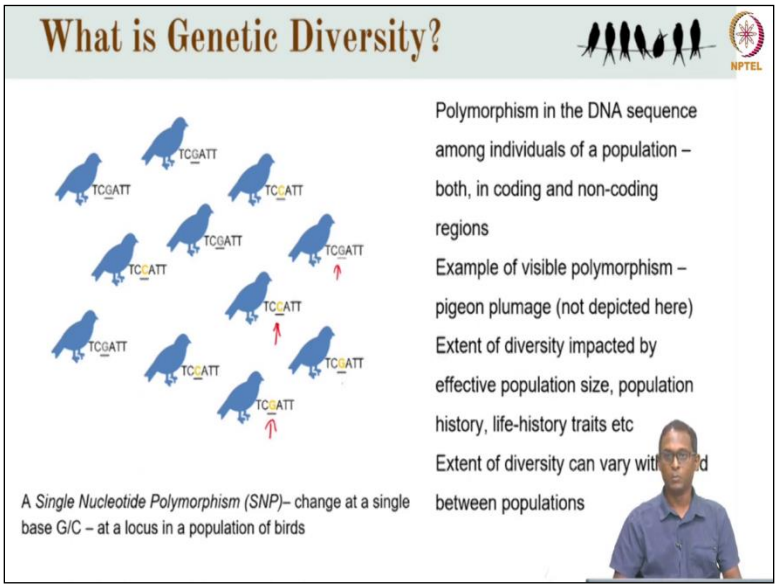
Basic Course in Ornithology
Dr VV Robin
Indian Institutes of Science Education and Research-Tirupati

Lecture 41
Molecular Techniques Part 2

Now, we go into different applications. What are the different applications? What can one do with this genetic data? How do you understand birds and the biology of birds once you have such data? and that is what we are going to focus on here.

(Refer Slide Time: 00:38)



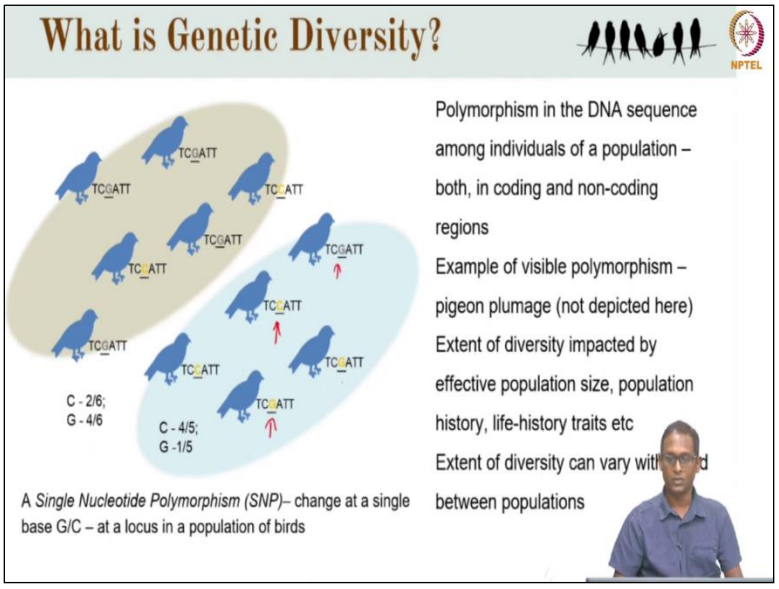
So, let us start by understanding what is genetic diversity? So, here is a set of different imaginary birds and these birds all have different sequences and now we know how we get these sequences if you remember the Sanger sequencing process. And you note that in all of them, they differ from each other at one single location at least that is the one that we are indicating to here right.

That one, that one, that one and that is essentially changing between different individuals, it could either be a G or a C, this is just an example. But such a location one point where there is a change this is called single nucleotide polymorphism or SNP in simple in short. So, at a single base locus there is there is a change in what base occurs there. And now you can use this data in various ways and we will talk about that.

So, this is at the heart of it if you can understand that there are SNPs how do these SNPs you know how do you interpret the data of these SNPs is essentially what we are going to discuss further. So, this polymorphism that exists it could be encoding or non-coding regions and some of it may be visual like you know there may be a morphological difference or a plumage difference but that is not necessarily true and it is not depicted in this example.

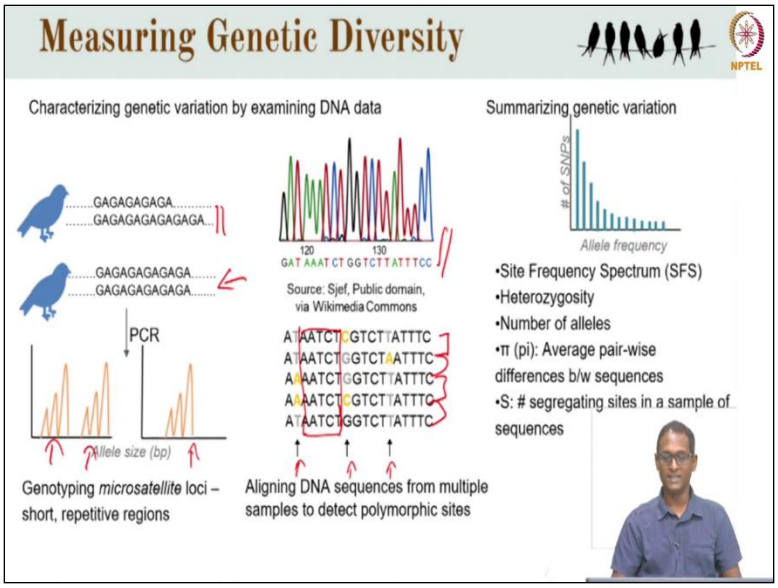
And this diversity that you see it can be impacted by various things including effective population size, how big the population is some kind of population history and life history traits and so on. And the extent of the diversity varies both within the population and between populations.

(Refer Slide Time: 02:53)



And as you notice in this example itself, you will see that these populations has a lot more C and this one has a lot more G. So, possibly the population these populations are kind of segregated separated from each other in some ways right.

(Refer Slide Time: 03:18)



So, that is that is at the at the core of it. Now you can also look at genetic variation in a different kind of genetic data what is called microsatellites. Microsatellites are nothing but repetitive regions. So, you have G A G A which is you know GAGA GAGA or. So, you know that that repeats. So, I will point to this example which is a heterozygote because it has different length of GAGA and that is represented here. So, that is 1 and that is 2.

On the other hand, this example has you know they are uniform. So, that is a homozygote. So, that gets represented here. So, you could just look at alleles this way and you could get some idea of genetic variation by looking at heterozygosity and the length of these repeats. So, you will find a lot of papers using microsatellite loci and that is how that works. You could also use the Sanger sequencing data that we talked about remember the you know the blues and the reds and the greens and so on.

They translate to different bases, this is for one specific example one individual one DNA. But you could put together such data for different individuals. So, let us say that you have a set of five individuals in this example and now once you align all of these data you notice that there are three points you know three SNPs if you remember the idea of SNPs that kind of differentiate all of these individuals.

