LESSON FOR THE ALL SPECIES PROJECT FROM THE HUMAN GENOME PROJECT

In the annals of big science - the Manhattan Project, the race to the moon, the supercollider, the Space Shuttle - sequencing the human genome is closest to our current effort to identify all species. There are three reasons for All Species to look to the Human Genome Project as a source of experience:

1) Unlike the other mega projects, The Human Genome Project featured a goal that probably would have happened anyway over time. This was not true of going to the moon or building an atom bomb. DNA sequencing was widely happening. The Human Genome Project took ordinary work and raised it to the level of a legend and myth by attempting to complete it "all" in a relatively short time. The Genome project then is primarily distinguished by its emphasis on "all."

2) The Genome Project is the only other big science project to be dedicated to biology.

3) Unlike the other mega-projects, which were entirely funded by government money, the Human Genome Project was greatly benefited by private foundation funding.

Because the All Species Project shares the three ambitions above, I found the short history of the Human Genome Project to be full of inspiration and counsel. The lessons I believe apply to All Species I've indicated in italics.

-- Kevin Kelly

Solid Foundations Are Worthwhile

The first person to request large-scale funding to sequence the entire human genome was Norman Anderson, a tool-maker at a Department of Energy (DOE) lab, who made the proposal in the late 1970s. His effort failed to raise any funds.

The idea was resurrected simultaneously and independently by four or five other groups coming together in workshops in the mid 1980s. Robert Sinsheimer, chancellor of UC Santa Cruz, was looking for a fundable idea to redirect a $30 million foundation bequest to the school that would otherwise be "lost." His solution was to gather a workshop in 1985 to consider the feasibility of sequencing the entire human genome. Later that same year, two different researchers at the DOE independently proposed a Human Genome Initiative at the DOE, which was already funding work in tracking heritable human mutations and needed a standard reference genome. In 1986 a roundup meeting on the Molecular Biology of Humans at Cold Spring Harbor run by James Watson added a last-minute rump session on mapping the human genome after hearing about previous meetings and proposals. Also in 1986, the Howard Hughes Medical Institute convened an international forum on a genome sequencing project, and Sydney Brenner urged the European Commission in Brussels to coordinate a large-scale program to sequence the genome.

Sequencing the human genome was clearly an idea whose time had come in 1986, but the exact shape of that idea, how it fit into the existing scientific establishment, was something that needed to be drawn out. The first two years of the Human Genome Project were spent in meetings, summits, workshops, writing editorials and white papers, and the like. While ordinary labs outside were merrily sequencing everyday, the mega project was not overseeing any work at all, beyond this essential framing.

All Species Project should spend its resources in the first few years to refining the idea, framing the terms of the discussion, gathering partners and supporters, and putting the concept out to society in the proper position. It should not expect to begin delegating taxonomic work until this architecture is in place.

Back to Top

An Unwavering Goal

The early gatherings of human DNA sequencing represented a great diversity of interests and reasons for wanting the human genome codified. The Department of Energy, of all agencies, was an early supporter of the sequencing initiative. Its funding was sometimes jokingly referred to as the "Unemployed Bomb Makers Act." The community widely differed on how far to broaden the investigation. Questions such as what other organisms to sequence, and how much mapping to do, or what to do with the information were unresolved. But the common center among the early proponents was a single, simple, clear goal: This project would compile a complete reference DNA sequence of all the 23 human chromosomes.

Because that goal was an understandable, worthy challenge, easy to say, clear to everyone, yet never moving, it was a goal that could command billions of dollars, millions of moving parts, thousands of workers, hundreds of laboratories, and tens of years. For the All Species Project to succeed it needs to fix on its own unwavering goal of identifying all living species on earth in a generation.

Back to Top

Resistance Is Internal

The proposal to sequence the entire human genome met with skepticism and resistance at every stage of the project (and still has its share of critics now). Lee Hood found the project "premature." James Watson said everyone he met was against the idea. Robert Weinberg of the Whitehead Institute said the idea was so nonsensical that he was "surprised consenting adults have been caught in public talking about it," Nobelist David Baltimore said, "the idea is gathering momentum… I shiver at the thought." Other comments by scientists at the time: "Absurd." "Dangerous." And of course, "Impossible."

