My name is Lynn Jorde. I'm chair of the Department of Human Genetics at the University of Utah, School of Medicine.

I was born in a little town about 30 miles south of the Canadian border right on the North Dakota and Minnesota border, probably the worst physical climate in the entire country.

Well I realized, I guess in my first year of undergraduate work, that I really enjoyed mathematics and computers. Somebody told me to learn about computers -- this is back in 1971 -- because no matter what you do, computers will be useful. So I learned as much as I could about computer programming and statistical analysis and that did turn out to be a very useful toolkit. Then in 1972 -- 1973 I took my first human genetics course. We used Curt Stern's 1972 Red Book and that was the book I couldn't put down. And that's when I knew human genetics was very likely what I would end up doing. But the quantitative training, along with the interest in human biology, human genetics, it seemed like a very good fit at the time and seemed like a field that had a lot of potential. And I think I was right about that.

Oh, well the Cumbria Project was something that started when I was a graduate student. We had a visiting speaker named Derek Roberts, who's a professor at the University of Newcastle. I showed him some of the work I was doing on a Finnish isolate population. He thought the same techniques I was using there could be applied to the Cumbrian data that he was collecting. So I visited him in 1978. We put our heads together. I did some analysis and the thing that we found really interesting was that there was a district in Cumbria called the Lake District where if we looked at the genetic data that were then-available, which is blood groups and protein polymorphisms, the people in the Lake District actually looked more like Norwegians than they did other people from England. And we think that that probably was a reflection of the Viking incursions in that area a thousand years ago that still had a genetic signature on the present day inhabitants.

Yeah, so we were -- as many people were back around 1990, or so -- interested in the competing out-of-Africa versus multiregional hypotheses. And there -- at that point I think there were more speculations than there were actual data, but that's been, I think, one of the great developments in human genetics. During the '90s and subsequently, we had access to much more data so we were really able to test those hypotheses more effectively. And in 1995, Alan Rogers and I published a review paper where we posited that probably neither of those hypotheses was exactly correct, and that the truth would lie somewhere in between, probably closer to out-of-Africa. And now with the genetic data we now can analyze, the sequence data and so forth, we know that we are mostly out-of-Africa, but there have been these smidgens of contributions from other archaic populations. So it is a bit of a blending of the two hypotheses.

Yeah, so if you look at where human populations genetics was, say in 1980, you know, we had an abundance of theory, we had a lot of great methods, we just didn't have data to apply them to. So the genetic variants that we could look at really consisted of several dozen blood groups, protein polymorphisms, all of which were ascertained originally in European populations. They all tended to have relatively high-frequency alleles. That's how you could find them. So the estimates that we got of population history were biased in a lot of ways. And some of them, it turns out, are under-selectioned so that further complicates the issue.

Another name that I'm sure you've heard would be Dick Lewontin. He had enormous influence on the field, not just in human population genetics, but in fruit fly genetics, just population genetics in general. Another very influential person was Newton Morton, who was in Hawaii for a long time and did both human population genetics and also gene mapping methods. He had an enormous influence on the field. Jim Neel at Michigan with his work in Japan and then on the Yanomamo had really a very lasting influence in our field.

Well the first time I was on a review panel with him, it was an NSF review panel, molecular evolution. And I have to admit I was intimidated. And we were both reviewing a proposal and I knew that his review was a lot more favorable than mine and I remember feeling some trepidation thinking that there might be a big argument here and I knew he was really good at arguing, so I laid out my argument why I thought the proposal was not really up to par. And he just said, "Well this is your area, I'll go along with it." I was relieved. We had a beer together that evening. We talked about linkage disequilibrium for two or three hours, had a wonderful time. So it was really a positive experience all the way around.

I did see him argue pretty strenuously with someone else and very effectively.

Well I think we are all products of our politics to some extent. I think that it's one of the things that we, as scientists, have to -- really have to watch, we have to struggle with because ideally our science would not be affected by politics or other prejudices. But that's -- I think that was one of our greatest challenges to try to maintain our objectivity.

