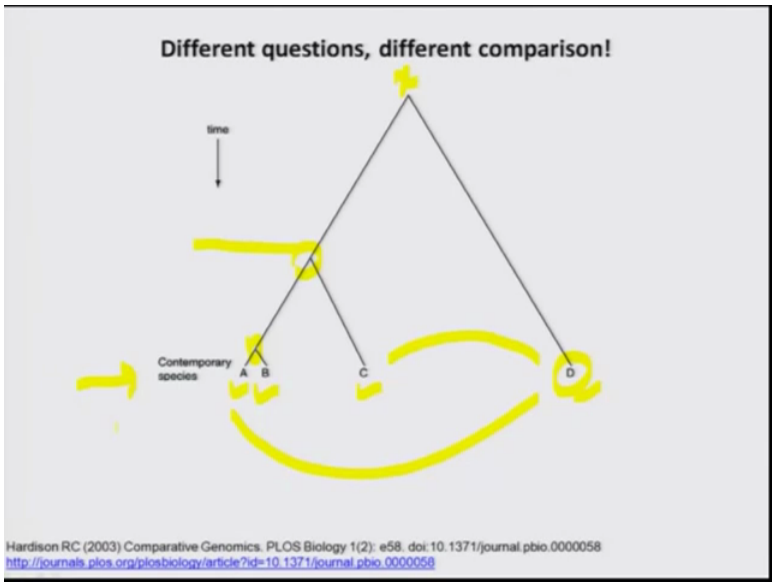
Functional Genomics
Professor S Ganesh
Department of Biological Sciences & Bioengineering
Indian Institute of Technology Kanpur
Lecture No 16
Comparative Genomics: Genome Sequence: Different Questions, Different Comparison

So welcome back to this course functional genomics, so in the previous lecture we looked into the power of comparative genomics. We looked into some of the examples as to how comparing various genomes of various species helped us to discover some of the changes that are having an impact as to how that species was successful we could get insight into certain unique feature we have seen examples how we can increase the the utility of a breed whether it is a milk whether it is a muscle mass and so on.

So now we are going to look into how really we look at the genome some of the examples as to how the genome has been, information has been used to decipher the changes and concepts or fundamentals behind that.

(Refer Slide Time: 1:08)

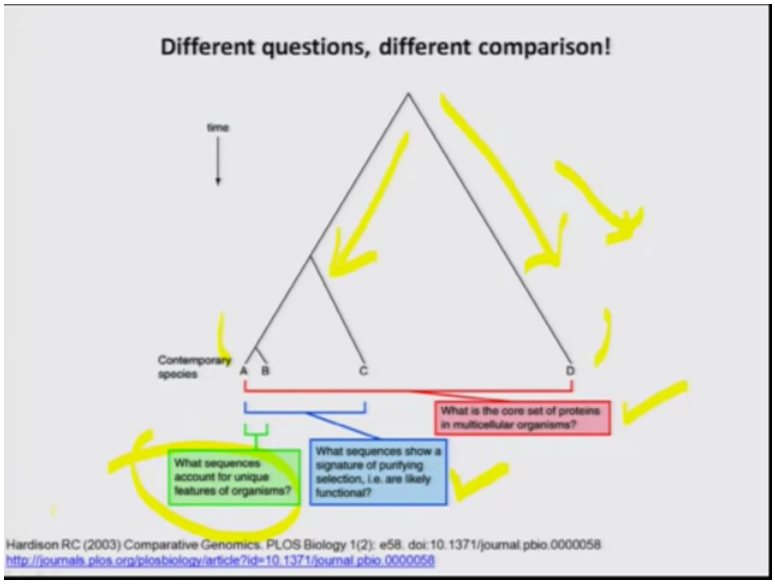


Let us look into this particular slide what we are talking about is that you have had a species X to begin with, with time the species has evolved and diverged into you know A, B, C and D right. So now it obviously from this plot what is shown here it is very clear that D is you know evolutionary far away from A right whereas C is more closer to you know D as compared to A

and so on. So it is clearly talks about at which time point possibly these two species could have evolved into two new species so that is you know relatedness so that is what we are talking about.

So what we have done in this process is that we basically look at the species that are contemporary meaning they are living today and try to decipher as to how the common ancestor would have been the connecting links are normally lost we really do not have for many of the such such events we do not have an evidence as to how could have happened we basically look at the genome and try to infer as to how similar or how dissimilar they are and try to plot and then sort of calculate when this species could have diverged right. So it all depends on what question you are asking and that determines how you are going to compare what kind of comparison that you are going to do.

(Refer Slide Time: 2:40)



Let us give some examples for example what is the course of the proteins in multicellular organism. So if this is a question for example you want to ask the question what are those proteins that are involved in say a very fundamental process like cell division or for example metabolisms. These are pathways that are highly conserved where you take either plant or animals these are very similar.

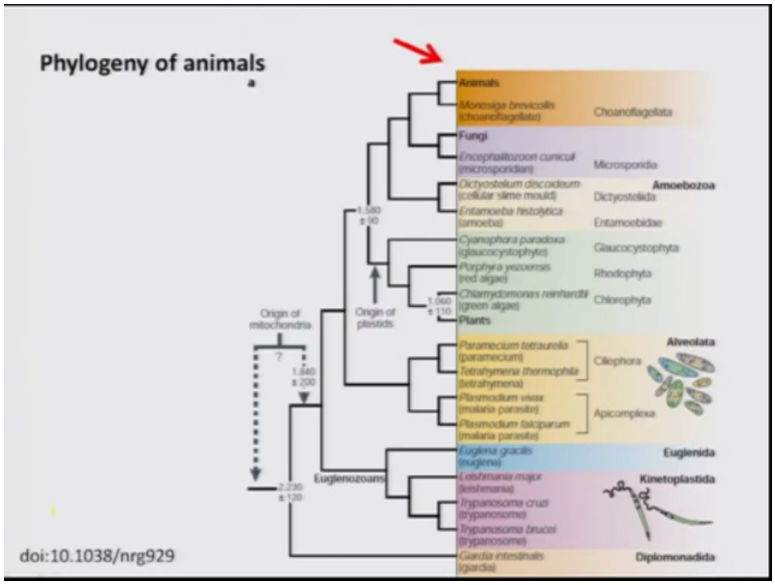
So if you look into genes that are highly conserved across species that are you know diverse for example you talk this as a human this as a grass and still there are genes that you can simply swap. Histone for example you know they have gene from the grazen or take a gene from human and put it in grass it possibly would work exactly like histone because they have set functions that is conserved for many years and frozen for whatever function they do and therefore they do not really change.

So it depends on what question you are asking or for example you asking a question what sequence shows a signature of purifying selection. There are likely to be functional, let us assume that this particular branch represent malign organisms and this particular branch represent terrestrial organisms so obviously this branch on the left represent certain changes that help the terrestrial organism to survive on land rest here you know they have selected genes that help them to live in the water so this is what we can we can ask what sequence show a signature of purifying selection likely functional and so on. So you sort of look into and so on or what sequence are conferred unique features of organisms for example humans.

We have higher order brain functions we have language abilities and others which is lacking in primates so what really made this change so you can compare with for example primates and ask the question what are the unique signatures that we have which helped us to have the higher order brain function as compared to primates right so that but if you compare it with cats then it is going to be far away because that would not tell you the difference it also talks about how they can hunt and many other you know characters that come in.

