Analysis of the Bioinformatics Industry

by

Carlos Benaim Jalfon

B.S. Computer Science, Pedro Emilio Col, 1987

Bachelor of Business Administration, Newport University, 1998

Submitted to the Sloan School of Management
in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Management of Technology
at the
Massachusetts Institute of Technology
June 2001

Signature of Author____
MIT Sloan School of Management
May 11, 2001

Certified by____
Edward Roberts, David Sarnoff Professor of Management of Technology
MIT Sloan School of Management
Thesis Supervisor

Accepted by____
David A. Webber, Executive Director
MOT Program

Analysis of the Bioinformatics Industry

by

Carlos Benaim Jalfon

Submitted to the Sloan School of Management on May 11, 2001
in Partial Fulfillment of the Requirements for the Degree of Master
of Science in Management of Technology

Abstract

The rise of the commercial genomic industry and the broadening application of genomic techniques in biology and medicine together with the growing availability of DNA sequence information have created a new industry: The Bioinformatics Industry. This thesis analyzes technologies, applications market and competitors in this industry and explores potential changes to the business models that are being used today. The technology and market information indicates that this is an industry in a very early stage. On the other hand, the business models being used are very similar to the ones used traditionally in the hardware and software industry: licensing, ASP (Application Service Provider), joint developments and hardware/software solutions. The actual market size is relatively small, estimated in no more than $300M.

Only by implementing strategies of horizontal or vertical integration, a company in this industry might be able to boost revenues in the long term.

Thesis Supervisor: Edward Roberts
David Sarnoff Professor of Management of Technology

Acknowledgements

Two years ago, I decided I wanted to continue my studies through a master program in a high-rank university. My search through the Internet took me to the Management of Technology Program at M.I.T.

I want to thank my parents, Leon and Rebeca, as well as my brothers Yaco, David and Eva for their constant and total support during this adventure and for the life lessons that you cannot learn in any university of the world. I also want to thank my uncle Alberto Jalfon, without his support I wouldn't have been able to attend to this program. And to my aunt Estrella Jalfon, who encouraged me to continue this path in moments when it looked difficult to follow.

Finally I want to thank Professor Ed Roberts, for his guidance and support in developing this thesis. I really admire him for finding a unique way to positively influence people surrounding him and society.

Shortly after applying to the M.O.T program for 2001, I was organizing some of my old papers and an old yellow letter with the M.I.T. logo appeared. It was the response I got from M.I.T. back in 1989 when I first thought in continuing my studies. To my surprise, the program I was asking information about, was also the Management of Technology program.

For the mix of feelings and thoughts that I have today, I can only think in one famous French sentence that represents part of it: La vie est belle!

TABLE OF CONTENTS

1. Introduction

"Using current procedures, a single research group might be able to sequence a human chromosome in about 1,000 years for $600 million". Michael Palazzolo, a geneticist with LBL's Human Genome Center, made this statement back in 1992[1].

The Human Genome Project is a national effort started in 1990 by the U.S. Department of Energy (DOE) and the National Institutes of Health (NIH) with the purpose of deciphering the human genetic information[2]. The priority given to this project, as well as to other similar projects around the globe, pushed the development of the related technology in such a way that today, February 2001, a draft of 90% of the whole human genome has been published[3].

As more data of DNA sequences were obtained, not only from the human genome but also from other organisms, new needs raised for managing and analyzing these data.

SEQUENCED ORGANISMS (Mb: Million of base pairs)			
Organism	**Genome size**	**Completion date**	**Estimated no. of genes**
H. influenzae	1.8 Mb	1995	1,740
S. cerevisiae	12.1 Mb	1996	6,034
C. elegans	97 Mb	1998	19,099
A. thaliana	100 Mb	2000	25,000
D. melanogaster	180 Mb	2000	13,061
M. musculus	3000 Mb	-	Unknown
H. sapiens	3000 Mb	-	35,000-45,000

Source: Elizabeth Penisi. The Human Genome. Science Magazine, vol. 291, no. 5507, 16 Feb 2001, 1177-1180.

In order to satisfy these needs, a new industry was born: the Bioinformatics Industry.

---

[1] Lyn Larrys. Machines and the Human Genome Project. Spring 1992.

[2] The Human Project website. http://www.ornl.gov/hgmis/project/timeline.html

[3] Elizabeth Pennisi. The Human Genome. Science Magazine. Volume 291, Number 5507, February 16, 2001.

The purpose of this thesis is to make an exploratory analysis of the bioinformatics industry. I will include some insights about how the technology is being used and how it is adding value to researchers. I will also include a business analysis, with complete profiles of the major players and a comparative analysis of them.

2. Defining Bioinformatics

In order to have an overview of the industry, it is important to select first a clear definition for the term bioinformatics. Literally, bioinformatics means the application of information technology to biology. However the concept has been narrowed by authors to the application of information technology to the analysis of DNA and protein structure.

Even under this framework, the definition is loose enough to allow misunderstandings in determining whether certain technologies are or are not bioinformatics.

In a more detailed approach, bioinformatics is defined as the set of computational tools and databases that support genomics research, which broadly encompasses the study of DNA/Protein structure/function, gene expression, and protein production/structure/function[4]. Three classifications are proposed by the National Center for Biotechnology Information:

- The development of new algorithms and statistics with which to assess relationships among members of large data sets
- The analysis and interpretation of various types of data including nucleotide and amino acid sequences, protein domains and protein structures
- The development and implementation of tools that enable efficient access and management of different types of information

From the previous definitions I derive that bioinformatics is concentrated in the algorithms and implementation of these algorithms required to manage genomic and proteomic

---

[4] Jason Reed. Trends in Commercial Bioinformatics. Oscar Gruss Biotechnology Review. March 13, 2000.

information. Some algorithms aim at finding similarities among sequences of nucleotides or amino acids, while others are oriented to optimize the database technology required to store, find and retrieve genomic and proteomic information. However all of them are considered to be under the bioinformatics framework.

