

GENOMICS: FROM MICROBES TO MAN

J. Craig Venter

2000 Sir John Crawford Memorial Lecture



**Consultative Group on International
Agricultural Research (CGIAR)**

CGIAR's Mission

To contribute to food security and poverty eradication in developing countries through research, partnership, capacity building, and policy support, promoting sustainable agricultural development based on the environmentally sound management of natural resources.

GENOMICS: FROM MICROBES TO MAN

J. Craig Venter

***2000 Sir John Crawford Memorial Lecture
CGIAR International Centers Week***

***Published by the CGIAR Secretariat.
March 2001***

Dr. Venter's address is part of the Sir John Crawford Memorial Lecture series sponsored by the Australian government at the CGIAR annual meetings. Sir John Crawford, a distinguished civil servant, educator, and agriculturalist, was one of the founders of the CGIAR and the first Chair of the CGIAR's Technical Advisory Committee.

It is a great honor to give the annual Sir John Crawford Lecture.

I do not know as much about agriculture as many of you do, but I certainly appreciate its central place in biology and in the genomic revolution. I will focus instead on some of the genomes I have characterized with my colleagues at The Institute for Genomic Research and at Celera. And I will talk about why characterizing genomes is important. I'll even try and tell you a little about your own genetic code and how that might be relevant to your health.

The history of genomics is remarkably short and yet we have accomplished a great deal. In 1992, Dr. Claire Fraser and I left the National Institutes of Health (NIH) to form the first not-for-profit research institute devoted to genomics, The Institute for Genomic Research (TIGR). We had funding of \$85 million for ten years from Wally Steinberg, a strong supporter and entrepreneur. In exchange, a parallel for-profit entity, Human Genome Sciences, would get access to the research to develop the discoveries made by TIGR into new therapeutics.

We set up the first large-scale DNA sequencing facility using what was thought of at that time as untried technology, automated DNA sequencers made by Applied Biosystems.

In 1994 we made a transition. TIGR began using the expressed sequence tags (EST) method I developed at NIH, a method for rapidly discovering genes without sequencing the entire genomes. We developed new mathematical algorithms to deal first with tens of thousands of sequences and then hundreds of thousands and now millions of sequences. And we decided we could go back and rethink how to approach genomics. Our colleague, Hamilton Smith, who shared the Nobel Prize in 1978 for the discovery of restriction

endonucleases, suggested we use a microbial pathogen. He suggested his favorite, *Haemophilus influenzae*, the major cause of ear infections in children and a major cause of meningitis. Virtually nothing was known about this organism, even though Dr. Smith isolated the first restriction enzymes from *Haemophilus influenzae*.

We decided we could use our new mathematical approach and sequence this genome relatively rapidly, given that the benchmark at that time was the *E. coli* genome, which is slightly larger than *Haemophilus*. The *E. coli* project was in its ninth year of federal funding and, in fact, it took a total of 12 years for the public effort to sequence the *E. coli* genome.

We thought we could do better. We wrote a grant and submitted it to NIH saying we could sequence the *Haemophilus* genome in one year at a tiny fraction of the cost. We named the new strategy for decoding genomes “whole genome shotgun sequencing.” We decided the chances of this grant getting funded were relatively low, so we went ahead and used some of the TIGR endowment to sequence this genome.

In 1995, when we had it 90 percent completed, NIH said that what we were proposing was impossible and they would not fund it. A short while later, the *Haemophilus* genome sequence was published in *Science*. It was the first sequenced genome of any free-living organism.