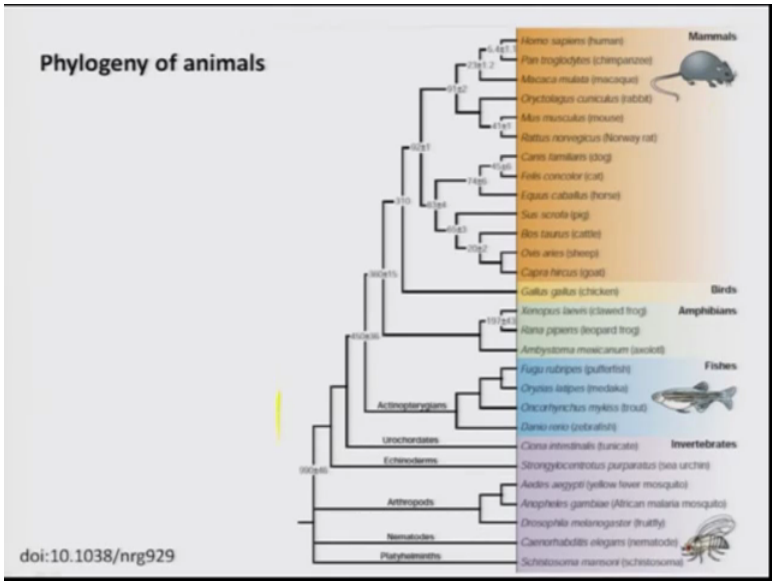
So you have to go for the close related species yet they do not have this unique feature and then look at you know what the differences is. So this all depends on what questions you are asking there such is you know approach.

(Refer Slide Time: 5:08)



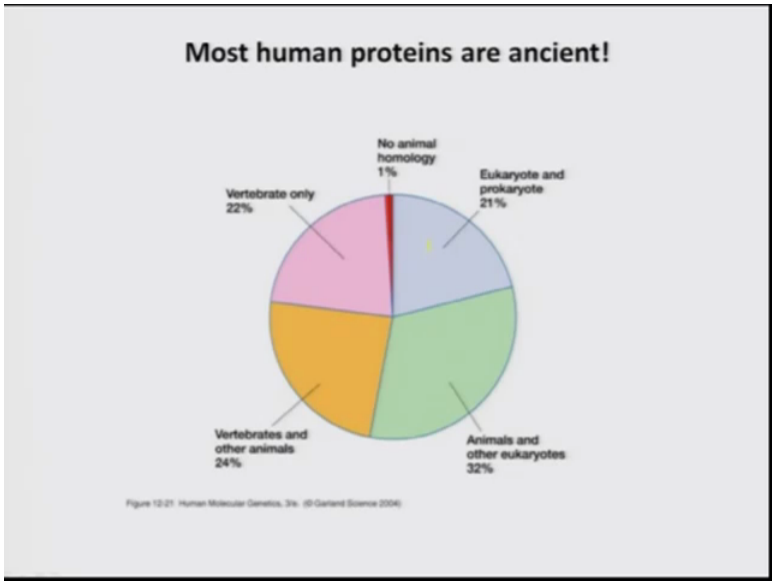
Let us look in here, so when you talk about you know comparative genomics you pretty much covering the entire spectrum of few species that live on the earth. So it goes on to this fellow genetic tree talks about the early evolution to what you have so we simply so we know we have this here so the animals represented here. The rest are all the other species we are talking about you know the primitive species that like lupinus and others amoeba and others fungi and and so on.

(Refer Slide Time: 5:41)



But if you go to more animals phylogeny that is what shown here so you have all the invertebrates here and then the vertebrates coming over here and then off course you have the homo sapiens here. So now it all depends on what questions you are asking right that that would help you to answers, so if you are trying to look at how for example you know the chimpanzee you know diversified from the humans or how similar they are or dissimilar they are. You know you are going to look into their genome sequence of these species but you are talking about how you know mammals evolved from for example birds as compared to (())(06:15) right or the ancestral species then we are going to go on and look at larger number of species.

(Refer Slide Time: 6:30)



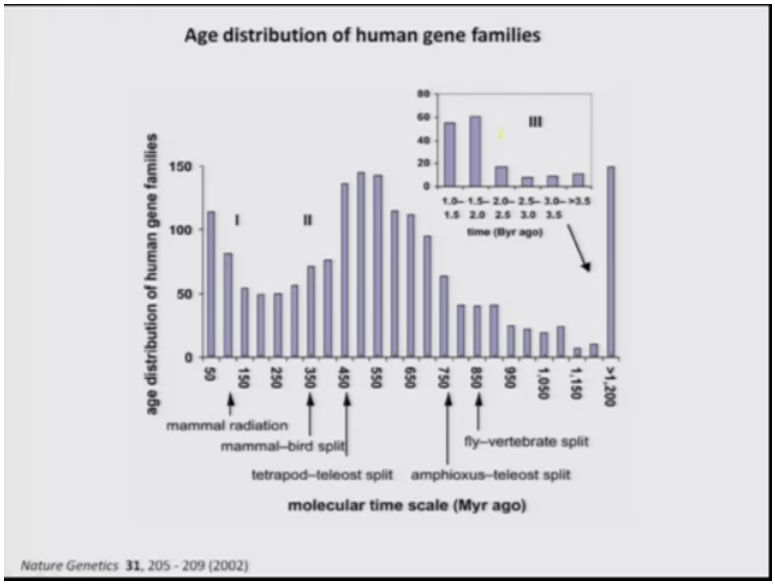
So this is the way people do it and what is interesting here is most of the protein that we have humans have have arranged it is not that we have something unique, majority that we have even you know fish would have right so you can see here if you look into the protein percentage in terms of coding potential, it about only 1 percent of our protein apparently has no homology meaning they do not show a similarity with other species they are unique to humans that is one percent.

In other words 99 percent of protein we share with other species closely 22 percent of the protein that we have is small fraction again is vertebrate only in other words if you combine 22 with 1 23 that means 77 percent of the genes are present even for example in star fish or in in in in earthworm and so on. So that means that you know we have carried the same gene some

variations in terms of functions and so on that's what it is and vertebrates and other animals animals and other eucariote, eucariotes and procariotes.

This includes even bacterium and for example even plants and other we share about 20 percent 21 percent of the gene they have very similar function that tells us how you know over time you know our genome have retained some of the genes because that are doing the critical function which cannot be altered but the same time there are other genes that either came in new set of genes or the existing genes have changed a bit and become you know took up the function that are either unique to your kingdom for example the vertebrate and so no so that is what happen right. But this information has come by looking at the genome so otherwise you would not have told you know somebody would have told you that we share 70 percent of the genome for example the rest of the animals including for example the fish and invertebrates like octopus and others you would not have believed right. So but this has come from such kind of analysis where you compared the sequence and try to classify them and tell it right.

