

**INTERNATIONAL REVIEW OF
INDUSTRIAL PROPERTY AND COPYRIGHT LAW**

**BEYOND BLUE GENE:
INTELLECTUAL PROPERTY AND BIOINFORMATICS**

DR MATTHEW RIMMER
LECTURER
FACULTY OF LAW
THE AUSTRALIAN NATIONAL UNIVERSITY

ACIPA, Faculty Of Law,
The Australian National University, Canberra, ACT, 0200
Work Telephone Number: (02) 61254164
E-Mail Address: RimmerM@law.anu.edu.au

**BEYOND BLUE GENE:
INTELLECTUAL PROPERTY AND BIOINFORMATICS
DR MATTHEW RIMMER***

Abstract

This article considers the challenges posed to intellectual property law by the emerging field of bioinformatics. It examines the strategies of biotechnology companies, information technology firms, and public sector researchers. Part 1 claims that the use of copyright law and contract law is fundamental to the protection of biomedical and genomic databases. Part 2 questions whether biotechnology companies and information technology firms are patenting bioinformatics software and Internet business methods, as well as underlying instrumentation such as microarrays and gene chips. Part 3 considers whether open source software and peer to peer technology will be able to counter this trend of privatisation. It raises important questions about integration, interoperability, and the risks of monopoly.

Introduction

A recent project called Blue Gene has come to symbolise both the promise and the hype of bioinformatics.¹ IBM has devoted \$US 100 million to build a super-computer which will seek to analyse protein folding. It boasts that Blue Gene is 1,000 times more powerful than Deep Blue, the machine that defeated world-chess champion Garry Kasparov, and can map the human genome.² IBM has also created a consulting division for its biotechnology and pharmaceutical customers. It has also set up a new organisation called Blueprint Worldwide, which will generate a public database of bioinformatics and biomedical data. Such grand ambitions herald the marriage of life sciences and information technology. They also highlight the importance of intellectual property to the field of bioinformatics.

Bioinformatics is the art and science of using computer systems to store, manage and analyse biological information. It brings together the diverse disciplines

* Matthew Rimmer, BA (Hons)/ LLB (Hons) (ANU), PhD (UNSW), is a Lecturer at the Faculty of Law in the Australian National University. The author is grateful to his informant Professor Simon Easteal of the John Curtin School of Medical Research at the ANU. He also gives thanks to Professor Brad Sherman of Griffith University for comments on this paper.

¹ Press Release. "IBM Announces \$100 Million Research Initiative to Build World's Fastest Supercomputer: Blue Gene To Tackle Protein Folding Grand Challenge", Yorktown Heights, New York, 6 December 1999.

² Such claims show a certain amount of chutzpah given that Celera Genomics relied upon a Compaq super-computer to map the human genome.

of mathematics, statistics, engineering, and computer science to map and model genes and proteins. Bioinformatics played a critical role in mapping the human genome in both the large public and commercial projects. Robert Cook-Deegan notes:

Databases, computers, and mathematical algorithms proved as important as DNA sequencing, cloning, and other more obviously biological techniques. As geneticists produced a deluge of data during the 1990s and beyond, those who understood hardware and software would play an increasingly important role.³

The public consortium relied upon cloning methods to map the location of genes, dividing the genome into small blocks. The private efforts led by Celera Genomics engaged in whole genome shotgun sequencing, fracturing the DNA of an organism into small fragments and then using powerful computer sequencing machines to identify the base pairs at the end of each fragments. The sequencing, storage and retrieval of genetic information has generated new possibilities for understanding the function and structure of genes and proteins.

Gene chip or microarray technology has also been vital to the analysis of genetic sequences. Taking its lead from computer technology, this technique has revolutionized genomic research.⁴ Kevin Davies comments upon this technology:

These DNA chips have generated the biggest buzz in molecular biology circles since the advent of the polymerase chain reaction - the technique invented by Kary Mullis that amplifies minute traces of DNA - some fifteen years ago.⁵

Microarrays are vast libraries of short DNA sequences attached to tiny glass or silicon supports and are being used to screen nucleic acid population. This revolution in miniaturization and high throughput screening is analogous to what has occurred in the computer industry over the past decades. Consequently, more data can be obtained more quickly than ever before. Affymetrix is the dominant player in the microarray or genechip technology. The patent position is complex and lawsuits that relate to the

³ Cook-Deegan, R. *The Gene Wars: Science, Politics, And The Human Genome*. New York and London: WW Norton and Company, 1994, p. 288.

⁴ Ekins, R. and Chu, F. "Microarrays: Their Origins and Applications", *Trends in Biotechnology* 1999, Vol. 17, p. 217; Sinclair, B. "Everything's Great When It Sits on a Chip?: A Bright Future for DNA Arrays", *The Scientist*, May 24, 1999, at 18; Sutherland, G. "The Human Genome Project", *Australian Journal Of Forensic Sciences*, 2000, Vol. 32, p. 19-24.

⁵ Davies, K. *The Sequence: Inside The Race For The Human Genome*. London: Weidenfeld and Nicolson, 2001, p 218-219.

patent rights in this technology are already in progress on both sides of the Atlantic.⁶ Other companies such as Motorola and Hewlett Packard have entered into the market for microarrays and gene chips.

Bioinformatics also plays an essential role in drug development. High throughput screening involves the use of fully automated robotic technologies to test compounds against a molecular gene target identified by genomic approaches. Combinatorial chemistry is used to generate vast on-line libraries of entirely novel chemical entities. Researchers predict that the time and cost it takes to develop drugs will be significantly reduced by biology-based approaches. It now takes about 12 years and \$500 million to bring a drug to market. William Haseltine, the Chief Executive Officer of Human Genome Sciences, declared:

We think we can reduce this by about six years. By both increasing the success rates and shortening the trial period, we think we can dramatically improve cost efficiency in drug discovery.⁷

The providers of bioinformatics services can aid and assist pharmaceutical companies who are engaged in drug discovery. They can also conceivably move into the business of developing and selling drugs themselves.

