Chapter 21 Genomes and Their Evolution

Genome sequencing has allowed close examination of the differences between species and larger groups. Also, we can use this to examine evolutionary history of genes and gene function.

* Many (thousands [>5000?]) species have been sequenced
* significant species to us
	+ human relatives, domesticated plants and animals, disease causing organisms of humans and crops and livestock, research organisms
* **genomics** is the study of sets of genes and their interactions
* bioinformatics is the field that applies computational methods to this data

Genome sequencing now

Human Genome Project started in 1990 to sequence the human genome

* 20 labs, 6 countries, mostly completed in 2003, last publication in 2006
	+ Not fleshed out in great detail but ‘virtually complete’

Progressively more detail from

1. Linkage maps
	* Karyotype banding patterns started the process of mapping. It showed that differences existed between areas of chromosomes
	* Linkage maps showed sequences of genes on a chromosome based on crossing over frequency, but not physical distances between genes
	* STRs, RFLPs, SNPs marked around 5000 genes
2. Physical mapping
	* Order of large segments determined by base pair sequences and overlapping restriction sequences
	* Labor intensive
		+ lots of restriction enzymes, cloned sequences, YACs and BACs,
3. Actual DNA sequences of chromosomes
	* In humans – 3 billion (3,000,000,000) base pairs daunting
		+ Developing computers powerful enough to process all the data was part of the necessary push to make human genome sequence known
			- 1980’s – 1000 base pairs/day sequenced
			- By 2000, 1000 bps/*second* , 24/7
				* Called high throughput computers
* Fig. 21.2 – simplified

Shotgun approach to whole genome sequencing

* J. Craig Ventner started it – his DNA BTW
	+ Celera Genomics is his company
	+ Started process in 1998, mostly complete by 2003
* Sequences DNA fragments from randomly cut DNA
	+ Computers assemble the sequence from overlapping sequences
	+ Fig. 21.3
	+ Shotgun approach used mostly today? (9th ed may need updating)
	+ Sequencing by synthesis (Chapter 20 method increases speed and decreases cost [Fig 20.12])
		- Still generates short sequences of known content that need to be assembled by a computer
* Sequencing the human genome took 13 years from the start and cost $100 million
	+ Someone sequenced James Watson’s genome in 2007 and it took 4 months and cost $1 million
	+ In 2010, a group sequenced 3 human genomes quickly for about $4400/ea
	+ Obviously in the decade since then, things have gotten faster and cheaper
		- 23 and me, etc
			* Btw, they don’t sequence the whole genome, just areas of the genome known to be related to genealogy and possibly genetic diseases
				+ Primers in PCR clone only the desired sequences
* Metagenomics is where a group of species’ DNA is sequenced
	+ Again need software to assemble the sequenced fragments
	+ So far used in microbial communities in a variety of settings
	+ Allows researchers to see specifically what’s there without trying to culture each individual species.

Bioinformatics

Reality is that there is a lot of data. Coordinating the storage of it necessary. Internet allows accessibility.

* Central resources by govt funded agencies doing it
	+ US – National Library of Medicine + NIH created National Center for Biotechnology Information (NCBI) – [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)
		- Database called GenBank
	+ Europe, Japan, China collaborate with us
	+ Smaller website also have databases
	+ GenBank – 119 million fragments = 114 billion base pairs
		- Updated frequently
		- Can compare sequences you’ve discovered with these known sequences, compare predicted protein sequences, search for protein sequences based on known amino acid sequences, show 3D protein shape,
	+ Is a Protein Data Bank also by another group ([www.wwpdb.org](http://www.wwpdb.org))
* Researchers can access huge amounts of data from anywhere in the world with internet

Accessing sequences presents a new difficulty. Predicting a protein function based on DNA sequence (determining phenotype from genotype)