Well I had two abiding interests. One was evolution, where things come from, where we come from, how we relate to one another because evolution is the great unifying theory of all of biology. My other abiding interest was human disease. And I was one of those people who couldn't decide whether to go to graduate school or medical science. So I maintained a strong interest in human disease. And what I really hoped to be able to do was to put those two interests together. And the whole field of linkage disequilibrium actually helped to do exactly that because suddenly people were interested in this previously-arcane evolutionary measure -- linkage disequilibrium -- in the context of disease gene mapping and discovery. So I was really happy to see evolutionary genetics and medical genetics begin to converge that way because originally in the 1970s, when I was being trained, they were two, very, very separate fields. But now we've seen that there is a really wonderful convergence of the two and they benefit one another because, of course, evolution tells us about all variation, including the subset of variation that causes disease. So we really can't understand disease and the genetics of disease without understanding the evolutionary factors that underlie it.

I think in many ways it was. And, of course, the convergence wasn't only due to linkage disequilibrium or the interest in linkage disequilibrium. But I think that that really was emblematic of the way in which the two fields could come together.

Well there's always been some interplay between medical genetics and population genetics for a lot of good reasons. And there are systems like the HLA System, where there is substantial medical significance, but also to understand the variation at HLA we have to understand the process of

natural selection, of diversifying selection. So there, there was a -- I think there were some common interests. But the reason that linkage disequilibrium received so much attention starting in the mid-1980s, was that for the first time with RFLPs, actually, we were able to identify polymorphisms that were close enough together so that there would be substantial linkage disequilibrium there in human samples.

Yeah, and we can essentially go in the other direction and make inferences about things like gene flow, population sizes, from patterns of linkage disequilibrium in the genome.

Well I think it's similar to the distinction we would make between just genetics and genomics. With population genetics we were always using specific lo-size, specific systems, sometimes for a specific kind of study, say mobile elements. But with population genomics, of course, we are looking at whole genome sequences. We can look at the totality of genetic variation in each and every subject. So it gives us just a much wider view of genetic variation.

So, well Henry was actually on my Ph.D. committee when I was a graduate student. And I think the -- his biggest influence on me was in telling me to take more math courses. [laughs] He was one of the best applied mathematicians I've ever known. He could look at a page full of equations and in a few sentences explain what it all meant. Or, if he wanted to model a process, an evolutionary process, he could write a page full of equations. So he had a really brilliant analytic mind and he urged his students to try to learn as much about analytic methods as possible. So he did early in my career have a substantial influence.

Well most of his -- much of his field work involved collecting blood samples and using, again at the time, blood groups, protein polymorphisms to look at genetic variation in Southern African populations. And at that time very little was known about the origins or affinities of those populations. And some of the methods that he applied were groundbreaking at that time. So he -- and he realized that a lot could be learned about the ancient history of those populations, potential migration patterns of Bantu-speaking versus other Africans and so forth. So in that way I think he had a substantial influence on the field as early as 1970 or so.

Well I was fortunate in having several people on my committee that were excellent scientists: Jim Spuhler, who had been at the University of Michigan for more than 20 years and then came to New Mexico I think some time in the '60s or '70s, was on my committee. He had a tremendous breadth and depth of understanding of genetics, of evolution. So he added a lot of real, I think, intellectual weight to my committee.

And a third member of my committee was a man named Peter Workman, who was trained in plant genetics at Davis and then became interested in humans. He was -- he became interested in Finnish isolate populations. So one of the reasons I chose to do my graduate work at New Mexico was that Peter had access to these wonderful, at the time, genetic data from Finnish populations that had high frequencies, for example, of Von Willebrand disease and also several recessive conditions. And we were interested in how genetic drift and founder effect may have contributed to that. And he had a great data set. I had quantitative skills, so that's -- bring those two together was actually a great opportunity for me.

I mean, anthropology subsumes so many sub-disciplines and I -- what I was attracted to was really the evolutionary genetics part. I have to confess I know very little about cultural anthropology or linguistics. I tried to be an archeologist for one summer. By the end of the summer I knew I didn't want to be an archeologist. So what attracted me was really the evolutionary genetics focus of some anthropologists and that's where people like my committee members focused. And that's where I thought there was fertile ground, not just for pursing evolutionary studies, but then applying that to hundred phenotypes, including disease phenotypes because that was something I was always really interested in. I had colleagues who could do theoretical population genetics all day long. And I could do that, but to me it was a little bit sterile. The mathematics is beautiful, but at the end of the day I want to be able to apply it to something that's real.