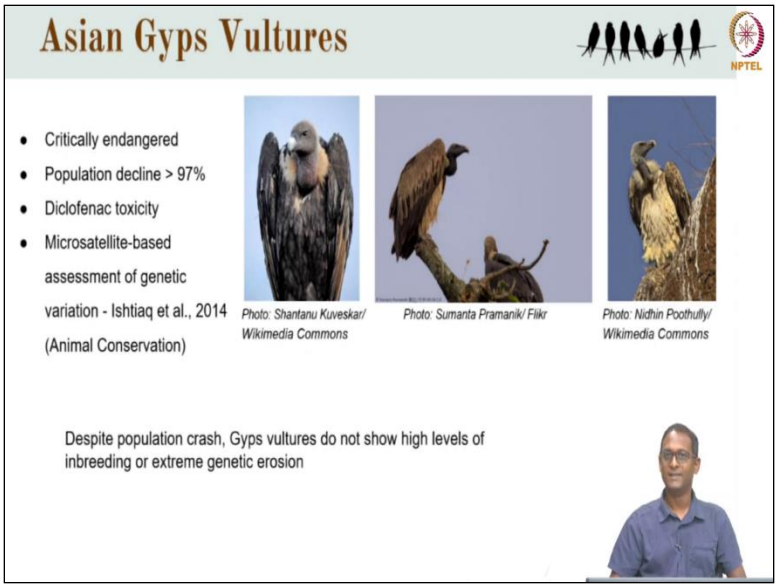
Now this is the valuable part of generating such data and what we could do with that is you could create you could summarize this genetic variation in something called the site frequency spectrum SFS for short. What SFS does is that it looks at the allele frequency and number of SNPs that represent that frequency. So, for example here you have two of these which differ in two. So, that is 2, 2 and there is 1 that differs in 1. So, that is 1, 1. So, that would go out here.

And you can look at a similar site frequency spectrum and understand how alleles are distributed across this population. It gives you an idea of heterozygosity, the numbers of alleles. You also you can also look at average pairwise distances. The pairwise distances are nothing but you know if you were to look at the number of distances in across each of these pairs and then you average them out, that is the average pairwise distances.

So, in these two there are you know two differences C to G and T to A. So, similarly you need to compute that for each pair and average it out between each of those sequences. So, that is a very intuitive thing that people have done they can then go on to create trees based on that data. Number of segregating sites that is this. So, is the same as the SNPs. So, in this data set there are three segregating sites and this is the informative part of the data.

So, all the other data that you have you know this whole section actually that does not count much. What you want is the segregating side. So, your data is really when something varies and that is what is used in you know all further analysis, whether it is phylogenetic building trees or looking at population level differences that is the goal you want to get to these segregating sites the SNPs all right.

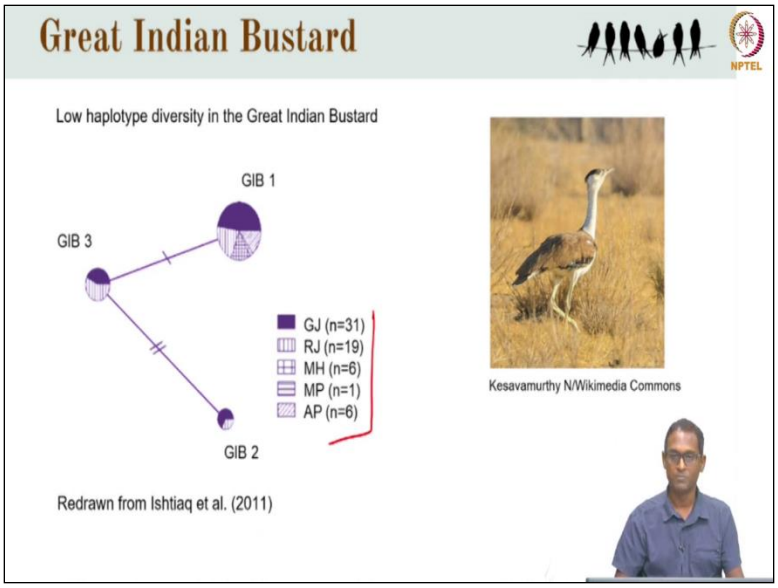
(Refer Slide Time: 07:53)



So, let us look at some examples a famous one is the Asian gypsy vultures. Everybody knows they're critically endangered population decline is greater than 97%. The diclofenac toxicity is one of the known causes and there was a micro satellite based assessment of this by Ishtiaq et al. So, despite the population crash the gyps do not show high levels of inbreeding or genetic erosion.

But this has to be taken with some caveats one is that the this is not entirely done in the wild these were from some birds that were found in captivity from maybe those birds were captured before the effect of this decline happened (I mean this is a possibility), studies of wild populations are ongoing. So, we will know about subsequent population differentiation with more genetic data in some time to come.

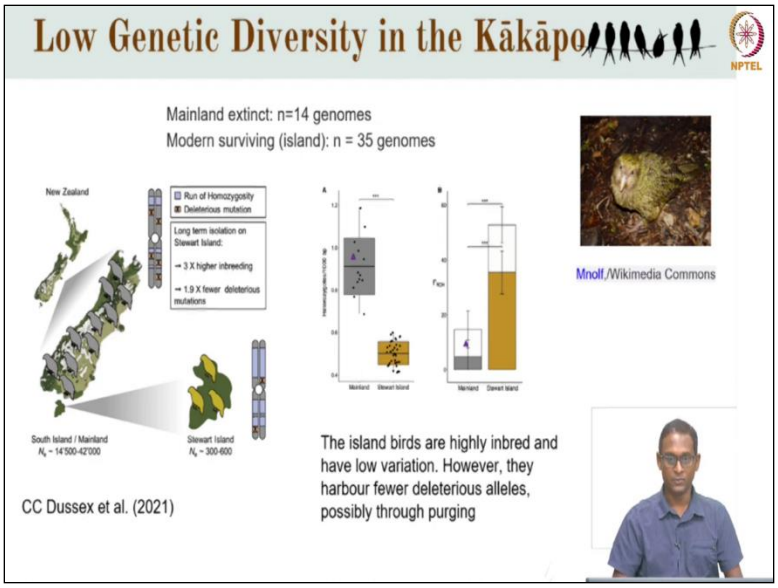
(Refer Slide Time: 09:04)



Great Indian Bustard this is another species that we know is is an important conservation species and in 2011 there was some genetic data from a lot of samples actually they had a lot of samples from all across its distribution and they found very few haplotypes. And these haplotypes are so, haplotype is basically so, if you say that Gujarat has a certain kind of haplotype and then you have Rajasthan which is this uh.

So, this shows a network of these haplotypes uh. So, there are very few haplotypes across the whole landscape which is surprising. But again, this data had limited you know genetic sequence data. So, more data would pro probably provide you a little bit more information and that I am sure will happen soon.