This has changed the paradigm of what can be done, and things have accelerated quite dramatically since then. A partial list of genomes that have now been completely sequenced by largely the same scientific team both at TIGR and at Celera includes a large

number of key pathogens, some very interesting environmental organisms that changed our view of the very nature of life itself. The first of these, *Methanococcus jannaschii*, came from a hypothermal vent a mile and a half deep in the Pacific Ocean. The temperature in the center of the vent is about 400 degrees Centigrade and the surrounding water is 2 degrees Centigrade. This organism makes everything it needs for life just from carbon dioxide, using hydrogen as an energy source. It is a true autotroph. Humans cannot do that. We need more than just CO₂ and hydrogen. At our body temperatures, this organism is frozen solid. At about 60 degrees Centigrade it comes to life. Its optimum temperature for growth is 85 degrees Centigrade. It's totally happy in boiling water.

Another sequence of an organism published by TIGR is the *Deinococcus radiodurans* genome, one of my favorites. It is literally like something from outer space. This organism can take three million rads of radiation and not be killed. Its chromosomes get broken apart with one or two hundred double-stranded breaks, but over a period of 12 to 14 hours it stitches its chromosomes back together and starts replicating again. Humans cannot do that either. Not even in Washington!

Francis Crick, one of the co-discoverers of the structure of DNA, had a theory, developed by others as well, called panspermia, that states life did not originate on Earth; it came from another part of the universe, and once it arrived on this planet, it then evolved further. It was not believed that life could survive in the environment of outer space. But *Deinococcus* is actually an excellent candidate for a panspermia organism. It can be completely desiccated. It has been found, in that state, on granite surfaces in Antarctica, where perhaps it's been for a very long time. It can absorb huge doses of ionizing radiation. Drop it in water and it starts to reassemble the badly

damaged chromosome and starts replicating again. So it could have very easily arrived on a meteor some time in our past, or in some organism related to us.

As we find these kinds of creatures of radiation resistance and temperature resistance, it becomes clear that they are not at all unusual, they're not even rare. So it changes our view of the focus of life. These organisms also have some tremendous tools for us.

The cholera genome was recently published by TIGR in *Nature*. Many scientists thought it was pointless to sequence the cholera genome because sixteen RNA studies showed it was exactly like the *E. coli* genome and there would be nothing to learn from it. Every genome that we have examined has held tremendous surprises. This was no different. *Vibrio cholerae* did not have one chromosome, as the field had assumed for years, it had two. And one of them did, in fact, resemble *E. coli*. The other didn't resemble *E. coli* at all, and that one is probably responsible for a lot of the actions and biology of cholera.

So whole genome shotgun sequencing has no preconceived notions. In every case some grand illusions have been shattered by it. In some of the first genomes of pathogens, one of the shattered illusions is that evolution does not happen just by random errors. We found built-in mechanisms for active change of the genetic code in essentially every human pathogen and most organisms. It is possible that everyone has *Haemophilus influenzae* in the airway, because in front of most of the genes that code for cell surface antigens are tetrameric repeats, and in every 10,000 or so replications of this bacteria, the DNA polymerase slips on these repeats and puts stop codons in the gene downstream, effectively knocking out the gene. This constantly changes the cell surface molecules to avoid the immune system. So it

is a real-time Darwinian evolution, not from random changes, but from pre-programmed evolutionary changes built forever into the genetic code. And that is why it was such a disaster when, in the late 1970s, the U.S. Surgeon General announced that we had won the war against microbes.

We have seen these mechanisms in essentially every pathogen that we have looked at. They have tremendous consequences. One company tried to develop new vaccines against *Haemophilus*, but ignored these mechanisms—taking every cell surface antigen and making an antibody against it, and they got great vaccines against that one strain of *Haemophilus*. But as soon as they went into the clinic, the vaccines didn't work anymore.

We could have predicted that. Chiron Corporation decided to listen to us. They funded a project at TIGR to sequence another major cause of meningitis, the *Neisseria meningitidis* genome. At the same time they wanted to use TIGR's ability to interpret the genetic code to tell them what candidates they might use for vaccine development.

Approximately a year later, the paper was published in *Science*, along with another study simultaneously describing two new vaccine candidates that are now about ready to go into clinical trials with Chiron, because they seemed to work against a very broad array of strains of *Neisseria meningitidis*, for which there is currently no vaccine. So vaccine development is one of the key areas. We can go directly from finding the genetic code, use the information we learn, and potentially have major preventative treatments for major diseases. This is a theme you'll hear over and over again.

