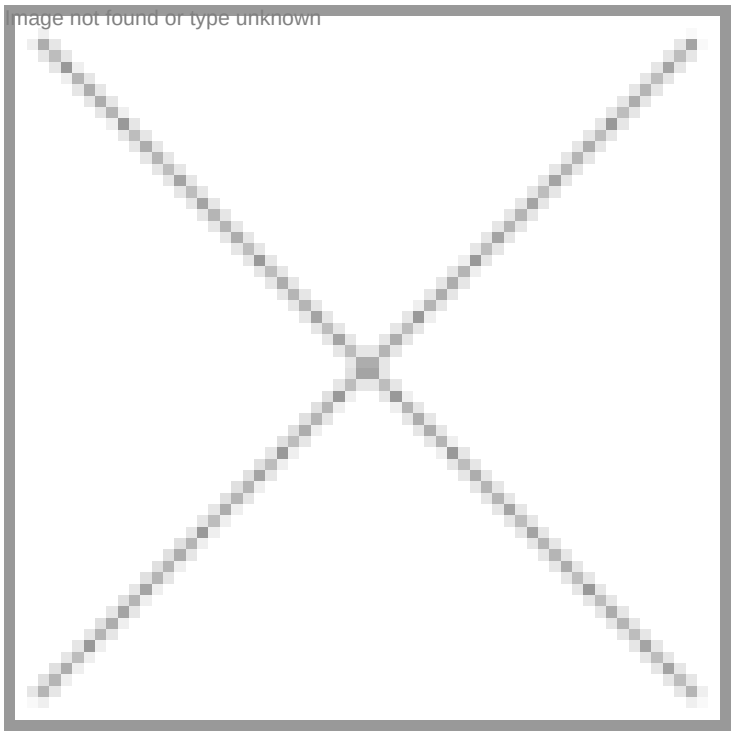
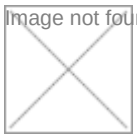
BioSpectrum the business of Bio & Health Sciences

Reader of the 'Book of Life'

10 April 2003 | News





In 2000, he lost narrowly to the then US president-elect, George W Bush, the title "The Time Man of the Year." But Dr Craig J Venter's influence will go beyond the immediate happenings of our life. For, called "Gene Mapper", by the Time magazine, his scientific work in completing the mapping of the human genome is an attempt to read the secrets of the humanity itself. This 56-year-old biochemist from University of California, San Diego, and then the chief of Celera Genomics, may have shared the fame with a government scientist Francis Collins when they shared the podium at the White House in June 2000 to announce their simultaneous mapping of the human genome. But the world knew it was Venter's "shotgun sequencing method" that made it all possible. One of the most cited biologists of modern times, Craig J Venter was in New Delhi in March to speak at the Millennium Summit -III organized by ASSOCHAM. In an interview with Executive Editor N Suresh, Venter unraveled some of the secrets of human genome

" You are not what your genes are"

What is the progress in the identification of the functions of human genes?
From the genome mapping, we have concluded that humans have only about 30,000 genes and not 140,000 or so as it was thought by scientists earlier. Now efforts are on to determine the functions of each of these genes. We know all about some 9,000 genes. We know some of the functions of some 5000 genes. We don't know anything about the remaining 42 per cent of the genes.

For each species, for each set of genes, there is a precise set of environmental conditions that impact them. You are certainly not what your genes are. When we are studying the genetic code, we are only studying at best half the question. Humans

have 100 trillion cells and they work together in mass with genes. So it will be some time before we even understand the basic functions, let alone how all of them interact effectively with the environment.

Where does this leave the theory of genetic determinism which has been popular for decades?
Each human has roughly two to three million of genetic code that are different from another person. But most of these codes probably have no biological significance whatsoever. They may have just some forensic and tracking applications, and maybe useful as mapping tools. Less than a percent of these codes occur in genes or regulatory regions. This means that at the biological level, at the gene level, we are all virtually identical twins. This clearly means that your life not determined by your genes. The fraction that is really different in a biological and meaningful way is such a tiny percentage it is stunning. The smaller number of genes and the similarity of human genes to that of a mouse hopefully put the nail in the coffin of the genetic determinants.