(Refer Slide Time: 10:07)

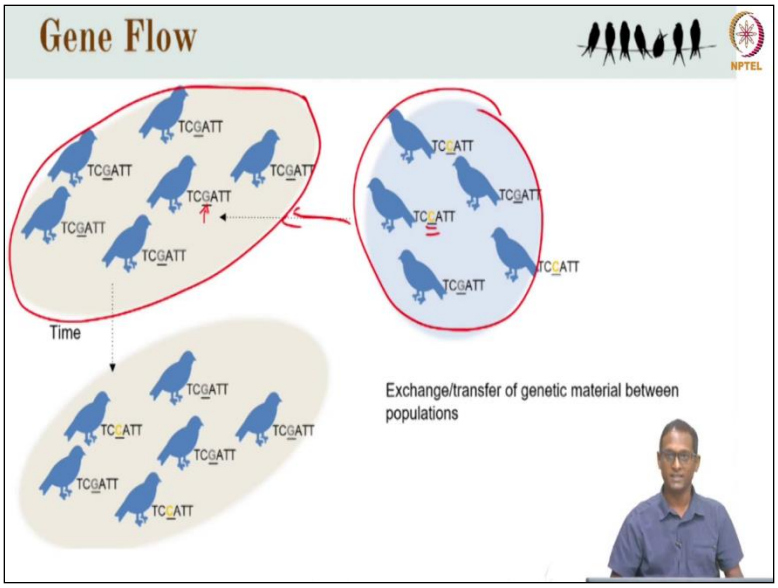


A third example that I am going to give you is about Kakapo again it is a very famous bird, in mainland New Zealand it had kind of gone extinct. And it was actually survived if I remember right maybe by one male or so on but very few individuals were remaining even if the exact numbers are unclear at least to me. But the main point here is that these researchers, they got genomes white data for 14 individuals that have gone extinct actually on the mainland.

And then of a lot of surviving individuals on the island on Stewart island and what they find is that there are effects of inbreeding this aspect called ROH which is Runs Of Homozygosity that is more in the isolated population which is highly inbred. But it never there was not as much deleterious mutation that they found in these Stewart island population. So, one possibility is something called purging where deleterious alleles are kind of purged out of the genome but that is just speculation.

But the point is that there are effects of inbreeding, there are other effects also (I think with that relate to the reproductive fitness itself). So, you could quantify genetic diversity in various ways and it has very significant effect in terms of conservation.

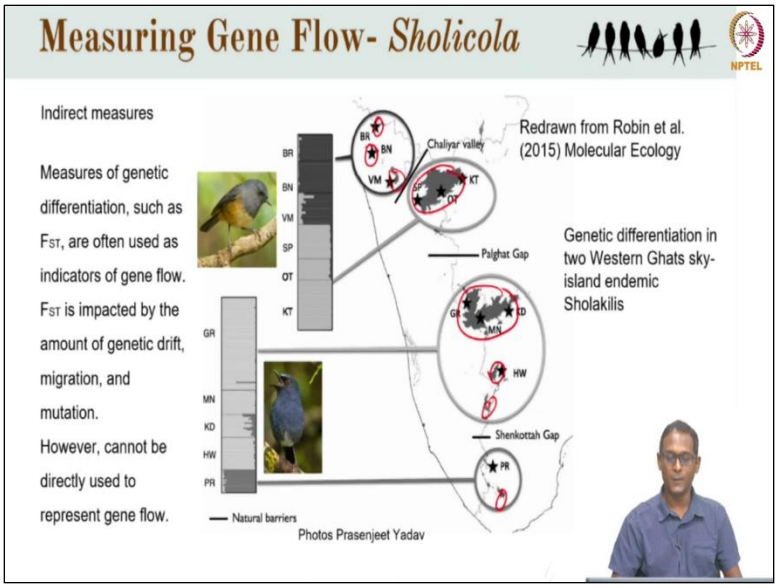
(Refer Slide Time: 11:58)



We also should you know keep in mind that you can look at this genetic data to examine gene flow. So, you want to keep your focus on this population here. So, if you consider this population, this has you remember we talked about this G and C variation. So, let us consider this population having just the G base at this spot and then there is some exchange or transfer of genetic material from another population this one into this.

And you will see that this population has some individuals with C in it. Now what that means is that then over time due to this mixing, you would have a population which has C and G. in this. Now, lot of methodological work goes into quantifying something like this that happens.

(Refer Slide Time: 13:05)



One example that I am happy to talk about is of the Sholicola which is a bird that I have studied for the last 20 years or so. Here, what we do is we let me explain first the graph and then I will go on to give you a little bit more detail. What you see here are these. So, this is the Western Ghats and these are mountain tops you know with isolated populations of the of the Sholicola and what we did was we collected blood samples from all of these birds from different mountains.

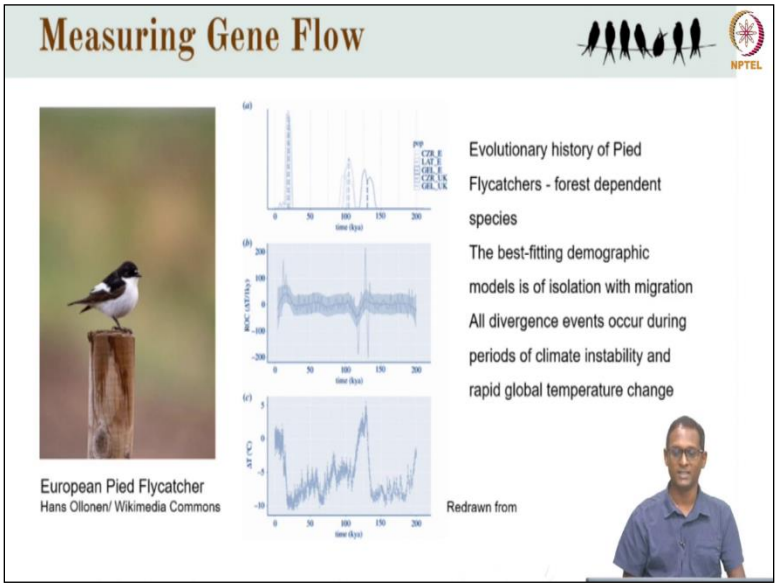
And we were we then used microsatellite data like the one that we described earlier to understand how alleles are shared between these populations and you can see here that there is.. so this is the north yeah all of these are the north and of course we know that Palghat gap is a big barrier. So, there is almost you know five million years divergence between the two. So, that is why they are represented in different plots altogether.

And you can see that across the Chaliyar valley which is the VM, you can see some some similarity some gene flow from the Nilgiris here and so, that is that is interesting. So, we suggested that these are one unit based on based on this data and also phylogenetic data actually. And so but there is some gene flow there and you can also look at FST which is an indirect measure.

And it is affected by amount of genetic drift, migration and mutation but it does not by itself it cannot be used to quantify gene flow. And then this is the Shenkottah gap and you can see that a similar differentiation across this. So, you find essentially that you can look at gene flow across

different populations once you have genetic data. How many of these alleles are shared between populations.