(Refer Slide Time: 8:45)

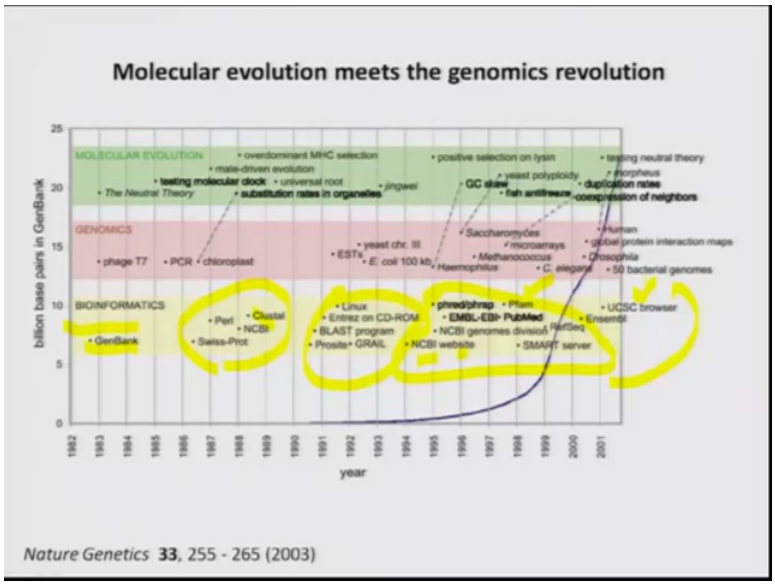


The other plot that also come from a paper is that the age distribution of human gene families what we look into is that when possibly the genes have evolved, so you can see that these are all you know molecular time scale. These are million year millions of years ago and we can see that this is possibly when you know this again talks about how these genes evolved. This is where the

vertebrate fly vertebrate split you know diversified and then this is where the animals the vertebrate started coming to the land and this is where the birds and mammals evolved.

And mammals radiation took place meaning they really flourished the large number of species evolved and they have sort of conquered the earth you know thats what happened, so you can see that majority of genes still up there there are some of them are pretty old you know set of genes still we have so that is what this ball diagram again represent.

(Refer Slide Time: 9:36)



This is just to tell you that molecular evolution you know how whatever understanding that we have has come from the genomic revolution that we discussed so far in this in this in this course. So what is shown here is the genomic revolution we talk about the birth of Bio-informatics you have the gene banks and then a large number of data bases for example for the protein for the and method to analyse and compare the sequence like for example cluster is one of such sequence. Comparison programme and then you have all these other powerful algorithms that really helped us to analyse the sequence right.

The blast programme and and so on right and then we have other powerful tools that have come in for example P families a tool that identifies the protein families and then you have this NCBA website everyday if you are a biology student you go on look into then you have the UCSC browser that we have looked at often as to how really it helps us to compare the genomes of

various species identified you know functional elements and other ensemble tools and this comes along with this is in this section it is strong is the genomics.

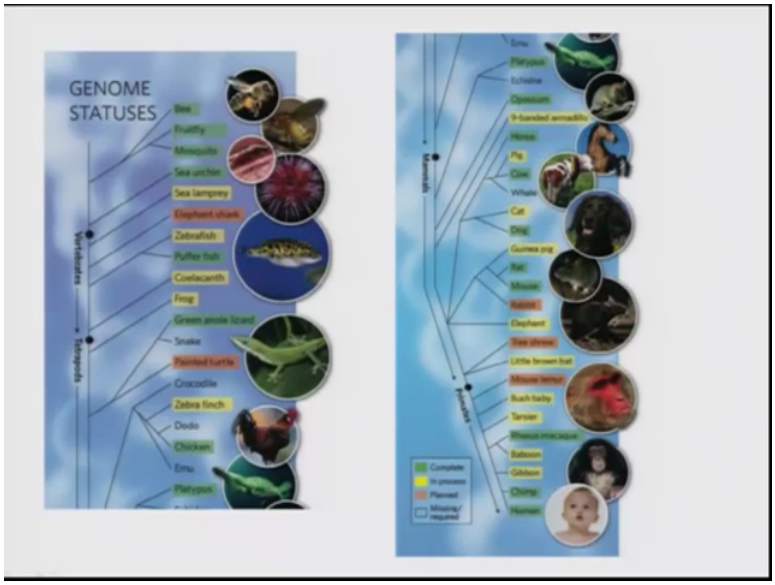
For example the forged T7 you know sequence is then PCS invented chlorofome genome was identified then you have all these genome information coming in EST equal genome, yeast chromosome, hemophilia you know hemophilus genome was sequenced up to the human right. So you have the explosion of the genome like ways it developed but at the same time we were unable to understand the molecular evolution meaning the possible event that took place during the evolution right.

So that is the you know a neutral theory, testing molecular clocks right substitution rates and organelle is another important thing. So we often compare the genome and tell these are the genes that are highly evolving right or they are evolving very rapidly right and these are the genes that are showing slow evolutionary rate. How do you really tell us so if you have more substitutions you know changes in the in the codon especially third base and amongst the members of the same species right wild species that means that the gene is still allowing changes what does it mean.

It means that the function of the gene is not well settled so still there is a you know change that is happening. Once that you know whatever function is set probably it will not change beyond that because then it becomes an essential for the organism. When a gene is not that much essential for the organism even if there are changes that is allowed. You will find an individual variants but if the gene becomes very essential for the organism to survive if there is a change in the gene then that individual may not survive. Therefore you do not see the variation so this is how people identified you know genes that evolve faster or genes whose functions has been you know sort of frozen right.

Then you have all other such you know remarkable discoveries few people talk about the fish anti-freeze mechanism even they can live in sub-zero temperature and then duplication breaks. How for example the genome has duplicated especially in the teleosts the bonny fishes and that led to the evolution of vertebrates right that we will discuss now a little later. So these are some of those important discoveries or hypothesis supported by the analysis from the comparative genomics aspects led to a newer understanding of evolution.

(Refer Slide Time: 13:31)



So this figure just to show you that what other different species that I have been sequenced or currently under sequencing so you can see somewhat dated but you can see that we are you know we are starting with bees you know honey bees and other such insects genome off course should fly mosquito seared change again important species because it connects the (())(13:50) terms with the fish because that how they have invertebrates to vertebrates have evolved.

And they have a fantastic you know genome that tried helps us to understand how the genome would have evolved and there are early fishes to zebra fish to puffer fish we will talk about this powerful animal puffer fish because it has got the smallest genome right. It has almost the same number of genes that humans have but it is one tenth of genome of the genome of the human, so it is like a compact version of you know species that is very unique and we can understand how the genome functions there and there are of course other species in the reptiles, some of the turtles, snakes and lizards and zebra-fish again these are birds that that Darwin has studied if you can understand what is process.

What you talk about now is adaptive radiations that have come from the zebra-fishes and then off course there are several examples when it comes to mammals. Starting with the primitive mammal what you call as platypus which lay egg and but still a mammal and to the opossum, horse, pig and and up to human right. So you can see that these are the varieties of species whose

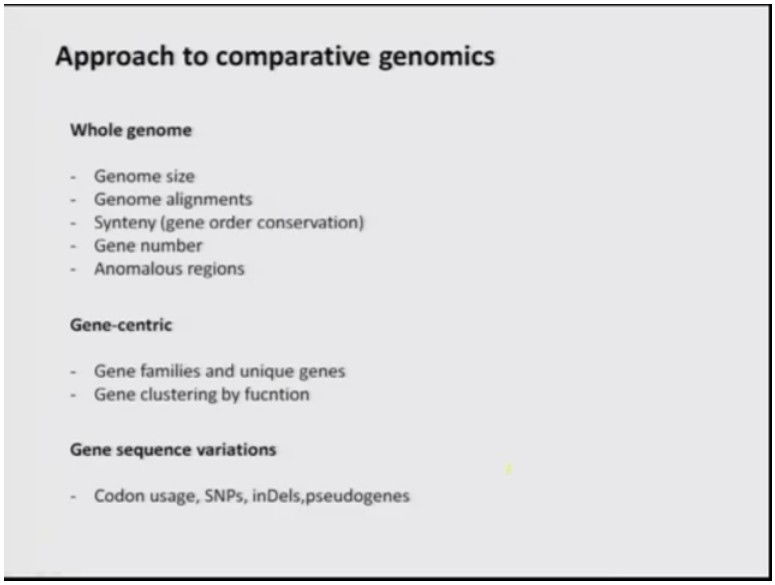
genome is being sequenced or completed and it gives you wealth of information to you know mine for look for and understand.