What is the next big thing after the genome mapping?
We have some 100 trillion cells in each of our bodies. This all interact with each other in many complex ways and produce a variety of proteins. With the information about the human gene sequence, it is possible now to measure these proteins. Because we can get at every protein in the genome. For the known useful genes of around 50,000 there are some 200,000 to one million proteins. We are building a large-scale protein facility to do roughly one million protein sequences per day. We are looking at using these proteins as markers for specific diseases.

The genome is important because with mass spectroscopy sequencing, the proteins get blown apart into small fragments and we can compare those sequences back with databases. Till now most of these did not match anything on the databases and so interpretation was difficult. Now every one of these will have a match and we can rapidly determine the sequence of the proteins in the cells and in the blood.

Does genomic information provide clues to our past?
Many scientists believe that evolution is adding on genetic information and adding on complexity. But we find that most human pathogens probably started from a much more complex organism and threw out genetic material during evolution. We tried to find out whether we could come out with a molecular definition of life. To test this we knocked out 200 or so genes from Mycoplasma pneumoniae and observed whether we still have a living organism. And with Mycoplasma genitalium (a small microbe found in the human genital area) we asked the question whether all its genes were necessary for life. And we found that the 200 extra genes in pneumoniae were completely dispensable and about 200 genes in genitalium too appeared dispensable.

We found out three stunning things. First of the 300-odd genes, 103 were completely new to science. We think we know a lot about biology, yet in this most minimal cell, we had no idea about what one third of the genes do, except that if we knock them out, the cell dies.

Two, we could not come up with a molecular definition of life. We found that life is context-sensitive, that is, the environment that the cell is in is equally important to any component of the genetic code. Third, the Darwinian evolution wasn't just random errors in the genetic code. With haeomphilus and every pathogen we have worked on, we found there was preprogramming in the genetic code to cause change in the structure of specific genes. Essentially a pathogen fools the immune system of its host. And each pathogen has a different way of doing it.

Essentially, the human genome contains a molecular clock as repeated sequences of letters between genes and mutates over time. Measuring the rate of these changes would some day allow us to predict the time at which it happened. This may help to pinpoint the epoch in which the traits that make us unique humans emerged.

You have been talking about the development of personalized medicine. What are these?
With the human genetic code, we find roughly two to three million variations in the chromosomes. We have about 2.8 million well characterized, so called snips (single nucleotide polymophisms) in our database that are now being used by scientists around the world to study linkage to disease.

For the first time, we can look at this genetic variation by chromosome. For example, you can discover genetic variation in the genome of individuals who have an increased risk of myocardial infarction. Scientists are using this information to find linkage to disease. The pharma industry is using this to find ways to improve clinical trials and drug effects. This leads to the development of personalized medicine.

Recently, there was a type II diabetes drug that had to be taken off the shelves because one out of 10,000 people had a severe liver toxicity to this drug. If we can find simple tests that predict toxicity, it will have a huge impact. Most drugs work only on 30 to 50 percent of the population but are given to everybody because in theory they are relatively safe and doctors

work on the principle that it is easier to give it everybody than find out whose disease it can actually cure. Personalized medicine would actually try to treat the patient and not kill him inadvertently.

What is the uniqueness of your 'shotgun sequencing' method?
All the sequencing technologies give us only 500 to 600 letters of genetic code. So how do you get the genetic code of something billions of letters long. The other problem is that genomes, including the human genome, have lots of repeats. The same sequence is repeated over and over again. In one method, scientists would sequence one piece, get 600 letters, make a little primer to read the next set of 600 letters and work sequentially down the clone with that. By this method, it would take 100 years to sequence the human genome. In the second method, genes of the first clone are mapped, then lined up and sequenced. This way, it took three years to just arrange the E.coli clones before sequencing could be done.