It was not just the greybeards either. Many young molecular scientists didn't like the idea of turning sequencing into a factory production model, with them doing "production sequencing," a task they imagined as boring and "more suitable for prisoners." There were legitimate concerns about big science funding siphoning money away from other worthy projects. Others felt that sequencing the entire genome was simply bad science since most of the genome was junk. Still, even some of these critics backed the idea as the politically astute thing to do, since "all" was easier to fund than "some."

The harshest criticism of the project came from within the ranks of molecular scientists, particularly at the onset. The molecular biologists' lobby, the Council of the American Society for Biochemistry and Molecular Biology, stated: "A large scale, massive effort to ascertain the sequence of the entire genome cannot be adequately justified at the present time… The Council wants to state in the clearest possible terms our opposition to any current proposal that envisions the establishment of one or a few large centers that are designed to map and/or sequence the human genome."

Three things changed this tune. Early successes in rapid sequencing made an impossible task seem less so. Being not so crazy liberated more money that flowed to many labs, blunting criticism. And the ease of contributing toward and exploiting from the successes of the project

via GenBank and other online resources gathered new allies.

We can expect that the greatest friction to an actual All Species Project will arise from within the ranks of taxonomists. There will be concerns about spending money to catalog every bacteria, or fungi, or mite. There will be arguments against wasting money on surveying anything less than a hot spot. There will be a case made against mapping species when the money could be spent on protecting habitats of species about to go extinct. We should be prepared for internal resistance, and overcome it via early successes and an open process of disseminating work and funding.

Back to Top

Feasibility Will Shift

The initial 1986 Santa Cruz workshop on sequencing the human genome surprisingly concluded, "a complete genome sequence was not feasible" at that time. A letter published in Nature about the emerging idea concluded "sequencing the genome would be about as useful as translating the complete works of Shakespeare into cuneiform, but not quite as feasible…"

At the time of the first conferences in the mid-80s, a state-of-the-art lab could sequence 500 bases a day, working non-stop. In 1985 Walter Gilbert, Nobelist, co-founder of one of the first biotech companies, Biogen, calculated that DNA sequences were accumulating worldwide at the rate of only 2 million base pairs per year. "At this rate," he figured, "a complete human genome sequence will take a thousand years." He then estimated that we could reduce that time to only 100 years "without any special effort." But that doing the genome would require an effort on the scale of the Space Shuttle to complete it in a reasonable amount of time. When he first announced that he estimated the total cost to be above $1 billion, there were gasps from the audience of scientists. Yet about this same time the Japanese embarked on a program to build a robot that could sequence 1 million base pairs per day. Nobelist David Baltimore, president of Caltech, was an early skeptic. He later said, "One of the things I didn't fully anticipate was the state of progress in automation."

To echo the Santa Cruz genome meeting, completing the inventory of all species today is not feasible at this time. It will require technology, innovations, resources, capacity, and funding not presently available. But the trick that technology entrepreneurs and the Human Genome Project have learned is that one can bank on shifts in feasibility occurring almost on schedule, and that one can leverage these future improvements by expecting them. All Species should plan on, and encourage in any way, these shifts in feasibility.

Back to Top

Tools First

Leaders of the Genome project constantly had to balance the job of making tools versus the job of using tools. Should the dollars flow to inventing new ways to speed up the process, or should the dollars be channeled to actually doing some work with imperfect tools? If you waited, the tools got faster, but if you kept waiting, you never got started.

This was the constant dilemma, addressed at almost every meeting. There was constant pressure from working scientists to push the deeper discovery of disease genes using the technology at hand, but Jim Watson, who lead the initiative for several years, was near obsessive in protecting the core, audacious vision of speeding up the discovery of all genes. In his metaphor the Human Genome Project was a cyclotron for "finding out what being a human is." It was opening up new vistas. In software terms, it was a new operating system, not new office applications. Resources should flow to perfecting the new system.

All Species needs new tools. It also needs to use those tools. The dilemma of funding tool development which postpones tool deployment, or funding work now instead of funding the acceleration of the work, is a dilemma that will remain for the entire duration of the All Species Project. To judge from the experience of the Human Genome Project the pressure to fund work now is natural, ordinary, and ubiquitous. The pressure to fund acceleration, on the other hand, is not ordinary and requires additional commitment and constant reinforcement.