New technologies like the DNA chip, a tool that enables comparisons of multiple genes simultaneously between two or more samples, have raised the need for other kind of algorithms for analysis of correlations between the activation of specific genes and pathologies or characteristics that might be caused by them[5].

Specific hardware could also be considered under the bioinformatics field. For instance Compugen, a company that provides bioinformatics tools, developed a parallel-processor hardware with proprietary software that allowed executing common bioinformatics tasks in a much faster way[6]. On the other hand, Timelogic offers servers with optimized software for bioinformatics analysis.

---

[5] Business Week Online. http://www.businessweek.com/bwdaily/dnflash/dec2000/nf2000128_877.htm December 8, 2000. By Amy Tsao.

[6] Compugen Homepage. www.compugen.co.il

3. Application of Bioinformatics in Drug Development

Although bioinformatics tools are being used in academic and research institutions as well as in biotechnology companies for different purposes, I will focus this section on the application of bioinformatics in drug development processes.

Pharmaceutical companies are the entities that can potentially generate more economic value to their businesses through the usage of bioinformatics methods (Jason Reed, 2000). A PricewaterhouseCoopers report, "Silicon Rally: The Race to e-R&D," cited recently by Business Week estimates that new computer-based technologies could save the pharmaceutical industry two to three years of effort and $200 million in development costs per drug[7]. Successfully integrating bioinformatics to drug discovery can improve target discovery and validation. This accelerates the drug development process by focusing on novel genomic targets[8] and therefore allows pharmaceutical companies to be more efficient in their allocation of research financial resources.

Pharmaceutical companies expend $200B per year in R&D. One of the major problems they face with this investment is the difficulty in detecting “bad drugs” in a very early stage of the project. In other words, if a new target is not going to work, the sooner they know it the better in order to avoid further investments in that particular drug. On the other hand if they reduce the proportion of bad targets vs. good targets, they get major efficiencies in their investments in R&D.

---

[7] Akja Agrawal. A Software Model That Fathoms the Human Heart. Business Week. November 21, 2000.

[8] Bellavance, L. L., Donlan. Bioinformatics Primer. Genetic Engineering News, 2000. 19:32-3

There are two basic approaches to drug development: the traditional screening approach and newer genomic-based approach.[9]. The traditional approach is organized around defined diseases and creates an animal model of a disease as the first step in the drug discovery process.[10] With this procedure, researchers can only elucidate the mechanism of drug action after it has demonstrated therapeutic value in the clinic. In other words target of the drug may be identified last. With the new genomic approach, drug discovery starts with the identification of potential molecular targets relevant to a particular disease. DNA chips for instance, are very powerful tools for gaining insight into the complexities of gene expression, detecting genetic variations, making new gene discoveries, fingerprinting and developing new diagnostic tools[11]. Using these tools, researchers are able to determine genes that have high probability of being activated in patients that have a particular disease, while having low or null probability of being activated in patients free of the disease. In other words they are able to predict which genes activation are causing a particular disease. Using this approach, the target of the drug is identified at the beginning of the process and the mechanism of action is evaluated very early in the discovery process.[12]

Bioinformatics plays a key role in this second approach. However pharmaceutical companies have been slow in adopting this technology. Instead of implementing a radical change in the way they develop new drugs, they have been using both approaches simultaneously (Gardner, 1999).

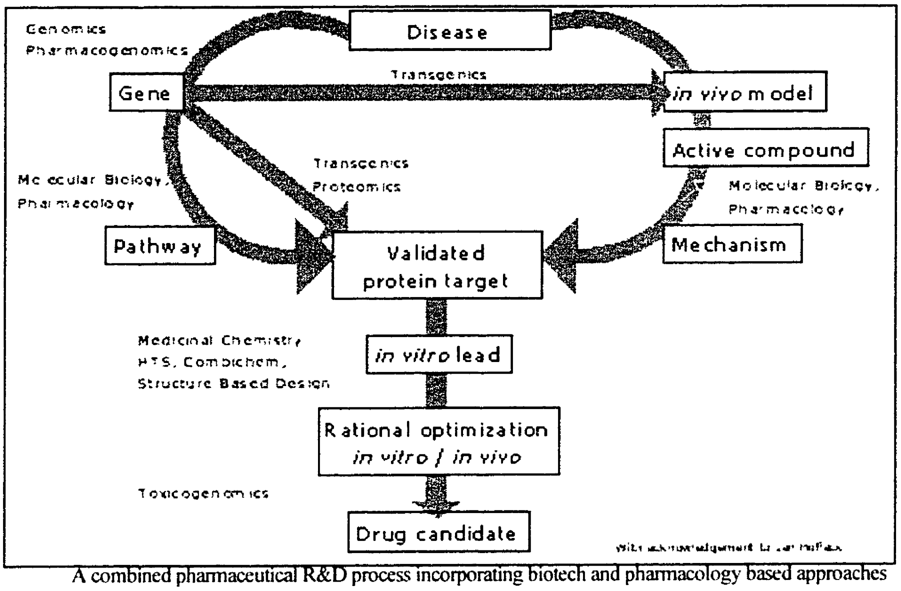
---

[9] Gardner, S. (1999). The Evolution of Bioinformatics. BITS Journal. http://www.bitsjournal.com

[10] Philip B.C. Jones. The Commercialization of Bioinformatics. EJB Electronic Journal of Biotechnology. Vol. 3 No. 2, August 15, 2000.

[11] Lisa Althoff. DNA Chip - Genetic Testing of the Future, 1999.