With 'whole genome shotgun'', we don't do any cloning first. We take the entire set of chomosomes, break those down into little pieces and use mathematical algorithms to try and solve the big jigsaw puzzles.

Why is it important to study mouse genome and not that of chimpanzees?
Studies show that the human and mouse genomes are almost identical genetically. This enhances the chances of predicting the regulatory regions in the genome and we will have comparative genomic studies. On the other hand, data generated at Max Planck Institute using the human genome data indicates that the difference between the two is just 1.27 percent in genome sequences. So it is better to study the mouse genome which evolved 100 million years ago and it will provide a better comparison on the evolutionary changes that had taken place.

We have to look closely at our own evolution and that of vertebrae. From the simple drosophila (fruit fly) to humans, the evolution has taken place in about 600 million years. And just four to five gene groups only expanded in this period. For example, humans have an immune system, while the fruit fly doesn't have a significant immune system. All the genes associated with our immune system expanded during this period. When we compare with homeostasis, all things associated with our vascular system expanded during this period. But we have virtually identical gene set with mice and other vertebrates. The key to our uniqueness are things like the transcription pattern in gene regulation, that turn on different sets of response to environmental conditions. Among all human chromosome, the chromosome 19 has the largest number of genes.

Because it has neurotransmitter receptors, olfactory receptors and transcription factors. In the future, we will be able to provide a date in some cases of species, where this evolutionary event took place.

What is your work related to synthetic cells?
We are now trying to develop a synthetic cell just to see how many essential genes are required to form a living thing. For example, mycoplasma lives on glucose or fructose. If you knock out the glucose transporter gene and you have both sugars in the environment, the cell will happily live. But if you only have glucose in the environment and you knock out the glucose transporter gene, the cell dies. So for each species, for each set of genes, there is a precise set of environmental conditions or a broad range of them.

Can you explain your work related to the use of microbes in the energy field?
There is a microbe, Methanococcus jannschii, which is totally frozen at human body temperature. It comes to life at 65 degrees Celsius and its optimum growth takes place at 85 deg. C. And it is very happy to live in boiling water. It uses just two sources for its metabolism: carbon dioxide for carbon and hydrogen as an energy source. This organism was found in the underwater thermal springs in the Pacific Ocean. We are trying to sequence its genetic code and see if it could be used to produce the organisms in large numbers . If that happens, it could be used to remove pollution by letting it use of all the waste carbon dioxide from industrial processes and also fix hydrogen, a clean fuel, directly from the atmosphere.

Should patenting of genes be allowed?
I think there should be a very high bar on patents on genes. Just a patent on genes alone is not enough. You need to know something about what the gene does that has a real purpose of going into to do something about medicine. Incyte has filed thousands of patents on genes but these are worthless because most these genes don't seem to have any useful function. Most genes have multiple proteins and multiple functions and unless this information is available, it is junk.

Do you support genetically modified foods?
We have to be very careful about things which have transfer of genes between organisms. We have to study the stability of the genes in the new environment not just the beneficial effects. Most of the lateral transfer of genes in the environment has taken place in nearby areas. Unfortunately, these issues are not discussed in the polarized environment existing on this issue. I have no reservations about the science. But I don't think all the questions that need to be asked have been asked. There are no intelligent discussions on this. Honestly, I feel that the benefits far outweigh the risks with GM foods.

In which areas of biotechnology should India invest?
The best investment is to improve the quality of education and research work. Keep Inder Vermas (a well know geneticists at Salk Institute, USA) in India itself.

"India needs to improve the quality of education and research work"

What is the progress in the identification of the functions of human genes?
From the genome mapping, we have concluded that humans have only about 30,000 genes and not 140,000 or so as it was thought by scientists earlier. Now efforts are on to determine the functions of each of these genes. We know all about some 9,000 genes. We know some of the functions of some 5000 genes. We don't know anything about the remaining 42 per cent of the genes.

For each species, for each set of genes, there is a precise set of environmental conditions that impact them. You are certainly not what your genes are. When we are studying the genetic code, we are only studying at best half the question. Humans have 100 trillion cells and they work together in mass with genes. So it will be some time before we even understand the basic functions, let alone how all of them interact effectively with the environment.