(Refer Slide Time: 15:21)

This is only to tell you how important it is because majority of such information with regard to sequencing the whole genome of various species have been published in journals like science and nature which really talks about you know the importance of you know such genome information including for example Panda you know that species not a good model but it is a unique species that gives us and and for example (())(15:44) you know the primitive the humans or the forerunner right, and for example the ancient human genome some of the fossilised samples which represents possibly the you know ancestor to homo sapiens.

People have sequenced we will talk about how for example even language other evolution probably link to this gene will give some examples right. So these are very important but it is not restricted to the animals its equally important also to look at that genome of plants because they are the primary producers our food mainly comes from you know the plants and we have rice, we have wheat genome, tomato. You talk about any commercial plant you know we do have a sequence coming out of such species right.

(Refer Slide Time: 16:41)



So what you do with this how really you do genomic or comparative genomics so one is that you look into for example the genome aspect you look into the different species with regard to the size of the genome right second you compare the genomes and look at alignments meaning what are the regions that are similar and what are the regions that are not similar. Third you go for synteny meaning you know you have a chromosome in human say chromosome1. You have a gene order in chromosome1 say A, B, C, D, E.

Now what is that you know the order for these genes in for example in chimpanzee is it chromosome 1 these five genes or together or they are split into many chromosomes so that is going to tell you how the chromosomes have evolved so we will look into one such example and second is gene number right so you may have the same kind of protein but for a given protein we have at least three different forms of it because the genes have been duplicated or the total number of genes have changed right that also tells you and what is called as anomalous region there are regions that are very rapidly you know changing right.

The evolution rate is very very high for example Y chromosome in case of mammals you know there is a huge difference when when you compare from one species to the other whereas X chromosome is highly conserved it is not much change happening there. So these are the you know information tells you what happens to a particular genome whether it is undergoing

evolution , whether it is conserved how it could have evolved and these are the information that come when you look at the whole genome into.

There could be also a gene centric studies so I am not interested in the entire genome but I am going to look into a particular group of genes, say I am looking into genes that are involved in higher order brain function cognitive abilities. So obviously you know if you you know higher order brain function or the genes are going to be enriched in organisms that have our cognitive abilities so I can now compare the apes with other mammals and compare it with humans and see what are the genes that are really you know makes our brain to function the way they are.

Therefore we are studying other species right or you know you know go on and look on gene cluster for example how the gene family has evolved so we will give you some examples on this and finally we also look into what is called as gene sequence variations. This is an important area which directly affect the humans because for example the sniff or single nucleotide polymorphism or insertion, deletion polymorphism or (())(19:21) genes you know the sequences that have gene structures but they are not functional.

These kind of signatures are very very important for the human health because you know while of us share 99.999 whatever percent of the genome sequence the small difference that we have between individuals makes a huge difference in terms of why you are for example able to tolerate a flew infection wherein I am more susceptible right and things like that you know there is everything.

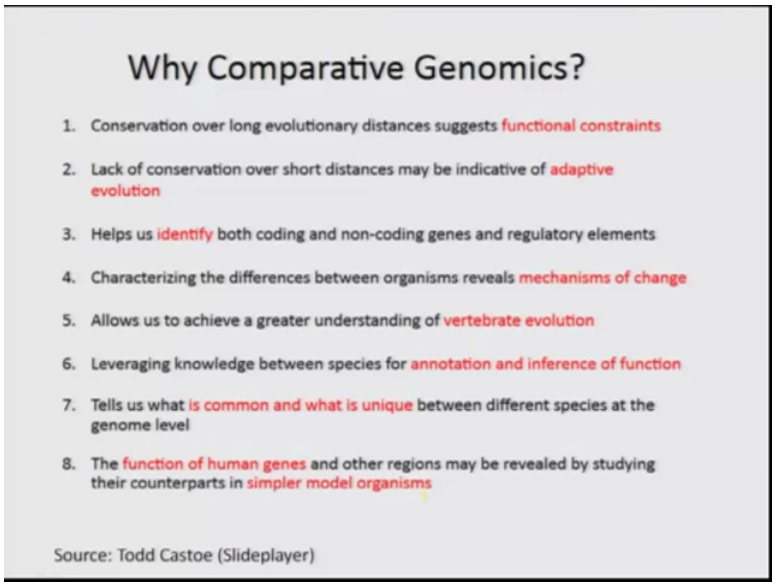
So that where now people are doing what is called as a whole genome association GEWAS what they call genome wide associations studies which clearly talks about the variations across a group of individuals compared to another group of individuals. What are these two groups one group could be normal so for example I am from Kanpur I would like to understand how pollutions here induce asthma in in the population right.

So it is not everyone is (())(20:30) some of them are still although living in the same environment still had not developed asthma, so it could be there are certain elements in your genome which makes you resistant for you know producing these allergic reaction right so what is that signature. So if have individuals from Kanpur population one that have developed you know asthma as compared to the other who did not develop asthma then I am going to compare

these two genome sequences and see what are the signatures that are unique to one group not the other that would tell you either genes that give you risk of developing asthma are genes that give you resistance right.

So this the last section the gene sequence variations and how the variations help in survival is something that we will not discussing in this course but this has been dealt with another course called human molecular genetics. I would encourage you to go on look at you know those videos that are available freely if you are not doing that course that will give you more understanding, so right now the remaining talk we are going to talk about the oligo-genome and gene centric studies with regard to how they help us in understanding the evolution.

(Refer Slide Time: 21:48)



So let us ask some simple questions why comparative genomics why at all you need to do this study. This is you know we have discussed some examples still I am going to go back and ask this question because that will help you to appreciate as to the importance of this field. Conservation over long evolutionary distance suggest functional constraints, this is statement what is it mean what does it mean by functional constraints right. The functional constraints is that your genome changes right your sequence change and the change is not directional meaning the process does not identify a region and changes here, it is random.

The minute happens because of mutation when the DNA replicates when the copying machinery is not faithful in copying or when the repair commissioner is unable to correct all the errors. Some regions you know you introduced some new change and this new change whether it would come to when I said introduce new change it happens during the meiosis in the meiotic cell that is one one come to the next generation and if so happen it becomes a gambit that fertilize an egg whether the new change would come in that individual whether the individual would survive depends on whether the change is affected the individual or not, right.

So suppose the change did not affect the survival of the embryo and and the birth of that individual then you have seen that mutation in that individual in the population because that comes guy comes in. So now you may have a problem for example you may have a sixth finger that is because of the mutation. Now the sixth finger is not affecting his survival, he is as good as anyone occurs. So you would see them in the population right that mutation has come in.

Some mutation would not affect anything he is absolutely normal just like you right and then you know it is there in the population it does not really affect anything, so it is called as the neutral variation. Some could be you know someone is born blind because of the mutation affected the way eye functions, so he may not transmit this gene to the next generation because he may not marry. So he you know stops there some affects even the embryonic developments so that is abortant the embryo is aborted.