Well there are a lot of different kinds of factors that can influence or bring about genetic diversity. Usually we think in terms of geographic factors. They isolate populations or they promote gene flow. But there can be social factors that influence mating patterns. And in the early 1990s, Mike Bamshad came to do his pediatrics residency in -- at our school of medicine. He had brought some DNA samples from India to do mitochondrial DNA analysis on. And he asked to have some space in my lab so he was doing lab work while he was doing his residency. And that got me interested in genetic variation in India. I hadn't worked in that area at all before. And as I learned more about the so-called caste system, I realized that this was another mechanism that influence mating patterns in up to a sixth of the world's population. And I think Theodosius Dobzhansky described it as "one of the greatest, if unintentional, experiments in human genetics ever carried out," where people did mate according to a prescribed stories. So we were curious to know what effects that system might have on genetic variation, if any. We thought it was entirely possible that we would see no genetic differences, but there were some, relatively small, but detectable genetic differences among the different caste groups that we had collected. And furthermore, we could see that different caste groups had greater or lesser affinities to other populations -- reference populations that we compared them to. And I think probably our understanding of Indian history wasn't as good as it could have been or should have been, but I think that the patterns that we did see in at least that sample, might reflect greater common ancestry among some caste groups with central or western Eurasian populations, a greater common ancestry with other populations for other caste groups. And the work that we published in the early 2000s was largely confirmed by the much larger study that David Wright published in 2009, where he saw differences between what he called a northern and southern Indian component. So I think the work that we did on a relatively small sample, although it was several hundred people, has been largely substantiated by subsequent work.

One of the things we saw was that there is greater differentiation for Y chromosome polymorphisms, which are entirely male transmitted and inherited, than for mitochondrial polymorphisms. And one of the -- what that appears to reflect is that women can move in the caste stories more easily than men can. So you would predict that there should be greater differentiation for Y chromosomes and that's one of the things that we saw.

Well I think certainly our studies of genetic variation can inform us about population history because every major event that has occurred in the history of a population, gene flow --bottleneck and expansion, leaves its mark on our genome. So, when we have a collection of genomes now from a present day population, if we're clever enough, we can make some, I think, good inferences

that have historical relevance. But I think it's also important for people who are doing genetics to learn as much as they can about written history, about the archeological record, the fossil record. I think sometimes we're not as conscientious about that as we could be.

Well, I--we were aware that the caste system is a sensitive issue, and, in fact, technically had been outlawed, but still existed. And it, in some ways has some similarities to the "race issue" in the United States, and there is unpleasant baggage associated with it. We tried to be very careful to specify that our work was simply looking at the consequences of a process of social differentiation. It was interesting, and a little bit disheartening to us, to see some popular articles in the Indian press that kind of turned it around, and said, "Your genes determine your caste;" which of course we didn't say, didn't mean, didn't intend, and would completely disagree with. But I think this is something that does happen in genetics research -- sometimes our results are misinterpreted, not matter how careful we try to be, both in our scientific publications and in explaining things to the lay public.

Well, I think sometimes we have to try to anticipate what the public may perceive from a statement we make, which means being doubly careful not to sort of, take them down the path to an inference that we would never intend, especially with things like race or ancestry. Because science is often misinterpreted, and I think sometimes we can head that off just with careful language. I think we have to be as true to the data as we can be. I think we have to be as accurate as we can be, but in our interpretation we, especially with issues like race which can have truly harmful consequences for people, we just need to exercise extreme caution.

I think with each new tool one of the things that we want to do is to validate it by comparing the results from that tool with what we've seen before. With mitochondrial data, one of the things that was reassuring was that the patterns that we saw with mitochondrial polymorphisms largely replicated what we had seen with other kinds of markers, autosomal markers, but they gave us a different kind of picture, especially of the maternal lineage, and they actually allowed us to make inferences about dates, for example, of the common mitochondrial ancestor of humans; out of Africa events, and so forth. But, I think, comparison with, not just other genetic tools, but especially in population genetics, population history, really evaluating the other kinds of evidence for history, can help us to make the best use of those, of the new tools, as they come along.

