

Open Source DNA?

Eugene Thacker

[eugene.thacker@lcc.gatech.edu]
School of Literature, Communication, & Culture
Georgia Institute of Technology

(Note: parts of text also to appear in *Biotech Hobbyist Magazine*)

Opening the Biomolecular Black Box

What follows here is a series of observations, comments, and reflections on the current intersections between computer science and molecular biology. In conjunction with issues pertaining to open source initiatives, the aim of this paper is to raise similar questions in the domain of biotechnology.

All of us have witnessed the media-hype generated by such biotech issues as the human genome, human cloning, and debates over the use of embryonic stem cells. But what often goes unmentioned is that the real generator of radical change in fields like biotech is not genome mapping, cloning, or genetic engineering – it is “bioinformatics.” Put simply, bioinformatics is a growing discipline which straddles computer science and molecular biology (here at Georgia Tech, where I teach, the first bioinformatics degree program was established in 1999). Currently, bioinformatics mostly means the use of computer technology to aid in the study of life (that is, new tools for molecular genetics and biomedicine). Already, over the past decade or so, numerous companies have formed which specialize in the application of computer science to solve problems in biotech research. The recent race to map the human genome is one such example: both the public and private teams made use of automated genome sequencing computers built by Perkin-Elmer. Without the aid of specialized software and hardware, research on the human genome would not have made the progress it claims to have made thus far. Last year, the investment firm Oscar Gruss & Co. released a study of the field, suggesting that bioinformatics may generate some \$2 billion over the next five years. As the *New York Times* put it, the human genome has, for better or worse, been “a technology-driven quest.”

But is that all that bioinformatics is? In other words, what other kinds of developments can emerge out of this intersection between computer science and molecular biology, between computer code and genetic code, between data and flesh? Could it be that approaches from computing (network theories, systems theories, parallel processing, a-life) might have something to teach us about the complexity of the organism? Could such approaches even transform the way in which molecular genetics and biotech has traditionally thought of the organism, the body, and biological “life”?

Download, Tweak, Upload

The title of this paper is more of a question than any sort of statement. What would it mean to have “open source DNA”? How might we define a group of heterogeneous activities under this name? What is open source DNA?

In the same way that the open source movements have raised issues concerning the production, development, distribution, and use of software systems, open source DNA could do something similar for the combined fields of molecular biology and computer science (including other areas, such as A-life, molecular modeling, telemedicine, complexity, network computing, and so forth).

Is there a need for open source DNA? From one perspective, DNA is already open source: the publicly-funded human genome project makes its findings available to the public through its GenBank database and website, just as other academic and government-funded projects do for proteins, SNPs, RNA, and other biological components. In addition, a number of software applications are available as freeware or shareware, along with the increasing number of research applications which function online (again, mostly from academic institutions).

But as we know, this is not the whole picture. Many datasets are privatized (such as those held by Celera, DoubleTwist, or Human Genome Sciences) and have exorbitantly high subscription rates (mostly intended for pharmaceutical corporations). In addition, a great deal of the computer tools which undergirds biotech research (hardware, software, and wetware) come at a great cost, with little or no low-tech or low-cost alternatives. Even when such tools are available, their learning curve is high enough that usually some background in either computer science, computer programming, database systems, statistics, or molecular biology is required. For individuals or groups working outside of specialized fields (artists, educators, activists, cultural theorists), and for those within such fields with only partial knowledge of biotech (scientists and engineers in other fields), the barriers to becoming active in biotech can be overwhelming.

Therefore the necessity for open source DNA is both political and technical. It is political because there is much critical and creative work to be done in relation to biotech’s general approach to the body, medicine, and perspectives on biological “life” itself (see Critical Art Ensemble’s book *Flesh Machine* for more). But it is also technical, because in order that any effective intervention in biotech can take place, the basic knowledges, skill sets, and tools of biotech must first be made available to individuals and groups outside of its specified disciplines, institutions, or corporations.

A Discipline or an Industry?

The interesting thing about bioinformatics is that, on the one hand, it is a new discipline, a hybrid of knowledges and techniques from computer science, as well as molecular biology. But on the other hand, bioinformatics has risen hand-in-hand with new companies, proprietary software, and a range of products and services.

Broadly speaking, bioinformatics includes several activities:

First there are the so-called “pick-and-shovel” companies. As the name indicates, these are companies that make the tools needed for biotech research, where research and product development are one in the same. Such tools can be software applications (such as Incyte Genomics’ “Lifeseq” software suite), they can be hardware (such as Affymetrix’s “GeneChip” microarray system), they can be database and networking tools (such as those offered by DoubleTwist), or they can be a combination of IT solutions for biotech research (such as those offered by Lion Bioscience).