[12] Alice Park, Dan Cray, Cathy Booth and Dick Thompson. Brave New Pharmacy. Time Magazine, January 15, 2001. Pg.64-65.

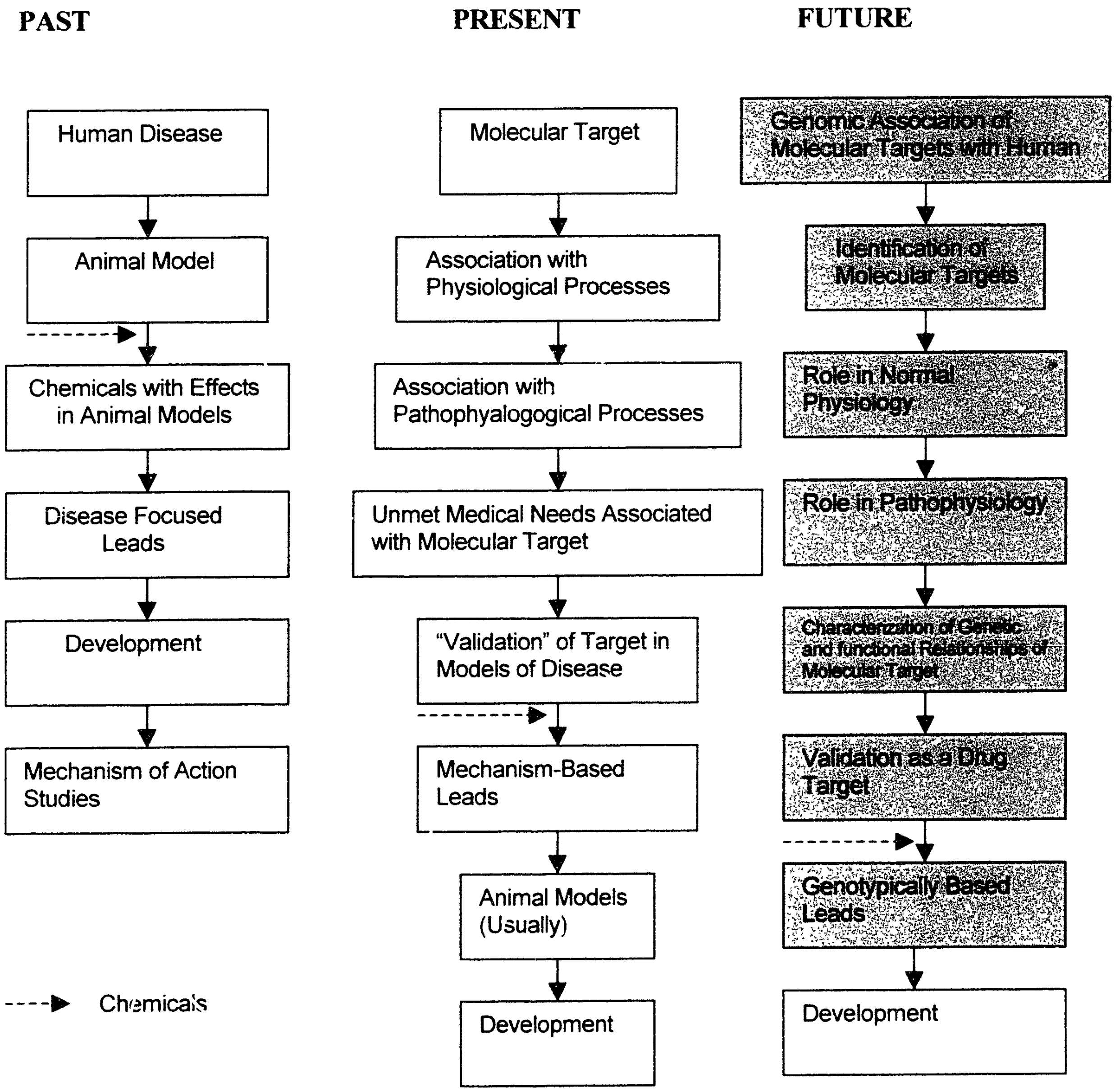


A combined pharmaceutical R&D process incorporating biotech and pharmacology based approaches

From a different perspective, below is a comparison of the drug development process in the past, present and estimated future according to Crooke[13] (Crooke, 1998).

---

[13] Stanley T. Crooke. 1998. Optimizing the Impact of Genomics on Drug Discovery and Development. Nature Biotechnology Magazine. Volume 16. Supplement. Pg 29-30.



Source Stanley T. Crooke 1998. Optimizing the Impact of Genomics on Drug Discovery and Development.

I shaded processes where bioinformatics is highly involved to illustrate the adoption of it in the future stage.

4. Activities Supported by Bioinformatics Tools

In the previous chapter I mentioned the general application of bioinformatics. In a more detailed analysis, the following activities are examples of the usage of bioinformatics tools (Jason, 2000):

1- SEQUENCING

Actual sequencing machines are able to decode nucleotides sequences from prepared samples through different sophisticated processes. However, the resulting sequence is broken into smaller peaces, usually 500 base pairs long. This opens the need for computational tools that apply algorithms to assemble all these pieces into the whole original sequence.

2- FIND GENES & FEATURES

New sequences are compared with previously known ones to determine where they express genes. Comparison with statistical information is required to determine whether the matches are evidence of the gene expression or if it is highly probable to find the same sequence even if the gene is not expressed. Bioinformatics tools assist in this process.

3- COMPARE

Once the researcher has identified a gene, he needs to find sequences among different genomes that might be similar to the gene he is analyzing. This activity allows him to identify other organisms that might have the same gene/function and there fore would work as a perfect target for in-vitro and in-vivo experiments. To do this they start with a specific section of the sequence and they search through databases to detect where this sequence is also present or at least a close approximation to it. Algorithms have been developed to determine how similar are two sequences. On the other hand the high volume of information requires database technology that enables this kind of operation in a more effective way.