Well, until really just the last few years, there was a huge class of variation we just really couldn't look at, and that was rare variation -- the variants that have arisen in the last five, 10, 15,000 years. We could look at subsets, we could look at samples of that variation, but with whole-genome sequencing we can literally look at it all. That gives us a much better window into things like population bottlenecks, expansions, and so forth, because rare alleles are very sensitive to those events. They give us really exquisite signatures of those kinds of events, so we have a degree of resolution that simply wasn't possible before.

Well, I think the real bottleneck now is more in the computational sphere. We, you know, we're amassing sequences now by the hundreds of thousands, by the millions. There are still populations that are unexplored, and I think there are interesting inferences to be made, but, you know, the big challenge now is a computational one. Even just storage for all of these sequences; we face this all the time. We get another 500 sequences, or 1,000 sequences, and how are we going to store all

of that, how are we going to process it. So we need clever methods for analysis; there's no question that that's where, I think, most of the real intellectual firepower is going be focused.

Well, certainly the whole process of informed consent has changed a lot in the last three or four decades. Consents that would have been acceptable 20 or 30 years ago now simply are not. And, I think, in general, although it does impose more burdens on the investigators, it does help to protect the people participating in the studies, to make sure that they do know how the data are being used, or will be used. And especially now when we can sequence whole genomes, we can look at every variant, every disease-causing variant, every susceptibility variant, it's important that people have adequate protection. When we're sampling populations, again the same principles of informed consent apply, and because of those, for those reasons, sometimes there are populations, certainly individuals, but sometimes whole populations, that just rather would not participate in these kinds of studies. And, you know, that's their right; it's, I think it's a lost opportunity for them in some ways, but still, we have to respect their preferences.

The two projects have very different sampling frames. So the Human Genome Diversity Project, or Human Genetic Diversity Project, really focused on very specific populations, often populations of specific historic or anthropological interest; sometimes populations that were really in danger of extinction. And part of the rationale was that these populations may soon be gone -- we should know as much about them as we can while we can, which, scientifically, makes some sense, but it was interpreted by some as saying, "Well, the real concern is just getting the DNA." I don't think that was the intent of Cavalli or any of the investigators involved in the HGDP, but that was an interpretation, and I think that worked against that project. The HapMap Project, very differently, sampled large, essentially cosmopolitan, populations. In the early discussions that I was involved with in planning HapMap, at one point some of the planners advocated no population labeling of the HapMap samples; the population geneticists in the crowd were opposed to that because we knew that linkage disequilibrium patterns [spelled phonetically] would vary among these major world populations, and to not know who was who would severely compromise what we could do with the information. So the, ultimately it was decided, and I think, I think wisely, to use basic labels so that we knew who was from which population. But I think because ethicists, social scientists, physicians, geneticists were all involved in the planning of HapMap -- those meetings involved hundreds of people -- I think there was a greater sensitivity to how the data could be misinterpreted, so a lot of the objections that arose with HGDP were kind of mitigated, just by the way the HapMap Project was planned. And it had the advantage of being planned later, and having seen some of the concerns that arose with earlier projects because of just a, maybe, a little bit of a lack of awareness of how things might be interpreted. I think HapMap was able to pretty much avoid that.

Yeah, I guess fitness is one of those examples of a term that is used in genetics and in population genetics in a very specific way; I mean, it can be measured numerically. But in the general population, I think in the lay population, when you hear a term like "fitness" it has a qualitative implication that some people are indeed more physically, or even cognitively, fit than others. And, really, in genetics, fitness simply refers to how many offspring you produce, and that, variation in that, can be due to all sorts of different factors. So that's a great example of the term that we use, with a very specific meaning, that can be grossly misinterpreted by the lay public.

The HapMap did have the very specific purpose of essentially informing our linkage disequilibrium studies to really help design the most effective genome-wide association studies. Now, the data have been used in a lot of different ways, but the real intent of the project, what the money was buying, was actually quite specific.