This paper would like consider the intellectual property implications of bioinformatics in light of significant private investment in the field.⁸ Part 1 will consider the protection of databases under copyright law. It will investigate the case of Celera Genomics. Part 2 will look at the patenting of bioinformatic systems. It will consider whether the popularity for software and Internet business patents will be translated into the life sciences. Part 3 will mention some of the strategies of public research institutions, such as the Sanger Institute, the European Bioinformatics Institute, and the Cold Springs Laboratory. It will consider whether open source software and peer to peer technology can counter the trend towards privatization.

⁶ Gilbert, P and Walter, C. "Patents and the Human Genome Project - New Claims for Old", *Trends in Biotechnology*, Vol. 19 (2), 2001, p. 49.

⁷ Philipkoski, K. "Genome Map Heralds Cheap Drugs", *Wired News*, 13 September 2000.

⁸ Saviotti, P., de Looze M-A, Michelland, S., and Catherine, D. "The Changing Marketplace Of Bioinformatics", *Nature Biotechnology*, 2000, Vol. 18 (12), p 1247.

PART 1
CELERA GENOMICS:
DATABASES

Historically, biological databases were the preserve of the public domain. Robert Cook-Deegan considers the institutional politics behind biomedical and genomic databases - such as Genbank and the European Molecular Biology Laboratory:

Both databases were struggling to secure resources, very much stepchildren in the biomedical research family. Molecular biologists had little notion of the difficult issues facing any large public database. They were impatient to use the data but unenthusiastic about paying for its storage.⁹

Stephen Hilgartner depicts genomic research as a heterogeneous process where the terms under which data access takes place are often contentious.¹⁰ He has documented the tendency for publicly funded researchers in the Human Genome Project to close off their research from others. Rather than engaging in the sharing of data for academic and commercial reasons, publicly funded researchers failed to make information available. Nik Brown and Brian Rappert examine disputes about the distinction between 'public' and private' in bioinformatics.¹¹ They identify the major sources of tension and discuss what these suggest for the purpose of public investment in bioinformatic research. The authors identify a trend of commercial companies developing private databases of genetic information, and selling access to subscribers.

J. Craig Venter established Celera Genomics in 1998, with the help of Perkin Elmer Inc. Celera Genomics used automated shotgun sequencing technologies to map the human genome in the year 2000. Although the company intends to make the data publicly available, profits will be generated from proprietary protection of the raw data and annotation parts of the database and the patenting of some sequences.

Celera Genomics denies any intention of setting up a pharmaceutical company. The company claims that it will only seek patents on 100-300 human

⁹ Cook-Deegan, R. *The Gene Wars: Science, Politics, And The Human Genome*. New York and London: WW Norton and Company, 1994, p 287.

¹⁰ Hilgartner, S. "Data Access Policy In Genome Research", in A. Thackray (ed) *Private Science*, Philadelphia: University of Pennsylvania Press, 1998.

¹¹ Brown, N. and Rappert, B. "Emerging Bioinformatic Networks, Contesting The Public Meaning Of Private And The Private Meaning Of Public", *Prometheus*, 2000, Vol. 18 (4), p. 437.

genes, less than 1 percent of the total genome. Craig Venter envisions that Celera Genomics will be an information provider, along the lines of Lexis-Nexis, Bloomberg, and AOL:

Our goal is to make the complex, sometimes overwhelming, and ever-increasing volumes of biological information more accessible and useful to researchers in academia and industry. Toward that end, we are creating an unparalleled library of genomic information in our databases. Annotation of the data by Celera scientists using an array of bioinformatics tools will act as the platform for developing a range of products and services. We will offer these tools in a manner similar to the models used by other information companies, such as Lexis-Nexis, Bloomberg, and AOL. The need for services such as these will only increase as the volumes of information and the complex interrelated nature of that information increase. Pricing for subscriptions to this service will vary appropriately, depending on the product, the customer, and the application. We will provide value-added information to academics and other non-commercial researchers at reasonable rates, naturally bounded by those customers resources and appraisals of the value-added.¹²

Celera has a number of large pharmaceutical companies as database subscribers - including Pharmacia Corporation, Novartis, Amgen, Pfizer and Takeda Chemical Industries Ltd of Japan. Furthermore, it has also enlisted major commercial life science companies and academic biomedical research institutions as subscribers. These subscribers can have access to all Celera databases, tools, and annotation, including the human genome.

Celera Genomics boasts that it will release the entire consensus human genome sequence freely to researchers on Celera's Internet site. It believes that it is in the best interests of both science and its company, since it will allow researchers to advance science and medicine, and be introduced to Celera's high quality data and software tools. The business statement of Celera Genomics highlights the fundamental importance of database protection:

Moreover, the Celera Genomics group may be dependent on protecting, through copyright law or otherwise, its databases to prevent other organizations from taking information from such databases and copying and reselling it. Copyright law currently provides uncertain protection regarding copying and resale of factual data.¹³

¹² Venter, J.C. "Prepared Statement", Subcommittee on Energy and Environment, U.S. House of Representatives Committee on Science, Congress, 6 April 2000.

¹³ PE Corporation. Annual Report. 2000, p. 52.

Celera Genomics emphasizes that it will place no restrictions on how scientists use this database. However it has indicated that it would seek database protection, as exists in Europe, to inhibit other database companies from selling the Celera database. However, Celera Genomics is uncertain about the level of protection available for databases under copyright law.

In the United States, there is some doubt as to the copyright protection of databases. In *Feist Publications Inc v. Rural Telephone Service Inc*, the Supreme Court criticised the 'sweat of the brow' or 'industrious collection' - the underlying notion was that copyright was a reward for the hard work that went into compiling facts.¹⁴ It stressed instead that the facts must be selected and arranged in an original way. This judgment has held sway - in the face of four unsuccessful attempts to persuade Congress to nullify the decision, and implement special protection for databases.

