends that require protection (Chapter: REPLICATION)
- instead of one site that helps to organize physical separation of genomes after duplication you need 23.... (Chapter: Cell Cycle 1 – MITOSIS)

so…by definition: 1→ >23 makes it repetitive, but beyond that "trivial" statement, the significance of "functional requirements" (chromosome separation and end protection) for "repetitiveness" becomes clear once you think about how to construct such sites

take chromosome separation as example: would you build a "large or small" target for the attachment of molecular machinery that physically separates the chromosomes during cell division?

Deducing Functional Roles of Repetitive DNA

how does the "low resolution" (1 vs several pieces) **knowledge about genome structure help with understanding the potential usefulness of repetitive DNA sequences?**



(a) Metaphase chromosomes

Answer: simple minded, but nevertheless true …going from 1 chromosome to 23 means....

- instead of having no end (circular genome in bacteria) you have 46 ends that require protection (Chapter: REPLICATION)
- instead of one site that helps to organize physical separation of genomes after duplication you need 23.... (Chapter: Cell Cycle 1 – MITOSIS)

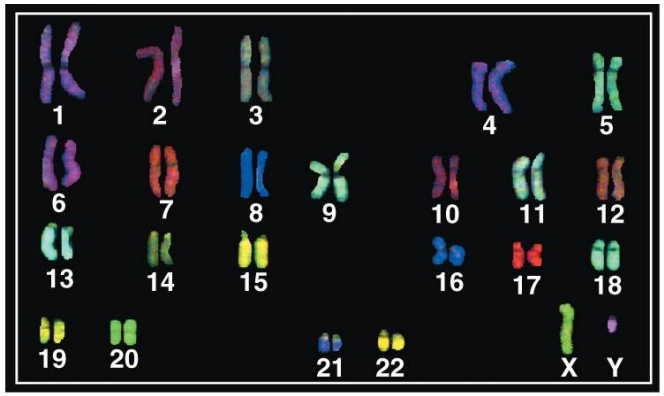
so…by definition: 1→ >23 makes it repetitive, but beyond that "trivial" statement, the significance of "functional requirements" (chromosome separation and end protection) for "repetitiveness" becomes clear once you think about how to construct such sites

take chromosome separation as example: would you build a "large or small" target for the attachment of molecular machinery that physically separates the chromosomes during cell division?

Answer: should be fairly large/significant size
→ would you make this "large target stretch of DNA" a unique sequence?

Deducing Functional Roles of Repetitive DNA

how does the "low resolution" (1 vs several pieces) **knowledge about genome structure help with understanding the potential usefulness of repetitive DNA sequences?**



(a) Metaphase chromosomes

Answer: simple minded, but nevertheless true …going from 1 chromosome to 23 means....

- instead of having no end (circular genome in bacteria) you have 46 ends that require protection (Chapter: REPLICATION)
- instead of one site that helps to organize physical separation of genomes after duplication you need 23.... (Chapter: Cell Cycle 1 – MITOSIS)

so…by definition: 1→ >23 makes it repetitive, but beyond that "trivial" statement, the significance of "functional requirements" (chromosome separation and end protection) for "repetitiveness" becomes clear once you think about how to construct such sites

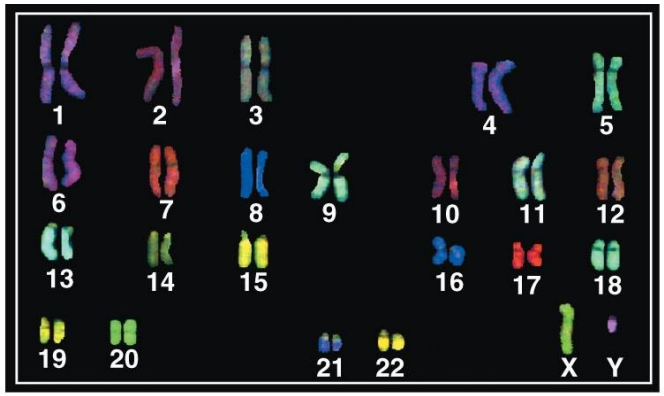
take chromosome separation as example: would you build a "large or small" target for the attachment of molecular machinery that physically separates the chromosomes during cell division?

Answer: should be fairly large/significant size
→ would you make this "large target stretch of DNA" a unique sequence?

Answer: no, that'd be pointless...because it would require a very complex counterpart to recognize it (= "how do molecules see?" (revisit that chapter to refresh your memory if you do not recall any of the details)
→ what you want is something ....**repetitive**...that "screams" ..."here, here, here, here, here ...." → these regions are called **centromeres**

Deducing Functional Roles of Repetitive DNA

how does the "low resolution" (1 vs several pieces) **knowledge about genome structure help with understanding the potential usefulness of repetitive DNA sequences?**



(a) Metaphase chromosomes

Answer: simple minded, but nevertheless true …going from 1 chromosome to 23 means....