(Refer Slide Time: 15:38)

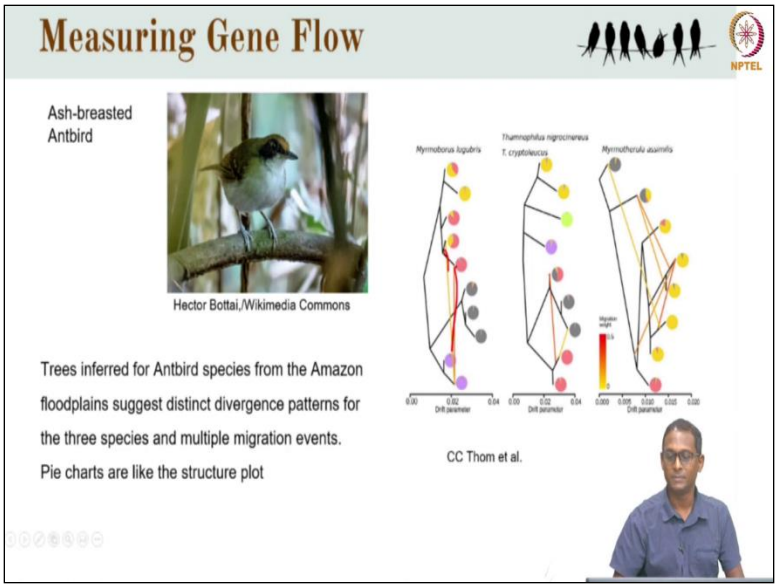


So, once you have genetic data like the kinds of data that we talked about earlier. You can use it to understand a lot more about the species ecology. For example, the European Pied Flycatcher, it you know these authors used it to understand the evolutionary history of the species. It is a fairly forest dependent species and you can you can look at different demographic models about how the population historic populations change with time.

And what they find is that at times of great climatic instability and global temperature change like here, you find changes fluctuations in the population itself. So, once you have data like this, you can then use it to understand other evolutionary aspects of different bird species. For example, in the European Pied Flycatcher these authors used genetic data to understand how these forest dependent bird species have actually been impacted over time with historic climatic reconstructions as well.

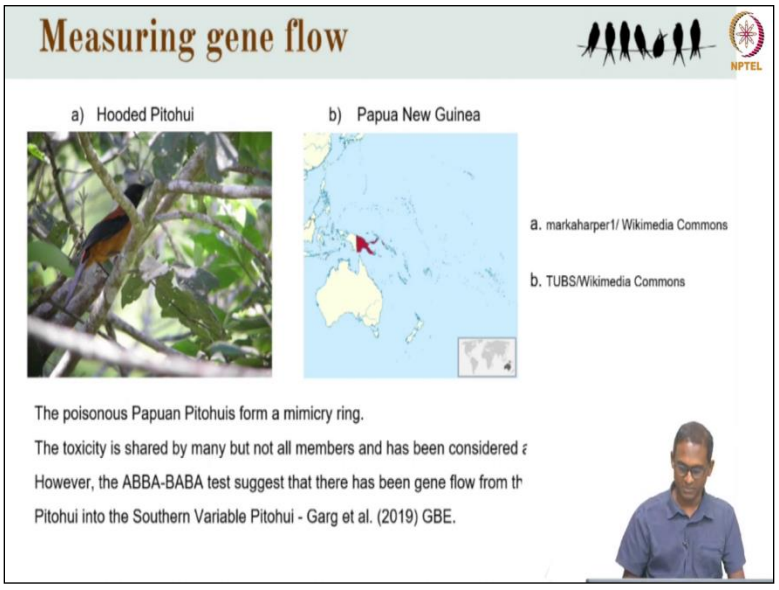
And they find that there are the divergences occur around the periods of climatic instability where the global temperature changed.

(Refer Slide Time: 17:11)



Another way to kind of another example of people studying gene flow this one is of Antbirds. For multiple species of Antbirds across the Amazon flood plain they find that for three species they find a different migration event which is shown in these lines. So, these are like the structure plots except that you find some alleles being shared across different populations because of migration.

(Refer Slide Time: 17:48)

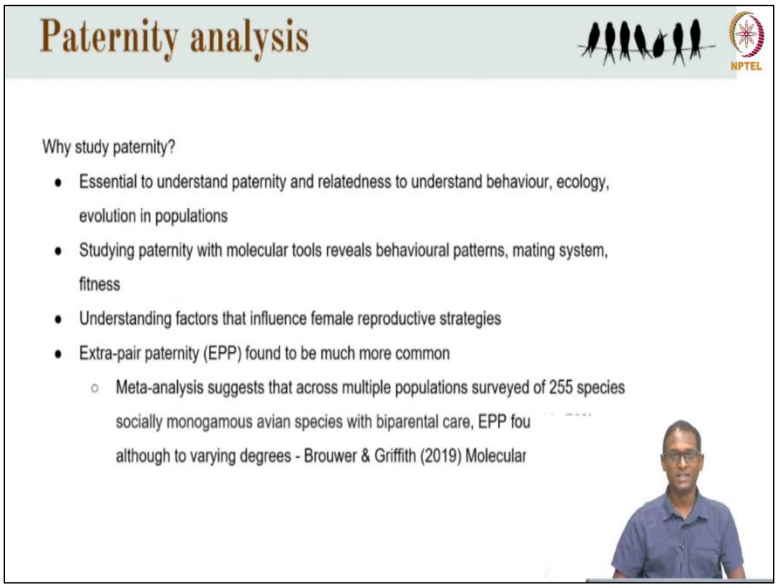


Another way another example and this one is really fascinating many of you may know about the poisonous Pitohui which is found in the Papua New Guinea area. The interesting natural history part is that these birds are thought to be poisonous but this toxicity is not is not there in all the members of the Pitohui. So, some have it and some do not and it is thought to be considered

ancestral. So, these researchers Kritika Garg et al did a very interesting test which suggests that there is a gene flow from the Hooded Pitohui into the Southern Variable Pitohui.

So, essentially you can use it to look for you know some kind of gene flow which can then answer questions about other ecological evolutionary ideas in this case it was about the toxicity all right.

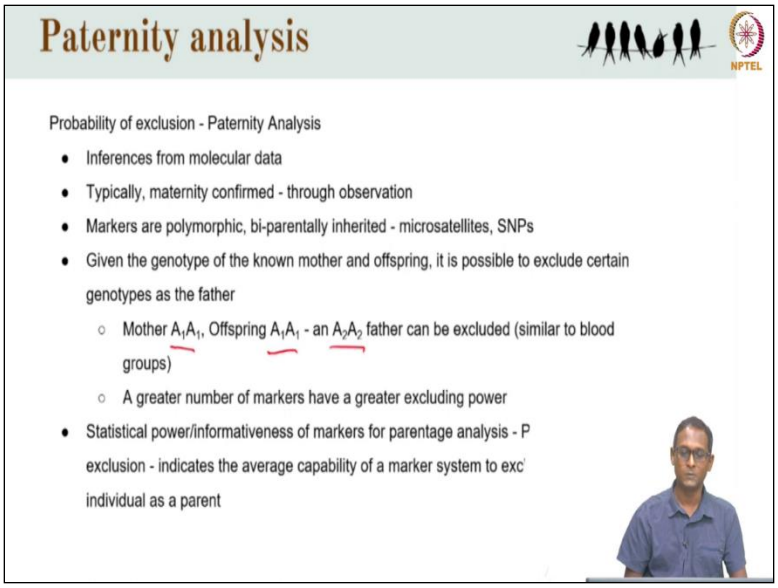
(Refer Slide Time: 18:52)



So, that was that was a lot about gene flow. Now, we move on to another idea which is on paternity. This is particularly useful for people who are working on various kinds of behaviours of species you want to understand relatedness between individuals. And this has been particularly revealing with molecular tools coming in. Because you want to understand mating systems and factors that kind of influence female reproductive strategy and so on.