Prior to the HapMap, our knowledge of disequilibrium pattern throughout the genome was, was very, very limited. I've sometimes compared it to a map of the world in, say, 1500. We just didn't know a lot about many regions of the world, just like we didn't know a lot about many regions of the genome, or genomic variation among the population. So, HapMap really changed that; it turned out to be absolutely essential in designing effective microarrays. I think you could estimate that it saved hundreds of millions of dollars in what would have been unnecessary genotyping. So, personally, I view the HapMap as a very wise investment that helped us to economize many, many hundreds of thousands of genotyping assays after that.

Well, one could always quibble about which populations are chosen for analysis, and, to some extent, it's always a matter of what is doable and what's, what's, in some cases, convenient, what's available. But I think both the HapMap project and the 1000 Genomes -- and I'm probably biased because I was involved in both of them -- but when you look at the returns, when you look at the thousands of studies that have used the results from HapMap, the thousands of studies that have used, and will use, the results of 1000 Genomes, as reference panels, I think they've had, they both turned out to be very, very sound investments. And, you know, it's something that NIH can actually do. It can fund a larger program like that, that an R1 [spelled phonetically] simply -- you couldn't get an R1 to do that. But in supplying a resource for the entire world's population scientists, I think both have been remarkably successful, more successful than maybe we would have predicted back in the early days of HapMap.

To my knowledge, there wasn't really a pivotal moment, or an inflection point. I think, at least what I observed, was a gradual buildup of resources and knowledge. But, I'm not, I'm not sure that, like, not even sure that I'm the best person to address that particular question. But that's my perspective. I, you know, I -- it's very different from, say, the whole genome sequencing story, where suddenly, next-generation sequencing came on the scene, and that just changed everything. I mean, that was revolutionary, really. There are few scientific methods, I think, that can compare to that.

We still know, for example, relatively little about genetic diversity in Africa, and I know there are efforts underway, consortiums, to really increase our understanding of African diversity. But to me, that's a clear target for more projects, just because there is so much human diversity there, and so much to be gained from learning more about it. And there are many, a number of other parts of the world where that's also the case. So, I think sampling more populations in more places, that also will give us a better understanding of the continuous nature of genetic variation. You know, one of the early criticisms of HapMap was that we had one European, one Asian, and one African derived population, set of samples, and that that would reinforce notions that there are basically three human groups. Of course, that was never the intent of HapMap, but I think this broader sampling will help to disabuse people of the notion that there really are three boxes, or four boxes, or whatever.

But, you know, ideally we would have a sampling, this has been suggested, a sampling grid, you know, every 100 kilometers throughout the world. But the problem is you don't have an IRB every 100 kilometers, you don't even have populations, or people, every 100 kilometers. But we still, I think working toward a more even and unbiased sampling of human genetic diversity is a worthy goal, because it will give us a better, more unbiased view of that diversity, and it will also help to, to convince people that we don't belong in boxes.

Well, I became interested in mobile elements, I guess, first when I realized that they compose possibly as much as half of our genome. And anything that constitutes that much of the genome, and is almost completely not understood, merits investigation. Now, we became interested in them, at first, in part because they are really good evolutionary markers. Virtually all mobile elements insert once, and are never deleted, so we know that if two people share an alu insert or an L1 insert, at the same place, they almost certainly shared a common ancestor at some point in the past. So, as evolutionary markers, they're essentially ideal; we know ancestral state, we know the derived state. So that was an initial attraction. But the other thing is that they are just innately cool, in that they be copied and insert elsewhere in the genome. We know of cases of disease that have been caused by the insertion of a mobile element. One of the first ones was a neurofibromatosis [spelled phonetically] case reported in the early 1990s. And we still don't understand fully that mechanism of retrotransposition; how exactly does it work, how do the alus piggyback on the L1s for retrotranspositional machinery. And we still don't know that much about their effect on the genome. And I think that, that may be the most interesting question of all; what do they actually do, to what extent do they change gene expression, to what extent do they affect things like recombination, the mutation rate. They're, they're a very strange thing to happen to a genome, and we still don't know what the consequences of their hopping around are.

And so another interesting question is, this is almost like a host/pathogen relationship. They do things; what do we do, in turn? How do we activate, for example, methylation [spelled phonetically] to silence them and to minimize their impact? Why was there a burst of alu insertions in the primate genome 50 million years ago? What did that do, what did it accompany, what effects did it have? How did it get muted?