So the process of introducing new variants in the genome is a random process but whether that new variant comes with the population depends on whether the variants is affected the function of the gene or not affected. So if you look into the normal population what you are looking at is variation that are allowed because they did not affect any of the gene function, so obviously regions that are not very critical for gene functions or the regions that have more variation because it is allowed because you are normal otherwise.

So that and the opposite of it is that there are regions in the genome in which variations are not allowed meaning if variations happens it affects the gene function therefore the person may not survive, may not transmit the gene to the next generation. Therefore there may be some reasons where there is a constraints there is a pressure not to allow changes meaning by natural selection right.

So that is what you call as a functional constraints so that tells you which segment of the genome is having a critical function, so if you compare a large number of genome of different species then you will find segments that are showing higher homology, similarity but not their genes they are not genes then that using information that this segment has got an important function for the cell for the organisms therefore it is not allowed to undergo changes right, so that is important information.

Lack of conservation over short distance may indicate of adaptive evolution, what do you mean by adaptive evolution right so you must have studied about mammals and other Darwin finches these are examples classic examples adaptive radiation when when there is a pressure like for example the same population of a species you know the members live in a locality or habitat if there is a sudden pressure on them change right then what they do they migrate to different you know region to survive.

So they wherever they go they may find a new challenge so whether the the individual of that particular species the members of that species that went there weather they are able to survive in that new place depend on whether you have add some variants that made you successful over there right. For example you know it could be that you are driven away from your habitat you went to a place where it is you know a region where it is close to the waters right seashore kind.

Then you have to have abilities to sustain that kind of you know humid condition or you may go to a desert region where it is there is no water at all. So you have to have adaptation or some abilities to survive with little water or you are gone to higher altitudes where it is much colder so you have to survive that cold temperature so if you had some changes which otherwise were not helping you in the original place.

Since you are driven out now if you go there now that particular you know change that you have which did not really help you earlier may give you an advantage right so then you populate you make a parent now if you isolate these species that originally diverse from a common ancestor very recently and then you do a sequence analysis you will find you know short regions of the genome that show lot of variations that talks about how the adaptive evolution took place that is another example.

Helps us to understand both coding and non-coding genes of unregulated elements again it is an important aspect you know the coding segment is you know we told you less than one percent the remaining is regulatory regions how will you identify what are the regulatory regions. Sequence wise it is going to be extremely difficult so if you compare the genome probably it will give you clue as to what are the segments you know that are highly conserved which gives an information as to these are you know important elements in functioning the gene or its expression or any of the other aspects. We will give some examples as to how people have understood that.

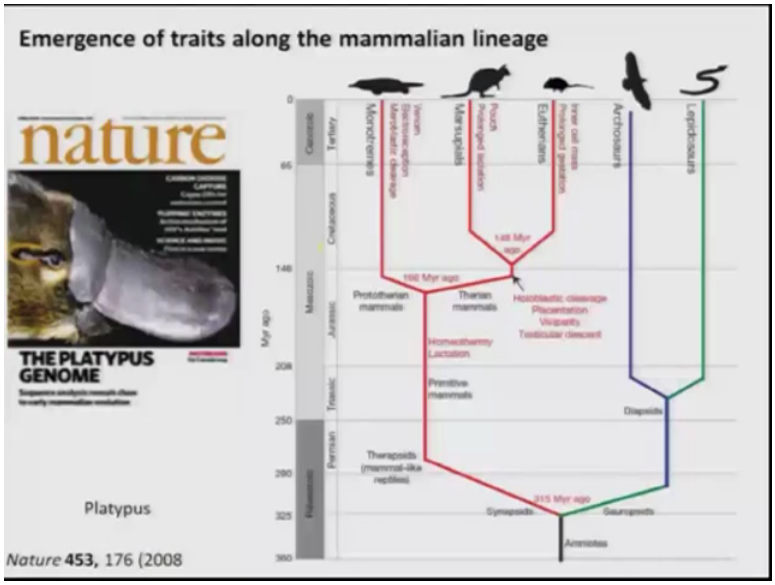
Characterising the difference between organism reveals mechanisms of change right so you know it it tells you like how certain organisms are successful in certain you know environment right so it really tells you like how certain species are successful and what are the changes that helped them to be successful and you can understand the vertebrate evolution. So all of us you know from fish to the mammals, we all evolved from you know a few set of species so we that is why we share the genome and we want to really reflect and see how we evolved. So that really helps we will show some example.

Knowledge between species for annotation inference of functions like we told you gene ontology is another example where you are studying yeast understanding a function of the gene that gene that we also have. So that information help us to understand what is the function of that gene in the human and tells us what is common and what is unique between different species and genome level. So for example why humans are very similar to apes, yet we are very different from the apes right it tells you what like you know we are successful because both of us can function, walk with leg by petal only two legs now we can walk.

So that made us to use our forearm which otherwise we are using it for walking now we are able to use it in a different way so we can make tools now because no longer hands are used for walking. So it was a gift given by the primates and this thumb we are able to put it across the four fingers again it helped us to hold the pen or chisel or whatever it is we are able to make tools because of this again this is a gift given by the primates to us right and then we have developed our own qualities for example our brain is far more functional more sophisticated in a cognitive abilities that is beyond what the apes can do.

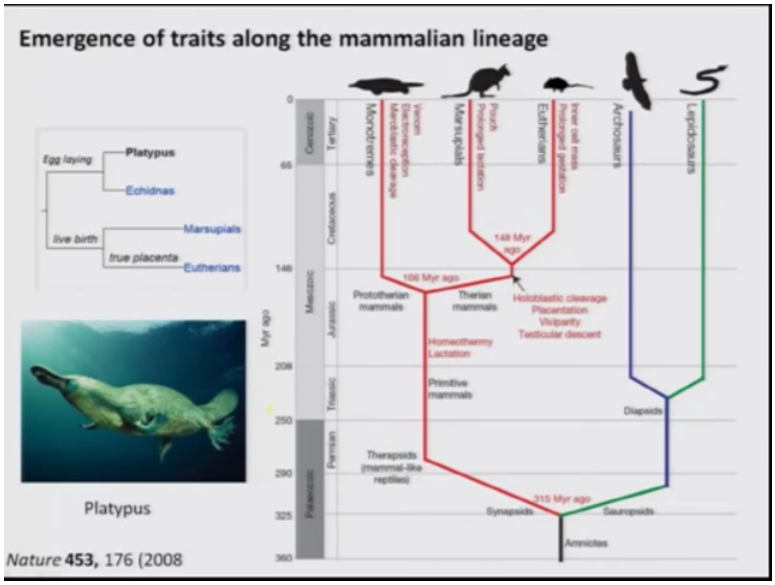
So we have acquired something new makes that unique so we can look into these these gene signatures and the functional of human genes another origins may be revealed by studying the counter parts in simpler moral systems. For example if you cannot understand the gene function in human we take this sequence engineer the animals put it over there and understand what could possibly be the function. So we are going to look into some of these examples how the model systems we have we are able to use to understand the human gene function.

(Refer Slide Time: 30:37)



So let us look into one good example that is the platypus ok this is sequenced this is nature issue that carried the genome sequence of platypus.