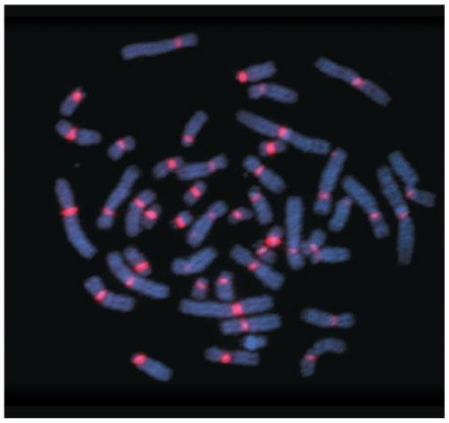
- instead of having no end (circular genome in bacteria) you have 46 ends that require protection (Chapter: REPLICATION)
- instead of one site that helps to organize physical separation of genomes after duplication you need 23.... (Chapter: Cell Cycle 1 – MITOSIS)

Answer: no, that'd be pointless...because it would require a very complex counterpart to recognize it (= "how do molecules see?" (revisit that chapter to refresh your memory if you do not recall any of the details)
→ what you want is something ....**repetitive**...that "screams" ..."here, here, here, here, here ...." → these regions are called **centromeres**

→similarly with end protection ...you don't need "sophistication" ...you just want a physical buffer at the end of chromosomes to lessen the impact of small deletions
(that happen during each genome duplication; Chapter: REPLICATION)
and chemical damage at the ends
→ these regions are called **telomeres**

Centromeres and Telomeres – Tandem Repeat DNA

centromeres and telomeres are build from short repeating units that are strung together in "**tandem repeats"**



(a) Centromeres 10 μm

- **centromeres:** mostly a 170bp repeat in humans making stretches of 0.2-7Mbp (depending on chromosome)

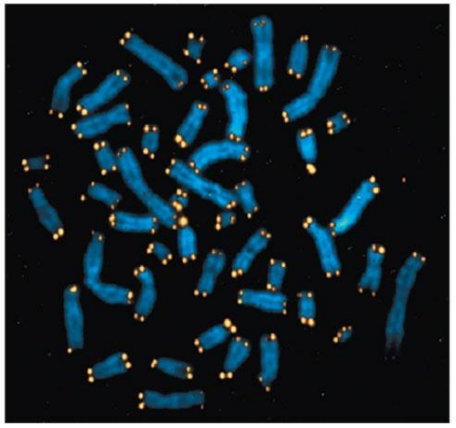
- **telomeres:** 250-1500 copies of "TTAGGG" at each end

adding it all up, admittedly, accounts for only ~100Mbp out of the ~500Mbp of **tandem repeats** in the human genome ... The rest is made from repeats that most often are ~10bp in length, and whose function poorly understood → much work left to do….

in addition to tandemly repeated DNA, the human genome also contains a very large amount – 54% - of "**interspersed repeated DNA".**

in humans, interspersed repeated` DNA comes in mostly two flavors, long (6-8kb) and short (~300bp) and is - as the name suggests - distributed throughout the genome.

….try to think about this for a minute … how is that even possible???
…having many pieces that are related to each other but with seemingly random distribution throughout the genome…..



(b) Telomeres

Centromeres and Telomeres – Tandem Repeat DNA

Centromeres and Telomeres are build from short repeating units that are strung together in "**tandem repeats**"

- **Centromeres:** mostly a 170bp repeat in humans making stretches of 0.2-7Mbp (depending on chromosome)

- **Telomeres:** 250-1500 copies of "TTAGGG" at each end

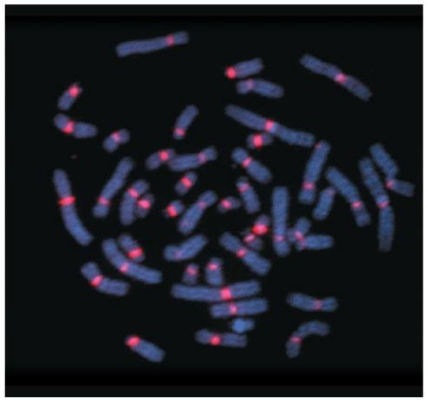
Adding it all up, admittedly, accounts for only ~100Mbp out of the ~500Mbp of **tandem repeats** in the human genome ... The rest is made from repeats that most often are ~10bp in length, and whose function poorly understood → much work left to do….

In addition to tandemly repeated DNA, the human genome also contains a very large amount – 54% - of "**interspersed repeated DNA".**

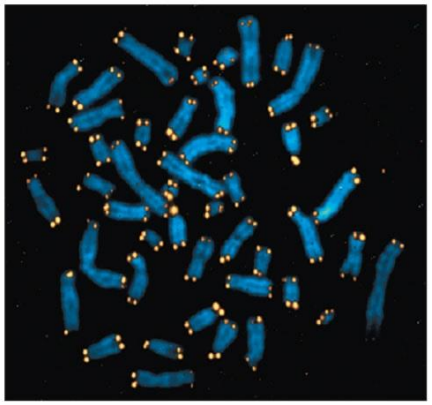
In humans, interspersed repeated` DNA comes in mostly two flavors, long (6-8kb) and short (~300bp) and is - as the name suggests - distributed throughout the genome.

- looking at these sequences revealed that they are mobile pieces of DNA (called transposons)!

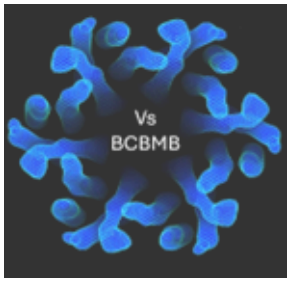
- more specifically: the long transposons code for proteins needed to excise and re-insert the piece; short transposons require assistance by the proteins encoded by the long ones. … yes …. sounds like a "parasite"….