Secondly, there are organizations which deal in handling biological data. The most familiar examples are the human genome teams, the public-funded groups (such as the National Center for Biotechnology Information, or NCBI) and Celera Genomics, a private genomics company. Both institutions house their own data on the human genome, the main difference being that while the NCBI offers its databases to the public, Celera charges for access on a corporate-level subscription basis.

Finally, there are research groups (many which exist at universities) whose primary interest is in developing novel ways of applying computer science to molecular biology research problems (Bioinformatics.org and Open Bioinformatics are examples). Research can range from the very practical (e.g., how to apply techniques in computer error detection towards genetic scanning) to the more radical (e.g., using AI or a-life to develop “intelligent” bioinformatics software apps).

Certainly, these are not definite boundaries, as nearly all biotech research requires computational tools of some kind. In addition, the past few years has seen a growing interest in computer industry and biotech industry mergers because of bioinformatics (e.g., Sun and DoubleTwist, IBM, Compaq and Celera, Motorola). Therefore, it is important to note that although bioinformatics may be an “emerging” discipline, in many ways it is already mature in its relationships with institutions, corporations, and academic disciplines.

This is worth noting because it means that any “alternative” approaches in bioinformatics and uses of biological data, will have to confront issues such as access to information, access to tools, development of skill sets, distribution of knowledge, and the challenges of trans-disciplinary work. The main question which is put forth is: How does an individual or group acquire the knowledge, skills, resources, and tools needed to work in a non-orthodox manner in biotech?

Not surprisingly, artists have been among the first to explore such questions. But the results are often less than satisfactory, even when art-science collaboration is involved; too often the resulting works operate only at the symbolic or representational level. However, such art-science projects have been instrumental in raising critical and political issues with regard to biotech, suggesting that a new type of serious research can co-exist alongside a critical and political consciousness.

We might begin, then, by elaborating a series of theoretical questions which bioinformatics raises. From there we can consider possible fields of research in biotech to look into, and then begin asking practical questions.

Soft Machines: Theoretical Questions

The human genome projects seem to suggest to us that flesh and data are equivalent: DNA can be extracted from an individual's body, then encoded into digital format (using fluorescence tagging), then sequenced and uploaded into an online database. That data can then be used in diagnostics, genetically-tailored drug design, gene therapy, or in regenerative medicine therapies. But is DNA really equivalent to binary code? Elsewhere, I have referred to this back-and-forth exchange between digital and analog DNA as "biomedia": the "translatability" of the genome between digital and analog. In the techniques of genomics, it is taken for granted that the wet DNA in a test tube is somehow "essentially" the same as the dry DNA spelled out on a computer database. But the larger implications of this technical assumptions are dangerous – it suggests that the true essence of the genome is not the material "stuff" out of which it is made, but rather some source code which exists irrespective of material instantiation (see Haraway, Hayles, and Kay for more). It seems that one of the questions which bioinformatics asks, is how much we can really claim to be uploading biological materials, and how much we are just cataloguing the body. Could a critical bioinformatics emerge from this, in which the situated, embodied character of the biological body is always taken into account, while never being totally divided from the informatic domain? If so, what are some of the challenges it would face?

In the same way that open source has contributed to a DIY computer culture and various types of hacker ethics, could the design of innovative bioinformatics software apps, combined with public access to the genome, spawn a DIY biotech culture? Could an increase demand on public access medical data, combined with advances in telemedicine, generate a new type of homeopathic health care? At the furthest reaches of the extreme, how might this "open source DNA" movement affect areas such as media art, education, body performance, regenerative medicine, body art, and wet computing?

Although there is a great diversity in biotech research, much of it has continued to focus on genes and the genome as their primary targets (as company patent portfolios illustrate). This single-minded approach has been countered recently by alternative approaches borrowing from systems theory and complexity. How might we think about the intersection between computer science and molecular biology be rethought as a hybrid discipline? What novel knowledges are produced in their intersection? What might the role of computer technology be, if it is to be more than a mere tool for bioscience research? How might each discipline not just aid, but actually transform the other?

While questions of ethics are always given at least a conciliatory nod in any discussion of biotech, ethics needs to be rethought with respect to biotech. One starting point may be the ethical debates generated by the discourses on open source, patent protection, and "hacktivism." Would an open source DNA movement confront the same ethical and political challenges that the open source software movements have? In this sense, how would a politically-motivated, open source DNA be different from forms of hacktivism? How would it be different from the controversial activities labeled "bioterrorism"? How might a genuine bioethical protocol be established, such that biotech resources are not used irresponsibly?

