Bones, Stones, and Genes: The Origin of Modern Humans
Lecture 2- Genetics of Human Origins and Adaptation
Sarah A. Tishkoff, Ph.D.

1. Start of Lecture 2 (0:00)

[Music]
[ANNOUNCER:] From the Howard Hughes Medical Institute... The 2011 Holiday Lectures on Science. This year's lectures, "Bones, Stones, and Genes: The Origin of Modern Humans," will be given by Dr. John Shea, Professor of Anthropology at Stony Brook University; Dr. Sarah Tishkoff, Professor of Genetics and Biology at the University of Pennsylvania; and Dr. Tim White, Professor of Integrative Biology at the University of California, Berkeley. The second lecture is titled "Genetics of Human Origins and Adaptation." And now, a brief video to introduce our lecturer, Dr. Sarah Tishkoff.

2. Profile of Dr. Sarah Tishkoff (1:11)

[Music]
[DR. TISHKOFF:] We're interested in learning about human evolutionary origins and modern African population history. We're also interested in the processes by which modern humans have adapted to different environments, particularly in Africa. And so, in order to address these questions, particularly in African populations, we first work together with international collaborators. Much of what I have done and many of my personal successes have been the result of working together with collaborators. I really feel that as a team, we can really tackle these important questions. And it's not always possible to be an expert in every single field. And that's why it's great to work together with other scientists. So, for example, we often work together with computer scientists, with statistical geneticists, and their research complements the type of research that we're doing, so that we can address these really interesting questions. I really think that nowadays, students need to be well trained in the biology and the wet lab component, and also in the computational analyses. It's absolutely critical in today's environment. So I really do encourage both, but there are some people who spend most of their day behind the computer, and they're analyzing data, and it can be a very slow and tedious process. But then occasionally, something comes out and it's just so exciting. When we do this analysis, you do it at different layers, where you assume that there were two ancestral populations, then three, then four, then five, then six, and so on. And I really felt like we were almost like archaeologists that were uncovering different layers of history, but tracing those layers through the DNA. And so, I do sometimes feel almost a bit like an archaeologist.

3. The human genome and how we differ from others genetically (3:15)

[Applause]

[DR. TISHKOFF:] So first, good morning, everyone. It is absolutely my pleasure to be here today and to tell you a little bit about my research. And as you heard in the video, I wear multiple hats. So today, I'm going to be wearing both my geneticist hat as well as my anthropologist hat. And the sort of questions that I'm interested in is to try to understand why do humans differ from our closest genetic relatives like the chimpanzees, and from each other? And I am interested in applying genetic methods to try to answer these questions. Our genomes are essentially all of the material, the genetic material that we inherit from our mom and dad. And if we're to just zoom in on a sequence of DNA, which is composed of the different nucleotides, and we sometimes call them bases, that's another word you'll hear, of G, A, T, and C, and if we were to look at how much do we differ-- if I were to compare genomes, how much do they differ? Well, if I compared identical twins, they don't differ at all. Now, what if we were to compare any of our genomes in this room to each other? Well, we would differ at about one out of every 1,000 of these nucleotides, and altogether, between genomes, about three million differences. Now, as we move further out in the evolutionary tree and look at more-distantly related species, like chimpanzee, we're going to see more and more differences. So now, one out of every 100 of these nucleotides will be different.

4. What is our place in the tree of life? (4:55)

So, looking at this tree of life, where is our position in the tree of life? Well, you can see the mammals in the lower right corner, and then we have the order of primates, and let's zoom in a little bit more. So now, zooming in on the primates, I've shown some numbers here, and these represent the times of divergence in million years. And that can be inferred based both on the fossil record as well as using genetic data. So the first lineage, primate lineage to diverge, around 63 million years ago, were the Prosimians; one of the representatives shown here are the lemurs. And they are currently located on the island of Madagascar, off the east coast of Africa. Then roughly 44 million years ago, you had divergence of New World monkeys found in the Americas; about 28 million years ago, Old World monkeys found in Eurasia and in Africa; then we have gibbons, 17 million years ago; orang, 14 million years ago; gorillas, seven million years ago; and our closest genetic relative, the chimpanzee, from which we diverged roughly six million years ago.

5. Genetic variation is greater among apes than among humans (16:05)

Now, here I'm showing you a phylogeny that is based on some genetic data, and if you look up in the upper right, the red little twigs there represent modern humans, and there are just these short little twigs at the end of that branch. Now, let's compare that pattern to what we're seeing for the chimpanzee shown in the lower right, or the gorillas, for example. And you see much longer branches, and that indicates that they have a lot of diversity along those branches. And what that's actually indicating, amazingly, that if you were to look at all the modern humans today, and look at all of our genetic diversity, we have much less, about a third less diversity, than chimpanzees in Africa. How could that be? So the reason, that you're going to learn at the end of this lecture, is that we all share a

relatively recent common ancestry, there hasn't been a lot of time for differences to accumulate amongst us, and also, probably at some time in our past, there was a rapid decrease in population size. We call that a bottleneck, a bottleneck event.

6. Fossil and genetic evidence inform us about human history (17:14)

So genetic evidence can tell you about evolutionary relationships among living species, whereas fossil and archaeological evidence can tell you about past diversity and intermediate steps. So what do we know about the origins of anatomically modern humans, *Homo sapiens sapiens*? Well, here, these red dots on this map of Africa are indicating the locations of fossil sites for the earliest anatomically modern humans. And the very oldest site is in southern Ethiopia, and it's been dated to about 150,000-190,000 years ago. We don't know that that's exactly where our species evolved, but that's the best evidence we have now. And so, we want to know, how can we trace the events that occurred after that time? What happened after modern humans evolved? How are we all related to each other today? And to do that, we're going to use evidence from the DNA.

