Just Mitochondrial DNA Episode Transciript

Introduction [00:00:05] Now that this recording, RTI International Center of Forensic Science presents just science.

Voice over [00:00:22] Welcome to Just Science, a podcast for justice professionals and anyone interested in learning more about forensic science, innovative technology, current research and actionable strategies to improve the criminal justice system and episode seven of the 2020 R&D season. Just Science Interviews. Kimberly Sturk Andreaggi, a research scientist at the Armed Forces DNA Identification Laboratory about the development of entire Marto genome reference data using an automated high throughput sequencing workflow. Disaster victim identification comes with a unique set of hurdles. Poor quality DNA, decades old samples and a lack of relative referential data can make the identification of victims from current and past conflicts challenging. This is the reason that the Armed Forces DNA Identification Laboratory uses mitochondrial DNA as their primary testing method. Stay tuned as our guest discusses the utility of mitochondrial DNA and a method for developing entire mighta genome reference data. In this episode of Just Science, this season is funded by the National Institute of Justice's Forensic Technology Center of Excellence. Here is your host, Dr. Megan Grabenauer.

Megan Grabenauer [00:01:39] Hello and welcome to Just Science, I'm your host doctor Megan Grabenauer with the Forensic Technology Center of Excellence, a program of the National Institute of Justice. Today, our guest is Kimberly Sturk Andreaggi, a research scientist in the emerging technology section at the Armed Forces, Medical Examiner Systems, Armed Forces, DNA Identification Laboratory. Kimberly, welcome to the podcast.

Kimberly Sturk Andreaggi [00:02:01] Thanks. I'm excited to be here.

Megan Grabenauer [00:02:03] I admit that I am not very familiar with the armed forces, medical examiner systems, armed forces, DNA Identification Laboratory. Can you tell me a little bit more about this?

Kimberly Sturk Andreaggi [00:02:12] Sure. The Armed Forces DNA Identification Laboratory, or AFDIL is a division of the Armed Forces Medical Examiner system, and AFDIL was established in nineteen ninety one as the only Department of Defense forensic DNA testing laboratory for the identification of human remains. There is a criminal laboratory for the military, but we are the only laboratory that focuses on the identification of human remains. AFDIL is an accredited laboratory according to international and U.S. standards. Some examples of how AFDIL has been involved in the past is Assal perform the identification of Michael Blassie in 1998, and that was the remains that were in the Tomb of the Unknown Soldier. And we've also helped with 911, numerous plane crashes, Hurricane Katrina and even the identification of the Romanovs in the past.

Megan Grabenauer [00:03:11] Wow. You get some pretty high profile cases that come through your laboratory and it sounds like, yes, both military, but also any federal cases that need identification.

Kimberly Sturk Andreaggi [00:03:22] And we also get to get involved in some historical cases as well.

Megan Grabenauer [00:03:26] So aside from that, the type of cases that might come through your facility. Are there any other key differences between how your laboratory operates compared to those that aren't associated with the armed forces?

Kimberly Sturk Andreaggi [00:03:37] The biggest focus that we have in our laboratory is that our DNA testing is focused on current and past conflicts solely for human identification. So like I said before, we're not a crime lab, so we're not really involved in testing the type of fluid that we're receiving. We're just typically getting single source tissue or bone from the medical examiners, which is a bit different than a lot of criminal casework because of the cases that we deal with. We are a pretty large laboratory. We have over 100 scientists and 20 laboratory. And because a majority of our samples are from those past military conflicts that involve remains that are decades old, our DNA samples are very low quality, although usually single source. Also, we don't typically have direct or close relatives for reference.

Kimberly Sturk Andreaggi [00:04:30] So that makes why mitochondrial DNA is our primary testing method at our laboratory.

Megan Grabenauer [00:04:36] So how did you end up at AFDIL?

Kimberly Sturk Andreaggi [00:04:38] When I was in graduate school at the George Washington University, one of my classes actually two of my classes were taught at AFDIL. One was taught by the medical examiner. So that was my pathology class. And then the other one was a DNA analysis class, and that was taught by a scientist at AFDIL. And at the time when I was there, they were advertising that they were hiring part time employees to work on an NIJ grant on mitochondrial DNA database thing. And I was able to get that job. And I that was back in 2004 and I've been there ever since.