4- DISPLAY

The amount of information collected requires a graphical interface in order to generate understanding. On the other hand, the possibility of viewing and managing 3D images of proteins enables researchers to detect patterns and make comparisons among different proteins. Bioinformatics tools assist this process.

5- INVESTIGATE

Once the researcher has a target, he might start analyzing how the gene expresses under different conditions. One way of doing this is through experiments with DNA chips that compare the gene expression among different controlled samples. Bioinformatics tools allow evaluating correlations between the expressed gene and the tested samples. Another way is to query databases of previous experiments.

In order to execute some of the activities mentioned above, the researcher has to perform a set of tasks. In an effort to classify these tasks and understand their frequency a study funded by AstraZeneca[14] surveyed 35 researchers. Among other results, they were able to obtain the

---

[14] Robert Stevens, Carole Globe, Patricia Baker, Andy Brass.A Classification of Tasks in Bioinformatics. Bioinformatics Magazine. February 2001. Pg.180-188.

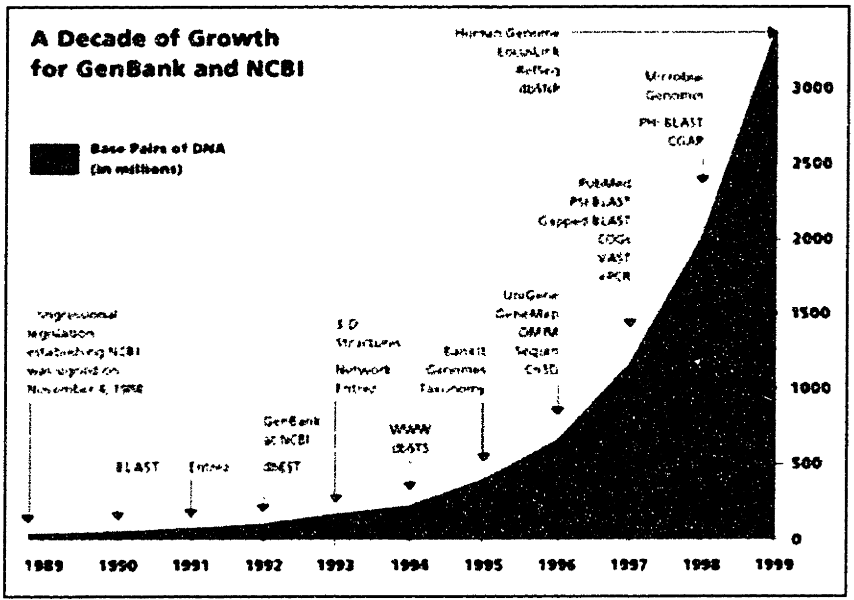
following table based on the most commonly executed tasks:

Question Class	Frequency
Sequence similarity searching	
Nucleic acid vs. nucleic acid	28
Protein vs. protein	39
Translated nucleic acid vs. protein	6
Unspecified sequence type	29
Search for non-coding DNA	9
Functional motif searching	35
Sequence retrieval	27
Multiple sequence alignment	21
Restriction mapping	19
Secondary and tertiary structure prediction	14
Other DNA analysis including translation	14
Primer design	12
ORF analysis	11
Literature searching	10
Phylogenetic analysis	9
Protein analysis	10
Sequence assembly	8
Location of expression	7
Miscellaneous	7

Four tasks accounted for 62% of reported tasks. And these tasks were similarity search, motif search, sequence retrieval and multiple sequence alignment. The major problem reported by the interviewed researchers was that the data were distributed heterogeneously and each of the tasks required searching through different data sources and formats.

5. Technology Evolution

The evolution of the bioinformatics technology has been tightly related to the evolution of the "reading" technology. I am calling "reading" technology the set of instruments and tools that allows determining the sequence structure of DNA or Proteins.



Source: The NCBI web page.

Although general opinion might be that this is a very recent technology, the first sequenced protein was bovine insulin in 1956 (Gardner 1999). Nearly a decade later, the first nucleic acid sequence was reported. One year later, Dayhoff created the first bioinformatic database gathering all the available sequence data. The protein DataBank was initiated in 1972 and the SwissProt protein sequence database began in 1987. These original databases were very rudimentary and usually flat files and simple indexes were adequate to access them.

These first steps in technology have a deep influence upon today's tools. Many of the sequences and protein structures are still accessed in flat file formats.

```
...
HELIX   32  32 PRO D  100  HIS D  116  5 17
HELIX   33  33 GLY D  119  GLU D  121  5 3
HELIX   34  34 PRO D  124  LEU D  141  1 18
LINK        FE   HEM A   1                 NE2 HIS A  87
LINK        FE   HEM A   1                 C   CYN A   2
LINK        FE   HEM A   3                 NE2 HIS A 229
LINK        FE   HEM A   3                 C   CYN A   4
LINK        FE   HEM B   1                 NE2 HIS B  92
LINK        FE   HEM B   1                 C   CYN B   2
LINK        FE   HEM D   1                 NE2 HIS D  92
LINK        FE   HEM D   1                 C   CYN D   2
CRYST1  102.540  115.170   56.700  90.00  90.00  90.00 P 21 21 21     8
ORIGX1      1.000000  0.000000  0.000000        0.00000
ORIGX2      0.000000  1.000000  0.000000        0.00000
ORIGX3      0.000000  0.000000  1.000000        0.00000
SCALE1      0.009752  0.000000  0.000000        0.00000
...
```

Extract from the PDB file corresponding to the structure of the human hemoglobin 1ABY.pdb. Downloaded from Swissprot database.