(Refer Slide Time: 30:56)



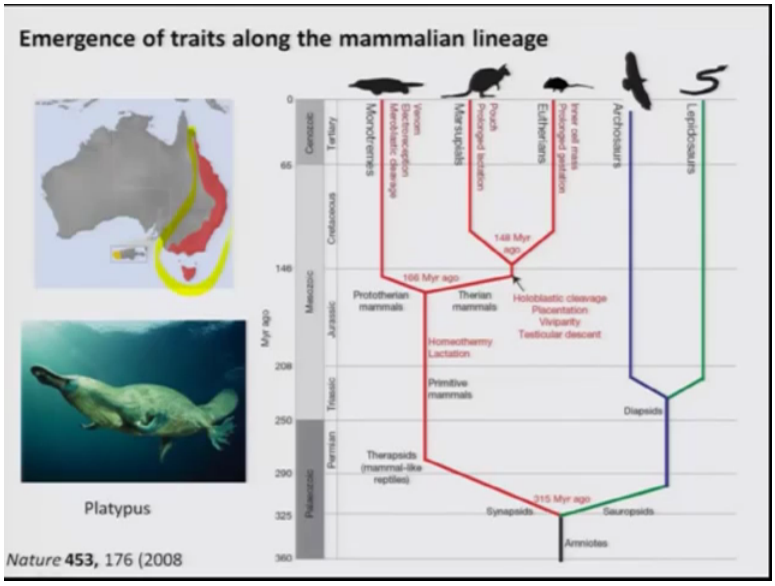
What is platypus platypus is a is a mammal monotreme what they call primitive mammals we can see that these are called as you know you have amniotes the amniotes are those in which the embryo grows in the there amniotic fluid whether it is birds or reptiles or mammals, we all have a amniotes fluid that is why it is called as a amniotic fluid. And then you know you have these two lineages that is birds and reptiles again they have egg but they are amnion and amniotic fluid and then you have this amelian lineage in which you have this prototherians primitive mammals which includes for example the monotreme example is the platypus that we have seen here. They are egg laying mammals ok they lay egg but they still they are mammals.

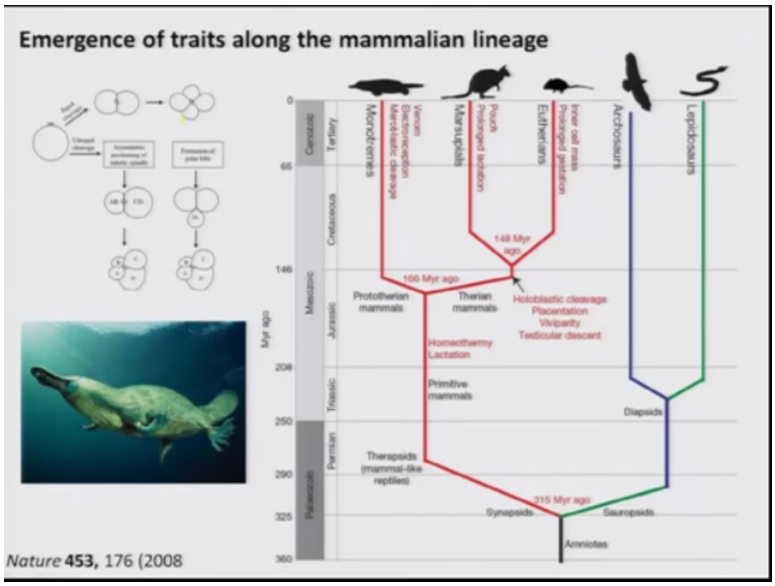
And then you have what is called as the morsupials which are kangaroos which are again the they do not have egg but new borns are not fully develop like humans so therefore they remain in the pouch they develop and then they come out. And then you have this eutherians or true mammals in which the live young one is completely developed you know so that that is the young one. We have these categories so there were theories which suggested that this platypus what you are talking about could possible a connecting link which connects us with the reptiles because they have egg and they have skin like but they have this scale and things like that right.

So that is what you know was the hypothesis therefore and and also they have venom this another important thing so they are just like the snakes and some of the lizards they can spike venom it can be lethal right so that that they have so therefore there was a one theory that they

are more closer to the lizard therefore their connecting link you know between reptiles and and the mammals. When they have sequenced the genome then altogether you found this kind of a lineage which was unexpected that is what it is.

(Refer Slide Time: 32:54)





So where they are this platypus lives in the eastern part of Australia so this is a region that you find platypus not in other parts right and what is unique about the platypus is that there are two ways by which they you know fertilize the egg one is that once the egg is fertilized the division is unequal like what you see here, so you have one smaller half one bigger half and go on

developing. This kind of cell division you find normally when there is yolk for example chick and many other which are characteristics of the species that are present on the right side the reptile and birds and so on whereas mammals you know follow this equal you know cleavage.

So one cell divides into two and four each one have same size that is why when at times when the embryo splits into then you will have identical twins so it is possible because you know all the cells are equal in size unlike you know in case of majority of reptiles and birds what is unique is that platypus followed this model that is they have a cell division in which you know one cell is larger than the other when this division happens. Therefore you know by looking at the embryonic development people always classified this is closer to the reptiles as compared to the mammals right.

So when they have done the genome sequencing it was understood that they are like what is shown here that the platypus is more closer to the mammals and to the reptiles right. We can see here they are separated probably about 350 millions of years ago and you have like what is called as a mammal like reptiles right though they are mammals and they are the common ancestors for all the three major groups of mammals monotremes, marsupiles and eutherian right.

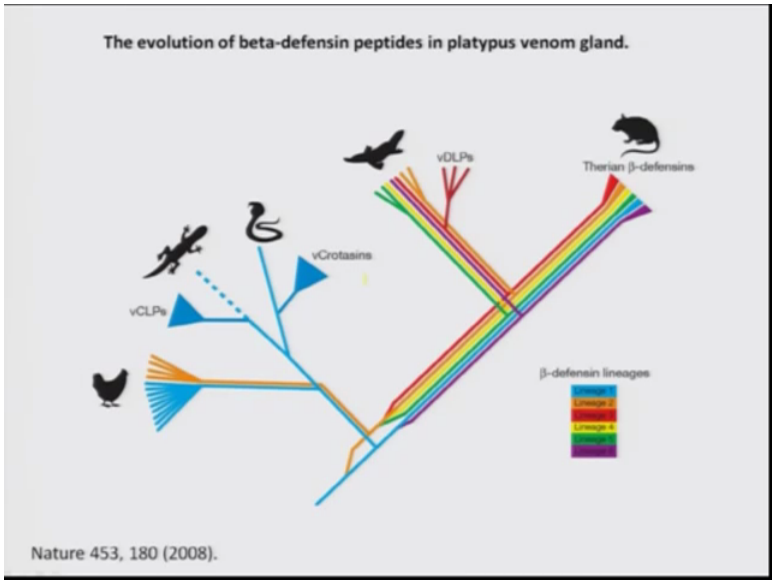
So it is important to know that both monotremes and marsupiles these are unique to the Australian continent whereas eutherians are present everywhere so but we do not know so but obviously they represent a different lineage about 166 million years ago that diverse the prototherian, therian mammals again you know you have what is called as a holoblastic cleavage this is what I am talking about, possibly this you know species sort of evolved a system wherein the egg is divided into two during development.

This is also could have happened because you know the egg no longer require a yolk because the embryo is bathed in the amniotic fluid and the fluid is always nourished by the mother because of that you know it need not have such unequal cleavage so the placentacion and possibly the growth of the individual within the mother is helped really to you know go into a different mode of you know development that is what you call as a placentacion view parity right. So these are some of the points that that really explains.

