Learning Objectives

- Define “gene” and describe a gene’s location on a chromosome
- List common forms of chromosomal mutations
- Describe two ways of using similarity to identify orthologous genes
- Define global and local synteny
- Compare and contrast why genomic neighborhoods may be evolutionarily conserved in prokaryotes vs. eukaryotes
- Use local synteny to distinguish/identify orthologs in different species
- Explain synteny using the β-globin gene as an example
- Use synteny to determine evolutionary relationships among species

1

Synteny

Genomics Education Partnership

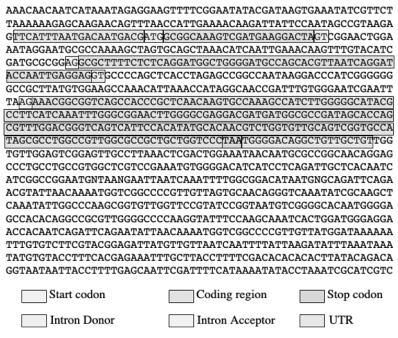
2

What is a gene?

- A gene is a heritable unit of information encoded in the DNA
- Many genes encode proteins
  - Protein coding genes include:
  - Open reading frames
  - Exons and introns
  - Alternative Splicing
  - Start and stop codons for translation
  - Promoter and cis-regulator modules
  - Transcription start sites
  - 5’ and 3’ untranslated regions
- Example: *Drosophila Rheb* gene (Located on chromosome 3R)

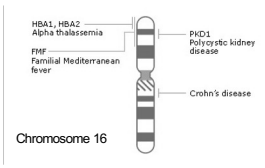
3

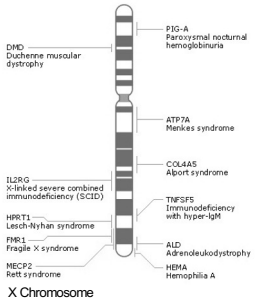
Genes are physical sequences



4

Genes Reside on Chromosomes



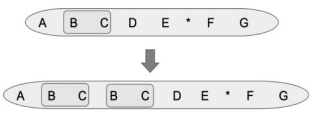


Genes have specific locations on a chromosome. Diagrams from the National Library of Medicine show the locations of genes implicated in human disease on chromosome 16 and the X chromosome.

5

Sometimes the order (or number) of genes on a chromosome can change:

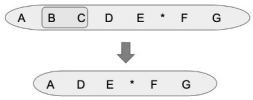
- **Duplications:** When a gene or region on the chromosome is copied



6

Sometimes the order (or number) of genes on a chromosome can change:

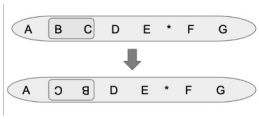
- Duplications: When a gene or region on the chromosome is copied
- **Deletions:** When a gene or region on the chromosome is removed



7

Sometimes the order (or number) of genes on a chromosome can change:

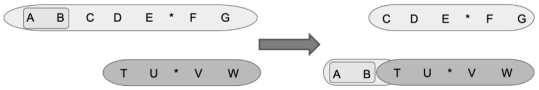
- Duplications: When a gene or region on the chromosome is copied
- Deletions: When a gene or region on the chromosome is removed
- **Inversions**: When a gene or region on the chromosome is reversed



8

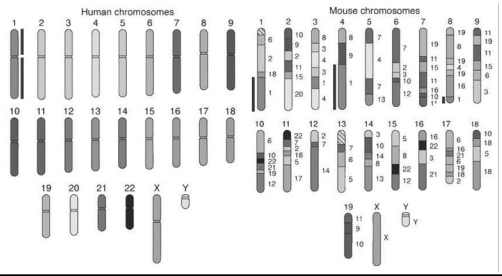
Sometimes the order (or number) of genes on a chromosome can change:

- Duplications: When a gene or region on the chromosome is copied
- Deletions: When a gene or region on the chromosome is removed
- Inversions: When a gene or region on the chromosome is reversed
- **Translocations:** When a gene or region on a chromosome is moved to another chromosome



9

Whole Regions of Chromosomes Can be Inverted, Deleted or Translocated



10

What can cause these changes?

- Double stranded breaks in the DNA that are not repaired properly.
  - Why do double stranded breaks occur?
    - Exposure to X-rays
    - Due to DNA replication in areas where single stranded breaks may be present
- Unequal crossing over
  - Due to repetitive sequences recombination machinery can cause the formation of a duplication on one chromosome and a deletion on the other.
- Transposable elements

11

What can cause these changes?