When we had the first genome and had the complete set of genes laid out in front of us, we decided it was too much for the human mind to interpret and comprehend even 1,800 genes – at least for my human mind and those of my colleagues at TIGR. And we went out looking for a simpler organism to see if there was something smaller and easier to comprehend.

Clyde Hutchinson, of the University of North Carolina, had been characterizing *Mycoplasma genitalium*, and argued that it probably had the smallest genome of any free-living organism. Dr. Fraser led a team at TIGR to sequence this genome in three months. It was the second genome done in history and it was published in *Science*. It had only around 475 genes, and so we just asked the simple question: Well, if *Haemophilus* needs 1,800 genes and this one only needs 475, is 475 the minimum, or could this get by with less? Could we come up with a molecular definition for life based on the gene content of a species?

A while later, another *Mycoplasma* species was sequenced, *Mycoplasma pneumoniae*, and it demonstrates a key feature of evolution that most people do not appreciate; that it is not only the addition of complexity and genetic material that leads to evolution, particularly with human pathogens. A lot of human pathogens evolve as pathogens by throwing out genetic material. So these two *Mycoplasmae*, around 500 and 700 genes, probably evolved from a 5,000 gene *B.subtilis*-type organism by throwing out things they didn't need because they could get those pathways and nutrients from us as a host.

All the genes in *Mycoplasma genitalium* were contained in the *pneumoniae* genome, but *pneumoniae* had 200 extra genes. So if we

were studying *pneumoniae*, we could say we could throw out 200 genes and get down to *Mycoplasma genitalium*.

To make a very long story short, we developed a method originated by Clyde Hutchinson called “whole genome transposon mutagenesis.” Basically, we use electrical power to get small genetic elements into the cell, and then they randomly insert into the genetic code.

If they insert in the middle of a gene, it disrupts that gene the same way as the mutants from the slip strand mechanisms do. The ones we deemed essential were ones that if knocked out, the cell dies. It's a pretty simple definition.

After analyzing all these data, we got down to around 300 genes that we calculated to be essential for life. Of the 300 genes, 103 genes are completely unknown to science. We have not a clue what they do except that if you remove them from the genome, the cell dies.

It's very humbling when we are now trying to analyze our own genetic code, to try to understand how 30,000 or 40,000 genes in 100 trillion different combinations lead to our own biology. The other notion that we had of trying to define life at a molecular level, to find the secret code for life, fell apart pretty rapidly because we discovered what most social biologists knew for a long time; that the environment is important. It's nice for a molecular biologist to discover that. In fact, it is even nicer to prove it. We proved it with this study. We could not define a set of genes outside of defining the environment. In a very simple example, a cell will grow on both glucose and fructose. And there is a gene for the transporter for each one of these.

If we knock out the glucose transporter, as long as there is fructose in the environment, the cell is totally happy and continues to live and replicate. If we knock out both transporters, the cell dies. But if you knock out the glucose transporter and there is only glucose in the environment, the cell also dies.

So you cannot define the molecular basis of any living entity without defining the corresponding environment. That is a major revelation for a molecular biologist.

We moved on to more complex species. TIGR wanted to characterize a number of species that affected a large number of people. Steve Hoffman, who is now at Celera, received the Legion of Merit Award from the Secretary of the Navy for initiating the malaria genome program. Malaria is caused by a mosquito taking a blood meal and at the same time injecting the malaria parasite.

The malaria genome was thought to be unsequenceable because of the high content of the nucleotides Adenine and Thymine. We found that it was readily sequenceable, but there was no map of the genome, and no way to characterize it. So we used some novel techniques.