7. Nuclear DNA and mitochondrial DNA (mtDNA) (8:14)

And so, what I first want to explain to you is that we have different sources of DNA. There is the nuclear DNA, which is contained within the nucleus of the cell, and it's in a compact form called chromosomes, and you inherit that from your mom and your dad. And then, within the mitochondria organelles, and remember, these are sort of the powerhouse of the cells, there is mitochondrial DNA, and I'm going to tell you a little bit more about those now. So in the nuclear genome, your nuclear genome consists of 22 autosomal pairs and the sex chromosome pairs. So for females, two Xs, and for males, an X and a Y. It is quite large, about 3.4 billion nucleotides, or bases, but it only codes for about 20,000 genes. So actually, only a very small fraction of our genomes codes for genes: less than 10%. Now, it's inherited from both parents, and every generation, we have recombination and shuffling that occurs, and that shuffling makes it a bit problematic. If we want to use that DNA to trace lineages, it's kind of tough because it gets shuffled up every generation. But the great thing about the nuclear genome is there is so... it's so large, and you can look at many independent regions, and each of those regions is giving you a snapshot of human evolutionary history. Now, what about the mitochondrial DNA genome? Very different: it's circular, it's small, only about 16,000 nucleotides, codes for 35 genes that are mainly expressed and have function in the mitochondria. It is maternally inherited: it's only passed on through the maternal lineage. There is no recombination: that makes it great for tracing back lineages. It has a very high mutation rate, and that makes it really useful for reconstructing recent evolutionary events. So what I'm going to do in this first lecture, I'm going to focus predominantly on the mitochondrial DNA evidence for tracing human origins.

8. Tracing lineages by mtDNA to Mitochondrial Eve

So what I want you to consider here, let's look at generation 11 down at the bottom. You see these red dots, and there are 13 of them. And let's say that they represent 13 mitochondrial DNA sequences from 13 individuals. At some point in the past-- let's go one generation back. These two that are highlighted coalesce back to a common ancestor. Let's go a little further: now we see these multiple lineages coalesce back to a common ancestor. And we can do this for all of the lineages that we sampled. Eventually, they will all coalesce back to a single common ancestor, and we call that the most recent common ancestor. Now, in the media, this common ancestor for all the mitochondrial lineages has been referred to as Mitochondrial Eve, and one of the big misconceptions is people thought, "Ah, there was just one founder of all modern humans: Eve." Makes sense. But that's not true. If you look at the top, look at generation one, and you can see all those different colors up there-- Eve lived in a population. There were other people there. But what happened is those other lineages died off; they didn't make it 'til today. But if you were to look at the nuclear genome, you would trace that back to many, many, many different ancestors. So that's something to keep in mind.

9. Using genetic variation to reconstruct lineage history (11:50)

Now, how do we do this? How are we going to use this information from the mitochondrial genome, for example, to reconstruct human origins? Well, one thing we can do is simply look at the mutations that have occurred over time. So if you look at the sequence on the top which represents a common ancestor, their sequence, and let's say that in the lineage leading to population A, there was a mutation on the left-hand side from T to A, it's highlighted in red, and then there was another mutation from C to G, and now everybody in population A has the A and the G nucleotides. Population B: they, by contrast, had this mutation from an A to a T, and everybody now has the T. So if you were to construct a phylogenetic tree, you would have all the individuals from population B that share that change are going to cluster together; all the other individuals are going to cluster together. And we can use something called maximum parsimony to construct trees, but the problem with that is that usually there's tens of thousands of trees that are possible to make; it can be very challenging. One other method we could do is, rather than looking at every single individual mutation, we can just summarize based on all the diversity, how close two lineages or two sequences are. So, for example, in sequence one and two, you can see in red that they differ at three out of a total of 20 nucleotides. Okay? So that's about a 15% difference. And you can see this little matrix, we're comparing now sequence one and sequence two in red, 15% difference. Now, if we compare sequence three and four, they differ at two out of 20, or 10% difference. So now, let's make a topology, and we can see that we're going to put one and two closer together, and we're going to put three and four closer together, and we can see that, if you look at that matrix, one and three are much more distant from each other. And that's how we construct these trees. Now, there's one other point I want to make. If you look at these black lines, and let's just assume that it's the mitochondrial DNA genetic tree, and it traces back to this common ancestor at some point in the past. We, if we have an estimated mutation rate, we can actually try to infer when the time of most recent common ancestry was for the DNA lineage. But an important thing to remember is that the DNA lineage will always coalesce

at a time that precedes the population divergence. So you could think of it as giving us a maximum time estimate.