* Gene annotation is the process of identifying all protein coding genes and the function of the protein
	+ Remember that many genes do not produce proteins but functioning RNA molecules
	+ Automated now
		- Scan a sequence for known start and stop codons, RNA splicing sites (in introns and exons) and they say other signs of protein coding which I can’t add to. They also know sequences of certain general (I think) mRNA molecules to look for
* Looking for undescribed genes in the genome involves maybe looking for sequences that are known to be genes in other organisms
	+ Again software does the search
* Sometimes a new sequence will partially match a sequence who’s function is known, indicating that sequence will do something like the other
	+ Think genes that participate in basic cellular function there – protein kinases, respiratory enzymes, ribosomal function, etc

Gene Expression at the Systems Level

One project (ENCODE) studied about 1% of the human genome, found that 90% of coding sequences produced RNA, about 2% produced protein.

Proteomics is the systematic study of full protein sets produced by genomes

* Protein do most of the things an organism does, so paying attention to them will yield understanding
* An example follows:
	+ Systems, rather than independent parts, describe the life of an organism
	+ Integrating the parts on a genetic and protein level the desire
	+ Proteins, thus genes, interact to make the organism/species
		- Protein networks important
			* Researchers have knocked out (disabled) certain gene pairs and watched what happened to the organisms fitness when the protein isn’t produced or functional
				+ If no change, the genes aren’t in the same network, aren’t related, don’t interact. If a change, they are related.
				+ Fig 21.5 for examples of the types of maps that are being examined

Again, software

Applications to Medicine

* Cancer Genome Atlas
	+ Looks at gene interactions
	+ One project looked for common mutations between lung, ovarian and a type of brain cancer.
* Human gene chips – analyze gene expression (DNA microarrays as in Fig 20.1 and 20.15)
	+ Had ‘modest’ success in certain cancer treatments so far
	+ Perhaps you all will be carrying this type of data round with you at all times in the future – a genetic bar code to help doctors personalize medicine
	+ At least now, we can address certain cancers specifically due to personalized medicine

Systems biology looking for the emergent properties due to gene/protein interactions

Genome size, gene # and gene density

Genome Size – Table 21.1

* Most archaeal and bacterial genomes = 1 – 6 million base pairs (Mb)
	+ We’re not sequencing many archaeal genomes
* Eukaryotic genomes larger = 12 – 3000 Mb (with a plant at 124,000 MB and an amoeba at 670,000 Mb)
* Weirdly, fairly closely related groups can have huge differences in base pair counts
	+ One insect species has 11x more bps than another
		- Wider ranges of genome sizes in protists, insects and amphibians than in mammals and reptiles
			* Birds are sadly forgotten

Number of Genes – same table

* bact and archaea have fewer genes than eukaryotes
* Prok= 1500-1700 genes Euk = 5000 – 40,000 genes
* Can have a big genome but not a bunch of genes
	+ *C. elegans* vs *D. melanogaster* in Table 21.1
* We expected humans to have way more genes (50 – 100K) than discovered (<21K) based on the # of different proteins we have
	+ Alternate RNA slicing of genetic sequences explains it
		- One ‘gene’ can have different introns removed and exons expressed to generate many different proteins.
		- The example we used earlier was something like 10,000 different proteins possible from one transcription unit in a *Drosophila* species
		- An average human gene has around 10 exons and over 93% of genes are alternately spliced to at least 2 different forms
			* Some human genes have hundreds of different ‘final copies’
	+ So a gene is a transcription unit to me. It is the sequence that produces the initial RNA before RNA processing, not the spliced final copy
	+ Also, protein processing after translation can be varied in different environments (cell types, developmental stages)
		- We have not discussed this but it is part of the explanation for lots of proteins and not many ‘genes’