- Double stranded breaks in the DNA that are not repaired properly
  - Why to double stranded breaks occur?
    - Due to DNA replication in areas where single stranded breaks may be present
    - Exposure to X-rays
- Unequal Crossing Over
  - Due to repetitive sequences, recombination machinery can cause the formation of a duplication on one chromosomal strand and a deletion on the other
- Transposable Elements

12

Large rearrangements confuse what DNA segments correspond to each other between different organisms' genomes

e.g., beta-globin gene

→ **Chromosome 7**

→ **Chromosome 11**

13

As a researcher comparing genomes, you likely don't know the following about DNA segments (e.g. genes) you're studying:

- On which chromosome your DNA segment is in your studied species
- What mutations occurred in this DNA segment over time
- What mutations occurred around this DNA segment over time

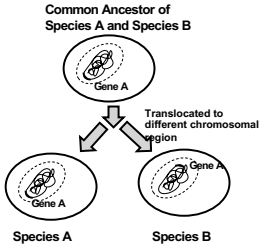
→ **You start with just several long sequences of As, Gs, Cs, and Ts**

14

The location of a DNA segment is in the genome is tied to its evolutionary history

Equivalent DNA segments in different organisms are known as **orthologs**

- DNA segments are related because of descent from a common ancestral sequence due to speciation
- To compare a particular DNA segment across species, need to always use the ortholog in each species
- Mistaken use of a non-orthologous region will give incorrect results!!!



15

How do we identify orthologous genes?
Main type of evidence is **genetic similarity**

***Within* the gene of interest**

Base pair sequence in putative gene very similar

Point mutations over time increase differences

Point mutations can greatly decrease sequence similarity given enough time

* ***Around* the gene of interest**

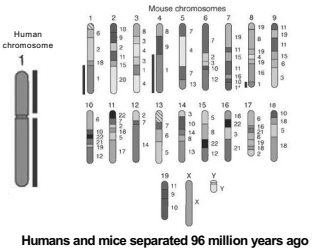
Neighboring genes are very similar

Large mutations (e.g. inversions, translocations) over time increase differences

BUT, large mutations rarely break up entire neighborhood

16

How do we identify orthologous genes?
Main type of evidence is **genetic similarity**



Humans and mice separated 96 million years ago

* ***Around* the gene of interest**

BUT, large mutations rarely break up entire neighborhood

Example: Large pieces of chr1 in humans remain together in mice.
→ Expect neighboring genes to mostly stay the same

Really good evidence for orthology

↳ This is called **synteny**

17

What is Synteny?

- Classic Definition ("Global" synteny): two "genetic entities" (genes, for example) on the same chromosome
- "Local" Synteny: conservation of the order and orientation of genes in the vicinity of a target gene - This is what the GEP uses!

18

What is synteny?

- As organisms evolve, changes occur in their genomes (deletions, insertions, etc.)
- Over time, these changes can disrupt synteny
  - Two genetic entities may no longer be on the same chromosome
  - The order and orientation of genes relative to a target gene is no longer conserved
- Generally, we hypothesize that no changes have taken place, because this is the simplest explanation
  - Evolutionarily, this concept is known as "parsimony"
  - More commonly, this is known as Occam's Razor: "the simplest explanation should be preferred over more complex theories"

19

Genomic Neighborhoods

Focusing down on a block of genes can determine if there is local synteny within a particular gene's neighborhood. Like human neighborhoods, we have seen that genomic neighborhoods change over time, as genes are "torn down", as others are "built" in their place, or as mobile elements change neighborhood locations.

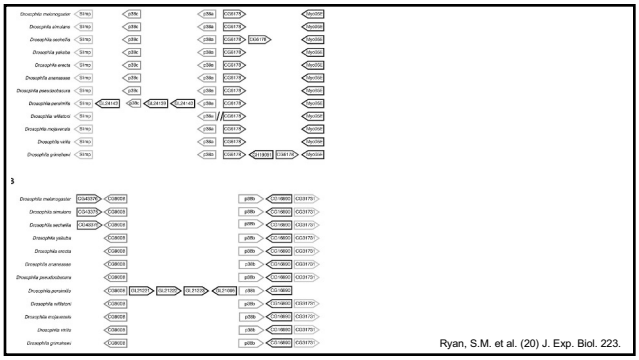
Prokaryotes vs. Eukaryotes

It is not unusual for more distantly related prokaryotes to display local synteny, but this often occurs for reasons that do not apply to most eukaryotes:

1) regulation of multiple genes in an operon using a single promoter
2) selective pressure for blocks of genes which work together to transferred together via horizontal transfer
3) two genes that produce two proteins which interact tend to cluster due to transcription and translation being coupled in prokaryotes

Because of this, before multiple eukaryotic genomes were sequenced, it was assumed that their individually regulated genes might be found in random locations. The next slide, however, shows two loci from a number of Drosophila species which have very similar genomic neighborhoods. Note that, while examples of various deletions and insertions of both genes and intergenic sequences are evident, the neighborhoods in question remain recognizable.