Megan Grabenauer [00:05:14] As I understand it, you are currently a graduate student pursuing your Phd. Correct? Yes. So what university are you attending for graduate school?

Kimberly Sturk Andreaggi [00:05:23] I am getting my Phd at Uppsala University in Sweden with Dr. Marie Allen.

Megan Grabenauer [00:05:30] So you're working full time at AFDIL while simultaneously pursuing your Phd. How does this arrangement work?

Kimberly Sturk Andreaggi [00:05:39] Yeah, it's I'm actually really lucky that I was able to get support from my supervisor and the staff at AFDIL, but I'm able to do a lot of the work for my dissertation while at AFDIL, so a lot of it coincide together. So I work on the research projects while I'm at AFDIL and then occasionally I do have to go to Uppsala usually once or twice a year to take required courses, but I love doing it. So and it's really enjoyable to go and get a little bit of that college atmosphere every once in a while. I'm usually only there for a week or two.

Megan Grabenauer [00:06:19] I'm curious for some of our listeners that are in various stages of their career progression. You had been at AFDIL for a number of years. You said you've been there since 2004. So what was your motivation to pursue a Phd at this point in your career?

Kimberly Sturk Andreaggi [00:06:37] I think it was just the opportunity to gain more experience and more knowledge maybe in areas that I wouldn't normally focus on. Also, I had a number of colleagues and friends. I had done something very similar while working at AFDIL. They went back and got their Phd. So that was a really motivating factor for me. Also, I think just learning new things, but also that in the future, if I have my Phd there certain opportunities that I might be given that I wouldn't have access to if I didn't have it.

Megan Grabenauer [00:07:12] Well, we're talking here today primarily because you recently presented results from an AIG funded grant at the American Academy of Forensic Sciences annual meeting. Your presentation was entitled Development of Entire Mito Genome Reference Data using an automated high throughput sequencing workflow. And that was presented as part of the NIJ Forensic Science R&D symposium. If listeners are interested in watching the archived recording, that presentation can be found on forensiccoe.org or the landing page for this episode. And before we get into the details, are there any other researchers in this project that you would like to acknowledge?

Kimberly Sturk Andreaggi [00:07:51] Yes, absolutely. So first off, Dr. Sharla Marshall is the primary investigator on the project. And then the real two workhorses of the project are Joe Ring and Kathy Taylor. These are the analysts that do most of the work on the project, including automation and the analysis, though the project really couldn't run as smoothly as it does without them. And then Dr. Walter Parsons, the big part of this, he is the head of IMPOP, which is NNAP mitochondrial DNA population database. And they're based out of Innsbrook. And this is a global mitochondrial DNA databank. And that's where actually where we end up submitting all of our new data. And it gets Q.C. check there as well.

Megan Grabenauer [00:08:38] Let's start with the very basics here. What is a mito genome?

Kimberly Sturk Andreaggi [00:08:44] So the mitochondrial genome or mito genome is the entire portion of the mitochondrial DNA genome. And this genome is a bit different than nuclear DNA.

Kimberly Sturk Andreaggi [00:08:54] The mitochondrial genome is circular and it's only sixteen thousand five hundred and sixty nine base pairs.

Kimberly Sturk Andreaggi [00:09:02] Historically, mitochondrial DNA testing focuses only on a small really variable portion of the whole Nido genome. So only about a thousand base pairs. But with the newer technology, it makes it possible to look at the whole mitochondrial genome instead of just a small portion.

Megan Grabenauer [00:09:21] How is the analysis itself different? Or is it different in looking at mitochondrial DNA compared to nuclear DNA or other types of DNA?

Kimberly Sturk Andreaggi [00:09:30] So mitochondrial DNA analysis typically is done through sequencing. And in the past, it's been primarily done with something called Sanger sequencing, which is just a method to be able to detect each base in the sequence. But this method is pretty laborious and time consuming compared to what's typically done for something like short tandem repeat analysis so that a bit more complicated. But there's added benefit to mitochondrial DNA analysis. The main benefits of mitochondrial analysis is that the mitochondrial genome is present in a much higher copy number and myself than the nuclear genome. So the micro genome is present in every mitochondria in each cell. So this could be hundreds of copies personnel versus