A single molecule of chromosome 2 from malaria that was stretched out on glass slides and viewed under the microscope showed that the DNA could still be treated with restriction enzymes. So we could get restriction digest maps of single molecules of DNA where the enzyme cuts the surface tension and the ends pull away, and we can just look down the chromosome and see a cut, measure the distance, and use this to verify the structure of the genome. This was the first malaria chromosome that was published in *Science* a few years ago in

a program headed by Malcolm Gardner, one of the top parasitologists at TIGR.

In fact, Malcolm is heading the team that is studying the East Coast fever genome, *Theileria parva*. East Coast fever is caused by a tick-borne parasite that causes cattle to die within three or four weeks of infection. The parasite enters the blood stream and enters lymphocytes and actually transforms them into a cancer-type phenomenon. It is the only case where the eukaryotic cell is known to transform lymphocytes to create a leukemia-type disease.

This has a tremendous impact in Sub-Saharan Africa, and TIGR was asked to help in sequencing the genome; in part to help develop new and novel vaccines against this species, even though very few parasite genomes were known.

It is a very small eukaryotic genome, in contrast to the pathogens, done in collaboration with ILRI. Dr. Gardner has now basically completed the sequencing of the genome and is in final closure and annotation stages, and will probably publish early next year. Already various antigens have been selected for vaccine development.

TIGR moved into the plant world and sequenced the first plant chromosome from *Arabidopsis*, a model plant organism, with work funded mostly by the U.S. National Science Foundation. And, again, there were a lot of surprises, even on the plant chromosomes. The gene density was mostly even, except in the region of the centromere, but it had been thought—and the centromere is the heterochromatin of our own chromosomes—that there were no genes. TIGR found there are, in fact, a lot of genes in these regions that were thought to be to be void of genes, many of them very important

essential ones, but about one-tenth the density that were found elsewhere.

So TIGR had sequenced genomes, including the tuberculosis genome and *Deinococcus*, that were a high Guanine-Cytosine (GC) content. We sequenced over half the chromosomes in *Plasmodium*, which are a very low GC content. And we knew these techniques could be applied on a much broader scale to larger species. What was lacking was the appropriate technology.

In 1998, Mike Hunkapillar of Applied Biosystems, contacted me about a new sequencing instrument he had developed. He thought it would be a tool that I needed to actually try to sequence the human genome. And, they were willing to consider investing \$300 million to fund the experiment.

By the end of the first day of looking at this new technology we had a plan for sequencing the human genome combining all the techniques we had developed at TIGR for the whole genome shotgun sequencing, the new capillary sequencer from Applied Biosystems, and a key component for high-end computing that was being provided by Compaq.

Following an expansion of the TIGR model, we built a new sequencing factory with 300 of these new \$300,000 machines. In contrast to roughly 3,000 scientists associated with the public genome effort, it took a team of only 50 scientists to sequence the human genome. As you can see, we substituted electrons for people.

The facility is a larger version of TIGR. It is a football-field-size room full of instruments. But now it only takes nine technicians to run the

instruments 24 hours a day, 7 days a week. So automation has changed the cost paradigm substantially.

The problem was in analyzing all the data, and we had to build the largest civilian supercomputer with the help of Compaq to be able to assemble and analyze the human genome.

It was a big leap going from a plant chromosome or the malaria chromosome to the human chromosome, and we also knew that even if we had the human genome sequence, it would be difficult to interpret without other genetic codes. So we chose to test this new technology on sequencing the fruit fly genome as a prototypical insect. And this is going to have a big impact now as various insect species, including the mosquito that carries the malaria parasite and others like the tick genome, are analyzed. We now have a basis of an insect genome.

This was done in a relatively short period. Once we had the sequence, a team of international scientists came to an Annotation Jamboree in Rockville, Maryland, and literally camped out for weeks while they analyzed the *Drosophila* genetic code. The *Drosophila* sequence was published in *Science* in March 2000, less than a year from the start of this project. The next largest genome was the *C. Elegans* genome, which took over eight or ten years to sequence from start to finish. So, again, this changed the time and the cost paradigm.