Here, I want to note that you know as our techniques and the tools of the trade keep improving, our understanding of bird biology itself has changed quite a lot. So, from 1960s 70s when birds were thought to be largely monogamous, I think there was a reference to about 90 odd percentage of birds being monogamous. However, today data suggests that extra pair paternity is found in over 76% of the species to very degrees. So, most birds are actually not thought to be monogamous.

(Refer Slide Time: 20:24)

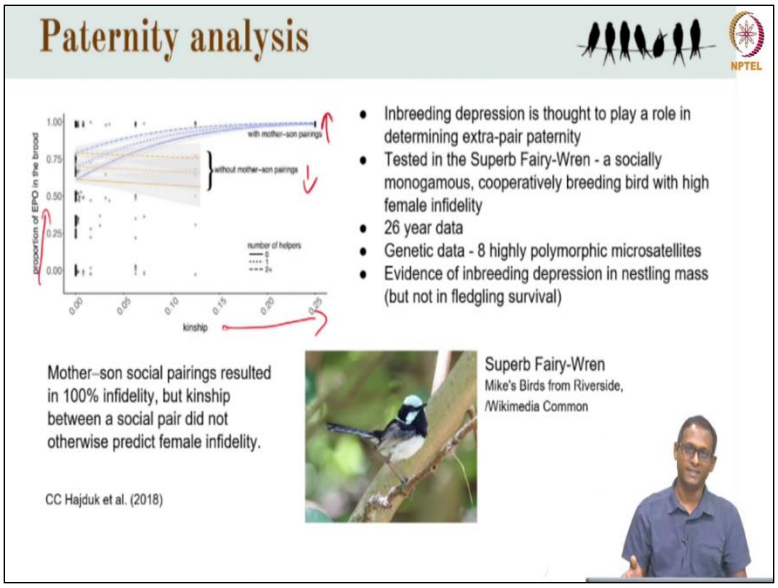


So, how do they find out things like this how do they get this kind of data with molecular tools. So, the overall idea of paternity analysis is by looking at this probability of exclusion okay and I will explain that. So, the inferences are from molecular data. So, you need SNPs data micro satellite data or some such and typically the maternity is confirmed through observation. So, that is usually you know who is the mother? It is you know it is with the young or something like that uh.

So, once that is confirmed and then you have markers which is the genetic data like microsatellites and SNPs and you know the genotype of the mother you know the genotype of the offspring and then you can exclude certain uh individual certain males as being the father. For example if you have A1 A1 the mother and A1 A1 as the offspring. So, you can immediately exclude any male that is A2 A2 because you know that you know that is not possible.

So, this is very similar to the blood grouping exclusions that are done but what you have to remember is that this is just an example with one set of alleles, what you need is large number of markers because then you have a greater you know power to exclude different individuals. So, that statistical power comes from how informative these markers are of excluding different individuals as a parent.

(Refer Slide Time: 22:13)

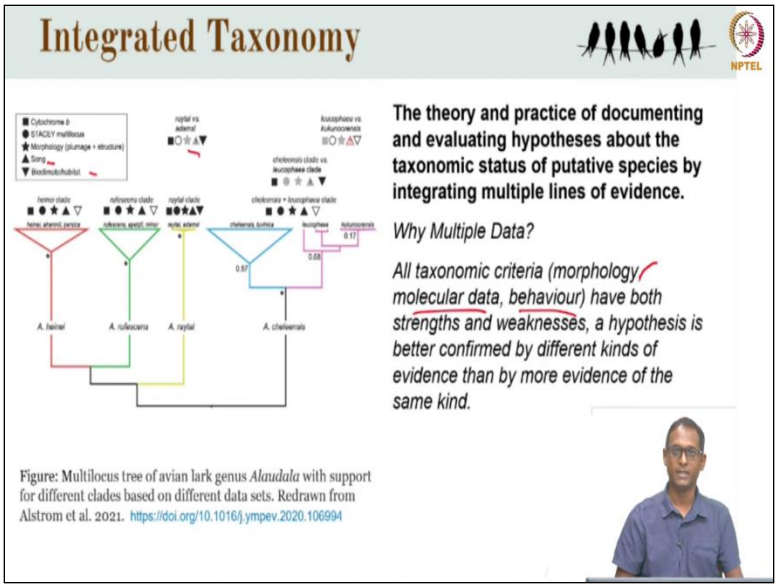


And I do not want to leave you with a dry kind of slide here is an example of a bird the Superb Fairy Wren and this is a really fascinating example and these birds are known to be socially monogamous. So, which means that if you were just observing individuals, you would find them together that is social monogamy. But they are thought to have high female infidelity and these authors collected data from 26 years.

And they had these eight polymorphic microsatellite markers and what they found was that of course you know one other thing was that there were a lot of mother-son social pairings. So, there were pairs which were mother-son but almost all the mother-son pairings resulted in 100% infidelity. So, so the way to read this graph is that kinship among parents as the as you are more and more related to your pair your extra pair paternity goes up.

And they looked at pairs with mother-son which is much higher than without mother-son pairings. So, that is fascinating. So, essentially the more related you are to your socially paired individual, the more extra pair paternity that is there. Now, I mean we are not discussing why and how that how this works. But at least we know that the genetic data that you collect with molecular tools can be used to answer questions like this.

(Refer Slide Time: 24:08)



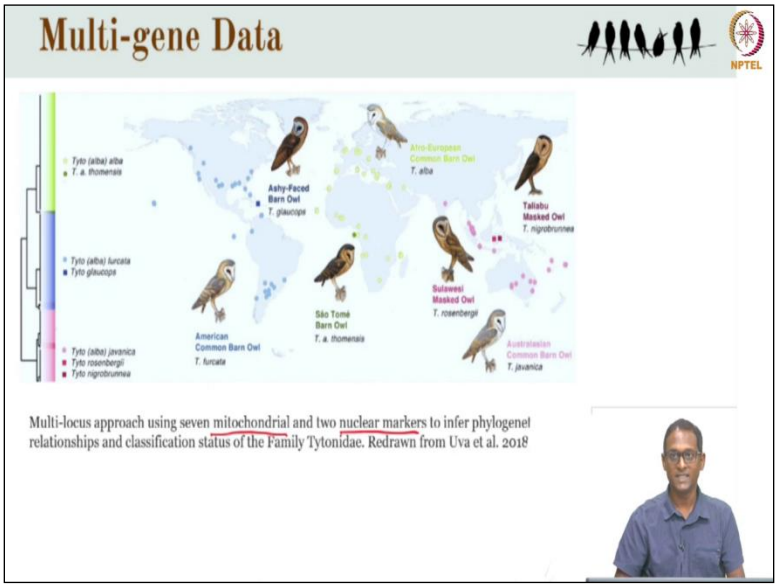
Now going on to taxonomy and biodiversity. So, how can we use genetic data molecular techniques to answer questions on taxonomy. Early days as we all know people used only a set of traits which is that you could look at plumage, you could look at you know bone structures or such and you would come up with a relationship between different organisms. Increasingly, what people suggest is that you need to have multiple types of data.