10. The evidence for African origin of all modern humans (14:29)

All right, so let's look at some real data here. So this was a study in which they sequenced the entire 16,000 base pair genome from mitochondrial DNA from a number of individuals from around the world, and from chimpanzee. Now, remember I told you that we diverged from chimpanzee roughly six million years ago. So what you can do is you can just count up the number of mutational changes since we diverged from chimpanzee. From that, you assume sort of a molecular clock, and if you have an inferred mutation rate, you can put some ages on this tree, on these lineages. And when we do that, we see that the time to most recent common ancestry of all the lineages is about 170,000 years ago. That corresponds really well with that fossil evidence I was telling you about. We can also see that the oldest lineages are the ones shown in orange, and they are all specific to Africa, so oldest lineages are in Africa. The non-Africans have a subset of the African genetic diversity and tend to have much more recent lineages. Now, the other thing we could do, if we want to get a sense of how much genetic diversity there is in different regions, is simply do a pairwise comparison of all the sequences. So just as an example, let's look at sequence one and sequence two, and you see that they differ at the positions highlighted in purple. Well, you can basically look at the whole genome, count up all the places where they differ, and look at the mean pairwise difference. If we looked amongst African individuals, it's considerably higher than amongst non-Africans. What's the reason for that? There has been more time for mutation to accumulate in Africa, and if non-Africans originated more recently, then there hasn't been as much time for that diversity to occur. So in summary, African populations have the oldest mitochondrial DNA lineages and the most mitochondrial DNA variability. This fact supports the hypothesis that they are the oldest populations and that they've had more time for mutations to accumulate. So basically, this supports the hypothesis of a recent African origin of all modern humans.

11. Human dispersal out of Africa and the founder effect (16:43)

Now, what happened after that origin, sometime around 190,000, 200,000 years ago in Africa? Well, for some period of time, there's been time for the populations in Africa to differentiate, for quite some time within Africa. And then around 50,000-100,000 years ago, small numbers, and when I say small, I mean it could have been in the hundreds, okay? Small bands of people migrated or dispersed out of Africa into different regions of the world. Now, when they did that, when you have a small group breaking off from a larger group, we call that a founder event: they migrate into a new region. And they are going to lose genetic diversity when you have a small group breaking off from a larger one. We think the earliest migration was into Southeast Asia and Australo-Melanesia where they arrived in that region by about 40,000-60,000 years ago. They didn't make it to Europe until about 40,000 years ago; into Asia, around 30,000-60,000 years ago; and into the Americas, around between 30,000-15,000 years ago. And then we have more recent migration dispersal events.

12. Other species of *Homo* left Africa earlier than *Homo sapiens* (18:01)

But were modern humans the first ones to make it out of Africa? Absolutely not. In fact, the species preceding modern humans, *Homo erectus*, made it out of Africa probably close to two million years ago, and they spread across the globe. And then, they gave rise to populations outside of Africa. For example, here are just a few of them, and I think you probably heard about Neanderthals. Well, they existed in predominantly Europe and the Middle East, a little bit into southern Siberia, for at least a couple hundred thousand years, until as recently as 30,000 years ago. So that means they overlapped with modern humans in that region because remember I told you they made it there by 40,000 years ago. Amazingly, on the island of Flores in Indonesia, there is another species, Floresiensas, that might have lived until as recently as 12,000 years ago, and you're going to hear more about that in Tim White's lecture. And then we have this kind of mystery called Denisova, and this is basically a single, a finger, a digit, that was found in southern Siberia. And they can't get really good date estimates on it. Based on some of the material with which it was found, they're thinking, could be 40,000, could be 50,000 years old, but we're not really sure. But as I'm going to explain to you, they got some genetic material from that. But what this is telling you is that until quite recently, there were different species hanging around outside of Africa. That's pretty remarkable.

13. Ancient mtDNA reveals human-Neanderthal-Denisova phylogeny (19:34)

So, what can we learn about that history using genetic data? Because I'd like to know, for example, what happened to those Neanderthals when modern humans came out? Did they admix with the modern humans? And amazingly, now we can start to combine genetic data with archaeological and fossil data, and to look at the question of the origin of the Neanderthals. Now, a couple of distinct features of the Neanderthals: they were really big, they were really bulky, they had big brains, they had these double-arched brow ridges, they had a really broad nose. Now, back in 1997, Svante Pääbo's lab in Germany were the first ones to get DNA from a Neanderthal, and this is showing some of the methods that they used. So they actually were able to extract the DNA, you have to do this very carefully, and on the right, they're showing some of the procedures. One of the biggest challenges, there's two, actually: one is that the DNA is chopped up; it gets broken down and degraded. And the second is contamination by modern human samples: if you breathe on that sample, you are going to contaminate it. They have to be extremely careful. So more recently, they have now collected DNA from a number of these species, these non-modern human species that existed outside of Africa, and they sequenced the mitochondrial DNA, and they made a phylogeny, and they compared it to modern human sequences. And what we can see, that's shown on the right, that the time to most recent common ancestry of the lineages in modern humans is around 200,000 years ago, as we had seen with the other study; the time of common ancestry between Neanderthals and modern humans, 500,000; and if we include Denisova, it diverged much more in the past, around one million years ago. But one thing I should point out is that you don't see any

evidence for any close relationship between the Neanderthals and the modern humans. They seemed to have been pretty divergent. Now, again, if we do this pairwise sequence comparison and plot that out, we're just going to count the number of positions at which the genomes differ when we compare among humans, or we compare between Neanderthals and humans, and so on. So in blue is among humans, and you can see we're pretty similar; we really don't have that many differences amongst us. If we compare to Neanderthal and human in red, we see more differences. There's even more differences between human and Denisova, and a lot of differences, relatively, between human and chimpanzee. And that fits really well with the timeline based on both the genetic and the fossil and archaeological data, that *Homo erectus* arose roughly two million years ago and dispersed out of Africa. Denisova split from humans at least a million... or more recently probably than a million years ago; that's the upper bound. Neanderthals split from humans around 400,000 years ago; origin of modern human, 200,000 years ago; and then dispersal out of Africa, within the past 50,000-100,000 years.