(a) Centromeres 10 μm



(b) Telomeres



Exploiting "Parasitic" DNA as Evolutionary Driver

functional role of tandem repeat DNA in centromeres and telomeres can be appreciated but: what is the "good" in the transposable elements that constitute most of the interspersed repeated DNA?

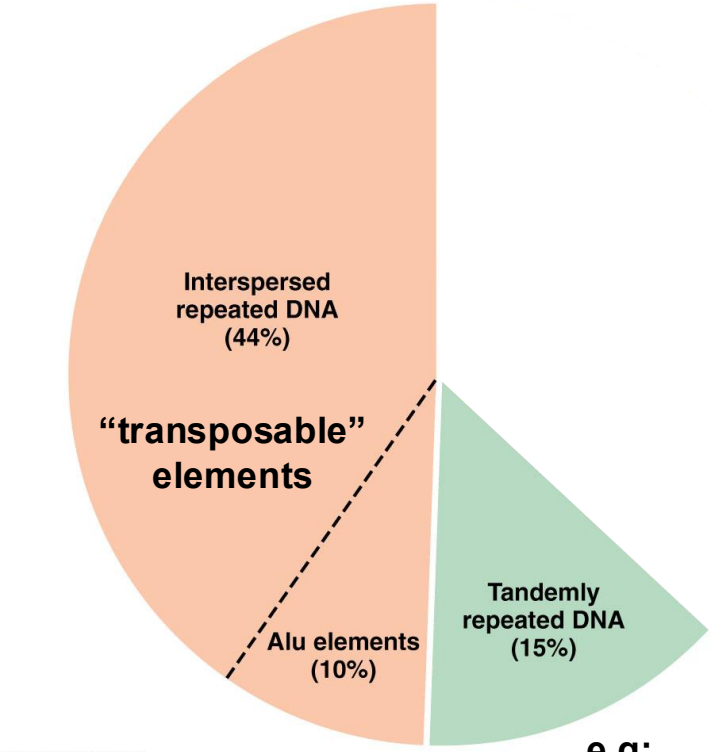
→ in fact …if these sequences account for >50% of our genome and do nothing but to copy and move themselves, transposons actually look like parasites or viruses....

→ Yes, transposons are selfish – but, by copying and moving themselves, inserting at random locations, these transposable elements provide an **evolutionary driver** – **why?**

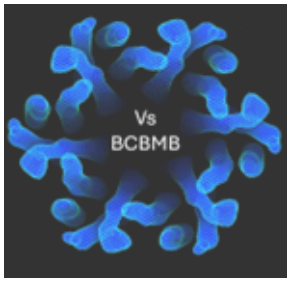
- long transposons carry start sites that regulate gene readout → if these move around, random changes in genome usage/dynamics are tested!
- adding new material to regions changes their spatial relationships => may lead to different structure/dynamics

→ if changes caused by insertion are advantageous or neutral pass it on

→ if insertion is harmful → organism dies.



e.g: centromeres, telomeres



Exploiting "Parasitic" DNA as Evolutionary Driver

Functional role of tandem repeat DNA in centromeres and telomeres can be appreciated but: what is the "good" in the transposable elements that constitute most of the interspersed repeated DNA?

→ in fact …if these sequences account for >50% of our genome and do nothing but to copy and move themselves, transposons actually look like parasites or viruses....

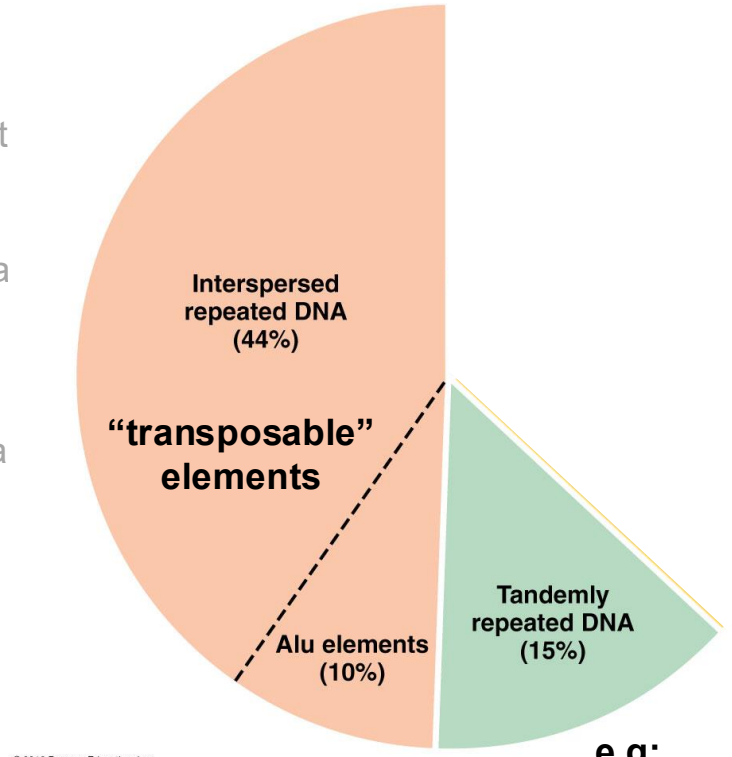
→ Yes, transposons are selfish – but, by copying and movin themselves and inserting at random locations, these transposable elements provide an evolutionary driver – why?

