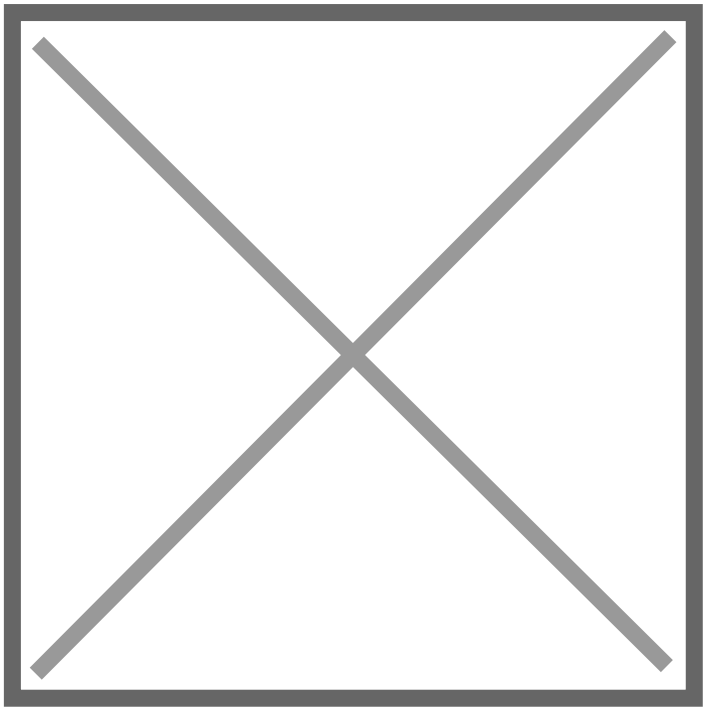
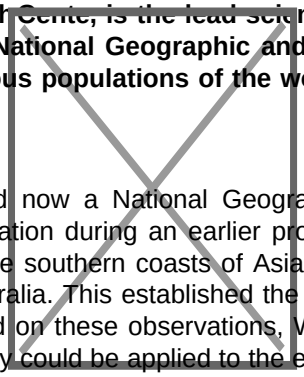
BioSpectrum
the business of Bi
24 years of leveraging the business of Bio
www.biospectrumindia.co

"The genographic project will allow individuals to trace their own migratory history"

07 July 2005 | News



Ajay Royyuru, head, Computational Biology Center at IBM's Thomas J Watson Research Cente, is the lead scientist for IBM on the Genographic Project. He is working closely with Dr Spencer Wells of National Geographic and the distinguished team of field researchers assembled to conduct DNA testing on indigenous populations of the world. In an exclusive interview, he speaks about the project.



How did the concept of the Genographic project take shape?

The Genographic project really began with Dr Spencer Wells, a population geneticist and now a National Geographic explorer-in-residence, when he was applying population genetics to understand human migration during an earlier project. Wells was then studying the genetic markers on the Y chromosome in populations across the southern coasts of Asia and was able to connect the markers with those occurring in the populations in Africa and in Australia. This established the land migration of people from the African continent to Australia some 45-50,000 years ago. Based on these observations, Wells wrote a book called "The Journey of Man". At that point of time he wondered if this methodology could be applied to the entire population of the world. Since it would be impossible to get data from all the six and a half billion individuals of this planet, so if data could be collected from a few million people, then how would that story be. That would be the story of migration of people in the eternity of time all the way from the center of origin of humans in Africa to present day in all corners of the world. It would tell a tale of how human beings have developed as a species in the last 200,000 years.

How is the project being executed?

The project was officially launched on April 13, 2005 at an event in Washington DC. We are now going around the world and telling people in each of the geographies about what the project is going to be, inviting their comments and listening to them. The Genographic project, an innovative five-year research partnership, will use sophisticated laboratory and computer analysis of DNA contributed by hundreds of thousands of people. The project is a partnership between the National Geographic and IBM, with support from the Waitt Family foundation and the public.