Gene Density and Noncoding DNA – same table

* Gene density is how many genes per length of DNA
* Gene density is higher in proks than euks
	+ Humans have 100-1000x the bps of bact but only 5 – 15x the genes
	+ Mammals have the lowest gene densities
* In proks, most of the DNA is genes that code for proteins, tRNA, rRNA.
	+ The small amount remaining are non transcribed regulatory sequences like promotors.
		- Remember the operon
* In euks, most DNA doesn’t code like in proks.
	+ Many more regulatory sequences
	+ Humans have 10,000x more noncoding DNA than bact
		- Some such DNA is in introns
	+ Avg human gene = 27,000 bps, bact gene = 1000 bps
	+ Euks have ‘vast amounts’ of non coding DNA between genes

Eukaryotic Noncoding DNA and Multigene Families

I’m calling transcription units genes. These code for proteins and all the RNA molecules that function in a cell.

Most multicellular eukaryotic genomes contain primarily noncoding DNA.

* Formerly ‘junk DNA’
	+ However, if it were junk, why would it be so common across a huge group of organisms, the Euks?
	+ It must have some function to be in so many groups
	+ Humans and rats and mice have about 500 regions of noncoding DNA that are identical
		- it must do something to persist unchanged in these groups that separated millions of generations ago
		- there’s more variation in the genes that produce related functional proteins than in this ‘junk’ DNA
			* it must do something
* we’ll use the human genome for our discussion
	+ 1.5% of our genome codes for protein or RNA
	+ 5% for regulatory sequences
	+ 20% (!) is introns
	+ The rest resides between functional genes. Fig 21.7
		- An interesting subcategory here is **pseudogenes**, sequences that were coding earlier in evolutionary history that have accumulated mutations and no longer produce proteins or RNA
	+ The book says the DNA that produces small noncoding RNA (e.g. siRNA, miRNA) would be considered part of the intron DNA % or the ‘unique noncoding DNA’ that they refer to like they have used that term before – more here when I figure it out
		- To me, it should be considered a gene. It produces a molecule that does something.
	+ Most intergenic DNA is repetitive DNA
		- Sequences of multiple copies of the same short DNA sequence
			* 75% of the repetitive DNA (= 44% of the total human genome!) is a type of repetitive DNA called **transposable genetic elements** and regulatory sequences related to them
	+ Transposable elements – move from on place to another in a genome
		- ‘Jumping genes’
		- Transposition = changing locations in genome
		- These elements do not detach from the DNA but the original site and the new site are brought together by DNA bending enzymes and substances
		- Barbara McClintock and maize showed us this is the 40s and 50s
			* The changes in the color of corn kernals between generations could only be explained by moving genetic elements within the genome
			* These elements must have moved to within a gene that controls kernel color, changing the gene function, changing the color of the kernel
				+ Another discovery that was met with skepticism but was later supported by boatloads of evidence
				+ She received the Nobel in 1983. It took a while for us to recognize the truth and importance of this idea.
		- How it works:
			* Transposons move within a genome via a DNA intermediate. Fig 21.9
				+ Move via a cut-and-paste mechanism

Removes transposon from original site

* + - * + Or moves by a copy-and-paste mechanism which leaves a copy at the original site and puts a copy in the new place
				+ Both require enzyme *transponsase* which is coded for within the transposon itself
			* Retrotransposons move via an RNA intermediate
				+ RNA intermediate a transcript of the retrotransposon itself, like the transponsase is
				+ Always a copy and paste situation – the DNA is copied into RNA and left at the original site. The RNA then moves sequence to site where reverse transcriptase encodes a DNA message that gets inserted into the new site.

Reverse transcriptase is coded for in the retrotransposon itself

This is one reason why we think retroviruses evolved from retrotransposons

* + - * + Another enzyme catalyzes the insertion of the transposed DNA
		- Sequences relating to transposable elements
			* Multiple copies of transposable sequences and sequences associated with transposition are located throughout eukaryotic genomes
			* One unit usually 100s to 1000s bps long
			* Some copies are of the actual transposable sequences themselves, other copies are of related sequences that lost the ability to move. Both produce needed enzymes and are part of transposition.
			* 25 – 50% of mammalian genomes are these sequences(!)
				+ Higher percentage in amphibians and plants

85% of corn genome!