Soft Machines: Practical Considerations

As a way of fostering some workshop-type thinking on this topic, we can form a beginning list of concerns for open source DNA:

1. What kinds of resources are currently available to the public, and in what kinds of formats? For instance, what types of data does the NCBI's human genome dataset make available? Is its format compatible with XML-based software apps such as those made by Rosetta InPharmatics? How much of this data is accessible online? How much of it depends on specialized software? What types of publicly available networks can be formed around such concerns?

2. What kinds of tools and applications are available for biotech research? What kinds of research do they make possible? How many of these apps – if any – are freeware or shareware? Are these applications open source? If not, what programming knowledge sets are they based on? Many bioinformatics apps are based on XML – could this open the way for an XML-based open biosoftware initiative?

3. How might open source DNA labs be set up in a manner that is compliant with safety, technical requirements, networking, and efficient access to resources? How can the computer lab and the molecular biology lab be integrated in innovative ways? How might further computer science – molecular biology collaborations be fostered in this area? How might various institutional bodies aid in the formation of such labs (grants, foundations, universities)?

4. What are some immediate practical and political consequences of open source DNA? At the policy level, at the health care level, at the research level, at the economic/corporate level, and at the industry level? How can cultural and political-activist projects be effectively realized in these fields?

Blood Music

As a cultural theorist of science and technology, this intersection of computer science and molecular biology has many significant impacts outside of the sciences. For one, biotech fields like bioinformatics are practically demonstrating the ways in which boundary between the body and technology are being transformed, and, in some cases, effaced altogether. No longer is the body the privileged domain of “nature,” just as our technologies are more than inert objects we simply control and use. It appears that biotech research is delving deeper into the carbon-silicon barrier, and finding not a barrier at all, but rather a permeable membrane that is constantly changing its shape.

This philosophical transformation has direct impacts in the political concerns over germline gene modification, DNA screening and privacy (bio-cryptography?), biopiracy and biocolonialism (population and ethnic genomes), and pharmacogenomics (genetically-tailored drugs). Ethical concerns over “tampering with nature,” biodiversity conservation, economies of biological materials, and other concerns all arise in part from the way in which the relationship between bodies and technologies is viewed.

Likewise, science fiction has, for a long time, imagined the extreme possibilities - both positive and negative - which biotech brings with it. Examples of such extreme biotech (BioX?) include: full-body regeneration (X-Men), replicant engineering (Blade Runner), next-gen horror movie efx ("bodies that splatter"), biotech telerobotics ("The Girl Who Was Plugged In"), biomolecular morphing (The Thing), bio-fashion (Schismatrix), and biomolecular consciousness (Blood Music).

However fanciful such visions may seem, they point to the need for alternative approaches for thinking about the biomolecular body. In actual science research, approaches such as systems biology, autopoiesis, self-organization, biopathways, epigenetics, and CAS (complex adaptive systems) are all pointing to different ways of thinking about biological life beyond the centrality of DNA or the genome.

Bioinformatics is the key to rethinking computer science & molecular biology across their traditional disciplinary divisions. While there are pragmatic examples of the ways in which computational approaches are advancing biotech research (such as the HGP), bioinformatics places flesh and data in such an intimate proximity that it challenges us to think of technology beyond the tool, just as it challenges us to think of biology as much more complex than a "master molecule" residing in nature.

References

Bear, Greg. *Blood Music*. New York: Ace, 1983.

Benton, D. "Bioinformatics: Principles and Potential of a New Multidisciplinary Tool." *Trends In Biotechnology* 14(8):261-72 (August 1996).

Bioinform (<http://www.bioinform.com>)

Bioinformatics.org (<http://www.bioinformatics.org>)

Critical Art Ensemble. *Flesh Machine: Cyborgs, Designer Babies, and New Eugenic Consciousness*. Brooklyn: Autonomedia, 1998.

Gershon, Diane. "Bioinformatics in a Post-Genomics Age." *Nature* 389 (27 September 1997): 417-18.

Haraway, Donna. *Modest_Witness@Second_Millennium.FemaleMan@_Meets_OncoMouseTM: Feminism and Technoscience*. New York: Routledge, 1997.

Hayles, N. Katherine. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago: U of Chicago P, 1999.

Howard, Ken. "The Bioinformatics Gold Rush." *Scientific American* (July 2000): 58-63.

Kay, Lily. *Who Wrote the Book of Life? A History of the Genetic Code*. Stanford: Stanford, 2000.

Open Bioinformatics (<http://www.open-bio.org>)

Palsson, Bernhard. "The challenges of in silico biology." *Nature Biotechnology* 18 (November 2000): 1147-50.

Persidis, Aris. "Bioinformatics." *Nature Biotechnology* 17 (August 1999): 828-830.

The Scientist. Special Issue: The State of Bioinformatics. *The Scientist* 14.3 (27 November 2000).