14. Comparison of human and Neanderthal genomes (22:37)

Now, quite recently, with the advent of what's called next-generation sequencing technology, we have the capability to sequence millions and millions of nucleotides of DNA, like that-- in one day, for example. And with that new technology, they were actually able to get nuclear DNA from the Neanderthal bones. Now, it was several years ago that they first got a pretty big chunk of sequence; it was several million nucleotides. But in 2010, they got the first draft genome sequence from Neanderthal. And when they compared that to human, to the human genome sequence, one of the remarkable things is how similar the Neanderthal is to us at the genetic level. So if we were to look at all the differences that have occurred in the hominin lineage since we diverged from chimpanzee, 92% of the changes are shared with Neanderthal. Only 8% are specific to humans. A fascinating question is to learn, what are those changes? What are those variants doing? Now, they had a few hints, some intriguing observations that some of them played a role in skin morphology and physiology, some seem to be expressed in the brain and may play a role in sociality, but there is a lot more to be done. We really don't know at this point. The other fascinating thing that came out of this study is that they showed some evidence for a small amount of admixture, on the order of 1-2%, between Neanderthals and modern humans outside of Africa. And what they think is that what happened is when modern humans originally left Africa, they think this admixture occurred in the Middle East, possibly around 80,000 years ago. And then we see evidence for this admixture in the genomes of all non-African humans. So in conclusion, Africa has the most genetic diversity. Human dispersions out of Africa populated the entire world, and we are the last of a series of hominin dispersal events out of Africa. And I am happy to take a few questions. Any questions? Yes?

15. Q&A: How can different species admix? (24:54)

[STUDENT:] I thought hybrids can't really, aren't very genetically viable, so how would the Neanderthal DNA have been transmitted to the modern humans?
[DR. TISHKOFF:] So you're asking how it is, if they're different species, how could they possibly have admixed? That is a fantastic question. And we can let the paleoniologist also talk about definitions of species because this can be very tricky. And some people have defined a species as it means that it can't possibly admix; they have to be so different that you can't have admixture. But Neanderthals clearly, they weren't that different in terms of the genetic makeup. It means that they were close enough, and when I showed you that figure where I said only differed about 8%, they were clearly close enough that they still could interbreed. Now, they were different enough based on morphology and probably behavior and other characteristics to classify them as a distinct species. But clearly, I think, or if you do believe... and I should say, you should keep in mind whenever people do these analyses, they are using often computational methods to try to infer these past events. And this DNA was really in bad shape, and they had a very... not-great quality of sequence, let's put it that way. So for me, I still am not, I want to wait for a little bit more data to say absolutely 100% there was admixture; I'd say at this point, it's suggestive.

16. Q&A: Why do ape species shoe more genetic diversity than humans? (26:23)

[DR. TISHKOFF:]Yes?
[STUDENT:] Why do you think that there is so much more difference between the apes than there is between humans? And why have they changed so much? **[DR. TISHKOFF:]** Yes, it's really remarkable that they're thought to be roughly three to six times more genetically diverse. And if you think about the number of chimpanzees, I mean, they're considered to be an endangered species, right? I mean, that's remarkable. The reason why is that our demographic history, and by demographic history, that means things like changes in population size, how big was the population in the past. That influences the pattern of genetic diversity in the present day. So for example, if a population rapidly decreases in size, you lose genetic diversity, and that's what I think happened. Somewhere along the line-- so we could have someone like Tim White maybe tell us later-- that sometime during that divergence, after we split from the chimpanzees, there must have been a decrease, I think, in population size. But actually, it did not occur to the chimpanzees. What we're seeing recently in terms of them being an endangered species, that's a very recent, much more relatively recent event. Any other questions? Yeah, in the back?

17. Q&A: Can you speculate on the future of human evolution? (27:35)

[STUDENT:] Considering a species' capacity to constantly evolve, its capacity and out of necessity, can you comment on the potential future of us as a people, both genetically, physically, mentally?
[DR. TISHKOFF:] Oh, boy, you gave me a really tough one: Can I speculate on the future evolution of all modern humans? What I'm... actually, you kind of led in just perfectly to

the second part of my talk because that is exactly what I'm going to talk about. So I'm going to talk about some examples of how humans are still evolving, and I absolutely believe that we are. And I think it would be naïve to think that we are not evolving. Now, at the same time, there are a lot of cultural adaptations, right? So some people have claimed, well, with medical interventions and cultural adaptations, we have overcome this, our process of natural selection. I disagree. I think it's still happening. But let me talk about that in the next section. In fact, I think we need to move on to that now.