When we sequenced the *Haemophilus* genome, we had to sequence 26,000 clones, and then it took four months at TIGR. TIGR could probably do this in a few weeks now. For *Drosophila*, we had to sequence over three million clones. This was during the scale-up period, and it took four months at the time. If we were going to

resequence the *Drosophila* genome, it would take roughly three and a half weeks. The *Haemophilus* genome that took so much time five years ago could be done now in two hours, and the yeast genome, which was a 12-year project in Europe, could be done in 24 hours.

We are dealing with a fundamental technology change. In terms of the computing, five years ago it seemed almost an impossible compute for *Haemophilus* – it took eleven days on a Sun computer to assemble the genome. Now, with the new algorithms and the new super-computer, it takes less than five minutes to do that same calculation. And all this technology is expanding. If anyone had asked five years ago whether the *Theileria* genome would be sequenceable in the short period of time that TIGR has done it, we would have said it could not be done. So it is a very short history with a very dramatic change.

In *Drosophila*, we found 13,601 genes; roughly only 2,500 of those were previously known. A comparison of the first few genomes that had been sequenced showed, not surprisingly, a large number of unknown genes. We thought once we sequenced a few more we would find matches to everything. That, in fact, has not happened. In every genome that we, TIGR, or anybody else does, roughly half of the genes in that genome are unknown to science. Half again of those are highly conserved; in other words, we find that same unknown gene in a large number of species. But we still do not have the slightest clue what it does.

To get a grant funded in the United States to study biology, a hypothesis is necessary, and without a hypothesis the grant will not be funded. The assumption is we know so much about biology that we should just be testing hypotheses at this stage. Genomics is clearly

saying we are still very much in the descriptive phase of biology and will probably be there for most of the rest of this century.

You cannot get a grant to study one of those 103 unknown genes that are essential for life in *Mycoplasma*, and you cannot get a grant to study any of the 18,000 genes from the various microbial species, or the 60 percent of the fruit fly genome that is completely unknown. So it is going to be a real challenge for biology. There are now approximately over 100,000 new unknown genes, far more genes than in our entire human genome, and biology has no idea what role they play. This is both a problem and a phenomenal opportunity, because these genes are going to represent unique points for intervention and therapy, new unique vaccine targets, and unique mechanisms for changing the biotechnology revolution.

All we know about the insect world now—and *Drosophila* is one of the most studied species—is about 40 percent of the genome. Most of these are new, but at least there are new families in the common categories of genes that we know about. In over half the human genome, the genes are new to science.

Seymour Benzer's group at Cal Tech characterized a gene that he named "Methuselah," which led to an increased life span in fruit flies. When we characterized the fruit fly genome, we found eleven Methuselah-like homologues. Everybody who was in that room—over 50 people—immediately took those to look in the human genome to see if we could find new longevity genes, and some of those are under investigation.

Gerry Rubin, now the Vice President of Biomedical Research at the Howard Hughes Medical Institute, was our key collaborator on the *Drosophila* genome. He characterized all the known human genes

and found that over 300 of them had counterparts in the fruit fly genome.

Our genes are nearly identical, or many of our proteins are nearly identical in structure to those in the fruit fly and other species. But now scientists studying these genes and their function in *Drosophila* are moving forward our knowledge of human disease genes that affect diseases like cancer.

That is a transition into human. At about the time we announced the completion of the *Drosophila* genome, we switched totally to work on the human. We took all the chromosomes together from five people, three females and two males, and made individual libraries. On June 26, 2000, we announced at the White House the successful assembly of the 3.12 billion letters of our own genetic code. The public genome effort announced at the same time. It was an exciting day for everyone.

We have now moved into trying to interpret this information. A lot of people in the press and on Wall Street have said, well, now, Celera sequenced the genome, genomics is finished, what's the next phase of life? Genomics is just getting started. This was a race to the starting line. Once there is a genome sequence of the malaria genome, research can begin for new treatments and new therapies. New malaria vaccines are being tested from the work of Malcolm Gardner and Steve Hoffman and their colleagues from the genome effort.