These early steps in building sequence and structure databases only offered access to the information based on exact matches either of names of proteins or short strings of nucleic sequences. Eventually, new tools were used to search these databases that allowed searches by approximation based on pattern matching or sequence alignment. One of these algorithms is known as BLAST and was introduced a decade ago. Five years ago new more sophisticated and slower algorithms were introduced. The most common of them are the ones known as FASTA and Smith-Waterman.

Even commercial tools offered today by some of the bioinformatics companies implement these algorithms in their tools.

As an example, if a researcher wanted to search a nucleic sequence database for occurrence of the sequence GCGCATGGATTGAGCGA, only an exact match of this sequence would be obtained in the original approach.

Today, an implementation of the Smith-Waterman algorithm is able to find matches by sequence alignment. So the above example could find

GCGCCATGGATGAG as a match even if they are not totally equal. The sequence alignment algorithms tries to find sequences that might have common ancestors or in other words that have been slightly changed with evolution but are essentially the same ones. The method would assume the following sequence alignment:

```
GCGC_ATGGATTGAGCGA
GCGCCATGGAT_GAG____
```

Where each underscore "_" indicates a possible insertion of nucleotides in the opposite compared sequence.

Hidden Markov Models (HMM) and Markov Chains are commonly used in bioinformatics tools to find genes among sequences and to try to simulate evolution.

Based on the mentioned basic bioinformatics operations and applications, in my opinion the real innovation in bioinformatics resides in the algorithm. The rest has to do with applying general information technology to solve specific problems in database access, modeling, interface and graphing.

According to Dr. Schnheider, CEO of Lion Bioscience Research, there is no expectation for major breakthroughs in algorithms that will enable operations that can't be done today. Innovative algorithms however might allow speeding up the processes that are already in place with actual algorithms (based on a meeting held with him on March 1, 2001).

However Dr. Craig Venter, president of Celera, argues that new algorithms might allow considering ever-greater amounts of biological interaction, to consider more genes, proteins, and cells in a calculation. Everyone has access to the BLAST, HMM, Smith-Waterman and many other algorithms, but they may not be scalable to consider the very large data sets that are now available with a whole human genome or a larger data set or the coming tidal wave of protein data. [15]

Parallel computing and innovation in hardware might also allow faster operations, so it seems that what we can expect in the future is an optimization in the way the technology is being use and innovation in the algorithms being implemented.

---

[15] Tom Headrik. Celera's Algorithmic Edge. COMMUNITY PERSPECTIVE. February 6, 2001

6. Dominant Designs

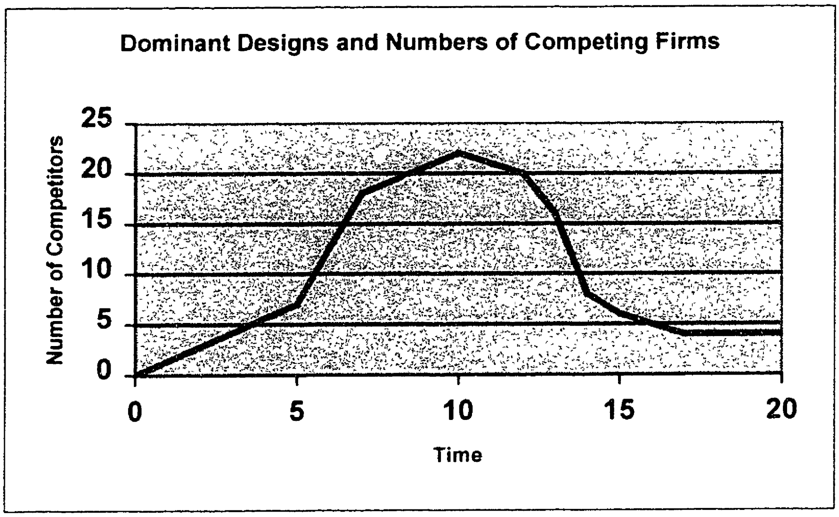
A dominant design is defined as a "new product or set of features synthesized from individual technological innovations introduced independently in prior product variants" that wins the allegiance of the market place.[16]

Several examples have been studied about the appearance of dominants designs in different markets. In the case of the typewriter industry, the use of QWERTY keyboards is considered a dominant design. This design was then incorporated in almost every typewriter no matter what company manufactured them.

Utterback relates the appearance of dominant designs to the industry structure and number of competitors. The analysis suggests that initially few competitors will start the industry developing the first products and new technologies. As the market is tested, dominant designs will appear by acceptance of product features. After this, innovators will try to figure out how to make that design as efficiently as possible. The number of firms will increase in the early stage, get to a peak when the dominant design appears and then decrease to only a few competitors after the industry is concentrated only in improving the dominant design. These few competitors will dominate the market with higher product technology and productivity. In the following figure, the dominant design appears in year 10, and effective competition takes place following that time on the basis of cost and scale as well as on the basis of product performance.

---

[16] Utterback, James M., Mastering the Dynamics of Innovation, (MA: Harvard Business School Press, 1994), pp.23-55.