In Europe, a database directive provides 15 years of protection for the contents of the database and each significant update, and permits database owners to prevent the use of substantial parts of the database. Bernt Hugenholtz reports that "five years after the adoption of the Database Directive the contours of the new database right remain obscure".¹⁵ Jasper Bovenberg speculates: "As the United States has not yet enacted corresponding domestic legislation, U.S. companies engaged in the biological data business might consider setting up their databases in Europe if they want to be eligible for this type of supplementary protection".¹⁶

In Australia, it is possible to obtain general copyright protection of databases. In *Telstra v Desktop Marketing Systems*, Justice Finkelstein of the Federal Court of Australia concluded that the old English law regarding originality had not been discarded. He did not think it possible to jettison such rules and replace it with the principles expressed in the United States case of *Feist*.¹⁷ As a result, it would be possible to gain copyright protection of databases because the level of originality is pitched at such a low level.

¹⁴ (1991) 113 L. Ed. 2d 358.

¹⁵ Hugenholtz, B. "The New Database Right: Early Case Law From Europe", paper presented at *Ninth Annual Conference on International IP Law & Policy*, Fordham University School of Law, New York, 19-20 April 2001.

¹⁶ Bovenberg, J. "Should Genomics Companies Set Up Database In Europe? The EU Database Protective Directive Revisited", *European Intellectual Property Review*, 2001, Vol. 23 (8), p 361.

¹⁷ [2001] FCA 612

Whatever the jurisdiction, it seems that Celera Genomics and its biotechnology industry rivals will have to establish that the databases and libraries are original works, in the face of allegations that they draw upon public databases.¹⁸ However, Jasper Bovenberg foresees complications:

More confusing in this respect and a potential legal battleground is the fact that DNA sequence databases are typically made of the contributions of multiple contributors; it took the results of a collaboration involving 20 groups from the United States, the United Kingdom, Japan, France, Germany and China to produce a draft sequence of the human genome. Both the collection of raw sequence data and the annotations or proposals for the functions of the genes in the database often represent substantial pieces of research in themselves. Even those created by commercial genomics companies are the result of the combination and extension of a mix of commercial and non-profit databases.¹⁹

However, the hurdle of originality is not an insurmountable one. As long as Celera Genomics and its contemporaries add sufficient value to the information, they should be assured of getting copyright protection in respect of the genetic information.

Celera Genomics supplements database protection under copyright law with contract law. It has developed a range of agreements to govern the educational and commercial use of its genetic information. The free public access click-on agreement allows an academic user to identify gene loci within the genome and download these results up to 1 megabase of Celera Data per week. The free public access agreement for Celera whole genome sequence allows a user to identify gene loci within the genome and obtain an electronic copy of the Celera Data. The material transfer agreement provides that Celera Genomics is willing to make the human genome sequence data as described in Science available to commercial entities for validation and verification purposes only. The devil is in the detail. The contractual terms seek to safeguard Celera Genomic's intellectual property rights. They seek to prevent the unauthorised commercial use of the database information. There is a confidentiality

¹⁸ Tim Hubbard from the Sanger centre claims that Celera Genomics sought to hijack the public Human Genome Project. He alleges that the genome that got assembled in the private domain was based on what came out in the public domain. Of course, Celera Genomics denies such claims. The leader of the bioinformatics project, Gene Myers, points out that the firm was able to map *Drosophila* without any assistance from the human genome project.

¹⁹ Bovenberg, J. "Should Genomics Companies Set Up Database In Europe? The EU Database Protective Directive Revisited", *European Intellectual Property Review*, 2001, Vol. 23 (8), p 361.

clause in respect of the material transfer agreement.²⁰ Such terms would seem to cut down the scope of fair use, and raise questions about the effects of competition.

Furthermore it is foreseeable that this use of contract will be augmented by technological controls on the access to the on-line databases.²¹

Monopoly Risks

There is a concern that monopolistic actors will take charge as the bioinformatic community struggles towards an acceptable degree of standardisation on hardware and software.²² Brown and others comment:

Small software developers, once characteristic of the bioinformatics sector, now tend to sell their products to much larger companies. The fear is that the area may become increasingly dominated by a small number of commercial actors providing highly integrated visualisation, search and design packages. Indeed, the dominance of the administrative sector by Microsoft is seen as one paradigm for the way in which computerised biological research will be increasingly served by monopolistic suppliers.²³

The authors conclude that monopolisation is an unrealistic scenario in the bioinformatics field. They suggest that the field moves too quickly for a monopoly to establish; academic and public participation in the field is higher than in other computational markets; complete harmonisation is relatively unlikely; CORBA and Java are expected to provide a degree of diversity within standardised protocols.

Nathaniel Heller reports that Celera Genomics has been sensitive to accusations that the company is the "Microsoft" of the life sciences.²⁴ He observes that the company has hired the lobbying firm Williams and Jensen in order to maintain good relations with the government, so as to forestall the possibility of an anti-trust action. Other companies such as Human Genome Sciences have followed

²⁰ Celera Genomics also emphasizes confidential information: "The Celera Genomics group relies on trade secret protection for its confidential and proprietary information and procedures, including procedures related to sequencing genes and to searching and identifying important regions of genetic information".

²¹ Davison, M. "Sui Generis or Too Generous: Legislative Protection of Databases, Its Implications for Australia and Some Suggestions For Reform", *University of New South Wales Law Journal*, 1998, Vol. 21 (3), p. 729 at 734.

²² Brown, N., Nelis, A., Rappert, B. and Webster, A. "Bioinformatics: A Technology Assessment Of Recent Developments In Bioinformatics And Related Areas Of research And Development Including High-throughput Screening and Combinational Chemistry", Final Report for the Science and Technological Options Assessment Unit, European Parliament, 1999, p 24-25.

²³ *Ibid*, p 25.

²⁴ Heller, N. "From 'God's Handwriting' To Corporate Profits", *The Public I*, 22 May 2001.

suit. There are two specific events, which have prompted genome companies to become politically active.