(Refer Slide Time: 35:57)

And what is again what is interesting is then if that is the case in a platypus has got venom you know this is a tip of an adult platypus spur somewhere close with the hand limb. The lower part of the leg so you have this spur which can eject you know venom which can be very very toxic and what they have done is an interesting study.

(Refer Slide Time: 36:21)

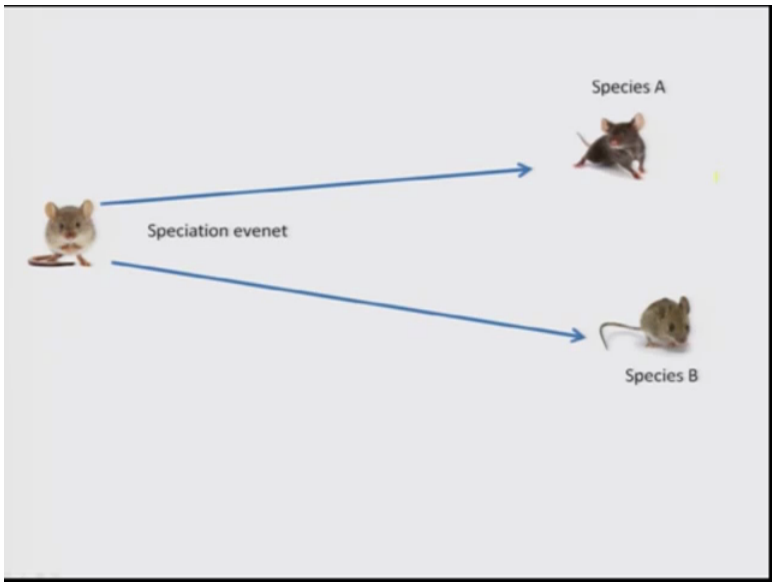


They have looked at the evolution of the gene that coades for the venom, its interesting is that you know there are different lineage and all these venoms are protein nothing but these are beta

defencing what they call lineage group of proteins. It is not that it is closer to you know the reptiles because they have you know you have the snakes and they have this venom again this proteins are more closer to the mammals and we also have these you know same protein, differencing which the which which the platypus uses a toxin.

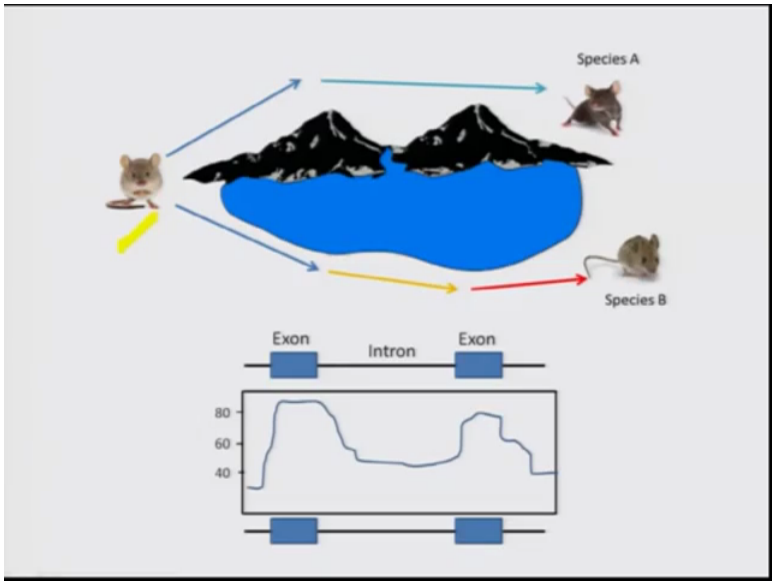
But here it is much milder in our body which is secreted as I told you before in the intestinal lining and in the skin which you know helps us to prevent bacterial infection and so on. So it is only the you know the way the protein is evolved in a given species made it more toxic whereas here it is lesser toxic so that is what it is but it is it is a what is called as a parallel evolution it is not something that that they acquired from the linear that led to the reptiles.

(Refer Slide Time: 37:27)



Now let us look into how the evolution takes place ok the speciation so how is that happening so let us say that this was this is species A and species B right and these are the contemporary species that you see today and probably they evolved from a common ancestor that is show here right.

(Refer Slide Time: 37:55)



So first of all the question is why should they become two different species right. There are several possible events that leads to new formation of new species but I am just showing you one. One could be geographical isolation so these species now they migrated say on either side of this the Killiland lake now this particular species now they remain here for long time so they are exposed to very different kind of environment likewise this species remain here for long time they exposed to different kind of environment as a result there are different changes in the genome that are selected to make them successful here as compared to the other species.

So over time if they are not breeding together therefore there is no gene flow then each would acquire a unique set of genomic signatures now they cannot you know breed have F1 therefore they become you know individual species that is what we talk about. So if you now look into these species in the genome we can understand now. So this is what it is they are together you know this is a ancestor they both belong to the same lineage but evolved over time now you are looking at this is say for example species A and species B and we are looking at a given gene and these are the exons two exons that are there we are comparing.

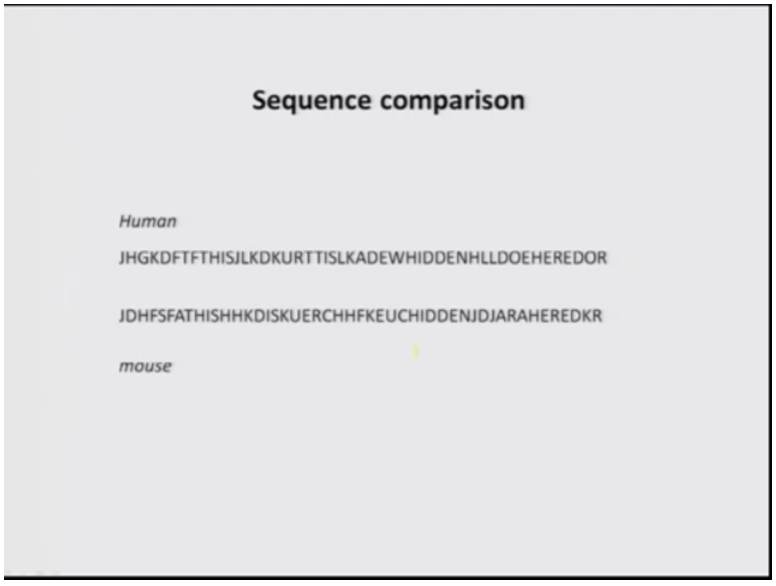
Now what is shown here is a person similarity, so you can see here when it comes to the exonic regions you have a higher similarity but it in intron again you have exon high similarity it goes down what does it mean. It means that in the intronic region you know you know with whatever number of generation in between before they became individual species but lot of changes that

happen and there is change is did not affect the gene or probably it helped in the species to survive in that particular habitat right.

Therefore it diversified but the coding region still have higher similarity because there is a pressure on what kind of amino acid that gene should code for so the changes are not allowed or they are not selected therefore you do not see in this population. Let us look into something else say suppose there is a region here right that region is very critical for the gene function. It could be the an enhancer remember we said that that there are elements within the gene on which some proteins comes and bind and then activate the transcription, help in the activation of the transcription.

Now there is an element and that element is certain DNA sequence right and if the sequence is altered the protein cannot come and bind therefore there is a pressure that that sequence is not altered so if you all of a sudden you find two exons which are very highly conserved plus there is a region that is highly conserved that immediately tells you that this is a regulatory sequence. It has some specific function so this is how we will try to look into genome and that is what it is.