Where does this leave the theory of genetic determinism which has been popular for decades?
Each human has roughly two to three million of genetic code that are different from another person. But most of these codes probably have no biological significance whatsoever. They may have just some forensic and tracking applications, and maybe useful as mapping tools. Less than a percent of these codes occur in genes or regulatory regions. This means that at the biological level, at the gene level, we are all virtually identical twins. This clearly means that your life not determined by your genes. The fraction that is really different in a biological and meaningful way is such a tiny percentage it is stunning. The smaller number of genes and the similarity of human genes to that of a mouse hopefully put the nail in the coffin of the genetic determinants.

What is the next big thing after the genome mapping?
We have some 100 trillion cells in each of our bodies. This all interact with each other in many complex ways and produce a variety of proteins. With the information about the human gene sequence, it is possible now to measure these proteins. Because we can get at every protein in the genome. For the known useful genes of around 50,000 there are some 200,000 to one million proteins. We are building a large-scale protein facility to do roughly one million protein sequences per day. We are looking at using these proteins as markers for specific diseases.

The genome is important because with mass spectroscopy sequencing, the proteins get blown apart into small fragments and we can compare those sequences back with databases. Till now most of these did not match anything on the databases and so interpretation was difficult. Now every one of these will have a match and we can rapidly determine the sequence of the proteins in the cells and in the blood.

Does genomic information provide clues to our past?
Many scientists believe that evolution is adding on genetic information and adding on complexity. But we find that most human pathogens probably started from a much more complex organism and threw out genetic material during evolution. We tried to find out whether we could come out with a molecular definition of life. To test this we knocked out 200 or so genes from Mycoplasma pneumoniae and observed whether we still have a living organism. And with Mycoplasma genitalium (a small microbe found in the human genital area) we asked the question whether all its genes were necessary for life. And we found that the 200 extra genes in pneumoniae were completely dispensable and about 200 genes in genitalium too appeared dispensable.

We found out three stunning things. First of the 300-odd genes, 103 were completely new to science. We think we know a lot about biology, yet in this most minimal cell, we had no idea about what one third of the genes do, except that if we knock them out, the cell dies.

Two, we could not come up with a molecular definition of life. We found that life is context-sensitive, that is, the environment that the cell is in is equally important to any component of the genetic code. Third, the Darwinian evolution wasn't just random errors in the genetic code. With haeomphilus and every pathogen we have worked on, we found there was preprogramming in the genetic code to cause change in the structure of specific genes. Essentially a pathogen fools the immune system of its host. And each pathogen has a different way of doing it.

Essentially, the human genome contains a molecular clock as repeated sequences of letters between genes and mutates over time. Measuring the rate of these changes would some day allow us to predict the time at which it happened. This may help to pinpoint the epoch in which the traits that make us unique humans emerged.

You have been talking about the development of personalized medicine. What are these?
With the human genetic code, we find roughly two to three million variations in the chromosomes. We have about 2.8 million well characterized, so called snips (single nucleotide polymophisms) in our database that are now being used by scientists around the world to study linkage to disease.

For the first time, we can look at this genetic variation by chromosome. For example, you can discover genetic variation in the genome of individuals who have an increased risk of myocardial infarction. Scientists are using this information to find linkage to disease. The pharma industry is using this to find ways to improve clinical trials and drug effects. This leads to the development of personalized medicine.

Recently, there was a type II diabetes drug that had to be taken off the shelves because one out of 10,000 people had a severe liver toxicity to this drug. If we can find simple tests that predict toxicity, it will have a huge impact. Most drugs work only on 30 to 50 percent of the population but are given to everybody because in theory they are relatively safe and doctors work on the principle that it is easier to give it everybody than find out whose disease it can actually cure. Personalized medicine would actually try to treat the patient and not kill him inadvertently.