The first was a joint statement in March by President Clinton and British Prime Minister Tony Blair stating that the human genome should remain in the public domain. That announcement crashed biotechnology stocks, bleeding the market valuations of companies such as Celera and Human Genome Sciences. Clinton and Blair clarified their positions soon after the announcement, maintaining that the patenting of specific human genes would still be legal and appropriate.

The second stimulus came in the form of the Justice Department's antitrust case against Microsoft Corporation. The antitrust investigations of Microsoft that began in the 1990s were based on concerns about Microsoft's use of copyright to block access to downstream markets. Microsoft had not been politically active, maintaining only a small office in Washington, until the Justice Department was well into its investigation of the company. A federal judge, Justice Thomas Jackson, ordered the breakup of the giant software company, dividing the company into an operating systems business and an applications business.²⁵ An industry observer commented: "A lot of these [genetics companies] CEOs saw what happened to Microsoft, not playing the game in Washington. If you're a highly regulated industry, you need representation in Washington. The phone companies get it. They know they need 10,000 lobbyists".²⁶ In response, Microsoft aggressively defended its position, and lobbied the new Republican President. In the end, the court of appeals allowed in part an appeal by Microsoft, and vacated the remedies of the federal judge.²⁷ The Justice Department reached a settlement with Microsoft.

The combination of the Clinton-Blair announcement and the Microsoft Corporation actions spurred the genome companies to seek to lobby and influence law-makers. The biotechnology companies will take appropriate measures to ensure that they are not the subject of antitrust actions by government regulators.

²⁵ *United States of America v Microsoft Corporation* (2000) 97 F. Supp. 2d 59; Heilemann, J. "Fear and Trembling In Silicon Valley", *Wired*, Vol. 8.03, March 2000; and Heilemann, J. "The Truth, The Whole Truth, And Nothing But The Truth", *Wired*, Vol. 8.11, November 2000.

²⁶ Heller, N. "From 'God's Handwriting' To Corporate Profits", *The Public I*, 22 May 2001.

²⁷ *United States of America v Microsoft Corporation* (28 June 2001, Court of Appeals, No. 00-5212).

PART 2
GENAISSANCE:
INTERNET BUSINESS METHOD PATENTS

In the wake of the human genome project, there was great controversy about the patenting of genes and gene sequences.²⁸ There was widespread concern about the thousands of patent applications being filed by ambitious biotechnology companies. In response, there have been a number of administrative, legal, and political responses. The United States Patent Office has laid down guidelines stressing the need for utility. The courts have been much more circumspect about the validity of patents regarding genes and gene sequences. Furthermore, the United States and European governments have expressed disquiet about the developments. The debate over the patenting of genes and gene sequences has monopolised public attention. It has lead to other important subjects being overlooked - in particular, control of biological software and hardware, and access to scientific information. The evidence suggests that biotechnology companies are not exclusively interested in the patenting of genes and genetic sequences. They are also patenting bioinformatics software and Internet business methods, as well as underlying instrumentation such as microarrays and gene chips.

Internet Business Methods

Rebecca Eisenberg comments: "As DNA sequence discovery has moved beyond targeted efforts to clone particular genes to large-scale, high-throughput sequencing of entire genomes, new questions have emerged. The DNA sequences identified by high-throughput sequencing look less like new chemical entities than they do like scientific information".²⁹ She draw analogies between the patenting of genes and genetic sequences with the recent controversies over Internet business methods. Eisenberg observes: "Recent decisions concerning the patentability of computer-implemented inventions may provide more guidance than prior decisions in the life sciences in

²⁸ Eisenberg, R. "Proprietary Rights and the Norms of Science", *Yale Law Journal*, 1987, p. 97, p 177; Eisenberg, R. "Re-examining the Role of Patents in Appropriating the Value of DNA Sequences", *Emory Law Journal*, 2000, Vol. 49, p 783; Forsyth, M. "Biotechnology, Patents And Public Policy: A Proposal For Reform In Australia", *Australian Intellectual Property Journal*, 2000, Vol. 11, p 203; and Shulman, S. *Owning the Future*. Boston: Houghton Mifflin, 1999.

²⁹ Eisenberg, R. "Re-examining the Role of Patents in Appropriating the Value of DNA Sequences", *Emory Law Journal*, 2000, Vol. 49, p 783 at 785.

predicting whether DNA sequence information stored in computer-readable medium may be patented”.³⁰ She goes on to discuss the cases of *State Street Bank And Trust Co v Signature Financial Group Inc*,³¹ and *AT & T Corp v Excel Communications Inc*³².

The United States has recognised that Internet business methods are patentable subject matter. In *State Street Bank And Trust Co v Signature Financial Group Inc*, the Court of Appeals for the Federal Circuit held that the use of a mathematical algorithm, formula or calculation to produce numbers will be patentable, as long as the result is “useful, concrete and tangible”.³³ Further, the Court expressly rejected the existence of a business method exception to patentability. In *AT & T Corp v Excel Communications Inc*, the Court of Appeals for the Federal Circuit reaffirmed the decision that Internet business methods were patentable.³⁴ In *Amazon Com Inc v Barnes And Noble. Com Inc*, the court applied this reasoning in relation to electronic commerce. One might add to this list the recent case of *Welcome Real-Time v Catuity Inc*,³⁵ in which the Federal Court of Australia affirmed the decision in *State Street Bank v Signature Financial Group* as persuasive in this jurisdiction.³⁶

Eisenberg makes an important insight that the patent protection of DNA sequence information is analogous to the protection of Internet business methods. She concludes from these cases “it is not obvious why DNA sequence information stored in computer-readable medium a product that requires human intervention and serves human purposes would be categorically excluded from patent protection”.³⁷ However, Rebecca Eisenberg fails to comprehend the full implications of this insight. Indeed she recoils from this position and reasserts the differences between DNA sequence information and Internet business methods: “Of course DNA sequence information stored is not the same thing as a computer-implemented business method, and it is certainly possible to define boundaries for the patent system that include the latter but not the former”.³⁸ The idea of Internet business methods is raised, but only

³⁰ *Id*, p 791.