Back to Top

Evolution, Not Revolution

Science magazine quotes David Botstein, who developed much of the new sequencing technology: "In the early days, it was believed that a radical new technology would be required. But it didn't turn out that way." Instead existing procedures were fiercely automated, and improved incrementally and ceaselessly. In fact very little of the blue sky technology that the early workshops imagined would be invented to get the job done was ever invented.

What worked were simply faster computers, and a radically different approach. Craig Venter and others pioneered a system called "whole genome shotgun sequencing" which greatly enhanced the power of existing technology. Ceaseless improvements, and new ways of using proven technology, won the race.

A reoccurring theme in the mission statement of the All Species Inventory is the need for radically different and new tools. All Species must be open to the possibility of succeeding using enhanced existing tools applied in new ways, or simply old tools automated to lightening speed. However because current taxonomic procedures are so low tech, almost any improvement may resemble radical technology. But we should encourage evolutionary improvements as a possible, or even likely, pathway.

Back to Top

Let Them Correct

To centralize or not to centralize, that is the question. The technical challenge for the Human Genome Project was how to coordinate, verify, synthesize and disseminate a vast amount of information from many sources. The first thought was to create a single large center.

The only person present at most of the initial workshops on the genome was Walter Gilbert, who had left Harvard to co-found Biogen. Gilbert viewed the genome project as the Holy Grail of human genetics. It would reveal divine secrets, "even if we don't understand them." Gilbert decided that the way to do this project was to start a company and make it the center of research, a company that would do all the sequencing on demand, as a business service to other labs. He pitched his start-up Genome Corp to investors. They would accumulate a central database, construct a physical map, and charge fees for access. Their market would be universities and pharmaceutical companies. They would first sequence the entire genome of the smallest self-contained organism, a bacterium found in goats, Mycoplasma capricolum.

Many of Gilbert's ideas lived on in the project, but his idea of a centralized research center did not. It was never funded. Research was distributed to hundreds of labs, in dozens of countries. Still, the issue of verification and reviewing the quality of the information remained. In the early days of the project there was an attempt to send all the sequence data through a review process. Only that data that passed inspection would be posted. As one could expect, the system overloaded and collapsed. The process reverted to GenBank, which accepted sequences without the usual peer-reviewed vetting process. Anyone could submit sequences and they could be junk. And a lot of it was. The quality was sorted out later. This pragmatic, though slightly inefficient process, had several advantages. It allowed rapid accumulation. It allowed many to participate. It solved the centralization issue (there was one database, but nobody was in charge of it, so to speak.)

A prime question for the All Species Project is likewise, how centralized should it be? Can we envision a method that would allow extremely speedy "publication" of new material, from many contributors of varying quality, without compromising the final reliability, and without having to vet each piece of information on the fly? Another metaphor is to imagine, like GenBank, a wider pool of peers (with appropriate tools) that can correct the raw information themselves. The model of the Human Genome Project is worth further study.

Back to Top

Judge By The Least

Throughout the Genome Project a steady debate raged over whether the main effort of the Project should focus on "mapping" genes and gene linkages, or on sequencing genes. Mapping meant analyzing the data to find markers and genetic links, being clever where you looked, and in some way prioritizing the effort. Sequencing came to mean almost a mindless, assembly line look at everything. Many wanted the project to scientifically prioritize the targets and spend the most resources on the mostly likely productive areas. Then later you go back and do everything.

During his reign early in the project Jim Watson emphasized sequencing. He knew mapping would occur no matter what happened. Sure you need to do mapping, he argued, but mostly you need to do bulk sequencing, which no one really wanted to do. "Saying you support mapping without sequencing is like saying I'll marry you but there will be no sex." This was somewhat in response to Craig Venter's pre-Celera group at the NIH who were sequencing primary protein coding regions of DNA which yielded results very quickly, and produced a very incomplete, but still useful index. Watson opposed this cherry-picking approach.

In one way the progress of the Human Genome Project could be evaluated on how much of the least interesting, least desirable sections of the genome were completed. Because the junk and barren parts were completed, the project was legendary.