- long transposons carry start sites for gene readout
  → if these move around, random changes in genome usage/dynamics are tested!
- adding new material to regions changes their spatia relationships => may lead to different structure/dynamics

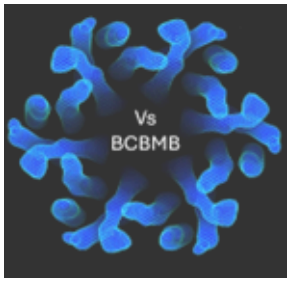
→ if changes caused by insertion is advantageous or neutra pass it on

→ if insertion is detrimental → organism dies.

→ the latter aspect causes an "asymptotic" behavior because each new copy has a higher likelihood to integrate in a location where it causes damage.



e.g: centromeres, telomeres



Exploiting "Parasitic" DNA as Evolutionary Driver

functional role of tandem repeat DNA in centromeres and telomeres can be appreciated but: what is the "good" in the transposable elements that constitute most of the interspersed repeated DNA?

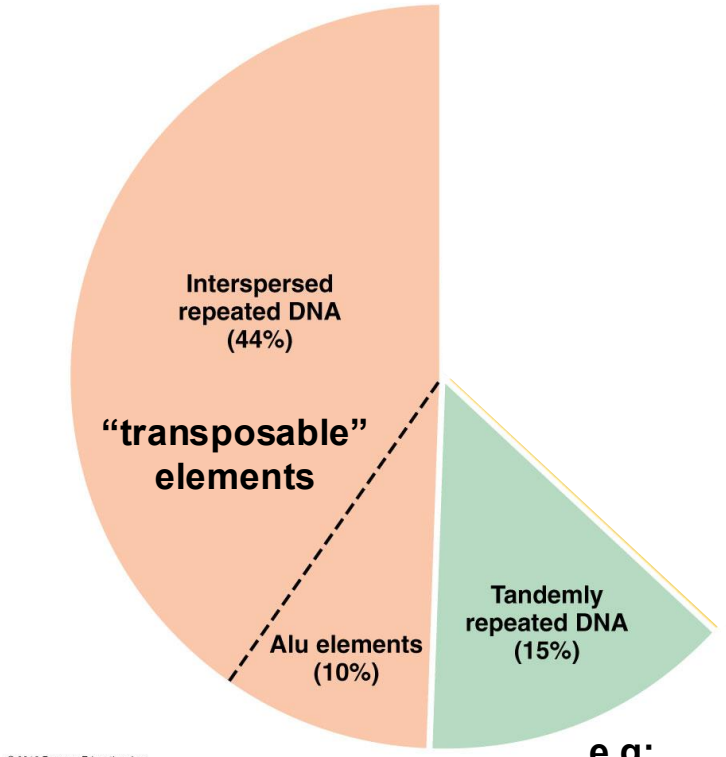
→ in fact …if these sequences account for >50% of our genome and do nothing but to copy and move themselves, transposons actually look like parasites or viruses....

→ the latter aspect causes an "asymptotic" behavior because each new copy has a higher likelihood to integrate in a location where it causes damage.

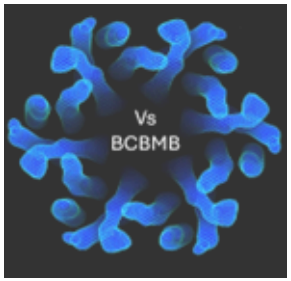
going beyond the asymptotic behavior for genome "saturation with transposons"

over time, most transposons lose the ability to move ….but ...there are still some active transposons in our genome, effecting change even within our own lifetime.

looking at the emerging picture ....what is the role of the remaining ~31% of DNA? How much of this is coding for molecular output?



e.g: centromeres, telomeres



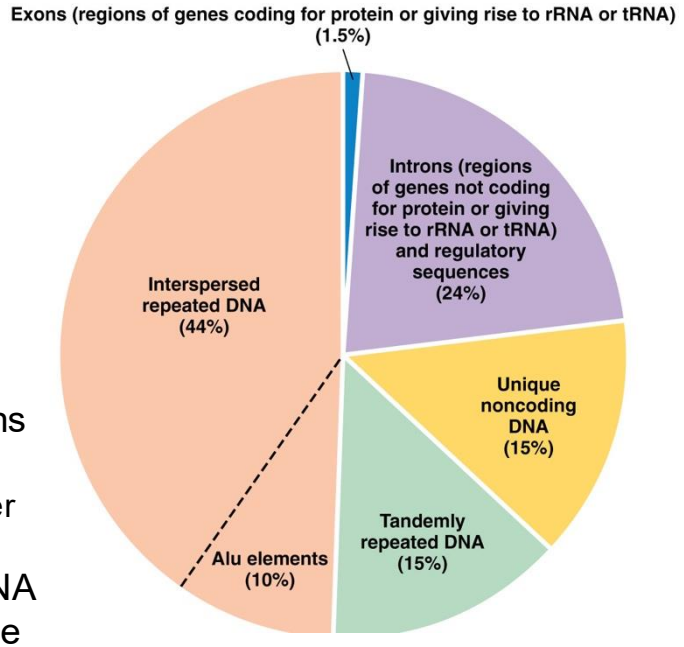
Only ~1.5% of the Human Genome Encodes Proteins/Functional RNAs

➔ besides repetitive DNA the genome harbors a lot of other "non-coding" DNA segments
what is their role?