³¹ (1998) 149 F 3d 1368

³² (1999) 172 F 3d 1352

³³ (1998) 149 F3d 1368.

³⁴ (1999) 172 F 3d 1352.

³⁵ [2001] FCA 445

³⁶ (1998) 149 F 3d 1368

³⁷ Eisenberg, R. “Re-examining the Role of Patents in Appropriating the Value of DNA Sequences”, *Emory Law Journal*, 2000, Vol. 49, p 792.

³⁸ *Id*, p 793.

as an analogy, and then it is dismissed. The possibility of database protection, which we have already discussed, is consigned to the footnotes.

Bioinformatics

However, the relationship between biotechnology and information technology goes beyond mere metaphor. There has been a significant convergence of the life sciences and the computer sciences. Biotechnology firms have been applying for patents over databases of genetic information, and other proprietary informatics systems for storing and analysing genomic variation data. They have been seeking patents for computer software and computer hardware related to the life sciences. They have been applying for patents over novel business methods that utilise technologies for providing genomic services to the pharmaceutical and biotechnology industry.

Patent attorneys and lawyers have hailed the decision in *State Street Bank* as opening the way forward for the patenting of bioinformatic inventions.³⁹ Ernest Buff is perhaps representative in his enthusiasm: "*State Street* and its progeny will likely change the way in which biotechnology and bioinformatics industries do business".⁴⁰ However, as Stephen Lesavich comments, such patents were well available before the *State Street Bank*.⁴¹ Most bioinformatic inventions - such as those related to software methods, software systems, data structures, the Internet and other software features - were capable of receiving patent protection with software patents under US patent law long before the *State Street Bank* and *AT & T* cases were decided.

A couple of studies have considered whether the advice of patent attorneys and lawyers has been acted on by biotechnology and information technology firms.

Silico Research

A recent study conducted by London-based consulting firm Silico Research found that only 50 software-related patents had been issued by the US Patent and Trade Mark Office between 1996 and 2001 to companies operating in the pharmaceutical,

³⁹ Sung, L. and Pelto, D. "Bioinformatics May Get Boost From 'State Street': Software That Can Manipulate Vast Libraries Of Genetic Data May Receive Patent Protection", *The National Law Journal*, Vol. 21 (8), 19 October 1998.

⁴⁰ Buff, E. et al "State Street Alters Landscape Of Biotechnology and Bioinformatics", *New Jersey Law Journal*, Vol. 157 (4), 26 July 1999.

⁴¹ Lesavich, S. "Bioinformatic Tools", *The National Law Journal*, 16 October 2000.

biotechnology, and genomics research.⁴² The study covered 317 companies. The firm searched for United States patents assigned under the international classification G06, which covers computers, databases, networks, and computing methods. The study excluded biological, pharmacological and chemical patents. It did not take account of pending patents. It also excluded patents issued to software companies like Microsoft, Oracle and IBM.

The senior partner of Silico Research Emmett Power found unexpected results regarding the number of patents over bioinformatics products:

We were genuinely surprised by the lack of technology and method patents issued to pharmaceutical and biotechnology companies. We had expected large pharmaceutical companies to be registering significant numbers of technology and method patents as a matter of course.

After all, they employ teams of computer scientists in original research roles and they are highly patent-focused as part of their business and value creation methodology. Add the fact that senior executives throughout the industry are preaching the importance of the convergence of life and information technologies and we had expected to see 'convergence patents' staked out across the landscape.

However, it appears that large pharmaceutical companies' intellectual property efforts are still exclusively focused on compounds and genes as research and development end-points. In the process, they are ignoring the value of any original methods and computer technology developed to get to those end-points.⁴³

The study concluded that the leading patent issuers in the sector were Incyte and Affymetrix with six patents each. They were followed by PE corporation with five patents, and Tripos, 3-Dimensional Pharmaceuticals, and Entelos each had three patents. A number of companies had registered one patent.

Nature Biotechnology Study

Another study published by *Nature Biotechnology* provides a sharper image of the changing marketplace of bioinformatics.⁴⁴ Paolo Saviotti and his collaborators sought to assess the commercial activity in bioinformatics by searching the Derwent

⁴² Power, E. "Pharmaceutical Companies Are 'Failing To Patent New Technologies', Silico Research, London, 2 May 2001.

⁴³ *Ibid.*

⁴⁴ Saviotti, P., de Looze M-A, Michelland, S., and Catherine, D. "The Changing Marketplace Of Bioinformatics", *Nature Biotechnology*, 2000, Vol. 18 (12), p 1247.

Biotechnology Abstracts (DBA) for patents containing the words "computer, computing, DNA chip, biochip, gene chip, bioinformatics or informatics".

Saviotti and his collaborators found that the number of bioinformatics related patents has been increasing steadily from 1979 to 1997 after which there was a notable boom in patent applications, with a peak of sixty patents in 1998.

The first group of patentees were companies and institutions interested in applying information technology to production processes. The patents filed were much older (1983-1992). The applicants were commonly Japanese or Russian.

The second group of patentees were pharmaceutical companies, bioinformatics startups and public research institutes interested in applying information technology to research and development processes. Examples include Affymetrix, Affymax and Human Genome Sciences. The date of patents issued to these companies range between 1994 and 2000.

The third group of patentees were companies and instrumentation firms interested in creating tools and solutions for research and development processes. The companies include firms such as Motorola and Kodak which have been involved in mainstream electronics, IT and telecommunications. The patents issued to these companies range from 1992 to 2000.

United States Patent and Trade Mark Office

Such developments will have implications for patent applications and patent examination. The United States Patent and Trade Mark Office reports that the actual number of pending bioinformatics patents is relatively small.⁴⁵ Anticipating a rush of patent applications in this emerging area, the organisation had the foresight set up a dedicated bioinformatics art unit in December 1999, which operates under the biotechnology centre. Jasmine Chambers, the director of biotechnology at the United States Patent and Trade Mark Office remarked: "We were a little surprised that we haven't seen the flood of applications that some people have predicted".⁴⁶ Currently the 11 examiners in the bioinformatics art unit are processing a total of around 200 patents. Of those, 160 are in various stages of prosecution and 40 are waiting to be examined. Only 11 have issued from the unit so far. With a turnaround

⁴⁵ Toner, B. "Bioinformatics Patents Remain A Rarity In IP-Heavy Biopharmaceutical Industry", *Genomeweb.com*, 4 July 2001.