Much of the wildly large genomes we find are like this. The outlier-large size is due to boatloads of transposons.

* + - * In humans and primates, most transposable DNA is part of a group called *Alu elements.*
				+ = about 10% of human genome
				+ About 300 nucleotides long

Which is actually shorter than most transposable elements

* + - * + Alu elements do not code for any transposition proteins

Many code for RNA of unknown function

So what the heck do Alu’s do?

10% of the genome with identical structure but no known function

* + - * 17% of the human genome is a type of retrotransposon called LINE-1, or L1
				+ About 6500 bps long
				+ Have a low rate of transposition

Certain sequences within L1s inhibit RNA polymerase function, necessary for a retrotransposon to move (makes copy of RNA mobile part)

* + - * + 80% of introns have L1 sequences in them

L1 sequences may help regulate gene expression

* + - * + Some think the differences in neural cell types may due to L1 effects on gene expression
	+ Other repetitive DNA and simple sequence DNA
		- Such repetitive DNA that isn’t transposable is probably due to mistakes in DNA replication or recombination
		- =14% of human DNA
			* 1/3 of it (about 6% of the human genome) is duplications of 10,000 – 30,000 bps units
				+ Seem to have been copied from one locus to another site and *may include functioning genes*
		- Simple sequence DNA contains many copies of tandemly repeated short sequences
			* About 3% of human genome
			* (eg GTTACGTTACGTTAC….)
			* Can be as long as 500 bps per repeating unit but usually are <15 bps
			* When 2 – 5 nucleotides present, they are called short tandem repeats (STRs) and we referenced these as targets in DNA analysis and identifying individuals in Ch 20 (Fig 20.25)
				+ The number of times the STR repeats can be variable between individuals. Those variable sites are targeted by PCR primers, amplified (cloned) and the number of repeats is used to identify individuals
			* Most of this is located by telomeres and centromeres and perhaps plays a role in chromosomal structure
				+ Centromeric DNA important in separating chromatids in cell division
				+ Perhaps this simple sequence DNA throughout the genome is part of chromatin structure (Interphase chromosome structure)
				+ Telomeric simple sequence DNA is the DNA lost when chromosomes shorten during normal replication. It’s not genetic. Lose it, no problem.
	+ Genes and Multigene Families
		- 1.5% of human DNA makes proteins, RNA etc (is genes)
		- If include intron and regulatory sequences (introns should be included in the gene definition, in my opinion, as they may be exons in a different situation), 25% of DNA is involved.
			* It makes sense to me to include this in your understanding of our genome. 25% of our genes are involved in making something. 1.5% of the DNA is the guide for making something functional, 23.5% of the DNA causes the expression of the 1.5%. Without the introns and regulatory sequences, genes produce no product.
				+ They say it this way: about 6% of the length of an average gene (1.5% out of 25%) is represented in the final gene product.
		- Many eukaryotic genes are present as one copy in the whole genome. That’s what we’ve talked about exclusively up to this point.
			* But < ½ of multicelled euk genes are such
		- More than half of many euk species’ genomes are in multigene families
			* Collections of two or more identical or mostly identical genes
			* Identical DNA
				+ Usually clustered tandemly hundreds or thousands of times at various places in the genome and produce RNA as a final product, (with identical multigene families that produce histones as an exception)
				+ Fig 21.11 (a)

Different rRNAs produced around the genome allow this commonly needed molecule to be present in huge amounts because cells need beaux coup ribosomes

They are tandemly repeated, or clustered, so the proteins are all available in the same area to make functioning ribosomes easily wherever this occurs

* + - * Nonidentical DNA
				+ Examples are the genes that code for globins that transport oxygen (α and β globins in us)

One family on chromosome 16 makes α, the other on #11 makes β. Fig 21.11(b)

α and β function differently and each have different forms. At different points in development, either or both are needed. The genes are expressed in response to signals from environment (other cells).