18. Types of genetic variability in nuclear DNA (28:28)

So in the first part, we had focused a lot on the mitochondrial DNA, and now let's switch to the nuclear genome and see what that tells us about patterns of genetic variation and about adaptation. Remember that each individual, every one of us in this room, unless you have an identical twin, has a unique genetic makeup, and these genetic differences are spread all over the genome. Different variants at any particular site in the genome, we're going to call alleles, so at any particular site, you might have a C or you might have a T. We're going to say those are different alleles. And another definition, single nucleotide polymorphisms, or they are often called SNPs, are one type of genetic variability, and they're just shown in this example to the right, where we have, say, two individual sequences, and they differ at the position shown in purple, so either T-G, or A-C. So a single nucleotide change, or polymorphism, we're going to call a SNP. Now, it's important to remember, remember I told you that there are about 20,000 genes, and I said that's a really small proportion of the genome? So if there are genetic differences that happen to be in those coding regions or in regulatory regions that are going to influence gene expression, and if they happen to cause a change, an amino acid substitution, for example, that causes a change in the protein, then they might be important to look at for learning about adaptation to different environments. But if we want to learn about reconstructing just our history, we actually want to focus more on the regions that aren't coding for anything, because they're not going to be influenced by natural selection. And I want to start by just telling you in addition to SNPs, some other types of genetic variability that we see. If you look at individual two in the middle, okay, and that's... start with that sequence, now let's look at individual one, and you can see that the portion that's in pink has been deleted: C-T-A-G. So that's called a deletion event or a deletion polymorphism. And then, at the bottom in individual three, we have an insertion of A-T-C-A. So this general type of polymorphism or variability, we call insertion-deletion polymorphisms, or in-dels for short. Now, these tend to be very stable, and for that reason, they don't revert back, you don't have back mutation occurring, and they can be informative for tracing more ancient evolutionary events.

19. Short tandem repeats (STRs) are highly variable genetic markers (31:01)

Okay, another type of variability is called short tandem repeats, or STRs. And these are simply short nucleotide sequences, between two and six nucleotides in length. So here, for example, is one with three, so C-T-A repeated over and over and over, and the number of

repeats can vary between individuals, so individual one has four, individual two has five, individual three has six. Now, these are often the markers used by the FBI and featured all the time in all these exciting crime shows, and the reason why is they have a really, really high mutation rate. And that means you've got lots of variability. So if you were to, say, combine ten of these markers and genotype them in an individual, and match it to the crime scene, you might be able to say pretty well if that person actually committed or was present when that crime was committed. Now, the fact that they mutate so quickly also makes them really useful as markers for reconstructing recent human evolutionary events.

20. Using STRs and other markers to study human genetic variation (32:06)

So now, if we're going to look at these different types of markers, this is just showing that you're going to see them spread all over the chromosomes, and they can be really useful for looking at evolutionary history and comparing patterns of variability. And I'm going to tell you now about a particular study that we published in 2009, and in that study, we looked at 773 of these short tandem repeats, and 392 of these in-del polymorphisms or variants, and they were spread across the whole genome. And we looked at them, genotyped them in over 4,900 individuals from all over the world, but with a particular emphasis in Africa, because Africa is a place that has been traditionally underrepresented in human genetic studies. All right, so this map on the right shows you the location of the populations that we included, and our goals were to reconstruct human demographic and evolutionary history, and if we're going to do that, I told you that modern humans evolved in Africa, we need to be looking in Africa, and also to better understand human adaptation.

21. African linguistic families and cultures (33:16)

Now, I've color-coded these populations based on the languages that they speak. And the reason that this is important is that knowledge of the language they speak actually is going to tell you something about their culture, because often they shared culture that goes along with shared language, and there may be shared genetic ancestry as well. Now, in Africa, there are over 2,000 languages spoken. That's a lot of variation, and they've been classified into these four language families. And in orange are those classified as Niger Kordofanian; they're thought to have originated in West Africa. And then there was a migration event within the past 5,000 years all across Africa that had a major influence on the genetic variation we see there. In purple are the Afroasiatic speakers. They're present mainly in North and Northeast Africa; there are some in Central. Also, I meant to mention that the Niger Kordofanian speakers very often practiced agriculture, and the Afroasiatic populations often practiced agriculture and pastoralism. In red are the Nilo-Saharan speaking groups; these are often pastoralist populations like the Maasai you might have heard of, and they tend to be in Central Africa and East Africa. And then, the last language family shown in green is Khoisan; those are the people who speak with the clicks, and I'm sorry, I can't do it, at least not properly. But that would include like the San hunter-

gatherers in southern Africa, and also two groups that exist in Tanzania, the Hadza and the Sandawe.

22. Challenges and ethics of human genetic studies in Africa (34:44)

Now, for over ten years now, I and my students and my post-docs and my collaborators in Africa have been doing fieldwork in Africa to collect DNA samples and to study both genetic and phenotypic variation. And here is one of the reasons that it's particularly challenging; I think this kind of says it all, that often these populations are living in very remote areas, lack of infrastructure, it can be very challenging to reach them. This is me talking to a group of Hadza hunter-gatherers, and one thing we're really careful about is to do this research in an ethical manner. And what that means is you have to get consent at every level, so you start out getting consent at your university through what's called an Institutional Review Board. Then I have to go through ethical review in each country that I go to; that can take a very long time. Once I pass that, you have to get community consent. And then ultimately, you have to get individual consent. Typically, we would get blood from individuals, and from the blood, we can extract DNA from the white cells that are in the blood. But there is no electricity in most of these places, and so, we have to use some clever methods like hooking up the centrifuge to the car battery. And this just gives you some idea of what the conditions are for processing the samples. We basically bring our whole lab; we take our entire lab, we bring it with us, and if someone can just give me a table, I'm fine. And then we can do... process these samples. Okay, so now, let's go back to that study I was telling you about,

23. Human Genetic diversity is greatest in Africa (36:22)

and the first thing I'm going to show you are levels of genetic variation. So the height of these bars, each bar represents a population studied, they're color-coded by geographic origin, and the height of the bar indicates the level of diversity. So we can see that the African populations that are shown in orange, they all pretty much consistently have the highest levels of genetic variation. This is what we see in almost every study. And then, as we move west to east across Eurasia, that's shown by the blue and the red, into East Asia shown in pink, and then Oceania shown in green, and the Americas shown in purple, we see decreasing levels of variation.