⁴⁶ *Ibid.*

goal of 36 months, Chambers expects a few more bioinformatics patents to issue over the next year.

Conclusion

There are a number of possible explanations for the rarity of bioinformatics patents at present.⁴⁷ First, the bioinformatics industry is relatively new, so emerging companies may have filed patents that have yet to issue. Second, the bioinformatics companies might be services driven rather than technology driven. Rather than create novel technology or data of their own, some companies simplify the data-gathering process for their clients by combining publicly available data sources and algorithms into one product. Companies such as Entigen and Doubletwise rely upon trade marks to protect their brands. Others depend upon the protection of Internet Domain Names.⁴⁸ Third, the bioinformatics companies might be wary of patenting their technologies because of the inherent complexity of software patenting. This has led to a number of vendors to keep their intellectual property as a trade secret. However, the entry of information technology companies into the marketplace may result in a greater activity in patenting. Thus the precedents in relation to Internet business methods patents have greater relevance to biotechnological inventions than has been previously been thought. They will have an important bearing upon whether bioinformatics - such as databases, computer software, and websites - can be patented.

PART 3

ENSEMBL AND DSAS: OPEN SOURCE INITIATIVES IN BIOINFORMATICS

The free software foundation⁴⁹ and the open source movement⁵⁰ have been a source of inspiration to public researchers involved in the human genome project. Many

⁴⁷ *Ibid.*

⁴⁸ There have already been disputes over the ownership of Internet Domain Names for genetic databases. Kleiner, K. "Domain Dispute: Hospital Sues Former Employee For Control of Gene Database", *New Scientist*, Vol. 172 (2318), 24 November 2001, p. 10.

⁴⁹ Moody, G. *Rebel Code: Linus Torvalds, Open Source, and the War for the Soul of Software*. London: Penguin Books, 2001, p 26-27.

⁵⁰ Raymond, E. *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. Cambridge, Mass.: O'Reilly and Associates, 1999; Wayner, P. *Free For*

researchers have been keen to ensure that scientific information and biological software remains in the public domain through the use of creative contracts. They have sought to live up to the impression that the scientific community is completely open and a place where ideas are shared freely. There are several main clusters of groups within the bioinformatics open source software community

The first group is the Bioinformatics.org and the Open Lab. This group offers web hosting and project support for a large set of projects relating to bioinformatics. The projects within the Open Lab are primarily end-user software tools for scientists looking to solve particular biological and bioinformatics problems. Bioinformatics.org are concerned that bioinformatics software has been extremely restrictive, with licenses reaching millions of dollars per institution. Third, the organisation campaigns against any notion of ownership of biological information and will work to develop a public or open licensing plan for information that has already been patented:

When genomics companies patent natural (not engineered) products, such as human genes, they act like prospectors or gold-diggers. They claim ownership of that which they haven't invented or produced: biological information. And they cannot purchase it from anyone who has. This was made very clear as companies filed patents on tens of thousands of human genes during the Human Genome Sequencing Initiative. Some companies went as far as to write software that automatically printed a patent application for each gene found.⁵¹

Bioinformatics.org is interested in the use of licensing and compulsory licensing to gain access to inventions, which has already been patented.

The second group is the Open Bioinformatics Foundation. Its purpose is to act as an umbrella organisation for a handful of projects called the bio projects. This group creates development libraries and tools for programmers in a variety of languages for bioinformatics generally, but mainly to facilitate sequence management and analysis. These projects grew out of the original BioPerl project. The goal of the foundation is to provide financial, administrative, and technical assistance for open source life science projects. Sun Microsystems awarded a hardware grant in support of the Open Bioinformatics Foundation.

All: How Linux and the Free Software Movement Undercut the High-Tech Titans. New York: Harper Business, 2000; and Moody, G. *Rebel Code: Linus Torvalds, Open Source, and the War for the Soul of Software.* London: Penguin Books, 2001.

⁵¹ <http://www.bioinformatics.org>

The third group is the Public Library of Science. This organisation is concerned that access to scientific information has been restricted to those who hold expensive subscriptions. It has circulated a letter proposing a boycott, beginning in September 2001, of journals that do not provide "unrestricted free distribution rights to any and all original research reports ... within six months of their initial publication date".⁵² They believe that the record of scientific research and ideas should neither be owned nor controlled by publishers, but should belong to the public, and should be freely available through an international online public library. The Public Library of Science has seized upon the strategies of open source software:

We have had extensive discussions with scientists, publishers and copyright experts about how authors who want to make their work freely accessible and useable can accomplish this while ensuring that they receive proper credit for their work. We have concluded that the best way to do this is for the authors and/or publishers to retain copyright on the work, but to irrevocably license the work to the public domain subject to the condition that proper attribution be given whenever the work is reproduced or redistributed. This practice is analogous to the way in which open source software is produced. By retaining copyright, authors and/or their representatives retain the right to enforce the terms of the license, but not the right to dictate how or by whom the work is used.⁵³

Establishment of this public library would vastly increase the accessibility and utility of the scientific literature, enhance scientific productivity, and catalyse integration of the disparate communities of knowledge and ideas in biomedical sciences.

The application of open source software and peer to peer technology to the life sciences can be illustrated in a number of case studies - Ensembl, DAS, and Blueprint.

Ensembl

The Ensembl project consists of computer programs for genome analysis and the public database of human DNA sequences. Ensembl is a joint project which is being run by the Sanger Center, the U.K. partner in the publicly funded Human Genome Project (HGP) consortium, and the European Bioinformatics Institute (EBI). It is funded by the medical welfare charity, the Wellcome Trust.