So, you that could be morphology it can be molecular data and it can be behaviour. For example, vocal behavior (vocalization itself) and this is a great example of Larks that were by Alstrom and group. And here what they show is that you know the gray shades are the ones which are which do not have support the filled dark shades are the ones that have support. And cytochrome B is genetic data there is morphology data which is plumage and structure.

There is song data which is the triangle and there is bioclimate or habitat data. Essentially, what I am trying to say is that in this i data set where they looked at various kinds of information some are consistent. So, genetic data may show a certain pattern but the plumage may not and song may not or may do that. So, how you reconcile all of these different lines of evidence is the new integrated taxonomy and taxonomic approach that most people seem to favour today.

So, gone are the days when you could just have a morphology based taxonomy or if you could just have vocalization based taxonomy or just based on DNA. Increasingly, people want an integrated taxonomic approach before putting out ideas of what forms a species or not.

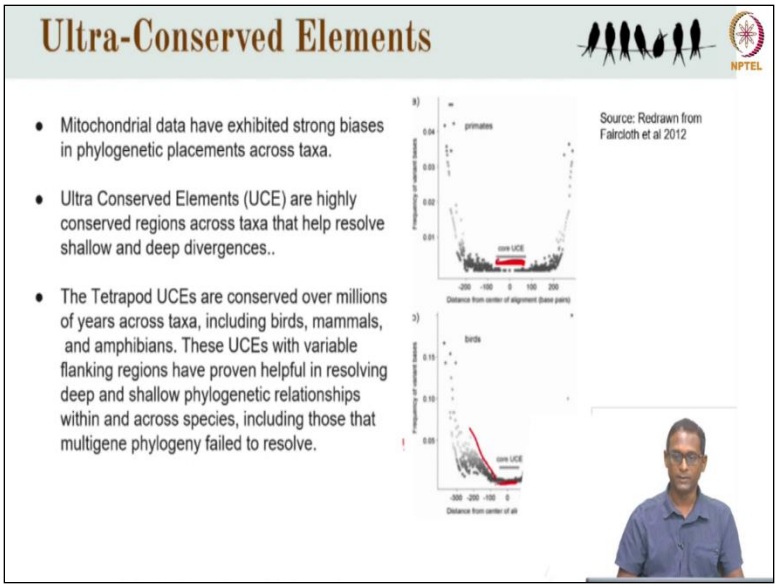
(Refer Slide Time: 26:30)



Multi-gene data this is this is just if you were to have genetic data. So, I am kind of going deeper into genetic data part. You want to have several lines of evidence because just having a few genes may not be sufficient. You want different kinds of genes, you want let us say, mitochondrial data like this study had seven mitochondrial genes and two nuclear markers and they then went on to describe relationship between the Tyto Owls right.

So, this is this is increasingly seen to be the robust way to do uh you know to assess relationships between different organisms today.

(Refer Slide Time: 27:20)

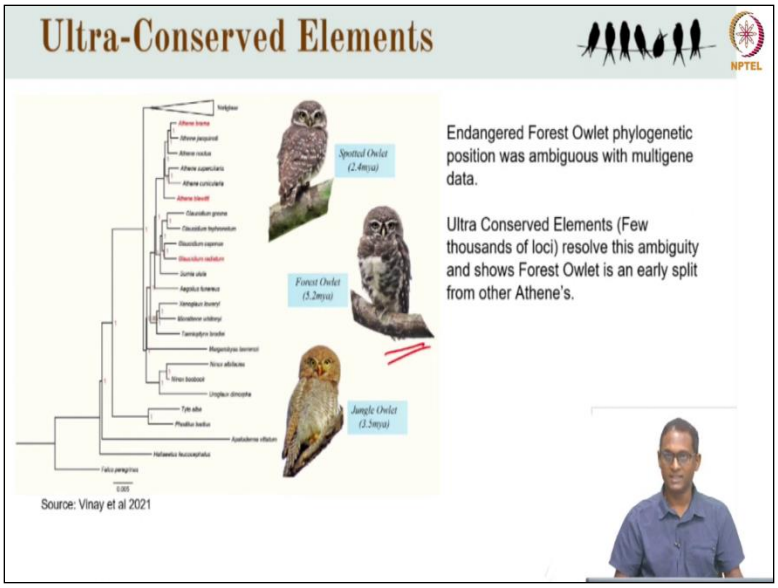


And here, we come back to Ultra Conserved Elements, if you remember in our previous section, we talked about UCE - Ultra Conserved Elements. This is another representation of how the UCEs kind of work. You have the core of the UCE which is where the.. if you remember from the previous slides you know that is where the capture primers are created the probes are created and those on the y-axis here is the frequency of variant bases.

So, as you can see that is not very variable the core is highly invariable which is why the capture works for multiple organisms, but as you go away from the core the variability kind of increases. So, the number of variant bases increases, which means that you get more and more SNPs as you go further away. And this is the idea that that people have used to understand relationships between different birds.

And with birds, a lot of birds have shown very fascinating results with UCEs. The point is that you can get thousands of loci, you can have that they have a probe set with 5000 loci and 3500 loci and you could use any of those for looking at relationships.

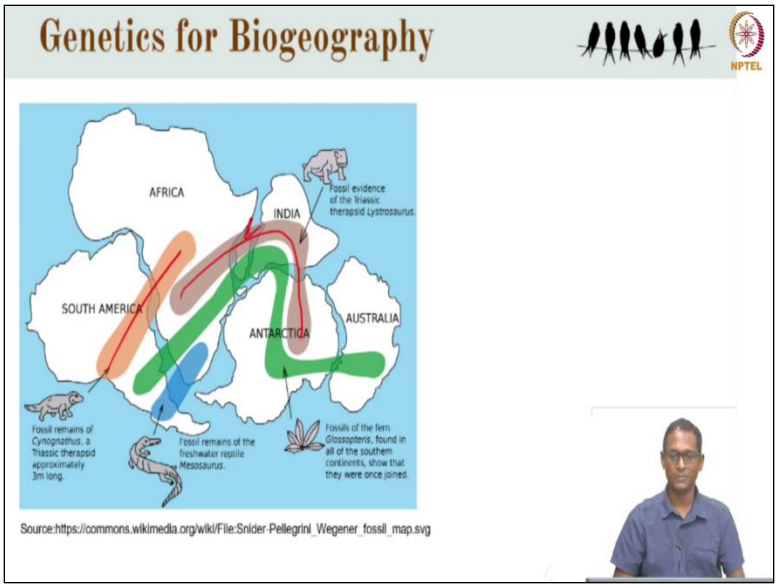
(Refer Slide Time: 28:48)



Here is a example again from my lab, Vinay created this data and this was in collaboration with several other authors. The point relevant for this example is that we wanted to look at the Forest Owlet and the previous iterations which used multi-gene data it was kind of ambivalent where Athene, the forest owlet kind of you know where it was with the other Athene.

What was its relationship with other Athene. And once we had this nearly 4000 loci data, what we can say more or less you know confirm is that this is an early split from the other Athene. So, it is a kind of early lineage with split from the rest of the Athene and we still suggest that it be retained as Athene as a member of the ethnic group to show its evolutionary relationship. So, this is probably you know one example of birds with UCE data from India which kind of tells us a little bit about the species that it is a member of the genus Athene unlike it was thought earlier.