24. Individuals in the same region are more genetically similar (37:04)

And that is very consistent with that model I told you about where modern humans left Africa, a small group, they took with them a subset of variation, and every time they moved into a new region, a smaller group broke off and you lost some genetic diversity. And then, that's exactly the pattern that you expect to see. This is using a method that's called principal components analysis, and you don't have to know the details; the only point I want to get across here is that each of these circles now actually represents a

person. I'm not looking at populations here; I'm looking at people. And I'm looking at how they are related, how they cluster with other people based on their genetic diversity. So if two of these circles cluster closely together, it means that they share a lot of genetic variation. They're more similar to each other. Now, the general patterns that we see is that, on the right-hand side, those are all the African populations, and they can be distinguished from the non-Africans who are on the left-hand side. And then we can see that individuals tend to cluster with others from the same major geographic regions, so we have Native Americans, Eastern Asians, Oceanians in green, Southern Indians in red, Europeans in blue, and then as I said, we have Africans on the right-hand side, and in yellow, those Hadza hunter-gatherers, they're pretty divergent from the other groups.

25. Different populations show characteristics of genetic admixture (38:29)

All right, now I'm going to tell you about a different type of analysis, and what we did is we used computational methods to try to infer the number of genetically distinct ancestral populations based on all those genetic markers that we typed. And when we did this on a global level, we were able to infer 14 ancestral populations, and they're shown by the different colors here. Now, the first thing you might note is that you tend to see not that many colors outside of Africa; that's because again, most individuals cluster by major geographic region. There's not quite as much diversity. But if you look in Africa, you see lots of colors representing a lot of variability in Africa, even in different geographic regions. Now, let's give some examples; let's look at some particular populations. And what we're going to do, these pie charts, the different colors are representing the proportion ancestry that they have from those 14 ancestral populations that I showed you. So general patterns: again, outside of Africa, individuals tend to cluster by major geographic region. But there is one other thing you should know. We see lots of different colors, often, in these pie charts. Do you know what that means? What do the different colors represent? Any ideas? Different ancestry, admixture, people coming together, right? We're not always just isolated populations in the world, particularly today. So as an example, if we look at African-Americans, you see they have predominantly orange ancestry; that orange represents ancestry from West Africa. The blue represents European ancestry. And there is on average around 20%, but it could be anywhere from zero to greater than 50%. Let's look at another population; this population is from Central Asia. They seem to have ancestry from East Asia and from western Eurasia. Let's look in Africa. These are pygmies, Biaka pygmies. The dark green represents the pygmy ancestry, but that orange represents the West African ancestry. So even in Africa, they're admixed.

26. Genetic diversity in Africa reveals admixture patterns (40:42)

And in fact, if we zoom in now to Africa and look at their pattern of variation, again, you can see that general trend that there's a lot of admixture. But we can also learn a lot about their history and past migration events. So for example, if you look at the orange-colored circles, and you see a lot of them on the left in western Africa, I told you that was the

homeland of the Niger Kordofanian speakers, and then they migrated across Africa, and we can see how they moved, for example, into East Africa and they mixed with the local people, and then they moved into southern Africa. In purple in the upper right-hand side, you have people who have Afroasiatic ancestry. And remember I said that originated in northeast Africa, and then they migrated down into Kenya and into Tanzania, and again, they admixed with the local people. And in red, if you look at southern Sudan, that's thought to be the homeland of people who speak a Nilo-Saharan language, and then they migrated into the east into Kenya and Tanzania, and another group went to the left. And remarkably, the hunter-gatherers also stand out. In yellow are the Hadza that I told you about; in the sort of light green in southern Africa are the San hunter-gatherers; in dark green in the center are the pygmies. And one of the really remarkable things was that we see some evidence that, you see those easternmost pygmies? They have some dark green? They have a lot of light green, showing possible common ancestry with the San from southern Africa. So that may represent that these groups had a common ancestor, and I think it could have been about 50,000 years ago, or more. All right, so to summarize the section on African diversity, people from different geographic regions are genetically more similar to each other, but in Africa, they've had a longer time to accumulate variation, and they've lived in diverse environments and are highly differentiated from each other. And ancestry, migration, admixture, and natural selection are going to define the patterns of variation that we see in modern populations.