⁵² <http://www.publiclibraryofscience.org>

⁵³ *Ibid.*

At the CODE conference at Cambridge University, Tim Hubbard from the Sanger spoke about Ensembl and the Human Genome Project.⁵⁴ He stressed the problems in distributing and integrating data about the human genome - namely, that while biological sequence data has been doubling every six months, computer speeds are doubling every eighteen months. Tim Hubbard believes that Ensembl is a means of making the project of genomics 'democratic' so that people can contribute information and share it effectively:

Something like the human genome is too complicated for any person, any group, any company, to have a monopoly on knowledge. On the other hand, if every organisation puts up on a web-site what they think is on the human genome, you have a terrible mess in terms of people trying to understand, comparing one website to another website. So the approach of this open software project is to be as open as possible. Very standard things - open CDS repository; open database; open discussion; everyone can get the software; everyone can do similar bits of work with similar interfaces. That does not address the overlapping of the annotation.⁵⁵

The Ensembl Project is working on client server development. It hopes to ensure that users can participate in the annotation of the human genome, in a democratic and constructive fashion.

The Ensembl project seeks to overcome problems of inter-operability in the field of bioinformatics. They are larger, industry-focused organisations forming within the community who seek to shape the direction of many of the standards for interoperability. Brown and others comment that there are many obstacles to interoperability, not least the historical development of the sector:

Bioinformatics systems were originally developed by relatively isolated research groups in response to local information handling problems. Established research groups have consequently exhibited a reluctance to part with their locally developed systems and their preferred vocabularies and terms of reference for compounds and genes. The degree of flexibility and openness to change by such actors is likely to be limited because of the financial cost of reorganising nomenclatures and data handling systems. If cross matching between data bases is to be as automated as is hoped, these difficulties have to be overcome.⁵⁶

⁵⁴ Hubbard, T. "Ensembl And The Human Genome Project", CODE Conference, Friday 6 April 2001.

⁵⁵ *Ibid.*

⁵⁶ Brown, Nik, Nelis, Annemiek, Rappert, Brian and Webster, Andrew. "Bioinformatics: A Technology Assessment Of Recent Developments In Bioinformatics And Related Areas Of research

Brown and others seize upon a remark by a public database provider that the private sector secured its market by using incompatibility (interoperability) as a main asset.⁵⁷ To the question whether interoperability is a technical problem or not, this respondent answered: "Linux, Emboss and other free and open projects clearly demonstrate that the problem is NOT technical, nor is it expenses. But just merely unwillingness of private companies to allow free competition".⁵⁸ Others claimed commercial soft-ware changed its data format so that interoperability would be reduced. For public databases this means buying new software might take up a rather large part of their budgets.

DAS

The Distributed Sequence Annotation System (DAS), developed at the Cold Spring Laboratory, is tightly integrated with Ensembl. They are both projects intended to bring the human genome into the public domain. DAS is a data transfer system, which runs on top of XML. Hence it can be used either in a peer-to-peer or client-server architecture. DAS is designed to solve the problems of data integration and third-party annotation. The designers Lincoln Stein, Sean Eddy and Robin Dowell explain the rationale of the system:

The pace of human genomic sequencing has outstripped the ability of sequencing centers to annotate and understand the sequence prior to submitting it to the archival databases. Multiple third-party groups have stepped into the breach and are currently annotating the human sequence with a combination of computational and experimental methods. Their analytic tools, data models, and visualization methods are diverse, and it is self-evident that this diversity enhances, rather than diminishes, the value of their work.⁵⁹

However, the main risk of third-party annotation is that it may fracture and fragment knowledge about the human genome. The solution that Ensembl proposes is to allow sequence annotation to be decentralised among multiple third-party annotators and

And Development Including High-throughput Screening and Combinational Chemistry”, Final Report for the Science and Technological Options Assessment Unit, European Parliament, 1999, p 53-54.

⁵⁷ *Id.*, p 31-32.

⁵⁸ *Id.*, p 32.

⁵⁹ <http://stein.cshl.org/das/>

integrated by client-side software. Multiple sites would act as third-party "annotation servers".

Lincoln Stein, an associate professor of bioinformatics at the Cold Spring Harbor Lab in New York, believed that peer to peer technology had the potential to allow scientists to share their discoveries of the genome. He was inspired by the controversial program Napster,⁶⁰ which allowed users to link directly to each other's computers to share MP3 files using a centralised server system:

I was very interested when I saw Napster. It has a similar architecture (to what we use now), but it allows for 'peer—to—peer' data exchange and it dawned on me that it would be marvelous for our annotation system.⁶¹

Such claims lead to credence to the case that Napster was protected by fair use, because it was capable of substantial, non-infringing uses, such as the distribution of scientific information and knowledge.⁶² Perhaps it was premature for the legal system to regulate peer to peer technology such as Napster at such a nascent stage.

Stein believed that Gnutella, a spinoff of Napster, would be an even better match for exchanging genomics research. That software was designed to create self-perpetuating networks that grow independent of the company's server. Users could connect to other "servant" computers, creating a chain of participating users – an architecture that would allow for one-to-one or many-to-many connections. Stein observed:

Gnutella works out over multiple servers which replicate the information -- that's a very exciting technology because what we have to deal with in the HGP is data that keeps growing. We now have about 5 terabytes of information and we're only done with two-thirds of the sequencing. It will grow an order of magnitude more. The larger genome centers have the wherewithal to put up servers to publish their data electronically. Smaller, independent biology labs are clients of that and they download the information.⁶³

⁶⁰ Philipkoski, K. "Napster: Not Just For The Music", *Wired News*, 29 July 2000.

⁶¹ Philipkoski, K. "Gene Research, Meet Napster", *Wired News*, 5 April 2000.

⁶² Association of American Physicians and Surgeons Inc. "National Physician Group Supports Napster Appeal", 29 August 2000.

⁶³ Philipkoski, K. "Gene Research, Meet Napster", *Wired News*, 5 April 2000.