(eg embryonic and fetal hemoglobin has a higher affinity for oxygen than adult hemoglobin so mom’s hemoglobin will more easily lose it’s oxygen to the fetal hemoglobin.)

Duplication, rearrangement and mutation of DNA cause genome evolution

Mutation the source of new alleles.

* Much of genome evolution involves incorporating positive mutations
* Early life certainly had a small genome
	+ Survival and reproduction needs for very simple organisms only
* How did genomes expand to provide the raw material for mutation and thus evolution?

Duplication of Entire Sets of Chromosomes

* Meiotic errors cause polyploidy sometimes
	+ Usually lethal, but on the odd occasion that the polyploidy is not, that’s a lot more genes around to potentially mutate
		- Between the extra sets, enough of the functioning chromosomes and genes can do the functions of the original set alone, the other genes can mutate and diverge in function without necessarily messing up the original genes
			* As long as one copy of an essential gene is expressed, the other is ‘free’ to diverge in its product, possibly giving an advantage to the organism
	+ In plants mostly, esp flowering plants
		- Rare in animals, African clawed frogs (*Xenopus laevis*) is tetraploid animal

Alteration of chromosome structure

* At time of evol divergence of chimps and humans, two chimp chromosomes merged in the human line making Chromosome 2
	+ n for humans is 23, chimps 24
	+ fine genetic sequencing evidence supports this
	+ Fig 21.12(a)
* Finer sequencing showed commonalities between our chr 16 and parts of mouse chrs 7, 8, 16, 17
	+ Fig 21.12(b)
	+ These genes stayed together throughout mouse and human evolution
* In looking at human and 6 other mammalian species, they found many duplications and inversions of large portions of chromosomes. Meiotic errors.
	+ Rate of mistakes increased around 100 mya
		- Dino’s going extinct opening up niches?
		- Rapid increase in the number of mammal species
	+ For the individual with polyploidy, repro can happen with a normal individual but offspring would have two nonequivalent sets of chromosomes, messing up meiosis for that organism
		- This could lead to two different mating populations, the normal and the polyploids
		- They can’t successfully mate with one another
		- Speciation is starting
* Chromosome breakage points weren’t randomly located
	+ Certain sites were used over and over
		- Certain of these correspond to locations of chromosomal rearrangements within the human genome that are associated with congenital disease
			* Other sites of common breakage being examined for such relation to disease

Duplication and Diversion of Gene-Sized Regions of DNA

* Meiotic errors can occur in smaller areas too
* Unequal crossing over and duplications and deletions
* Fig 21.13
* They also mention something they call template slippage during replication
	+ where part of the template strand is either skipped or copied twice
	+ as a result, a segment of DNA is either deleted or duplicated
	+ this is thought to be a part of why different people have different STRs in areas of repeats
* multigene families are evidence of this sort of unequal crossing over and template slippage
	+ human globin gene family
		- Fig 21.14
		- α and β globin genes split 450-500 mya
		- each duplicated several times (and gave an advantage when duplicated)
		- this same ancestral globin gene gave rise to myoglobin and to a plant protein leghemoglobin
		- the idea is that at least one copy of the globin gene was doing what it should do, the other copies were free to accumulate mutations, with the odd mutation providing an advantage and being selected for
			* presence of ‘pseudogenes’ within these sequences supports the random mutation idea
				+ these genes have been rendered unfunctional by mutations
* Evolution of genes with novel functions
	+ In globin gene family example, differences were subtle causing differences in function but still had the same gen function (carry oxygen)
	+ An entirely new function could also happen
		- Lysozyme and α-lactalbumin genes for example
		- Lysozyme an enzyme that protects animal against bacterial infection by hydrolyzing the bacterial cell walls
		- α-lactalbumin is a nonenzymatic protein that is part of milk production in mammals
			* similar in structure (primary – 3d shape)
			* both genes present in mammals but only lysozyme gene present in birds
				+ sometime after the lineage leading to birds and mammals split, the α-lactalbumin gene formed via duplication then mutation in the mammalian line