What is the uniqueness of your 'shotgun sequencing' method?
All the sequencing technologies give us only 500 to 600 letters of genetic code. So how do you get the genetic code of something billions of letters long. The other problem is that genomes, including the human genome, have lots of repeats. The same sequence is repeated over and over again. In one method, scientists would sequence one piece, get 600 letters, make a little primer to read the next set of 600 letters and work sequentially down the clone with that. By this method, it would take 100 years to sequence the human genome. In the second method, genes of the first clone are mapped, then lined up and sequenced. This way, it took three years to just arrange the E.coli clones before sequencing could be done.

With 'whole genome shotgun'', we don't do any cloning first. We take the entire set of chomosomes, break those down into little pieces and use mathematical algorithms to try and solve the big jigsaw puzzles.

Why is it important to study mouse genome and not that of chimpanzees?
Studies show that the human and mouse genomes are almost identical genetically. This enhances the chances of predicting the regulatory regions in the genome and we will have comparative genomic studies. On the other hand, data generated at Max Planck Institute using the human genome data indicates that the difference between the two is just 1.27 percent in genome sequences. So it is better to study the mouse genome which evolved 100 million years ago and it will provide a better comparison on the evolutionary changes that had taken place.

We have to look closely at our own evolution and that of vertebrae. From the simple drosophila (fruit fly) to humans, the evolution has taken place in about 600 million years. And just four to five gene groups only expanded in this period. For example, humans have an immune system, while the fruit fly doesn't have a significant immune system. All the genes associated with our immune system expanded during this period. When we compare with homeostasis, all things associated with our vascular system expanded during this period. But we have virtually identical gene set with mice and other vertebrates. The key to our uniqueness are things like the transcription pattern in gene regulation, that turn on different sets of response to environmental conditions. Among all human chromosome, the chromosome 19 has the largest number of genes.

Because it has neurotransmitter receptors, olfactory receptors and transcription factors. In the future, we will be able to provide a date in some cases of species, where this evolutionary event took place.

What is your work related to synthetic cells?

We are now trying to develop a synthetic cell just to see how many essential genes are required to form a living thing. For example, mycoplasma lives on glucose or fructose. If you knock out the glucose transporter gene and you have both sugars in the environment, the cell will happily live. But if you only have glucose in the environment and you knock out the glucose transporter gene, the cell dies. So for each species, for each set of genes, there is a precise set of environmental conditions or a broad range of them.

Can you explain your work related to the use of microbes in the energy field?
There is a microbe, Methanococcus jannschii, which is totally frozen at human body temperature. It comes to life at 65 degrees Celsius and its optimum growth takes place at 85 deg. C. And it is very happy to live in boiling water. It uses just two sources for its metabolism: carbon dioxide for carbon and hydrogen as an energy source. This organism was found in the underwater thermal springs in the Pacific Ocean. We are trying to sequence its genetic code and see if it could be used to produce the organisms in large numbers . If that happens, it could be used to remove pollution by letting it use of all the waste carbon dioxide from industrial processes and also fix hydrogen, a clean fuel, directly from the atmosphere.

Should patenting of genes be allowed?
I think there should be a very high bar on patents on genes. Just a patent on genes alone is not enough. You need to know something about what the gene does that has a real purpose of going into to do something about medicine. Incyte has filed thousands of patents on genes but these are worthless because most these genes don't seem to have any useful function. Most genes have multiple proteins and multiple functions and unless this information is available, it is junk.

Do you support genetically modified foods?
We have to be very careful about things which have transfer of genes between organisms. We have to study the stability of the genes in the new environment not just the beneficial effects. Most of the lateral transfer of genes in the environment has taken place in nearby areas. Unfortunately, these issues are not discussed in the polarized environment existing on this issue. I have no reservations about the science. But I don't think all the questions that need to be asked have been asked. There are no intelligent discussions on this. Honestly, I feel that the benefits far outweigh the risks with GM foods.

In which areas of biotechnology should India invest?
The best investment is to improve the quality of education and research work. Keep Inder Vermas (a well know geneticists at Salk Institute, USA) in India itself.