Answer:

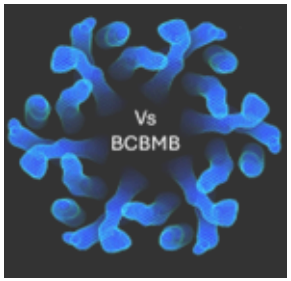
- some of the remaining non-coding sequences are "remnants" of once active genes ("**pseudogenes**"),
  = useful "close to functional" playground for developing new "functionalities"
- and then there are "introns" – regions that interrupt the DNA sequences that code for proteins.

→ of all the non-coding DNA segments….the full functional significance of introns is still elusive ... but here are a few thoughts on why they are useful:



- introns (and repetitive DNA) provide a safeguard against mutations in coding regions (radiation, chemical, replication, transcription);
  perspective: radiation induced DNA double strand breaks: 10-50 per cell per day! = some 100 trillion per day in your entire body
- introns provide a safeguard during recombination events when DNA is repaired, or parental DNAs are mixed; also allow for emergence of new genes.
- introns allow molecular diversification by "alternative splicing"
  (Chapter: TRANSCRIPTION-GENERAL PRINCIPLES)

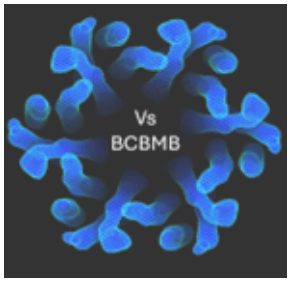
➔ **overall take-home message: (a) only 1.5% of sequences are correlated with a functional gene product, (b) non-coding, repetitive DNA appears "wasteful" but does seem to have functions as well, some of which we know about, and others that will emerge as time goes (and possibly: new functions will emerge over millions of years).**



Genome Complexity Necessitates Physical Sequestration

contemplating the functional heterogeneity and complexity of the human genome – does it have any consequence for how the cell accommodates the genome?

…what do you think?....

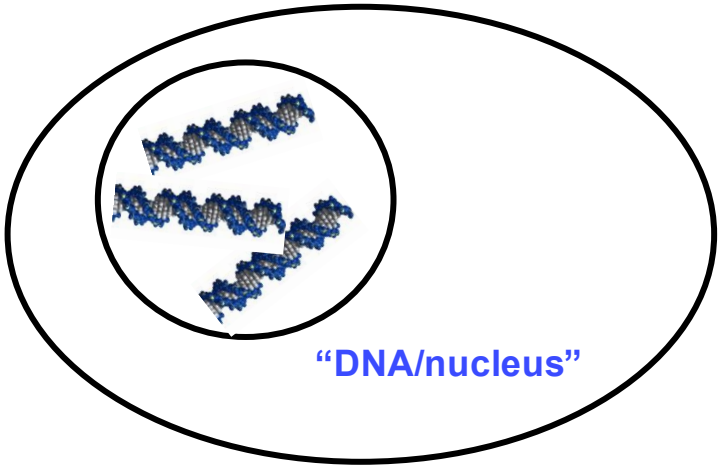


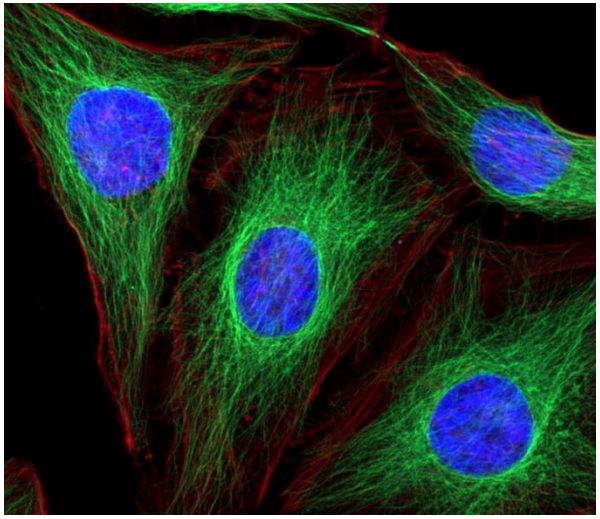
Genome Complexity Necessitates Physical Sequestration

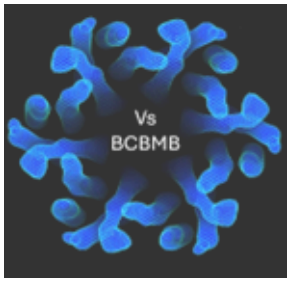
contemplating the functional heterogeneity and complexity of the human genome – does it have any consequence for how the cell accommodates the genome?