where each cell only has one nucleus. So the nuclear DNA is at a much lower coffee number than mine. So this is really beneficial for us for degraded samples like age, bones or teeth, because there's so many more copies available for testing. And it's also really helpful for hair where there's little nuclear DNA and so might go into being the only way to get DNA and data out of a hair sample. And the other main benefit that really differentiates mitochondrial DNA from other types of DNA is it's inheritance pattern. So mitochondrial DNA is inherited from the mother, is passed down from the mother to her children. And what this does for us is that any maternal relatives will actually share the same mitochondrial DNA. So for decades, old cases like the human identification efforts that we do for past military complex, the fact that we can have an individual who might be a fairly distant relative, they may share the same mitochondrial DNA and we can actually use them to identify the individual that's missing.

Megan Grabenauer [00:11:37] OK, so you've sold me on the benefits advantages of mitochondrial DNA to the point that now I'm wondering why would you ever do anything else?

Kimberly Sturk Andreaggi [00:11:49] Yes, the main negative of mitochondrial DNA testing is because everyone who is maternally related shares the same mitochondrial DNA. The discrimination power is not as good as autosomal markers. So where in when you're watching CSI, they'll say, you know, it's one in a trillion chance that it's this person and they're using short tandem repeat with mitochondrial DNA. You might get a statistic more like one in a thousand. So the discrimination power is much lower, but there are other advantages that make it beneficial for certain applications.

Megan Grabenauer [00:12:29] That makes sense.

Megan Grabenauer [00:12:30] You mentioned previously mitochondrial DNA was analyzed using Sanger sequencing, but now you're doing something different. What is the new method that's routinely used?

Kimberly Sturk Andreaggi [00:12:42] We're currently now using something called next generation sequencing or it's sometimes called massively parallel sequencing. And this just described. Some new sequencing methods, a new being probably around 2005 where these new technologies started to be commercially available, but these methods involve Cronal amplification and then sequencing by synthesis. So all this means is that you actually collect the DNA sequence as the DNA is actually being replicated. So Sanger sequencing is done differently. You you synthesize the DNA and then you run it through a machine and collect the sequence. However, next generation sequencing, we actually collect the DNA sequence as it's being made. And this is done on a massive scale. So millions to billions of DNA sequences are collected in each run.

Megan Grabenauer [00:13:37] So I guess it ends up being a much faster process. Is that in results?

Kimberly Sturk Andreaggi [00:13:42] Yeah, faster and at a much greater throughput. So what was having this massive amount of data being able to be generated? You can combine many more samples together and many more markers together. So you can do mitochondrial DNA. You can do audism. More short tandem repeats. You can do Y chromosome all shorthand and maybe you can do this all in one single reaction. So it becomes much more cost effective to generate this data.

Megan Grabenauer [00:14:13] So I understand that AFDIL was one of the first laboratories in the U.S. to validate and implement next generation sequencing for your work samples. You've been doing it since 2016. So can you tell me, do you know since you've been there for that whole time, what influenced the decision to adopt this new methodology and then what the impact has been to your casework?

Megan Grabenauer [00:14:34] Yes. Is this the reason why we started looking into next generation sequencing, which started probably in 2009 because we had a set of chemically treated samples. So these were a set of about eight hundred individuals from the Korean War that had been processed through the Kokoro March and Japan before they were brought back to the U.S. for burial at the National Memorial Cemetery of the Pacific, which is located in Honolulu, Hawaii. And in this particular sample set, we had a success rate of only about 6 percent, and that's compared to over 90 percent for all of our other samples for mitochondrial DNA testing. So the main issue with these samples is that they had been chemically treated in formaldehyde, which left the DNA extremely degraded. So less than one hundred base pairs or fragments of less than one hundred base pairs and extremely damaged. The other issue with these samples is that there almost was no human DNA. There was tons of bacteria. So it made it really hard to use typical processing methods. And so what next generation sequencing offered us is a sensitive method that was amenable to very small fragments. These methods had already been used by the ancient DNA community and we used them and applied them to these really poor samples. And we're able to start getting DNA from samples that had been previously on successful with traditional methods. So we back in 2016, we started out with only seven individuals that were trained. And now we have over 20 people trained in NGS. And as of this January, we've processed roughly sixteen hundred samples, which we process everything and duplicate. So this actually represents about 3000 reactions. And now we're seeing a success rate with these samples of about fifty three percent, which doesn't seem like a ton, but compared to six percent, this is a huge increase. And these analysis have also contributed to over 140 identifications.