27. Lactase persistence as an example of human adaptation (42:43)

Now, in the last part of my talk, I want to move on to the question of how have we adapted to so many diverse environments? Think about it: when modern humans left Africa, 50,000-100,000 years ago, think of all the diverse environments that they have moved into. Very diverse climates; we now have very diverse diets, agriculture, and people who domesticated cattle, we call them... and they drink milk, we say they're pastoralists; so we see, and many different, exposure to different infectious diseases. Now, there have certainly been cultural adaptations, right? You can see here these Eskimos, and they have the really heavy clothing on and the fur; that's a cultural adaptation. But there have almost certainly been genetic adaptations as well, and I'm really interested in trying to find those. So I'm going to give you an example today, a classic example of adaptation, and that is the ability to digest milk as an adult, referred to commonly as lactose tolerance. And we are going to compare, for example, on the left is an African population that has... practices cattle-herding, we refer to them as pastoralists; and on the right, we have a European farmer who practices dairying. So milk contains very high quantities of the sugar lactose, and the molecular composition is shown on the milk bottle here. Now, in order to digest the sugar, lactose, you need expression of an enzyme called lactase. And lactase is going to chop it up, that lactose, into its component sugars, glucose and galactose. And if you're able to digest the milk and if you have the active enzyme, that's going to be rapidly taken up into your bloodstream. And I also want to mention that the lactase enzyme is expressed specifically in your small intestine. Now, almost every baby is able to drink milk, right? So everybody's got an active form of this enzyme. However, shortly after weaning in all mammals, and in many human populations, this enzyme is shut off, okay?

So that means that as an adult, you don't have an active form of the enzyme and you cannot digest the sugar lactose that's present in milk. So often, that's referred to as lactose intolerance. But in some people, due to a genetic mutation that I'm going to tell you more about, they maintain the expression of this gene at high level as adults, and they can digest the sugar lactose, and we say that they are lactose tolerant. We can also say that because the enzyme is maintained as active, we could say they're also lactase persistent; they have the lactase persistence trait.

28. Global distribution of lactose intolerance (45:29)

Now, for quite some time, anthropologists have noted this very interesting distribution. In sort of the brighter-colored purple here, that is showing the prevalence of the lactase persistence or lactose tolerance trait around the world. And it's long been known that this trait is at highest prevalence in northern Europe, so in the Finnish population, something like 99% of the population is able to digest milk. As we move into Europe, into southern Europe, and then into the Middle East, we see less and less and less of that trait. As we move into East Asia, very rare; the Americas, very rare; Africa, very rare. So one take-home point you should remember is that the ability to digest milk is a recently derived trait in populations whose ancestors practiced dairying and they drink milk. And most of the world cannot digest milk. Keep that in mind.

29. A simple test for lactase persistence (46:29)

Now, there had been a mutation identified in Europe, and I'm going to tell you more about that in a moment, but in Africa, we knew nothing about how people digest milk. And in East Africa, you have a lot of these groups like the Maasai that you might have heard of, who are pastoralists: they have lots of cattle, they drink lots of milk. So we hypothesized that probably they have adapted as well, and they probably have the mutation. So to figure this out, we gave a test called a lactose tolerance test, and what you do is you basically give them the sugar lactose in a powdered form, you add some water, it tastes like Kool-Aid, and everybody lines up, drinks it at the same time. And then you're going to take a baseline measurement of glucose in the blood, and then every 20 minutes, you're going to do a finger prick and you're going to use one of those sugar blood monitoring kits, just like a diabetic uses, to measure the amount of glucose in the blood. And you're going to look at the rise over a one-hour period. People who have an active version of the enzyme, they're chopping down the lactose, the glucose is going into their blood, and they're going to have a rapid increase, okay? And that's how we're going to classify them as being lactose tolerant, or having the lactase persistent trait.

30. The genetic basis of lactase persistence in Europe and Africa (47:40)

If we look at the lactase gene, it's a really big gene, it's like 50,000 nucleotides, and it's on chromosome two. It's shown in yellow here. Not in the gene, but near the gene, upstream

of the gene, about 14,000 nucleotides away, is where we and others have found the mutations that regulate the expression of this gene. In green is shown the mutation that regulates this trait in Europeans, so people who have a T at that particular position can digest lactose in milk, and those who have a C cannot, if you have two copies of the C. In Africa, we found a different variant; it was about 100 nucleotides away from the European variant. And if you have a C, you could digest milk, and if you have two copies of the G wild type, you cannot. And what it appears is that these two mutations occur in a regulatory region, a region that's going to be important for regulating gene expression, and we call that an enhancer. And what happens is people who have the lactase persistent associated mutation, shown here, they can more strongly bind a protein called a transcription factor, and that transcription factor is what's going to turn that gene on or keep it on, okay? And so, they're basically able to bind it more tightly. Now, I'm going to just show you Tanzania as an example. If we measure the lactase activity in that region, and that's shown in the top in green, high is darker color, and then medium you can see, and you can see that most of the groups in that area, they have a lot of very high lactase activity, and that corresponds to the fact that there's a lot of cattle, a lot of pastoralism. In the bottom part, I'm showing you the frequency of individuals who are either homozygous, meaning they have two copies of the C variant that is associated with lactase persistence, or they're heterozygous, they have one copy of wild type, one copy of the C; that's shown in orange. And then, in blue are individuals who are homozygous for the wild type; they have a G-G. And you can see that if you look at that prevalence, it's really generally well correlated with lactase activity; it explains much of the trait. But you might have noticed this one exception. And this is a group, the Hadza. We keep talking about the Hadza, and they're unusual in a number of ways. They're hunter-gatherers. They don't have the mutations associated with drinking milk or digesting milk. Okay, that's not a surprise. The surprise was when we did the lactose tolerance test, that it appeared that they could digest milk, about 50%. Now, why is that? Well, it was a small sample size, so maybe it's a mistake, and we're actually going back and testing more of them. Maybe they had cattle in the past; maybe that's a possibility. Or maybe the enzyme plays a role in breaking down other things that they're eating, so it turns out it might play a role in digesting some of the material present in barks and roots in the area.