Rearrangements of Parts of Genes – exon duplication and shuffling

* duplications and movement of exons, functioning parts of certain genes, facilitated evolution
	+ recall that exons often code for certain domains of a protein (CH 17), these often have certain structural roles in the protein or functional roles in the protein
	+ a particular exon could be duplicated or deleted
		- the gene with the duplicated stretch would code for two copies of the domain
			* this could increase the protein’s stability, enhancing it’s ability to bind to a ligand, or change some other property
		- collagen gene is an example.
			* It has a repeated pattern of exons that produce the repeated pattern of amino acids in the collagen molecule
* Also, occasional mixing and matching of completely different exons within a gene or between genes occurs caused by meiotic errors
	+ Obviously could lead to proteins with novel functions
	+ Tissue Plasminogen Activator example Fig 21.15
		- Protein helps control blood clotting

Transposable Elements and Genome Evolution

* Transposable elements such a big part of some eukaryotic genomes they must be important
	+ They promote recombination
		- Their presence increases length of DNA
	+ Disrupt cellular genes and/or control elements
		- Jump could occur to or from within a gene or control sequence
		- Could decrease or increase gene expression
	+ Carry individual exons or entire genes to a new location
		- Looks like this is the story in the globin family
		- One way that an exon from one unconnected/unrelated gene gets to a different gene causing truly novel function
* Reminder that most of the changes that occur this way are negative, but this is a source for the odd positive change that gets passed on to future generations

Comparing genome sequences to establish evolutionary relationships

Technology has allowed us to sequence genomes in a much faster and more functional way

* Advances in sequencing itself, in data collection, assessing gene activity and gene function, in understanding how genes and their products work in complex systems have made this a “truly new world”
* Comparisons between species show relationships more surely than ever before
* The genetics of development show such relationships

Comparing Genomes

* Obviously, the more similar two genomes are, the more closely related the species are
* Comparing closely related groups reveals recent evolutionary changes, more distantly related groups reveals more ancient evolutionary changes
	+ Duh
	+ Distantly related species
		- Which genes have remained similar, i.e. are highly conserved?
		- Bact, archaea and eukaryotes diverged 2 – 4 **B**ya
		- Highly conserved segments can be used for research into like segments in humans
			* Yeast and people for example
	+ Closely related species
		- Genome organization similar due to recent divergence
			* Can use similarities to ‘scaffold’ one species’ genome from a related known genome
				+ Used human genome to speed sequencing the chimp genome
		- Usually only a small number of genes different between two closely related species
		- By comparing genomes of a variety mammals, we can find genes important to being a mammal, for example
		- Comparing genes of chimps and us we could find what it takes to be a primate, etc
		- Fig 21.16
		- chimps and humans diverged about 6Mya
			* 1.2% genetic differences, when looking at single nucleotide substitutions
			* Looking at longer stretches of DNA, 2.7% difference due to insertions or deletions of large regions of one genome or the other
			* 1/3 of human duplications are not present in the chimp genome
				+ Some dups associated with human diseases
			* We have more Alu elements
			* They have way more retrovirus provirus in their genome
			* We don’t know how these differences account for the different characteristics of the two species
				+ Researchers are looking at specific genes and types of genes that are different between the two species

Also looking at these genes in other species

Have found some genes that are changing faster than others in humans and are changing faster compared to those genes in chimps and mice

Genes that defend against malaria and tuberculosis and one that regulates brain size

Gene groups that code for transcription factors are evolving rapidly, too

These regulate gene expression and overall control of the ‘genetic program’

One rapidly changing transcription factor is FOXP2

Thought to be part of vocalization in vertebrates

Mutations in the gene cause severe speech and language impairment in people

It’s expressed by songbirds, when as young birds, they are learning their songs

Fig 21.17 – for an experiment where the gene was knocked out. Then we looked at phenotype.