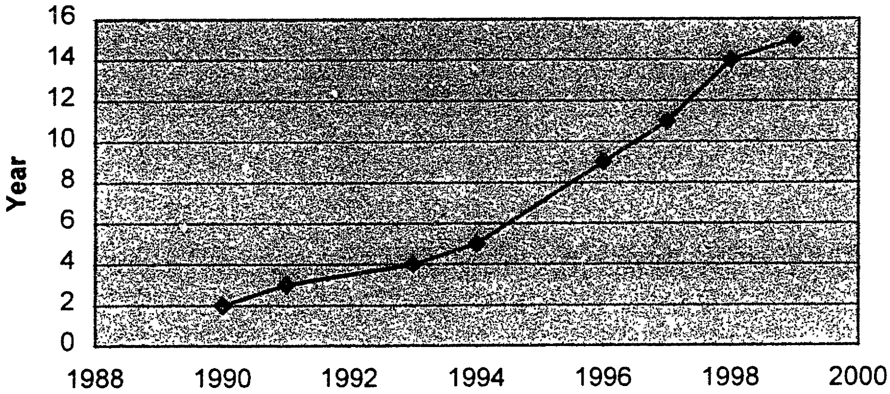


Source: Mastering the Dynamics of Innovation (Utterback, 1994)

This pattern was proved to be true over and over in Utterback's results. He confirmed this pattern with historic U.S. data in the typewriter industry, the automobile industry, the television industry, the transistor manufacturing industry, the electronic calculator industry, the integrated circuit industry, the Winchester disk drive industry and the massively parallel supercomputer industry.

In a selection of competitors, based on a report of Oscar Gruss (Jason 2000) and my own scanning of the market, I took incorporation dates of each company and determined the following number of concurrent bioinformatics companies in time.

Number of Bioinformatics Firms



The following firms were considered in the analysis:

Firm	Symbol	WebPages	Incorporation	# of Firms
Timelogic		www.timelogic.com	1981	1
Informax	INMX	www.informaxinc.com	1990	2
Incyte	INCY	www.incyte.com	1991	3
Compugen	CGEN	www.compugen.co.il	1993	4
GeneLogic	GLGC	www.genelogic.com	1994	5
Genomica	GNOM	www.genomica.com	1996	6
Netgenics		www.netgenics.com	1996	7
Spotfire		www.spotfire.com	1996	8
Structural Bioinformatics		www.strubix.com	1996	9
DoubleTwist		www.doubletwist.com	1997	10
Lion Bioscience	LEON	www.lionbioscience.com	1997	11
Entigen		www.ebioinformatics.com	1998	12
Silicon Genetics		www.sigenetics.com	1998	13
Celera	CRA	www.celera.com	1998	14
Molecular Mining		www.molecularmining.com	1999	15

In bioinformatics, some of the algorithms used today can be considered dominant designs. They are widely accepted and implemented. Smith-Waterman and Blast are examples of algorithms that are being incorporated by many of the products of the companies mentioned above, according to their website.

Is difficult to predict what is going to happen in the near future. One argument could be that the dominant designs already appeared with the commonly used algorithms (BLAST, Smith-Waterman, FASTA, etc.), and therefore the number of competitors from now on will shrink leaving only a few dominant firms in the market. However the increasing number of firms in the past and the fact that these technologies have not yet been well adopted by pharmaceutical companies, could impulse the bioinformatics industry to keep growing in technology developments, number of competitors and end-user adopters.

7. Competitors

In this section I will include some basic information about each of the competitors I analyzed. I used information published in their websites and in some cases the information reported by The Wall Street Journal online. In the section "Business Model" I utilize the term ASP (Application Service Provider) for those cases where the company is hosting applications or data and offering to customers for a fee or per contract.

7.1. Timelogic

Webpage: www.timelogic.com

Locations:

Incline Village, Nevada.

Background:

According to their website they were founded in 1981. They don't add any additional information about their trajectory. They have an agreement with Sun to incorporate their hardware accelerators into Sun Servers.

Product Offering:

Decypher: They offer a combination of hardware and software to solve bioinformatics problems. Their servers are customized to run specific algorithms common to bioinformatics operations and they use parallel computing to enhance the performance of these operations. They include implementations of widely used bioinformatics algorithms as BLAST, Smith-Waterman and Hidden Markov Models. Each server uses as much as four processors but placed in a network they can work in a collaborative way. Timelogic's software is enabled to distribute processing tasks among any connected server.

Business Models:

- Hardware: They charge for their hardware, which is delivered to customer's installations, and they also charge an annual fee for maintenance service.

Customers:

Abgenix, AGY Therapeutics, AlphaGene, Bristol-Myers Squibb, Chiron, CuraGen, DKFZ (Germany), European Bioinformatics Institute, Genome Therapeutics, Hoffman-La Roche,Japanese Patent Office, Los Alamos National Laboratory, National Center for Genome Resources, National Institute of Radiological Science (Japan), National Institute of Genetics (Japan), Novartis (Basel, Switzerland), Novartis Institute for Functional Genomics, Novartis Pharmaceuticals (Summit NJ), Rikei Corporation (Japan), Paradigm Genetics, Sanger Center, Schering-Plough, SUGEN, Stanford University, Synaptic Pharmaceutical Corporation (upgrade), Syngenta (Novartis Agricultural Discovery Institute), Teijin Systems Technology Ltd. (Japan), Washington University, Whitehead Institute MIT, University of Minnesota.

7.2. Informax

Webpage: www.informax.com

Locations:

North Bethesda, MD. Golden, CO. South San Francisco, CA. San Diego, CA. Cambridge, CA. Oxford, UK.

Background:

Dr. Alex Titomirov founded the company in 1990. According to their website they have a user base of more than 22,000 researchers in more than 1300 research organizations. Their products enable the analysis and interpretation of genomic, proteomic and related information.

Products:

- Vector NTI Suite: It allows data management, mapping and illustration, primer design and analysis, and strategic recombinant design. It offers multiple sequence alignment algorithms as well as sequence assembly options. It also allows analysis and plotting of protein sequences and structures. The product runs in a PC.
- Genomax: data mining platform designed to run on enterprise networks. It integrates multiple genomic data types and allows research teams to coordinate efforts and distribute computing resources. It implements public available bioinformatics algorithms and allows the incorporation of customized ones. They claim they are developing additional modules that will be incorporated to this platform in order to allow genomic viewing, protein-protein interaction and SNP analysis.