Megan Grabenauer [00:16:46] Wow, that is a big difference. Yeah. Yeah. Thank you. I appreciate the background information about DNA sequencing and testing and mitochondrial DNA. So let's talk a little bit more about the research that you presented at the academy meeting. Can you just start with some kind of overview of what you've been working on with this NIJ grant?

Kimberly Sturk Andreaggi [00:17:08] Sure. So like I mentioned with for our casework, we're now transitioning from only looking at a really small portion of the mitu genome. And now with these new technologies, we're able to look at the whole Nido genome. But one of the main limitations to doing this is that we need reference data to be able to compare the data that we obtain to find out how frequent that particular DNA sequence is observed in the population. And that's what we use to create statistics and to give a value to a match. And so at the time when we applied for the grant, there were only about fifteen hundred mito genome in EMPOP and as I mentioned EMPOP is the EDNAP Mitochondrial DNA population database. And that's one of the global. That's probably the main population database that's used by the forensic community for mitochondrial DNA. And in comparison, at the time, there were over forty thousand profiles of that smaller part of the Mito genome. So we wanted really to be able to take full advantage of using the mito genome and being able to use NGS. We needed to increase the number of mito genomes that are available for this statistical analysis. So the grant is focused on generating five thousand mito genomes using NGS.

Megan Grabenauer [00:18:38] All right. As I understand it, you're hoping to get about 4000 of those from the U.S. population and 1000 from other parts of the globe. So what level of diversity can you expect or are you targeting with that number of samples with with 4000 samples from the US?

Kimberly Sturk Andreaggi [00:18:58] We're hoping to look at five different sub populations within the U.S. So Caucasians, African-American, Hispanic, as well as Asian Americans and Native Americans. And by looking at these different sub populations, we can look and see if there is a particular haplogroups, which we know there are particular profiles that we expect to see in these populations. And so it's really important to have databases that are set up for each population because a particular profile might be rare in one population, but would be more more common and another. So we want to make sure that we have a good enough variation both geographically. So we're targeting different states as well, but also different populations so that we can ensure that there's data generated for any casework search that needs to be performed to create these statistics.

Kimberly Sturk Andreaggi [00:19:54] And then for the global population, we're looking at generating data from countries of interest both on AFDIL side, but also that are pertinent to the US like Central America or other parts that might be underrepresented EMPOP.

Megan Grabenauer [00:20:13] I know that that's the size of your database has to be pretty important, right, for getting an idea of how strong your identification is. Do you have any idea? Can you give an estimate of how large a database you need?

Megan Grabenauer [00:20:26] Yeah, that's actually a really good question. As far as I know, we can never have enough in the database for happily markers like mindo and Y chromosomal markers because they're always suffering from underrepresentation. So we're never able to get every profile that's in the world. So for us, essentially we would just want to try and collect as much as we possibly can. So as I was mentioning, EMPOP has over 40 thousand of that smaller mightve profile in their database. So ideally we could we would like to get closer to that number. But as of right now, there's only about five thousand in and pop of the micro genome. So we're just trying to get as many as we possibly can.

Megan Grabenauer [00:21:12] So with your global samples, are you targeting any specific regions or populations that you know, there are a lot of people currently living in the U.S. that may be descended from a specific region or area? Is that factor into your sample?

Kimberly Sturk Andreaggi [00:21:27] Right now, we're just trying to hit the different regions. So, you know, Asia, we're trying to get, you know, a few populations from Asia, a few populations from the Middle East, from Europe, as well as Central or South America and then also Africa. So Africa is pretty underrepresented in terms of mitochondrial DNA. So trying to get more samples from those kind of underrepresented regions is one of our goals.

Megan Grabenauer [00:21:55] All right. You explain in your presentation that there may actually be more discriminatory power in the coding regions versus the control regions that are typically analyzed using standard mitochondrial sequencing methods. Can you expand on this a little and explain what this means for using these regions for human identification purposes?