20

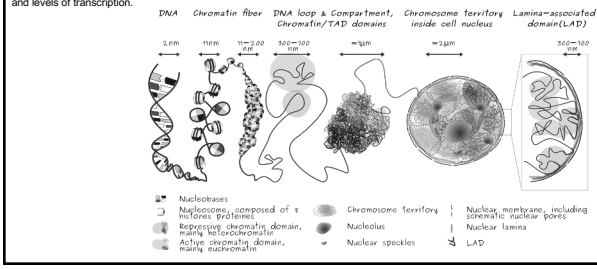


Ryan, S.M. et al. (20) J. Exp. Biol. 223.

21

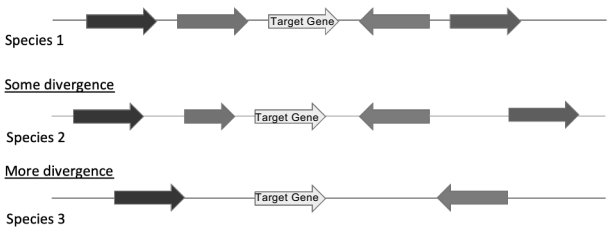
Conservation of Genomic Neighborhoods in Eukaryotes

Although what drives local synteny in Eukaryotes is still not well understood, it has been noted that, despite the absence of traditional operons, genes with a close proximity in a given eukaryotic genome often share a similar level of expression. What is more, genes that contribute to the same metabolic pathway are often in close proximity, despite the rarity of horizontal gene transfer among eukaryotes. The presence of chromatin domains within eukaryotes may help explain the correlation between gene proximity and levels of transcription.



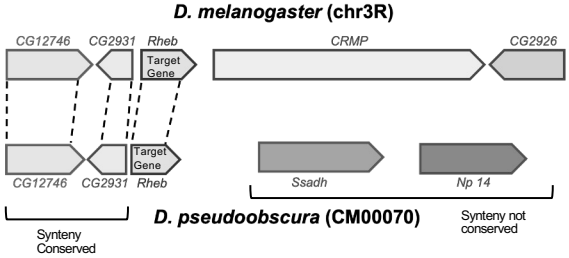
22

What does synteny look like in the GEP?



23

What does synteny look like in the GEP?

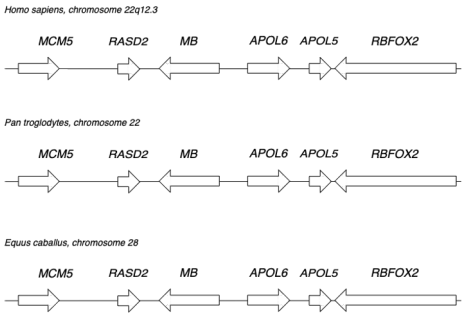


24

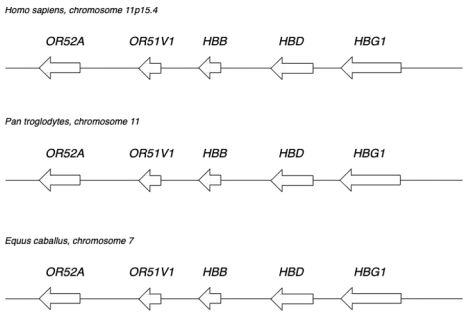
Why is understanding synteny important?

- Confirmation of gene identity
- Tracking of large-scale genome changes: evidence for evolutionary relationships
- Evidence for co-regulation of nearby genes

25



26



27

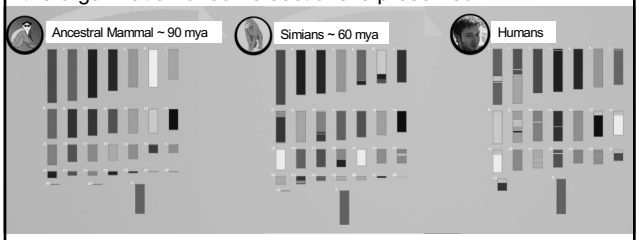
It helps in identifying gene orthologs:

- Imagine you've identified a globin gene from centaurs.
- Its sequence shows roughly equal similarity to myoglobin and β-hemoglobin.
- Is your gene the centaur version of the gene for myoglobin or for β-hemoglobin?

Looking at the surrounding genes is an important piece of evidence to help you decide.

28

When we compare chromosomes from different species, the organization of some sections is preserved.