In the same way, All Species should judge its own progress on how much of the non-hot-spot parts of the world it discovers. Focusing on hot spots is a natural and logical attraction for doing a global inventory, but there must a steady, relentless press toward exploring the least likely places, the neglected species, the hard-to-reach biomes. Hot spots will have a natural constituency. The cold spots won't. Yet the value of "all" can only come if the cold spots are pushed. As the most barren places, and least valuable taxa, are completed, the project will attain the legend and import it deserves.

Back to Top

Inefficiency Is Overrated, Competition Is Under-appreciated

Competition is inefficient. Why have two organizations, with their attendant overheads, replicating the work of each other? In a perfect world, teams would cooperate, overlaps be rooted out, and duplications outlawed. Yet there is no doubt that the frenetic race between the NIH and Celera to be the first to sequence the human genome enormously sped up the process of discovery. Celera re-did all the sequencing the NIH/DOE had done, and then completed the project in a total of three years, and all for about $300 million, or about one tenth the cost of the government's project.

All Species should not be afraid of some inefficiency. Inefficiency due to competition or lack of centralization is overrated as a problem. This kind of inefficiency is simply a cost of flexibility and innovation. Racing two different technologies, or teams of researchers, while inefficient in the short run produces long-term gains in know-how, ambition, and speed.

Back to Top

Global is Difficult

The first meetings convened to sequence the entire human genome took place in the US, yet from the start, the pioneers wanted the effort to be an international one. This proved difficult. The Japanese, for example, never increased funding to any large extent. And most of their relatively modest contributions came from and were invested into industry. Japanese companies such as Seiko, Hitachi, Mitsui, and Fuji invented and manufactured the speedy robot devices to do automatic sequencing. Europe had early and eager funding. Britian especially was instrumental in initial funding and research. There were other small steps in the right direction towards a global effort. In 1988 a Workshop on International Cooperation for the Human Genome Project lead to a UNESCO program to fund fellowships for young scientists visiting from developing countries to learn techniques to bring back home. A more ambitious program, HUGO, was founded to promote international cooperation, but eventually failed to get serious funds.

GenBank was the most useful and powerful globalization move. Permitting anyone anywhere to post information, or use information, did more for globalizing the project than any other organization. (Of course GenBank is used for more than the human genome and its operation predated the Human Genome Project.)

The difficulty of global can be stated simply. Most of the funding for this work came from governments, and governments usually have to limit their funding to domestic organizations. It is difficult to fund an amorphous international non-profit, without setting up an office in each country. Furthermore as the US government stepped up funding, the project increasingly began to be perceived in some areas as a US project. Even the race was a US-US race. In later years significant funds from other countries were invested into the project, but the aspirations of the founders to develop a truly international effort never materialized.

Doing all species necessarily means this must be a global project. It must be more global than the Human Genome Project. The All Species Project should take deliberate steps to avoid being seen as primarily a US-government funded project. It may want to incorporate itself in other countries. As in the genome project, the most useful step it can take is to develop an open process that accepts material from all over the world, and delivers information and knowledge all over the world.

Back to Top

Private Funding Is a Catalyst

While the bulk of funding for the Human Genome Project came from governments, chiefly the US, critical early funding came from private foundations. Two institutions played prominent roles.

The Howard Hughes Medical Institute was vital in the late 1980s. It funded several important international conferences on the launching of the initial grand idea. It also picked up funding of the still-young Human Genome Mapping Library in New Haven, It funded some French work on unifying genetic linkage maps globally, and it provided critical funding for HUGO, the Human Genome Organization, which was an international effort to coordinate worldwide efforts to map the gene. In total the HHMI contributed $16 million in the first 5 years of the project.

The Wellcome Trust, based in England, altogether has devoted over $300 million to the human genome effort over the course of the project. They also funded HUGO in its early years. In the1990's Wellcome's contribution to the genome effort exceeded the total of the British government's funding through its Medical Research Council.

And just as private money was instrumental in launching the genome project, private commercial money from venture capitalists was instrumental in completing the project. Celera used about $300 million in venture-raised funds to fast-forward the completion of the project. This amount purchased about 300 high-tech sequencers running around the clock, massively parallel processing sequences. There were other biotech companies who also funded hundreds of millions of dollars toward work on the human genome.