31. The genetic footprint of recent natural selection (50:40)

All right, so the last thing we wanted to do is see, if this mutation was so important, and if it went to high frequency so rapidly, did it leave behind a genomic footprint of natural selection? Well, there's a way we can do this. Let's assume this is the region of DNA near the lactase gene where these mutations are present. And I'm showing five individuals here, and they each have two chromosomes, and each of these colors represents a SNP, a single nucleotide polymorphism. Okay, sometime in the past, this mutation occurs; this mutation regulates gene expression, keeps the lactase enzyme active, and it increases the fitness of the people who have it. What do I mean by that? I mean that they are more likely to survive, they're more likely to have kids, their kids are more likely to have kids. When that happens, it's going to increase in frequency in the population. And what it's going to do, it's going to drag with it those variants that are flanking it, the ones shown in orange

and yellow. They're going to rise up to high frequency as well. And we call that, that's due to positive selection. Over time, recombination is going to shuffle things up a bit. Now, if we were to simply highlight the regions of the genome where they are identical because remember, the mutation occurred in red, flanking variants rose to high frequency, but then over time, recombination breaks things down. Now let's look at some real-life data. On the top, I'm showing you individuals from East Africa who have two copies of the lactase persistence allele, and we're looking out three million nucleotides flanking that, and we're asking, how far out are they identical? This is pretty amazing: they were identical up to two million nucleotides on average. Now, if you compare that to people who have two copies of the wild-type allele, we don't see that at all. They're only identical about 1,800 nucleotides. That is a whopping signature of natural selection, a real genetic footprint. Now, what about if we compare to Europeans? They show a very similar pattern. So what's neat about this is we have two totally different mutations, they arose in totally different regions of the world due to this common selective force of being able to drink milk as adults.

32. Lactase persistence: An example of gene-culture coevolution (52:56)

Now, we can use computational methods. If we know something about the recombination rate, then we can infer by looking at how far those tracks are, we can infer how old the mutation is. And in Europe, we inferred it to be about 9,000 years; in East Africa, about 3,000-7,000 years old. And what was really remarkable is that it correlated perfectly with the archaeological record for the origins of cattle domestication, which is thought to have originated in the north of Africa and/or in the Middle East, those are the circles in orange and red at the top, somewhere around 9,000 years ago. Then there was a rapid migration into Europe, shown by the red arrow. But in Africa, cattle were not introduced south of the Sahara until after 5,000 years. So that corresponds perfectly with that date estimate, a great example of gene- culture co-evolution.

33. Other examples of natural selection in humans (53:53)

Now, there are other examples of natural selection: for example, our ability to taste bitter substances is one example. Also, the sickle cell anemia trait, there is a very high prevalence of the mutation that causes sickle cell anemia; you have to have two copies of that mutation to get the disease. And we think it's very common, because people who are heterozygotes, so they have one copy of the wild type and one copy of the sickle cell allele, have resistance to malaria. So it rose to high frequency in East Africa, and now people of African descent have a high prevalence of that disease. Other examples: we think that skin pigmentation may be adaptive because of UV protection and vitamin D production. We don't have that entirely worked out, there's about five to seven genes known, but they show these signatures of selection as well. And lastly, I'm often asked the question, are humans still evolving? And I think that this example of lactose tolerance, the evolution of that, is a great example that yes, we definitely are, because that trait arose very recently,

just in the past few thousand years. And with that, I am happy to answer any questions that you may have. Yes?

34. Q&A: Why is lactose intolerance so low in my community? (55:07)

[STUDENT:] You mentioned that going to the east from Africa across the world, that lactose tolerance decreases. But in general, I was just trying to think back, I don't know many people who are lactose intolerant. Is that because in the States, we've had a lot of European immigration, or is it a different...? **[DR. TISHKOFF:]** Okay, so you were saying that you don't know many people, personally perhaps, who are lactose intolerant, and could it be because there is a lot of admixture in the United States and introduction of the mutation? I think there's a couple things going on. One, I bet you there's a lot more than you know. So people either may feel uncomfortable talking about it. Also keep in mind that it's, that there are a lot of products out there that allow people to be able to consume dairy products, so they can have... the enzyme lactase can be added, for example, or you can have lactose-free, there's a lot of lactose-free products out there. Also, if you culture milk, you can make it into cheese. You can eat cheese, that's no problem; yogurt, you can eat yogurt, no problem. So I think it's more prevalent that you realize. But yes, you are absolutely correct that in places like the United States, there's intermarriage between different groups, that it could be introduced through gene flow, or through intermarriage as well. Yeah?

35. Q&A: How does gene expression affect the lactose intolerant trait? (56:23)

[STUDENT:] So I've heard that genes can be turned on and off throughout an individual's life. How did you deal with that when you were looking at the lactose intolerance in people?
[DR. TISHKOFF:] Yes. So the question was that she's heard, correctly, that genes can be turned on and off throughout an individual's life. That's absolutely true, and in fact, lactose, that lactase enzyme, the regulation of that is a great example because normally, during development, it's actually turned off normally, right? I was saying in most people, and it's only in people who have those mutations that it's kept on. Generally, for that particular trait, it stays on. There is some evidence that as people get much, much older, they move into, say, 60s or so, I believe there can be a slight decrease in the prevalence of the gene expression activity. But it's not huge for that particular trait. But there are other examples where there are developmental regulators of gene expression, and it can shift over time. And I've just been told that unfortunately, that's my last question. [Applause] [Music]