Answer: Yes – big time!

in fact …having 46 different pieces of DNA inside the cells (except germline cells that only have the haploid set of 23 chromosomes) **requires** a seperate (membrane bounded) compartment – called **"nucleus"** - to host the genome – **why?**





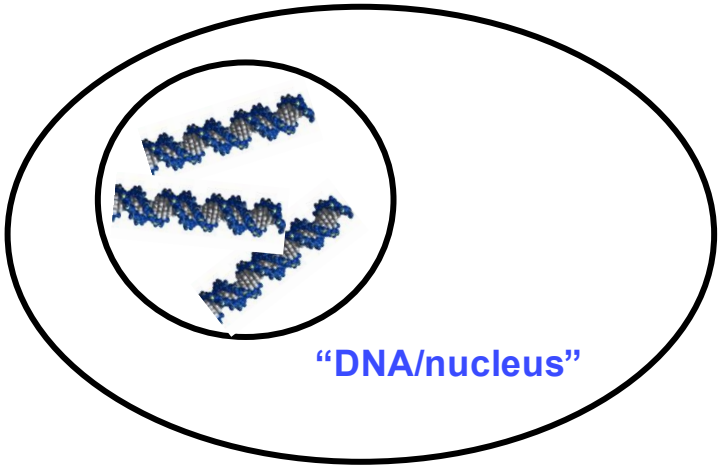


Genome Complexity Necessitates Physical Sequestration

contemplating the functional heterogeneity and complexity of the human genome – does it have any consequence for how the cell accommodates the genome?

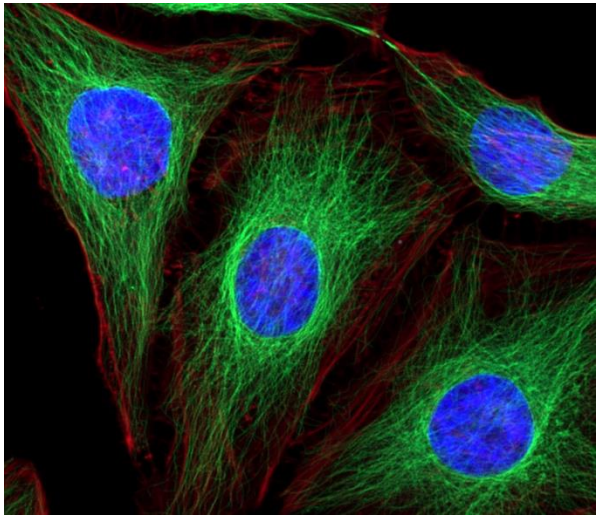
Answer: Yes – big time!

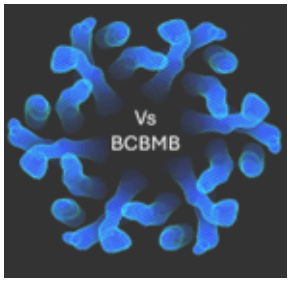
in fact …having 46 different pieces of DNA inside the cells (except germline cells that only have the haploid set of 23 chromosomes) **requires** a seperate (membrane bounded) compartment – called "nucleus" - to host the genome – **why?**



Answer:

- [conc] of reactants in DNA-dependent **chemical** processes would be too low if dispersed;
- **biological** coordination and regulation of chromatin dynamics would be impossible if dispersed because communication between these scattered elements would be difficult to manage *[note how this involves two disciplines...third one about to show itself .....]*



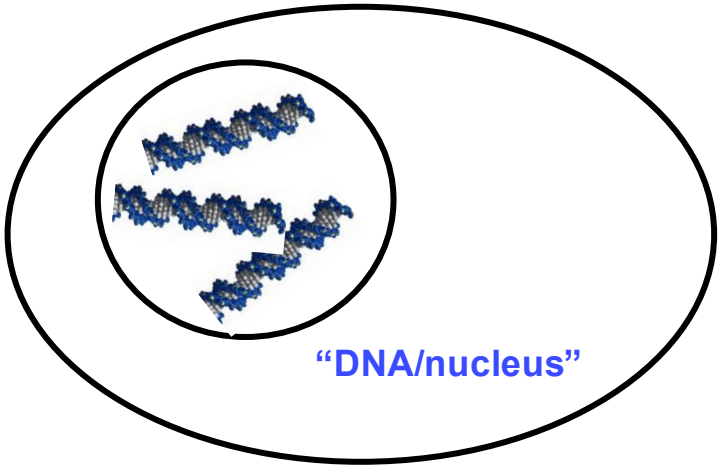


Genome Complexity Necessitates Physical Sequestration

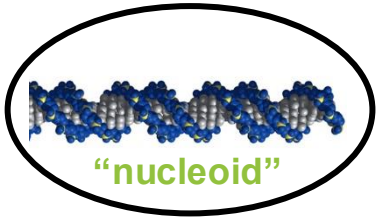
contemplating the functional heterogeneity and complexity of the human genome – does it have any consequence for how the cell accommodates the genome?