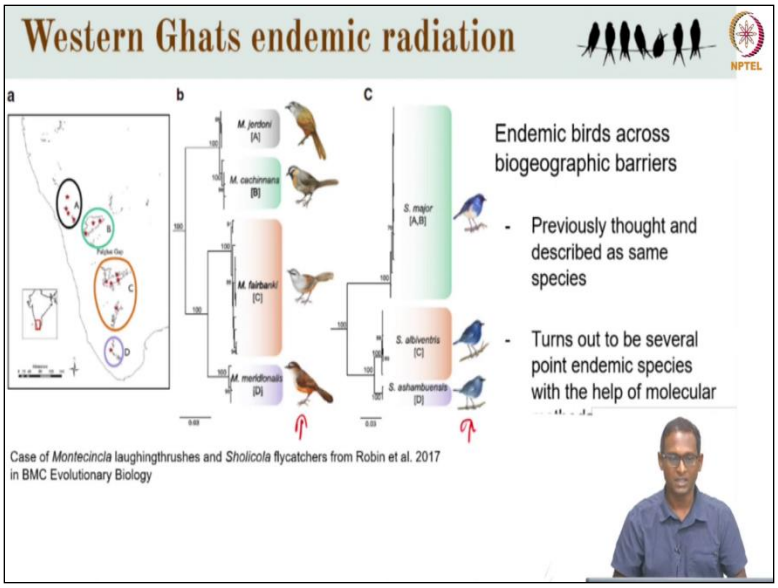
(Refer Slide Time: 30:20)



You can also use genetic data for understanding biogeography and by biogeography you mean the relationship between organisms based on their genetic samples that you get from different areas. For example, you could sample between South America and Africa or across Africa, India and Antarctica. There may be some relationships that show up which can be that kind of show up with interesting genetic data.

One example is the Purple Frog which is thought to have ancestor in Madagascar and that would not have been easy to find out without genetic data. So, this was one of the early examples of use of genetic data to understand biogeography.

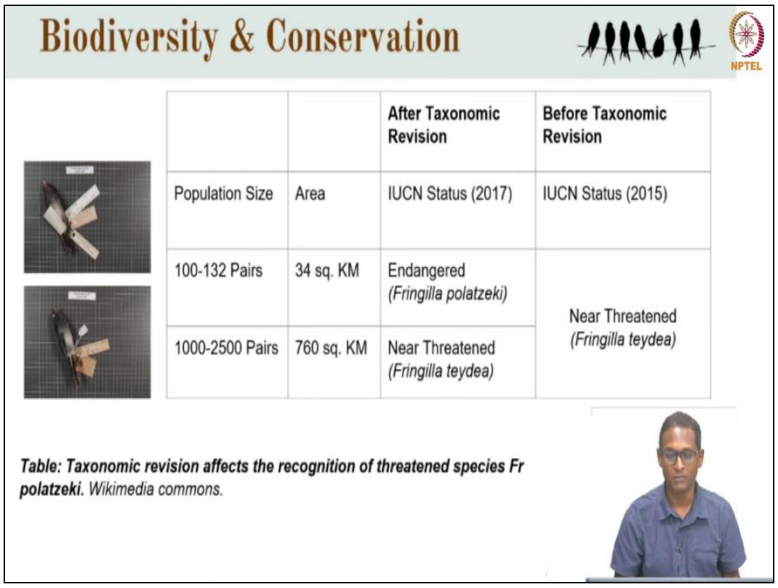
(Refer Slide Time: 31:18)



An example of birds is again from my lab, we used genetic data to understand an endemic radiation of birds in the Laughing Thrushes which are here these are the Western Ghat Laughing Thrashers or Montecincla and this is Sholicola found again only in the mountain tops. So, these are the mountain tops which is shown by A, B, C and D here. And our genetic data shows that some of these lineages are endemic to some of these mountains and we recommend them to be distinct species.

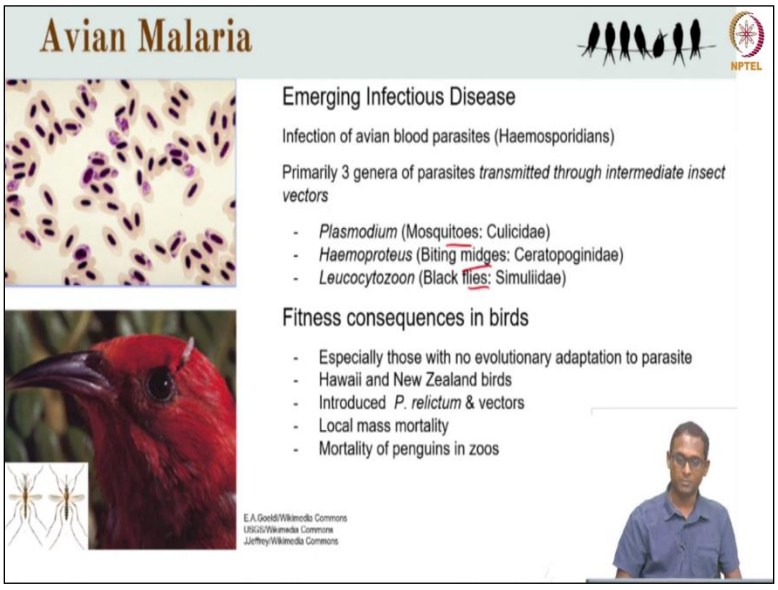
So, now we have several endemic species as opposed to a species that was thought to be just widespread uh. So, it provides information on true diversity you could say of a place. So, we went from having three species to having seven species just because we had the power of genetic data to help us understand a little bit more.

(Refer Slide Time: 32:28)



Biodiversity and conservation, now this is people have used various ways of looking at conservation with genetic data. Especially, with taxonomic revision just like the previous example, you know the area changes in the example of this species which is the *Fringilla* species, you find that before taxonomic revision, it was a near threatened species but because it was split up, it became one of them is endangered actually because it has just a 34 square kilometer area.

(Refer Slide Time: 33:08)

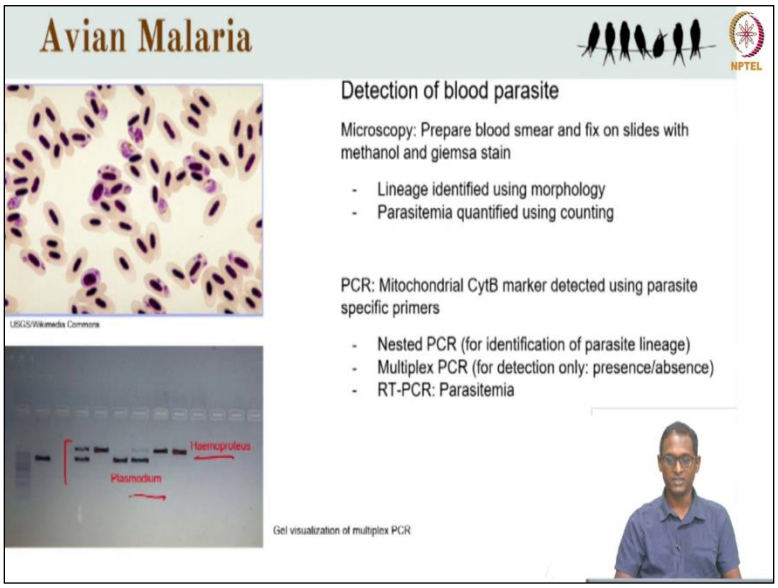


Molecular techniques can also be used to study disease. There are various emerging infections disease, avian malaria is one of them and this is primarily Haemosporidian which are of three genera *Plasmodium*, *Haemoproteus* and *Leukocytozoon* from different vectors, they can be

mosquitoes - Biting midges and black flies and it is important because it has a big consequence on birds.