Kimberly Sturk Andreaggi [00:22:13] Sure. So the control region is a pretty small portion is about twelve hundred base pairs. So when we look at our population, oftentimes there are common profiles that we see. So for example, in the U.S. Caucasian population, there is particular profile that scene and roughly 7 percent of individuals. So this means that if you get that particular profile and you have a match to a missing individual, that statistic or that match isn't as strong as it could be. And in fact, only about. Seventy five percent of those control region profiles are unique. So twenty five percent of them are seen more often. However, if we're expanding out to the whole Mito genome so that additional almost sixteen hundred based sixteen thousand base pairs, nearly 95 percent of mindo profiles are unique. So this gives us a much greater descrimination power and allows us to have more identification to be made using mitochondrial DNA because that statistic, that frequency of the profile will be much less.

Megan Grabenauer [00:23:27] And you also say that you're using a PCR free library preparation. Can you describe that a little and what advantages that gives you?

Kimberly Sturk Andreaggi [00:23:35] So eliminating PCR, which is just a way of replicating a particular DNA sequence, lots and lots and lots of time. The benefit of removing this from library preparation is we can reduce error and bias so often during PCR. Mistakes can happen as the DNA is being replicated within the reaction. And this can even result in the wrong nucleotide ending up in the DNA strand. And then if this gets replicated over and over again, we can see that PCR error in our data. And then for bias, there's certain types of DNA that may amplify better than others. Certain regions. So you might see a loss in coverage in one area of the mito genome compared to another. Although we do use PCR to enrich our samples, so we make sure that there is mitochondrial DNA present in our sample prior to library preparation. There's also some PCR during actual sequencing, but we want to try and eliminate as much PCR as we can that we can minimize any error or bias that might be introduced during these steps.

Megan Grabenauer [00:24:45] So you still do PCR, you're just minimizing the extent to which you do those amplifications?

Kimberly Sturk Andreaggi [00:24:50] Correct? So most library preparation, which is just what has to be done to the samples in order to sequence them on these next generation sequencing platforms. Most of them require an additional PCR step. And basically the method that we're using move that from the method. And it allows us to just eliminate one additional step where we could be introducing error.

Megan Grabenauer [00:25:15] All right. Now, I don't know a lot about next generation sequencing, but from what I've heard anecdotally, it's much more laborious and requires a lot of hands on steps. As I understand it, though. You've got some automation in place. Can you talk a little about what your strategies are for automation and how they have impacted your implementation?

Kimberly Sturk Andreaggi [00:25:36] So many of the steps that we use in in this particular method or for for next generation sequencing methods aren't that different from traditional methods that we use either for short tandem repeat typing or even Sanger sequencing. However, probably the biggest difference is that next generation sequencing has a lot of purification steps that require speed washes. So these are little beads that are used to collect the DNA and then we're able to clean the DNA and get rid of anything we don't want before moving on to the next step. And these are really not fun to do when you're processing an entire plate of samples. It can take, you know, two to three hours and

you're in there the whole time. Hands-On doing these bead washes and also requires pretty precise pipetting. And the other steps that are pretty laborious are also the number of normalization steps, though, where we try and get all the DNA to be the same concentration and quantification steps. And so what we've done with our automation is we've tried to automate as much as we possibly can. So that includes all of these small normalization quantification steps, all the bead washes, in addition to the more reactions that are required for library preparation, which are pretty simple, these are just either PCR or just adding summary age entry for an incubation. But what we've tried to do is automate the entire process to allow us to have it nearly 100 percent hands off as soon as our initial amplification is completed.

Kimberly Sturk Andreaggi [00:27:11] And we currently have validated a methods for our family references in casework, but this method is low throughput, so we only allow them to do about 24 samples. And that's because it's done manually and because it is so laborious and time consuming. But now we've taken the method that we're using for the NIJ grant and we're in the process of validating it currently for casework, for the family reference processing, and that will allow them to have a high throughput method for next generation sequencing of the mito genome.

Megan Grabenauer [00:27:45] Nice. So then a lot more throughput and larger portions being analyzed I'm assuming that generates a lot more data now than you were previously working with. So can you explain what AFDIL is using to analyze the data and then how the analysts are adapting to this different form of analysis?