Comparative genomics is one of the most important tools going forward. Understanding the fruit fly genome will really help us understand the human. About eight years ago, we were character-

izing some of the early ESTs that we did from human, and we compared the human sequences to those from *E. coli* and from yeast. Mismatched DNA repair enzymes had been well characterized in yeast and *E. coli*. I do not think anybody would have predicted that the genes that cause colon cancer would have been discovered by characterizing brewer's yeast and the *E. coli* that lives in our gut. But it was because of the close similarity of genes throughout all species that these matches immediately leapt forward. It was very clear these human genes, by their close similarities to the counterparts in bacteria and yeast, were mismatched DNA repair enzymes. Bert Vogelstein's lab at Johns Hopkins, our collaborator on this, quickly showed that changes in the genetic code in these genes were the cause of non-polyposis colon cancer.

So characterizing a variety of species will have a huge impact. The *Theileria* genome will help us understand things in our own genetic code. Understanding the human genome will help interpret the *Theileria* genome.

We have recently announced that we finished sequencing three strains of mice to help interpret the human genetic code. If you chop up the mouse chromosomes and lay them on top of the human chromosomes, they're virtually identical, just some of the order is slightly different.

There is a gene called the Pax-6 gene. If this gene is knocked out in fruit flies, it leads to what is called an eyeless phenotype. The fruit flies have no eyes. If it is mutated or knocked out in mice, it leads to blindness.

A disease called *aniridia* has mutations in the same gene. Children are born without an iris. They go blind at an early stage because they

cannot regulate the light going into their eyes. It is possible to take the human or mouse gene, put it in a fruit fly, and it will rescue the phenotype. The parts are conserved through billions of years of evolution and are largely interchangeable.

In the sequence of a single human genome there will be about two to three million variations. The chromosome set passed down from each parent differs from each other in roughly one out of 1,200 letters. So there are roughly two to three million differences in the genetic code from person to person out of over three billion letters.

We have recently announced our database of over 2.8 million of these single nucleotide polymorphisms that are being used to characterize human disease. They are helping the pharmaceutical industry do better clinical trials, find drugs that are more effective, and look for hints in the genetic code for detectable toxicity. For example, a major Type II diabetes drug was recently taken off the market because it caused liver toxicity in one out of 10,000 patients taking the drug. If those kinds of toxic events could be predicted in advance before the drug is taken, it would fundamentally change medicine. It would change the economic condition of one pharmaceutical company that lost a billion-dollar-a-year drug, and roughly 100 people, I've been told, have died from this severe liver toxicity. They would have liked to have known whether that could have been predicted in advance. That is the promise that this work has.

But as much as people want genetics to be deterministic, it is not, with rare exceptions. It is not in the single cell with perhaps 300 genes. It is not even in the cancer genes I mentioned. If we find these mutations in any person, it can tell whether he has a greatly

increased risk of getting colon cancer, but it won't tell if he will get colon cancer. It won't tell if he won't.

So the next stage is interpreting the genetic code in terms of understanding the protein world. The hope is that the genetic code will give us predictions. It will tell us who has an increased likelihood, a propensity for different diseases. If we can find specific protein markers for breast cancer, like the PSA antigen for prostate cancer, we think that breast cancer could be detected earlier than by using mammography and before a lump would be detected.

If there are 50,000 genes, there may be as many as a million different proteins from different combinations. Our complexity comes not from the genetic code, but from what happens after that.

We are now building the world's largest protein-sequencing facility in Rockville, Maryland, which will allow us to sequence on the order of a million proteins a day, comparing things from a large number of clinical situations. This is just one of the areas in which genomics is essential. Without having the human genome sequence this could not be done, because when we sequence proteins with mass spectrometry, the proteins get blown apart into small pieces that then get compared back to the genetic code for interpretation.