Stein hopes the technology will be used by more than just human genome researchers. “My hope is that it would be used by all biologists. I would hope that even bright high school students would be able to contribute”.⁶⁴

Of course there is a possibility that such a peer to peer system would violate copyright law. Companies such as Celera and Incyte charge from thousands to millions of dollars for their databases containing genomics information. Whether a Napster-like technology could infringe on revenue from the sale of such database subscriptions remains to be seen. Cyrus Harmon, president of a genomics firm called Neomorphic, comments:

It comes down to how the companies with genomic information make money. For companies like Incyte, or partly [for] Celera I think the answer is yes. I think this is somewhat inevitable - the information is going to be out there.⁶⁵

However, there has been resistance from some quarters of the scientific community to the open source and peer to peer distribution model. The DAS model is not universally beloved. NCBI director David Lipman is concerned that the human annotations may be full of rubbish because they will not be peer-reviewed.⁶⁶ He is concerned that one of the norms of scientific publishing will be undermined by the open source model. Lincoln Stein acknowledges the possibility but hopes that good annotation will drive out bad. He fears that the spirit of volunteerism will flag when faced with personnel changes and the vagaries of funding.

Are comparisons between Ensembl and Linux apt? Are parallels between DAS and Napster appropriate? Thinking of public and commercial annotation products as rivals misses the point, observers say. In the words of Sean Eddy of Washington University, who is working on DAS: “The human genome is too big for anybody to look at alone. We’re going to have to figure out ways for the public and private sectors to work collaboratively rather than competitively.”⁶⁷

Blueprint

⁶⁴ *Ibid.*

⁶⁵ *Ibid.*

⁶⁶ Powledge, T. "Beyond the First Draft: Making the genome data useful may depend on the public project Ensembl", *Scientific American*, 2000.

⁶⁷ *Ibid.*

The open source software movement and peer to peer technologies have been co-opted by commercial interests in the field of information technology. It is conceivable that the same pattern will occur in relation to bioinformatics.

Recent developments in the life sciences suggest that commercial entities are being influenced by some of these strategies. Notably, IBM and MDS Proteomics has set up a new organisation called Blueprint Worldwide, and the Biomolecular Interaction Network Database (BIND). The database builds upon other public databases such as NCBI's GenBank, generating a database of all bioinformatics and biomedical data. The director designate Francis Ouellette comments:

The study of biomolecular data is essential to developing better medicines to treat diseases. The management of this data is of critical importance. Blueprint will play a key role in supporting researchers worldwide by providing a centralized collection for growing volumes of scientific data on proteins, RNA and DNA interaction data. While GenBank provides researchers with the crucial list of parts (with some annotations), blueprint provides the assembly diagram - a 'blueprint' of how all the parts fit together. Using Blueprint's BIND database, researchers will have easier access to critical biomolecular data in one place, fostering accelerated discoveries worldwide.⁶⁸

Blueprint promises to be a non-profit organisation. It will generate revenue through training and educational programs, advertising, marketing, and consulting. Blueprint will never charge for data access or for the use of the software used to manage and mine this database.

Blueprint is supported by a consortium of leading international health organisations - including the National Center for Biotechnology Information, the European Bioinformatics Institute, the Samuel Lunenfeld Research Institute, the Canadian Genetic Diseases Network, the National Research Council, and some of the Canadian Institutes of Health Research. The organisation is a combination of the scientific community, non-government sponsors, and commercial sponsors.

This flies in the face of competitors - such as Celera Genomics and Myriad Genetics - that are relying upon hospitals, academic centers and companies to pay for genetic data.⁶⁹ IBM is looking to test a more co-operative model among scientists. Caroline Kovac said: "We're really at the beginning of a lot of the discovery that's

⁶⁸ IBM. "Science and Technology Leaders Launch Blueprint Worldwide, Giving Researchers Unprecedented Access To Vital Biomolecular Interaction Data", Vancouver, Ottawa, 30 May 2001.

⁶⁹ Couzin, J. "IBM's Biotech Ambitions", *The Industry Standard*, 11 June 2001.

going to happen in biology, and we believe very much that's going to require new computer tools. We can't do that by ourselves".⁷⁰

CONCLUSION

There needs to be a consideration of the future development of intellectual property and bioinformatics. It raises important questions about access to information, integration, interoperability, and the risks of monopoly. Recently, a number of mainstream information technology companies - such as IBM, Microsoft, and Compaq - have invested in bioinformatics. Commentators wonder what effect the entrance of these new players might have on the market for bioinformatics:

The movement of these IT-based entrants into the market is important because they are very large and powerful firms capable of shaking up the industrial structure of bioinformatics. It is tempting to speculate that the expertise of these companies in other industries might be rapidly translated to software solutions that provide the kind of standardization, integration, and analysis of the data so sorely needed.⁷¹

Furthermore, the information technology firms will have an important impact upon the field of bioinformatics. Such companies have shown great talent in fully exploiting both copyright law and patent law in managing the protection of computer software and hardware.⁷² They may be able to translate such tactics and strategies in the management of intellectual property to the field of bioinformatics. Similarly, public researchers and even private organisations have much to learn from open source software and peer to peer technology. Such strategies may provide the means to resist the privatization of genetic information. Furthermore, there should be a consideration of the implications of intellectual property and bioinformatics for related fields of information science.⁷³

⁷⁰ *Ibid.*

⁷¹ Saviotti, P., de Looze M-A, Michelland, S., and Catherine, D. "The Changing Marketplace Of Bioinformatics", *Nature Biotechnology*, 2000, Vol. 18 (12), p 1247.

⁷² Likhovski, M., Spence, M. and Molineaux, M. "The First Mover Monopoly: A Study On Patenting Business Methods in Europe", *Oxford Electronic Journal Of Intellectual Property Rights*, November 2000.

⁷³ Reichmann, J.H. and Uhler, P. "Promoting Public Good Uses of Scientific Data: A Contractually Reconstructed Commons for Science and Innovation", Conference on the Public Domain, Duke University School of Law, 9-11 November 2001, <http://www.law.duke.edu/pd>