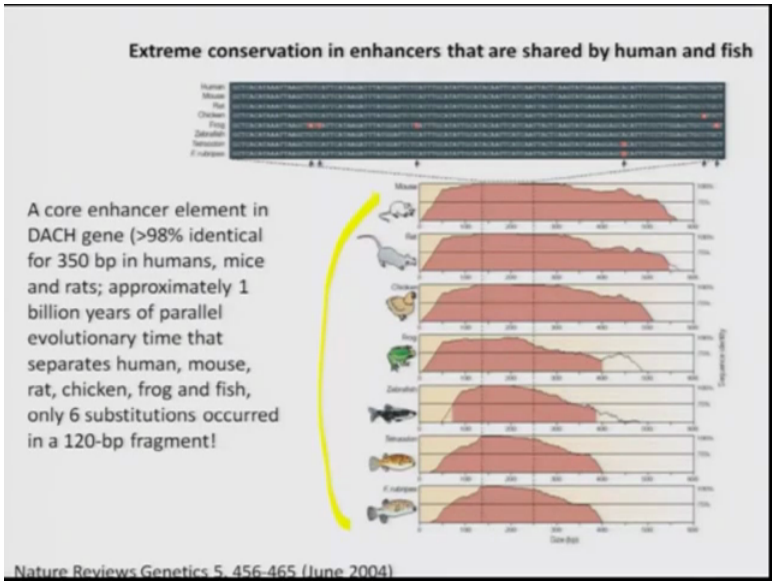
(Refer Slide Time: 40:48)



So basically you look into the sequence if you look for example human and mouse the sequence look like this and you may not really make any you know difference but you otherwise very carefully you will find certain signatures that are identical in these two sequence which tells

these are the reasons possibly there is some functional significance. Therefore they did not change much right. So that would help you to understand further, so I am giving you some examples this is a gene called DACH lets not worry about what is the gene but what is interesting here is that we are comparing from fish to mouse.

(Refer Slide Time: 41:52)

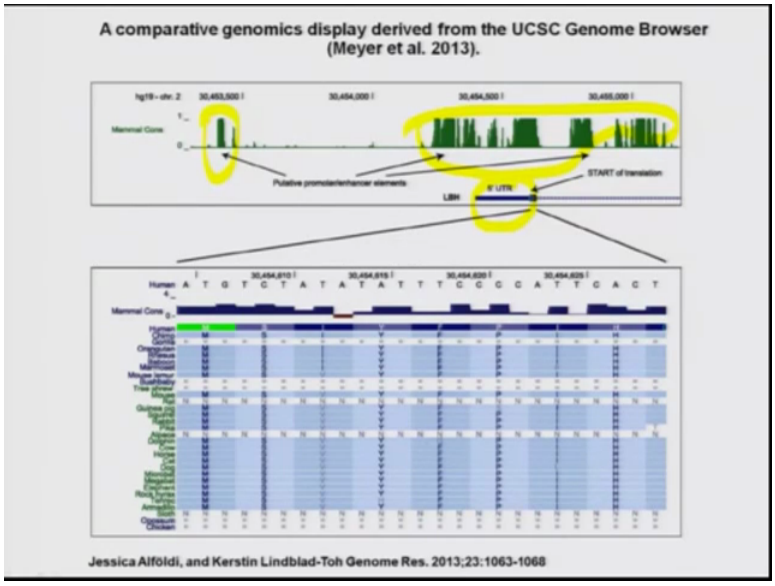


A small segment of the gene the core enhancer element ok its present in one of the intronic region and this is not coding for any you know protein therefore you know there is no constraints on protein sequence but it is purely the basic one. Since which is 98 percent identical for 350 base per in humans, mice and rats approximately 1 billion of parallel evolutionary time that separates humans and other fish you find that this 220 base per fragment remained frozen they did not change at all, that clearly tells this sequence is so critical for the gene function. If there is any change that would not allow the organism to survive so you never see it in the population so always you find this sequence to be present which clearly tells that it is a critical region not coding for any protein but still it is very critical.

So this you would not have understood if you have not compared this many you know genome sequences that again gives you glimpse as to how comparative genomics can help you. So you can go and look at you know I am sure you would have seen couple of lectures by (())(42:30) where they have given you some introduction on how to use various Bio-informatics tools you can go on analyse the sequence, you should go on try it out this is one of them is the University

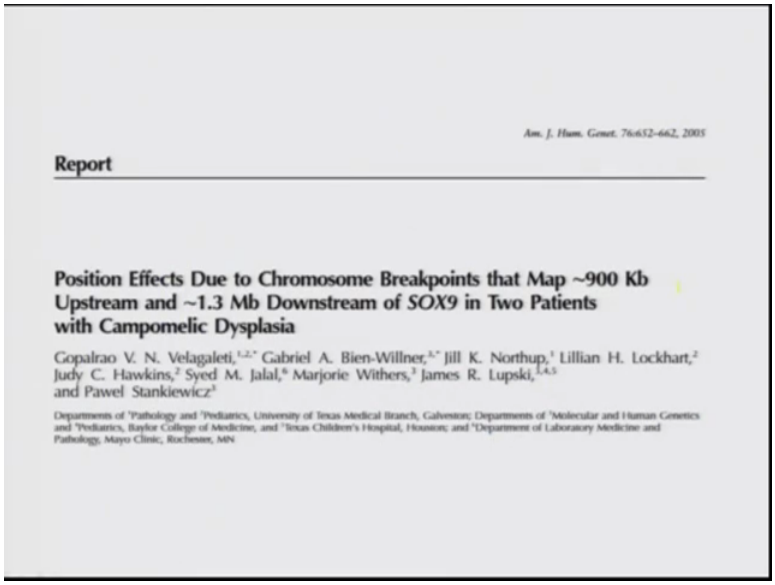
of California genome browser we have introduced many a times in this lecture. So you can go on and look at any region so again I am giving you one particular gene sequence right.

(Refer Slide Time: 43:09)



So you can see this is a UTR right un-translated region and we are finding it you know there are regions that are highly conserved right and these are the regions that are you know regulated some proteins comes and bind you know so that you can see that you can compare with a large number of species you find that that region is present right the sequence we can see here it is very very very critical for function. So this is information that you can get by comparing the sequence right.

(Refer Slide Time: 43:24)



Am. J. Hum. Genet. 76:652–662, 2005

Report

Position Effects Due to Chromosome Breakpoints that Map ~900 Kb Upstream and ~1.3 Mb Downstream of *SOX9* in Two Patients with Campomelic Dysplasia

Gopalrao V. N. Velagaleti,[1,2,*] Gabriel A. Bien-Willner,[3,*] Jill K. Northup,[1] Lillian H. Lockhart,[2] Judy C. Hawkins,[2] Syed M. Jalal,[4] Marjorie Withers,[3] James R. Lupski,[3,4,5] and Pawel Stankiewicz[3]

Departments of [1]Pathology and [2]Pediatrics, University of Texas Medical Branch, Galveston; Departments of [3]Molecular and Human Genetics and [4]Pediatrics, Baylor College of Medicine, and [5]Texas Children's Hospital, Houston; and [6]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN

So we will stop here and then the next lecture we will see few more example as to how you test this hypothesis that is if you have identified an element which you believe is a regulatory element because its conserved how will you test whether indeed it is doing that function, see you again.