Results support the idea that FOXP2 functions in vocalization and we can assume it has similar function in other mammals

In another experiment, researchers replaced mouse FOX gene with a human one (that is different than the chimp one by two amino acids and may be responsible for our ability to speak)

Mice were healthy and had different vocalizations than normal mice and activity in their brains in areas associated with human speech

* + Comparing genomes within a species
		- Shows differences between people
			* 200K years as a species = not long
			* Not a ton of variation between humans
				+ Mostly SNPs

Usually at a rate of 1/100-300 bps

Found ‘several’ million SNP sites, more being found

* + - * Inversions, deletions and duplications aplenty
			* Also, a widespread occurrence of copy-number variants (CNVs)
				+ Locus where people have one or multiple copies of a gene or genetic region, instead of the usual two, one on each homologue
				+ These areas have been duplicated or deleted in a manner inconsistent with the whole population

Studied 40 folks, 8000 CNVs, involving 13% of genome and only looked at a small subset of genes

* + - * + Since these are not SNPs, but longer stretches, it is more likely that CNVs will have phenotypic consequences

Play a role in genetic diseases and disorders

Book mentions that the high number of CNVs make it hard to say ‘a normal human genome’

* + - Comparing genomes of two Africans
			* Showed the high diversity of genomes among African pops
		- We can use this stuff to establish migratory patterns from our distant past

Comparing Developmental Processes = evo-devo

* The aim is to understand how developmental processes have evolved and how changes to them can change species features
* Genomes or species with wildly different forms may differ in only minor ways regarding gene sequence or gene regulation
* Widespread Conservation of Developmental Genes Among Animals
	+ Remember *Drosophila spp*. and its homeotic genes (Fig.18.19 – 18.22)
		- The all include a 180 nucleotide sequence called a homeobox
			* Produces a 60 AA long protein called a homeodomain in the resultant encoded protein
				+ Identical or extremely similar sequences exist in the homeotic genes of people and fruit flies
				+ These genes have even kept the same chromosomal arrangement in these varied creatures (Fig21.18)

Same sequence on a chromosome

* + - * + Similarities extend between larger groups

Plants and yeast

* + - * + Homeobox DNA evolved early in evol history and was valuable to org’s development so it was retained relatively unchanged in many orgs –

Mutations away from this must not be as functional

* + - Homeotic genes in animals called *Hox* genes
			* Homeobox containing genes
			* Some homeobox sequences are not in homeotic genes
				+ Most such genes are associated with development, in some way
		- Homeobox encoded homeodomains is the parts of a protein that binds to DNA when the protein is a transcription factor – it’s active site
			* Shape allows it to bond to any DNA segment, not just one
				+ Other, more variable domains determine where it attaches, which genes it controls
			* Homeodomain containing proteins recognize specific enhancer sequences
				+ Coordinate transcription in development by switching batteries of genes on or off
				+ Central to pattern formation
		- Genes other than homeotic genes that are involved in development are also highly conserved between species.
			* Cell signaling pathways have very similar proteins
		- With so many genes in common, how do different creatures result?
			* It’s small changes in regulatory sequences of developmental genes that can cause large scale changes in body form
				+ Fig 21.19
				+ Differences in hox expression leads to largely different body shapes
				+ Small differences in master regulatory genes causes large differences in body
* Comparison of Animal and Plant Development
	+ Last common ancestor probably a unicellular eukaryote alive over 600 mya
	+ Plants – cell wall, etc differences don’t allow same dev as animals
		- Cells not mobile during development
		- Similarities exist, though
			* Both rely on a cascade of transcriptional regulators turning on and off genes in a well coordinated manner
			* Plant have MADS-box genes that are their homeobox genes