Participation from the indigenous populations and the general public forms the core of the project. Wells and a consortium of scientists from prominent international institutions will conduct the field and laboratory research. It is aimed that least 100,000 blood samples will be voluntarily collected from the indigenous populations of the world, who are a key aspect of this project. They have a story to tell about ancestry. It is their distinct existence, cultural and genetic distinctness that provides the clearest signpost in this historic view of human migration. Their DNA information will be collected via the blood samples. For this purpose, ten research laboratories have been identified around the globe, each affiliated with a university or similar institution which would sample indigenous population in a particular geographic area. An international advisory board will oversee the selection of indigenous populations for testing as well as adhering to strict sampling and research protocols.

The public can take part by purchasing a participation kit and submitting their own cheek swab samples. The public participation samples are being processed by Family Tree DNA Inc, in association with the university of Arizona and the

results will be posted on the website in about four to six weeks time. Then one can log on the website and track the overall progress of the project as well as learn their own migratory history. As more and more data is gathered the patterns get richer and richer. In all the samples the individual results focus on a small number of genetic markers in the non recombinant Y chromosome of males and the mitochondrial DNA in females.

What does the project aim to achieve?

The objective of the Genographic project is to trace the migratory history of all people on the planet. The project is actually trying to connect many pieces of evidences, the genetic evidence being only one of them - anthropological, cultural, linguistic, geographic; we have had all this kind of data as handed down information. Most of us have the knowledge about who our ancestors were going back to a few generations. Beyond that we have information through oral history, our culture, our scriptures, etc. But we do not have any way to trace our ancestry to any migratory event that has populated our subcontinent. The question that we would like to answer through this project is 'which was that migratory event that has populated my ancestors in this land and where did they come from? And how did my ancestors populate other parts of the world?'

Essentially the whole world is one big family and we only have to go back in time to recognize this fact. The diversity that we see crossing this map is the result of this migration and how the paths have crossed of various people. Merely 15 thousand years ago we had a common ancestor and today we are all in different parts of the world and speak different languages. This project provides an opportunity to recognize the similarity that exists between us and to appreciate how related we all are.

What is the role of IBM in the project?

IBM is a partner in the Genographic Project. It is contributing very heavily both towards the design and the execution of the project. IBM is putting the technology infrastructure in place in all the 10 regional centers. We are providing computer systems, softwares, basically the capability to gather the data. Then we are using very sophisticated communication technology, encryption packaging, for sending the data from all the regional centers to the center that we have set up in Washington DC, National Geographic headquarters for storing the data and a good amount of computing infrastructure to analyze the data. The challenge for IBM is in the data analysis because of its scale. What we are really trying to tell is that of all the diverse populations that we see, what is our shared ancestry and how did that diversity come upon. We should be able to weave the fabric of human history of migration and the journey that our ancestors traveled by getting the genetic threads and tying them all together.

If one thinks about the genotype as a set of markers that are being used on an individual and the phenotype as all the cultural traits: languages that you speak, geographic location that you are in today and so on. We are looking for correlation in this space and they constitute an interesting observation in this exercise.

How is it being ensured that the data generated from the samples will not be misused?

As we are asking people to volunteer perhaps their most personal information - their own DNA, so we have to make sure that it is held with the highest degree of security and privacy. We ensure this through three means. First we gather only that information which is useful for the project and we are not asking for any information other than that. Like for example we are not asking for any medically relevant information. We are not genotyping volunteers for markers that might be indicative of disease traits or any medical outcome. We are only looking for markers that speak of deep ancestry. These essentially are markers on the Y chromosome and on the mitochondrial DNA. The only other information that we know about the volunteer is their gender and that too is self-confessed.

The second aspect is the anonymity of participation. We do not ask volunteers for their name. Each kit has a random code assigned to it on the inside and we have no clue which kit went to whom. So when a person receives a kit, the code is the identifying number, which the person will use to track his data in the project.

The third is our commitment to the project of being very public about anything we do. We will publish research results as and when we arrive at results and will disseminate them through the National Geographic website, channel, magazine and all other means of communication available to us. As the project progresses, we will put all the gathered data in the public domain and will not be patenting any information that is generated out of this data. Our being very public will prevent any accidental abuse of the data by any organization or individual.

Lastly in all the regional centers, the IBM thinkpad provided to collect the data is equipped with a biometric fingerprint scanner in the machine enabling only the authorized people to enter and have look at it. This is another way by which we are ensuring that the data is secure. In this way we are going to address the concerns that people as to how are we going to protect their most personal data.

Rolly Dureha