So, in populations that are not used to parasites these parasites and when they are introduced there especially past *Plasmodium relictum* in New Zealand and Hawaii there was a lot of mortality local mass mortality that is seen.

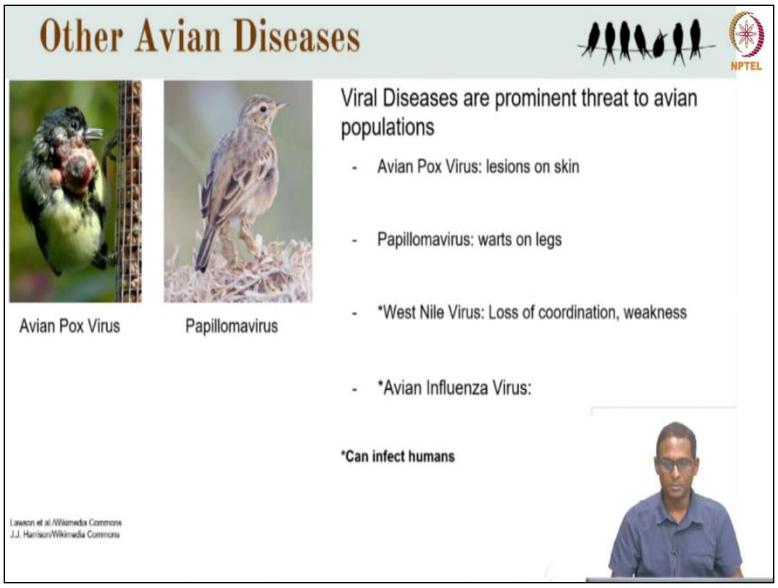
(Refer Slide Time: 34:04)



Now, you can detect this with microscopy which is the with the traditional way to do it. You can look at the morphology and you can identify lineages and you can quantify Parasitemia by counting the parasites. Alternatively, you could use PCR techniques, there are some primers that are designed and these depends on the purpose some are only for detection like multiplex primers which can tell you if it is present or absent. But others you can also sequence and get the sequence data and here is an example of *Haemoproteus* and *Plasmodium*.

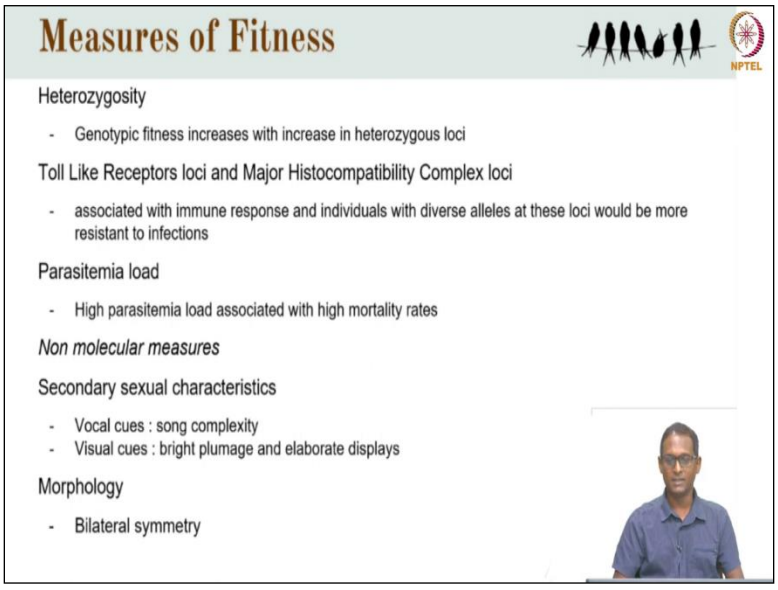
So, this is a primer that can work for both *Haemoproteus* shows a band at a specific size and Plasmodium shows it another. So, that is how you can visualize differences in different species. What is that species infected with and you will notice that some actually have both the parasites in in them in fact (multiple infections).

(Refer Slide Time: 35:10)



There are also other avian diseases that you can detect like Avian pox, other is Papillomavirus like you know which have some physical deformities, West Nile Virus and Avian Influenza as well and these two can of course infect humans. So, it is it is worth kind of being a little careful about that.

(Refer Slide Time: 35:35)

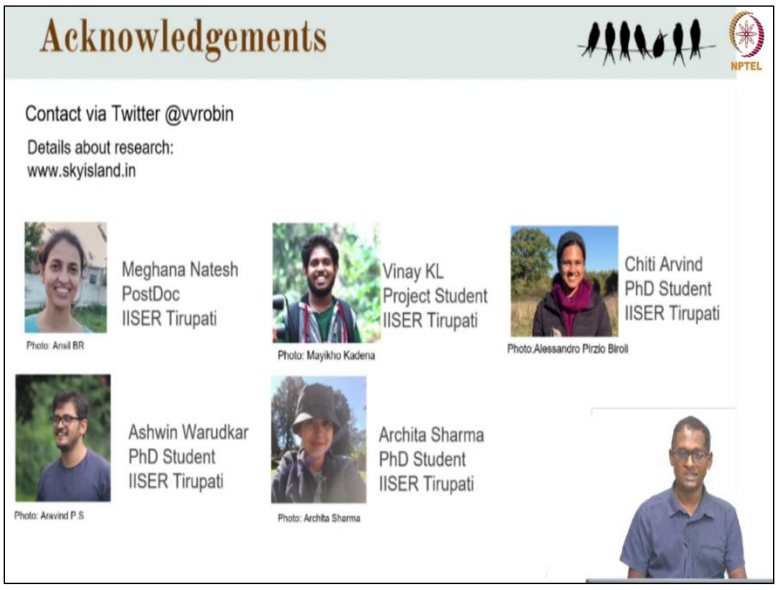


You can also use genetic data to as a measure of fitness. There are various measures that go there heterozygosity which is we discuss heterozygosity with the microsatellite alleles. So, genetic fitness increases with increase in heterozygous loci. There are other loci like the toll-like receptors,

MHC which is a Major Histocompatibility Complex which is thought to have some immune response effects.

Parasitemia loads of course can also reflect fitness. There are non-molecular measures as well it is not that everything has to be molecular like you could look at vocal cues like song complexity other visual cues like plumage, brightness, elaborate displays and so on. Morphology, like bilateral symmetry but essentially what we are trying to say here is that you can use molecular tools to aid other tools where you detect fitness.

(Refer Slide Time: 36:48)



Finally, I want to end by acknowledging a number of students who helped put together a lot of the data that you have seen here Postdoc Megna Natesh who is done a fantastic job with all the gene flow slides, Vinay is a project student, Chiti's PhD student and Ashwin and Archita are also PhD students working on different aspects using molecular tools. You can reach out to me via twitter @vvrobin and more details about the research is at the website indicated here, thank you.