Before we finished the genome, most of the pieces did not match anything and the protein structures could not be found. That is now changed, and more changes are coming fast. With new technology that is being developed by our sister company, Applied Biosystems, these machines should be able to do on the order of 10,000 samples an hour versus on the order of 100 or so today.

We received the first prototype machine just a few weeks ago. This is a new "Time of Flight - Time of Flight" mass spectrometry instrument, known as the "TOF-TOF" for short. It uses two mass spectrometers in a row to sequence over 10,000 proteins an hour, and we are building out a whole floor to be filled with these machines.

So, whether it's a parasite or whether it's a cancer that grows in our own systems, we think this will lead to new approaches, not only for diagnosis, but for cancer-specific vaccines by finding the proteins that are expressed specifically in different tumors the same way that malaria vaccines—and the same way that we hope *Theileria* vaccines—will be developed. We think we have a chance to develop new vaccines for diseases such as cancer.

So we are moving from evolving levels of genomic information to study the protein world, to a fuller study of medicine and the complete spectrum of biology, of which agriculture is a very key environmental counterpart.

J. Craig Venter, Ph.D.

Celera Genomics
45 West Gude Drive
Rockville, MD 20850
(240) 453-3500

J. Craig Venter, Ph.D. is the President and Chief Scientific Officer of Celera Genomics Corporation and the Founder, Chairman of the Board and former President of The Institute for Genomic Research (TIGR), a not-for-profit genomics research institution.

Between 1984 and the formation of TIGR in 1992, Dr. Venter was a Section Chief, and a Lab Chief, in the National Institute of Neurological Disorders and Stroke at the National Institutes of Health (NIH). In 1990, he developed expressed sequence tags (ESTs), a new strategy for gene discovery that has revolutionized the biological sciences. Over 72 percent of all accessions in the public database GenBank are ESTs from a wide range of species including human, plants and microbes. Out of new algorithms developed to deal with 100,000's of sequences TIGR developed the whole genome shotgun method that led to TIGR completing the first 3 genomes in history and a total of 21 to date.

In May of 1998, Dr. Venter and Perkin-Elmer (now known as Applera) announced the formation of Celera Genomics. Celera's goal is to become the definitive source of genomic and medical information thereby facilitating a new generation of advances in molecular medicine. Celera is building the expertise and information that will enable scientists to transform the way in which human and health problems are diagnosed and treated. On June 26, 2000, Celera announced that it had completed the first assembly of the human

genome, which has revealed a total of 3.12 billion base pairs in the human genome. On February 16, 2001, Celera's manuscript on the sequencing of the human genome was published in *Science Magazine*.

Dr. Venter has published more than 160 research articles and is one of the most cited scientists in biology and medicine. He has been the recipient of numerous awards, including the 2000 King Faisal Award in Science and was recently selected as a runner up for TIME Magazine's Man of the Year and was selected as Man of the Year for the Financial Times. In addition to receiving honorary degrees for his pioneering work, he has been elected a Fellow of several societies including the American Association for the Advancement of Science and the American Academy of Microbiology. He received his Ph.D. in Physiology and Pharmacology from the University of California, San Diego.

Scientific papers published include:

- Complementary DNA Sequencing: "Expressed Sequence Tags" and the Human Genome Project. *Science* 252, 1651-1656 (1991).
- Potential Virulence Determinants in Terminal Regions of Variola Smallpox Virus Genome. *Nature* 366, 748-751 (1993).
- Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd. *Science* 269, 496-512 (1995).
- Initial Assessment of Human Gene Diversity and Expression Patterns Based Upon 52 Million Basepairs of cDNA Sequence. *Nature* 377 suppl., 3-174 (1995).
- The Minimal Gene Complement of *Mycoplasma genitalium*. *Science* 270, 397-403 (1995).