Kimberly Sturk Andreaggi [00:28:06] Yeah, I think this is actually the biggest challenge to implementing next generation sequencing in casework laboratories. There is just so much more data. So, you know, as an analyst, you're used to being able to look at every peak and be able to make a decision. However, with the next generation sequencing, that's really not an option. There's just so much data. And as you mentioned, looking at the mito genome, instead of looking at six hundred base pairs, you'd have to look at sixteen thousand base pairs, which would take forever. So what we've done is we partnered with CLC, which is a company now under QIAGEN and developed our own specialized tools for the analysis of mitochondrial data. So we perform all of our analysis in the field, see genomics work bench and then apply the special tool to apply international conventions. So the conventions that are used across the U.S. and globaly to analyze mitochondrial DNA, it also allows user edit and other QC measures, quality control measures to ensure that the data is of the best quality.

Megan Grabenauer [00:29:20] All right. So you've got your analysis methods worked out. It sounds like you've you've settled on a robust data analysis algorithm in how you're handling that. How is the research itself going?

Megan Grabenauer [00:29:32] Where are you in your goal of collecting 5000 entire mito genomes for your database?

Kimberly Sturk Andreaggi [00:29:39] We're more than halfway done as of right now. We have a number of samples that are actually at an pop rate now getting their final QC. So we are anticipating to be on target to have all of the data generated by the end of the year to meet all of our deliverables for the grant.

Megan Grabenauer [00:30:01] How is that data going to be used by the community? Where's it going to go? How is it going to be accessed? Like, how is this going to impact day to day casework for other entities or AFDIL?

Kimberly Sturk Andreaggi [00:30:13] Yeah. So our our goal at the end of all of this or our plan to disseminate all the data is first off, it'll all be uploaded into pop, which is pretty much where most laboratories do generate those frequency statistics. So it'll be available there. But we'll also make it available through publication where individual laboratories wanted to take that data and put it into a database, their own internal database. They would be able to do that. And in fact, we're right now in the process of putting together a publication that'll include a number about sixteen hundred of the NIST population samples.

Kimberly Sturk Andreaggi [00:30:54] Another three hundred samples from Colorado and then another three hundred samples from around the US. And all of this will be added into a single publication that would be available to other laboratories to be able to have access to that data.

Megan Grabenauer [00:31:11] What do you see as the next step?

Kimberly Sturk Andreaggi [00:31:14] We're hoping just to continue to keep generating more and more mito genome. One thing I think is that as laboratories start implementing next generation sequencing for other markers, the mighta genome now becomes more accessible. So laboratories in the past haven't really been able to unless they had very specific applications. Mito was something different and would require more step in the laboratory. So a lot of labs didn't implement it. But now with next generation sequencing, you can now get that large number of other markers, but you can also be able to access the micro genome. So with the generation of these reference data set now laboratories we'll have there won't be any limitations to being able to implement mitochondrial analysis in their laboratory where there might have been limitations in the past.

Megan Grabenauer [00:32:08] Well we're running near the end of our time together. But before we conclude, are there any other final thoughts or aspects of your research that you want to share with our listeners?

Kimberly Sturk Andreaggi [00:32:18] No, just thank you for letting me talk about the NIJ grant that we're working on and also about mito usually get to and it gets to be the lost genome I think sometimes. But you know, at AFDIL we we really rely on on mitochondrial DNA and it's it's power for our application.

Megan Grabenauer [00:32:39] Well that's all the time we have for today. So I'd like to thank our guest, Kimberly Sturk Andreaggi for sitting down with just science to discuss her NIJ funded Grant. Thank you, Kimberly.

Kimberly Sturk Andreaggi [00:32:50] Thank you.

Megan Grabenauer [00:32:51] And I'd also like to thank you, the listener, for tuning in today. If you enjoy today's conversation, be sure to like and follow just science on your podcast platform of choice. And for more information on today's topic and resources in mitochondrial DNA field, visit forensiccoe.org. There you'll find additional webinars, guidance documents, reports and conference information.

Megan Grabenauer [00:33:13] And also, please follow the Forensic Technology Center of Excellence on Facebook, Twitter, LinkedIn or sign up for our newsletter for release dates on resources. I'm Megan Grabenauer and this has been another episode of Just Science.

Voice over [00:33:29] This episode concludes our 2020 R&D season, be sure to stay tuned and keep an ear out for our next season. Digital evidence, opinions or points of views expressed in this podcast represent a consensus of the authors and do not necessarily represent the official position or policies of its funding.