Governments will have to play a role in funding the All Species Project. But just as early and steady funding from private sources, both philanthropic and commercial, elevated the Human Genome Project away from the ordinary, so too can early and steady private funding catalyze the All Species Project. Private funding, properly applied, can bring such a project to the level where it is able to attract needed commercial, industrial and governmental funding. And it can still attract further private funding that can continue to accelerate its completion.

Back to Top

The Press is Vital

From nearly the first meeting in Santa Cruz to the last announcement of the completion of the last gene sequence, the press was invited to participate in just about every phase of the genome project. And they did. The early workshops had little consensus. The press witnessed and wrote about the screaming between participants as they battled each other to set the

agenda. The press wrote about the doubts and internal resistance to the idea. They covered the political backstabbing (of which there was much) and the in-fighting. And they covered the actual scientific news.

In the first few years, during the workshop phase, the heavy lifting needed to make the idea known was done in editorials in Science and Nature, and sometimes in general circulation magazines (Newsweek, U.S. News and World Report) and newspapers (The New York Times).

The constant press coverage accomplished several things. First, it made the idea important. It told everyone that this idea was so important that even the disputes and disagreements about it were important. Secondly, the press served as a recruitment device that communicated the concept to other scientists outside the genomic field, so that when their input was solicited for making scientific priorities, they at least were aware of the project and some of its issues. Thirdly, as the project took off and specialties increased, the press also communicated findings to those in the field who couldn't keep up with distant edges of news. The science press, in that respect, formed a small part of the entire project.

The All Species Project should welcome the press into its genesis and development. While the press can bite and be nasty at times, the value they can bring is worth the trouble. If All Species is as important as we say it is, the press should be covering this in full. By that coverage they will convince others it is important, including other scientists, if not other taxonomists.

Back to Top

The Public Counts

While criticism may peak internally, support can be sought outside. In the late 90s, a survey by UCLA reported that the greatest support for the Human Genome Project was much more likely to be found among government and industry workers, rather than in academic circles. When asked by the Industrial Research Institute to rank major science projects, industry and government administrators put the Human Genome Project up front.

All Species has an even better chance to become a project the public cares about. Far more people can identify a known species than can identify a known DNA sequence. Far more would be thrilled to find a new species, than to find a new sequence. The leaders of the Human Genome Project were forever striving to come up with public-friendly ways to convey the immediacy of their project and the closest they came was to invoke the possible cure of disease. The spokespersons for All Species should have an easier job, if we do it right. Amateur naturalists can and should play a part in this inventory. Biodiversity and All Species Days where communities of citizens count up all the species they can find in one day could be fashionable. This is a project that can relate to everyone: All species for all people.

Back to Top

Have a Memorable Five Year Plan

The first 5-year plan for the Human Genome Project developed in the early workshops could be written on the back of a business card. It was easy to see, remember, and keep in mind. It was:

1. Mount a pilot program or sequence a model organism.
2. Reduce sequencing costs by half (from $1 per base pair to 50 cents).
3. Sequence 10 million base pairs.

The All Species Project should have an equally measurable and memorable goal for the first five years.

Back to Top

Now Is the Time

As chancellor Robert Sinsheimer struggled to convene the first conference on sequencing the entire human genome in all it enormity, he reflected: "The human genome surely would someday be sequenced, once and for all time. The achievement would be a landmark in human

history and the knowledge would be the basis for all human biology and medicine of the future. Why not now?"

It's a good question, why not now?

Someday all the species living on earth will be identified, although surely not all the ones alive today. Why not now?

I have focused primarily on the genesis of the Human Genome Project, within its first five years. I have not dealt with the many issues and lessons that came up later, such as patents for genes, or the question of sharing and commercializing public domain information. Indeed there is much I haven't mentioned. I can refer readers eager for more to a great book on the early history of the Human Genome Project called Gene Wars, by Robert Cook-Deegan. There is also a brief seven-page history published earlier this year in the February 16, 2001 special issue of Science on the completion of the project (pp. 1182-1188). I also conducted some interviews of participants via phone for this research. The interpretation of, and comments on, this history are mine and not the All Species Foundation.

Back to Top