- Pedant Genome Database: it includes a collection of annotated and classified genomes. Through the database the researcher can explore and compare more than 60 genome databases.

They also distribute related products developed by other companies:

- Prochart: a product co-developed between AxCell and Informax that allows comparisons of multiple proteins through an integrated interface. It allows correlation of information about particular proteins with the related information on sequence, expression, tissue distribution and bibliographic information.
- BIORS: A product developed by BioMax, AG. This product allows integrating heterogeneous external databases with in-house proprietary databases. Through a single screen, the researcher can perform a unique search across a selection of databases of biological information.

Business Models:

- ASP Model: They offer access to some of their tools through their website for a fee.
- Distribution: They act as distributors of some of the products offered. They charge the end user and pay the manufacturer.
- Software Licensing: they offer their Vector product under software licensing agreements.

Customers:

Merck & Company, Genzyme Corporation, Proctor & Gamble, Johnson & Johnson, Bristol-Myers Squibb, Pfizer, AstraZeneca, Diversa Corporation, Novartis Agribusiness Biotechnology Research, The Whitehead Institute for Biomedical Research, M.I.T., University of Tokyo, National Institutes of Health.

7.3. Incyte

Webpage: www.incyte.com

Locations:

In the US: Palo Alto CA, St Louis Missouri, Fremont CA, Santa Clara CA. In the UK: Cambridge.

Background:

Incyte Genomics was incorporated in Delaware in April 1991in order to acquire technology and assets from Invitron, a company who had developments in the proteomics area. They sequence and identify genes of known sequences and their corresponding proteins. They compare partial human genes or protein sequences to genes or proteins of known sequence in order to predict their biological or therapeutic function. At the beginning they use a similar approach but only to identify specific white blood cell proteins. They became public in 1993 through an IPO on the NASDAQ stock exchange.

Product Offering:

- LifeSeq: provides access to a sequence database and a view of available human genes. They claim they have more than 120,000 gene transcripts where 60,000 of them are proprietary and are not available anywhere else. They also offer access to gene information on more than 3000 genes patented by them.

- Workbench: a product that allows scientists to keep track of their research creating links to specific LifeSeq pages. It also allows customers to save under private folders results of queries that might be consulted in the future and won't require to execute and wait for the database query.

- Reagents and Services: They offer their sequencing installations capabilities in order to execute customizing tasks related to sequencing but adjusted to specific customer requirements. They also offer reagents that are used in the sequencing and related processes.

Business Model:

- ASP: They offer access to their database through a website. The website requires a username and a password that is offered to users for a fee.
- Intranet: They offer a version of their product that can be installed in the customer's network in order to improve performance. They install a copy of the database as well as their tools in the customer's servers and users can access it from the intranet of the company.
- Professional Services: they offer professional services mainly through the customization of the sequencing process in order to satisfy specific customer needs.
- Reagents: they sell and deliver chemicals and articles required in genomic laboratory research.

Customers:

Not reported in their website.

7.4. Celera

Webpage: www.celera.com

Locations:

Rockville, MD - US.

Background:

The company was founded in May 1998 by the PE Corporation. The PE Corporation already owned PE Biosystems who was and still is the dominant manufacturer of DNA sequencing machines. Celera claims to have the major sequencing installation of the world. They sequenced and published the genome of the Drosophila on September 1999 and are currently sequencing the human and the mouse genome. They published a draft of more than 90% of the whole human genome on February 2001. Celera claims that their business model will be to offer their own-generated biological information and bioinformatic tools in a similar manner to the way information companies such as Lexis-Nexus and AOL do.

Product Offering:

- Celera Discovery System: this is their major service offering. It includes a set of bioinformatic tools that enables access to three major data sets: the human genome, the mouse genome, their SNP Reference Database and more than 30 integrated public data sources. The bioinformatics tools allow users to view and analyze gene structure and function, protein classifications and genetic variation.
- Celera Laboratory Services: Technical professional services offered to develop and implement custom laboratory programs. Among other services they include under this

category contract sequencing (usage of their sequencing capability per contract), protein classification, SNP discovery and gene expression profiling.

- Celera AgGen: This product enables genomic discovery and DNA testing for agriculture. This is a combination of professional services, sequence information and analysis tools.

Business Models:

- Technical Service: they offer professional services to implement the usage of genomic information in laboratories, and the usage of genomic information in drug discovery.
- ASP: They offer access to their tools and databases for a fee. Their tools also access more than 30 public databases and enables integration through their interface.

Customers:

Pfizer, Takeda Chemical Industries, American Home Products, Inmunex Corporation, Valigen, Vanderbilt University, Harvard University, University of Texas Southwestern Medical Center at Dallas, University of Cincinnati/Children's Hospital of Cincinnati, The Ohio State University, The Government of Australia, The Institute for Genomic Research, Weizmann Institute of Science, Hospital for Sick Children in Toronto, California Institute of Technology, Genset, Max Planck Society, Karolinska Institutet, Center of Excellence Program at the University of Tokyo.

7.5. Lion Bioscience

Webpage: www.lionbioscience.com

Locations:

Heidelberg, Germany, Cambridge UK, and Cambridge MA, USA.

Background:

LION was founded in March 1997 by six scientists from the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany and the University of Heidelberg, together with entrepreneur Friedrich von Bohlen. I had the opportunity to interview one of them, Dr. Schnheider, who is now CEO of Lion Bioscience Research. I add this interview at the end of this section.

At the beginning Lion Bioscience offered analysis services. They later evolved to provide bioinformatics tools to assist drug discovery and drug validation activities.