-
- Complete Genome Sequence of the Methanogenic Archeon, *Methanococcus jannaschii*. *Science* 273, 1058-1073 (1996).
 - The Complete Genome Sequence of the Gastric Pathogen *Helicobacter pylori*. *Nature* 388, 539-547 (1997).
 - The Complete Genome Sequence of the Hyperthermophilic, Sulphate-Reducing Archaeon *Archaeoglobus fulgidus*. *Nature* 390, 364-370 (1997).
 - Genome Sequence of the Lyme Disease Spirochaete, *Borrelia burgdorferi*. *Nature* 390, 580-586 (1997).
 - Shotgun Sequencing of the Human Genome. *Science* 280, 1540-1542 (1998).
 - Complete Genome Sequence of *Treponema pallidum*, the Syphilis Spirochete. *Science* 281, 375-388 (1998).
 - Chromosome 2 sequence of the human parasite *Plasmodium falciparum*: Plasticity of a eukaryotic chromosome. *Science* 282 (5391), 1126-1132. (1998).
 - Global Transposon Mutagenesis and a Minimal Mycoplasma Genome. *Science* 286, 2165-2169 (1999).
 - Sequence and Analysis of Chromosome 2 of *Arabidopsis thaliana*. *Nature* 402, 761-767 (1999).
 - Complete Genome Sequencing of the Radioresistant Bacterium, *Deinococcus radiodurans* R1. *Science* 286, 1571-1577 (1999).
 - The Genome Sequence of *Drosophila melanogaster*. *Science* 287, 2185-2204 (2000).
 - Sequencing of the Human Genome. *Science* 291, 1304-1351 (2001).

The CGIAR Family. Created in 1971, the Consultative Group on International Agricultural Research (CGIAR) is an association of public and private members that support a system of 16 international agricultural research centers known as the Future Harvest Centers. The Future Harvest Centers work in more than 100 countries to mobilize cutting-edge science to reduce hunger and poverty, improve human nutrition and health, and protect the environment. The CGIAR's budget in 2000 was US\$340 million. All new technologies resulting from the Centers' research are freely available to everyone.

CGIAR-Supported Future Harvest Centers

- Centro Internacional de Agricultura Tropical (CIAT), www.ciat.cgiar.org
- Center for International Forestry Research (CIFOR), www.cgiar.org/cifor
- Centro Internacional de Mejoramiento de Maiz y Trigo (CIMMYT), www.cimmyt.cgiar.org
- Centro Internacional de la Papa (CIP), www.cipotato.org
- International Center for Agricultural Research in the Dry Areas (ICARDA), www.icarda.cgiar.org
- International Center for Living Aquatic Resources Management (ICLARM), www.cgiar.org/iclarm
- International Centre for Research in Agroforestry (ICRAF), www.cgiar.org/icraf
- International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), www.icrisat.org
- International Food Policy Research Institute (IFPRI), www.cgiar.org/ifpri
- International Institute of Tropical Agriculture (IITA), www.cgiar.org/iita
- International Livestock Research Institute (ILRI), www.cgiar.org/ilri
- International Plant Genetic Resources Institute (IPGRI), www.ipgri.cgiar.org
- International Rice Research Institute (IRRI), www.cgiar.org/irri
- International Service for National Agricultural Research (ISNAR), www.cgiar.org/isnar
- International Water Management Institute (IWMI), www.cgiar.org/iwmi
- West Africa Rice Development Association (WARDA), www.cgiar.org/warda

CGIAR members. The CGIAR partnership includes 22 developing and 21 industrialized countries (South Africa has been a member since 1997), 3 private foundations, and 12 regional and international organizations. The Food and Agriculture Organization of the United Nations (FAO), United

Nations Development Programme (UNDP), and the World Bank serve as cosponsors.

*The CGIAR Secretariat
The World Bank
1818 H Street, NW, MSN G6-601
Washington, DC 20433, USA
Tel: (1-202) 473-8951 Fax: (1-202) 473-8110
E-mail: cgiar@cgiar.org or cgiar@worldbank.org
www.cgiar.org*