Chromosomal organization changes over evolutionary time spans. As more time passes, changes accumulate.

Figure created on: https://woovthan.github.io/synteny-explorer/

29

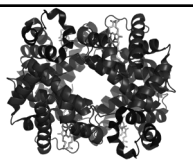
What can we do with synteny?

1. Synteny tells us a lot about evolutionary relationships. After a change in organization occurs, all descendants of that organism will share that change for millions of years.
   - We can group species with shared features and know they are related by a common ancestor.
2. When a group of genes is held together in specific configuration long after other similar groups have been reorganized, it suggests that those genes are co-regulated for a reason.
   - That tells us we should look for some important aspect of function and/or expression.

31

Hemoglobin genes

- *Hemoglobin* genes are good examples of how synteny and loss of synteny can reconstruct evolutionary relationships and inform about gene regulation.
- *β-globin* combines with α-globin in red blood cells to form hemoglobin.
- Hemoglobin binds and transport oxygen from the lungs to the body.
- α and *β-globin* genes are found in clusters.
- Both α and *β-globin* genes have been duplicated several times since the dawn of mammals.
- Species that are closely related share the same gene organization within the cluster. They are syntenic.



Structure of human hemoglobin. **α** and **β** subunits in red and blue, respectively.
Source: https://en.wikipedia.org/wiki/Hemoglobin

Human globin gene clusters

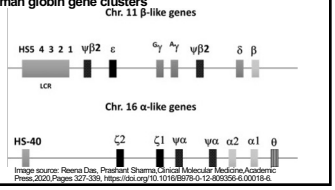
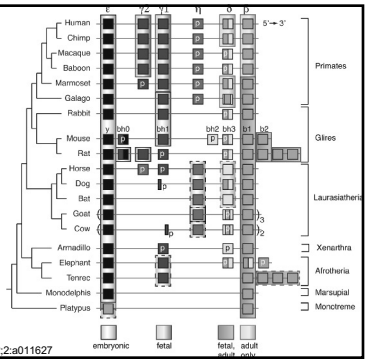


Image source: Reena Das, Prashant Sharma,Clinical Molecular Medicine,Academic Press,2020,Pages 327-339, https://doi.org/10.1016/B978-0-12-809356-6.00018-6.

32

β-globin gene duplications

- In mammals, embryogenesis begins inside the mother, so the embryo needs globins with higher O_2 affinity to draw O_2 from the mother's blood.
  - All mammals have at least two versions of beta-globin – one for the embryo, one for the adult (and sometimes the fetus)
  - As gestation evolved, the globin genes *duplicated* in some mammals and were adopted for use in the fetus



***Mammalian β-globin* gene clusters**

Image source: Cold Spring Harb Perspect Med 2012;2:a011627

33

Beta-globin genes

- Comparison of *β-globin* clusters between species reveals synteny between close relatives (e.g., Humans and Chimps)
- Comparison of less related species reveals non-synteny and evolutionary change.
- Sometimes a duplicated gene is not needed, so it can mutate and become a non-functional "pseudo-gene" (indicated by ρ in the figure).

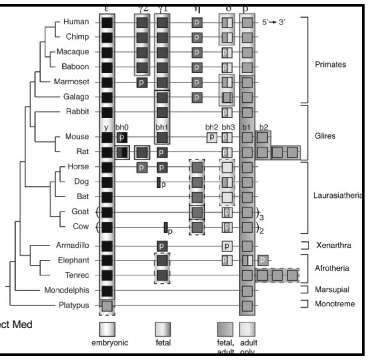


Image source: Cold Spring Harb Perspect Med 2012;2:a011627

34

Synteny, at the gene level, suggests co-regulation

- Hemoglobin genes are clustered in most or all vertebrate species.
- Why have those genes stayed together over ~500 million years in ~70,000 species?
- All genes in a cluster are regulated by a core element that ensures that the genes are expressed only in red blood cells.
- If the cluster is broken, some of the genes will not be properly expressed, and that would likely be fatal.

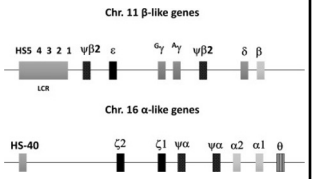


Image source: Reena Das, Prashant Sharma,Clinical Molecular Medicine,Academic Press,2020,Pages 327-339, https://doi.org/10.1016/B978-0-12-809356-6.00018-6.

35

Conclusion

- Synteny is when two species have the same genes in the same order and orientation.
- When two species initially diverge, their entire genomes are mostly syntenic.
- Over millions of years, the chromosomes accumulate structural changes and the blocks of synteny get smaller.
- Sometimes conservation of synteny can indicate co-regulation of related genes.

36