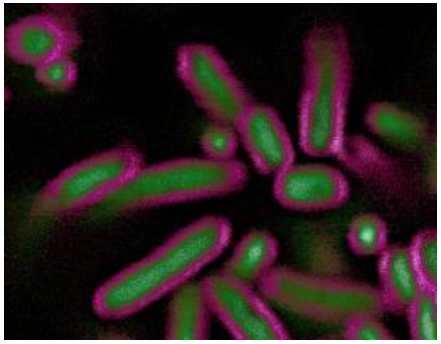
Answer:

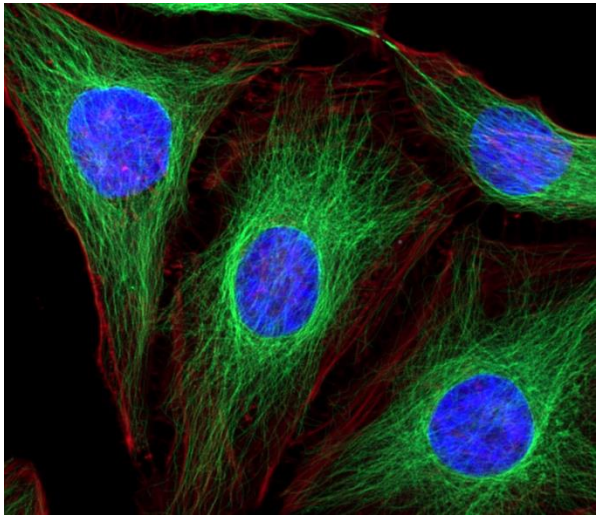
- [conc] of reactants in DNA-dependent **chemical** processes would be too low if dispersed;
- **biological** coordination and regulation of chromatin dynamics would be impossible if dispersed because communication between these scattered elements would be difficult to manage *[note how this involves two disciplines...third one about to show itself .....]*

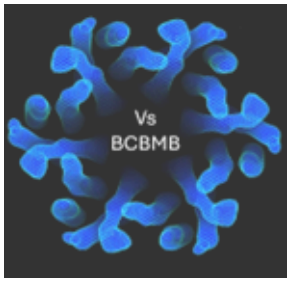


the nucleus represents a cytological marker that divides **eukaryotes** (eu = true, karyon = nucleus) from **prokayotes** that lack an enveloped nucleus.



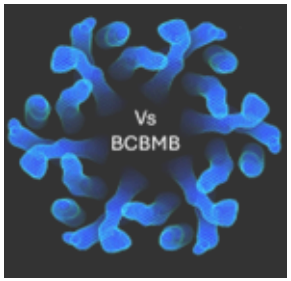






How to Fit 10,000 Miles of Spaghetti into a Basketball?

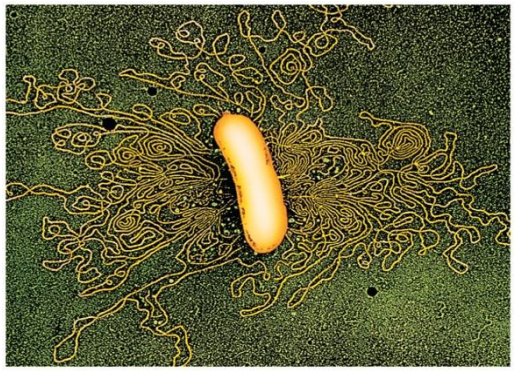
having obtained a glimpse of why complexity requires sequestration ....isn't the need for a nucleus a catastrophe??



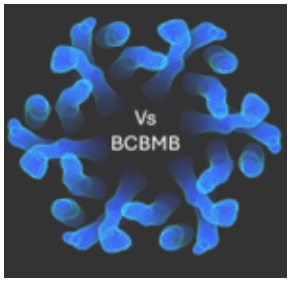
How to Fit 10,000 Miles of Spaghetti into a Basketball?

having obtained a glimpse of why complexity requires sequestration ....isn't the need for a nucleus a catastrophe??

Answer: that thought is completely logical if you remember the visceral image of a bacterium spilling its chromosome….. and even more so if you consider that



- human genome is almost 1000x larger ... yet **chemistry and biology require it to be compacted to fit into something that is significantly smaller than the entire cell** (at ~6μm average diameter, the nucleus occupies only ~10% of the cell).

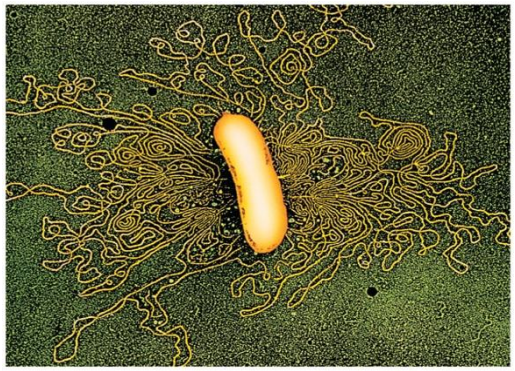


How to Fit 10,000 Miles of Spaghetti into a Basketball?

having obtained a glimpse of why complexity requires sequestration ....isn't the need for a nucleus a catastrophe??

Answer: that thought is completely logical if you remember the visceral image of a bacterium spilling its chromosome….. and even more so if you consider that

- human genome is almost 1000x larger ... yet **chemistry and biology require it to be compacted to fit into something that is significantly smaller than the entire cell** (at ~6µm average diameter, the nucleus occupies only ~10% of the cell).

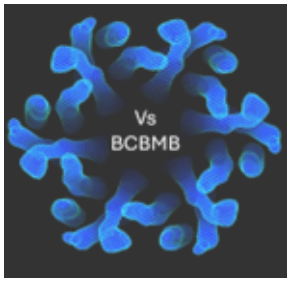


Translated into volumes:
the chromosome of **E coli** has ~1.6mm circumference with an **unconstrained** volume of 200µm^3 that are accommodated in a nucleoid of ~0.5µm^3 → 400-fold compaction or if you prefer this measure: ~1,000-fold shortening (~1.6mm → ~1.6µm)

the diploid **human genome** is ~2m long with an unconstrained volume of ~3x10^7 µm^3 that is fitted into a nucleus with volume of ~200 µm^3 → that is a compaction of 150,000-fold ….
HOW IS THAT POSSIBLE??

5m **unconstrained** (left) vs
10m **constrained** (right)

(easy to tell which one takes up more space)

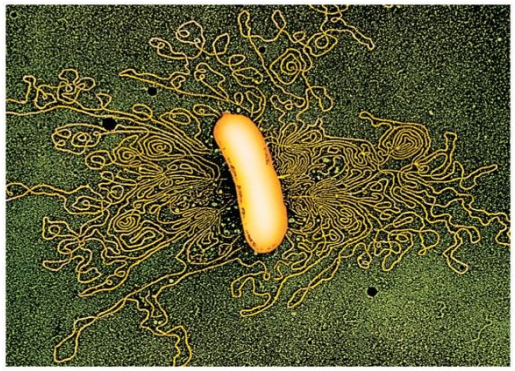


How to Fit 10,000 Miles of Spaghetti into a Basketball?

having obtained a glimpse of why complexity requires sequestration ....isn't the need for a nucleus a catastrophe??

Translated into volumes:
the chromosome of **E coli** has ~1.6mm circumference with an unconstrained volume of 200μm^3 that are accommodated in a nucleoid of ~0.5μm^3 → 400-fold compaction or if you prefer this measure: ~1,000-fold shortening (~1.6mm → ~1.6μm)

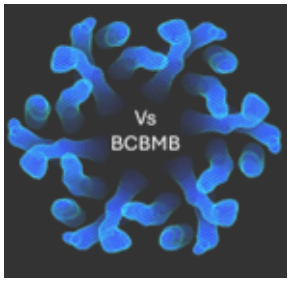
the diploid **human genome** is ~2m long with an unconstrained volume of ~3x10^7 μm^3 that is fitted into a nucleus with volume of ~200 μm^3 → that is a compaction of 150,000-fold ….**HOW IS THAT POSSIBLE??**



Answer: ..many layers to this answer – but an important aspect of it is the very fact that the human genome is distributed over 23 chromosomes → segmentation allows the enormous compaction to happen by "simply" **shortening the unconstrained length of each segment by a factor of 10,000-fold** (called packing ratio). That shortening seems more reasonable than the ~150,000 fold shortening that'd be required to fit 2m into a 6μm nucleus, but nevertheless is still quite impressive ➔ question remains ... **how can that be done?**

→ here comes the third discipline**: physics** → extreme compaction is possible by exploiting the mechanical idiosyncracies of the double helical DNA polymer.....

.....that, together with proteins, allow the formation of a special structure called "chromatin" (which we will explore in the next Chapter)



Summary Genomes

starting out, these were the explicit goals you should know and be able to explain

what a genome is
why genome size does not scale linearly with complexity of organism
why "non-coding" is not a synonym for "non-functional"
why large genomes require subcellular confinement and fragmentation

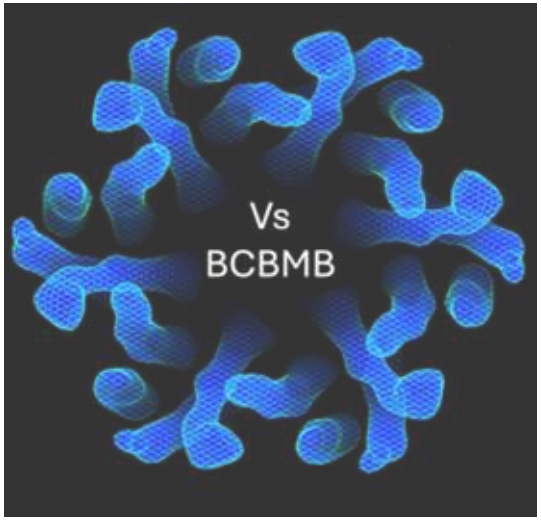
the short answers are:

- **What** a genome is: *minimum complement of DNA that defines an organism (haploid set of "chromosomes")*

- **Why** genome size does not scale linearly with complexity of organism: *lots of non-coding DNA, some of which consists of mobile elements (transposons) that keep making copies of themselves*

- **Why** "non-coding" is not a synonym for "non-functional": *while some of the functions of non-coding DNA remain to be discovered, many functions have ALREADY emerged – eg.* ***centromeres, telomeres, pseudogenes, introns****, protection against mutation, error reduction during recombination events, alternate splicing, regulatory sequences*

- **Why** large genomes require subcellular confinement: *[conc] of reactants for DNA dependent processes, physical separation, regulation of access/read out*



Thank You for Working Through This
if you found this Tutorial/Silent Lecture useful – recommend it to friends and consider downloading other the free tutorials (with clickable links) from this site

your comments and ideas for improving these tutorials are welcome

pdf-comments@vsbcbmbstudy.com