Their offer includes integrated solutions for genomics and Life Science informatics to the Life Science industry. Encompassing the fields of molecular biology and information technology LION creates an integrated platform for accelerating its customer drug discovery and LIFE Science efforts. LION bioscience AG (ticket: LEON), is publicly traded since August 2000.

Product Offering:

- SRS: This is a data integration platform for sequence retrieval, providing access to a wide range of different genomic and proteomic databases. They claim to integrate more than 400 different databases through their interface. The top screen asks for the string of characters that the user wants to find and offers a list of databases that the user can select from. Once executed, the system goes to each of the selected databases, searches for the string and brings to the interface a list of hits, all in the same format. A variation of this product is SRS Objects, an Application Program Interface (API) that enables customers to use SRS functions in other applications in an integrated way.

- BioScout: This tool includes some bioinformatics methods to discover genes and the function of these genes. Among other activities, this tool enables the researcher to make sequence analysis and comparison, protein family analysis and 3D Homology modeling. It also offers methods to analyze Microarray results and to add and query annotations made by the researcher.
- ArrayScout: This tool is exclusively oriented to execute statistical operations on gene expression data. It also allows the researcher to keep track of previous experiments and query results of gene comparisons made in the past.

Business Models:

- Drug Targets: They are delivering drug targets to Bayer, using their own tools.
- Software Licensing: The software is installed in the end user computer and can be used under a license agreement.
- ASP model: Users are offered access to the tools through Lion's website. Each user is required to have a login and a password and Lion charges a fee per user per year.

- Technical Service: They offer IT professional services generally related to their products. These services include installation and customization of Lion's products as well as technical training.

Customers:

Aventis, Bayer, Boehringe Ingelheim, Sumitomo Pharmaceuticals, GlaxoWellcome, Ganssen Pharmaceutica, Celera, Dupont, British Biotech, Daiichi, eBioinformatics, SmithKlime Beecham, Novaritis, Novo Nordisk, Pharmacia & Upjohn.

Summary of Interview with CEO of Lion Bioscience Research: Dr. Reinhard Schnheider

The following includes my interpretation of comments made to me by Dr. Schnheider in a meeting in his office that took place on March 1, 2001.

According to Dr. Schnheider, Lion Bioscience is trying to integrate a horizontal IT solution for Pharmaceutical companies that will enhance and support the processes involved in drug development from drug discovery to clinical trials and health validation.

Their original products helped pharmaceutical companies in the drug targeting process. However, to achieve their strategy of adding value in the whole drug development chain, they acquired key companies and signed some alliances.

For the toxicology part of the process they acquired Trega, a smaller company with developed technology in this area. For the health part, they acquired 16% of the German company GMD who has developed technology for this part of the process. Lion has a call option to acquire an additional 9% of the company in 2002.

According to Dr. Schnheider, they will probably try in the near future to acquire companies with developed IT products that support Clinical Trials processes as well. This way Lion Bioscience will be able to provide to their pharmaceutical customers the integrated solution for drug development.

About the deal with Bayer, Dr. Schnheider explained that the deal for $100 M includes the promise of delivering 500 targets in 5 years. The planned milestone indicated that they had to deliver 200 targets by the end of 2000 and they have exceeded this milestone by delivering almost 250 so far.

As part of the deal, Bayer has the option to spin-in the Lion Bioscience Research subsidiary at the end of the 5-year period.

Asked if this deal could threaten future deals with other pharmaceutical companies that compete with Bayer, Dr. Schnheider explained that the opposite effect has been observed. Companies are pleased to be able to have access to software that some how has been enriched with Bayer's experience.

Dr. Schnheider doesn't see major competitors in the integrated IT solutions for drug development. There are however some competitors in specific parts of the process such as Informax, Double Twist, Compugen, etc.

Regarding the technology involved in bioinformatics, Dr. Schnheider said that there is no major breakthrough innovation expected. The major benefit in the future will come from the ability of these tools to seamlessly incorporate fragmented information and data into an integrated interface.

7.6.Compugen

Webpage: www.compugen.co.il

Locations:

Tel Aviv – Israel, Jamesburg - NJ USA, Sunnyvale – CA USA, Tokyo – Japan.

Background:

Compugen was incorporated in 1993 in Israel by a group of mathematicians from the Talpiot program of the Israeli Defense Force. This program groups only very talented people with outstanding skills in mathematics and related fields. In 1994 Compugen sold their first product -the bioaccelerator- to Merck. This product was a hardware unit that accelerated up to 1000 times gene analysis and related activities. In 1997 they incorporated in the US and started offering a software platform named LEADS. In the year 2000 they went public through an IPO in NASDAQ.

Product offering:

- LEADS platform: An integrated set of bioinformatics tools that enables the analysis of genomic and proteomic data. These tools are oriented to drug target discovery and allows scientists to identify genes in genomic sequences. Compugen claims to have discovered 4000 new genes using their own set of tools.
- Gencarta: This product is a proprietary Compugen database with sophisticated annotations on genomic data. The product includes the database access, a graphical interface and a set of query tools. The whole product is installed in the customer location and the database is updated regularly.

- Geneguide: Through this product a researcher is able to generate complete reports of specified genes. The product combines Compugen's proprietary information with public available databases.

- DNA Chip & Analysis: This product assists the design of DNA chips. Motorola bought this system in order to design new DNA chips. Compugen claims a proved improvement in DNA chip's quality through the usage of their software.

- C-Gen Oligolibrary: This tool allows redundancy elimination by selecting only one or few oligos (small sets of nucleotides) per gene cluster. This process allows the representation of more genes per design.

- 2dGel Image Analysis: this is a set of proteomics tools that allows the image analysis of proteins as well as a set of comparisons and alignments among proteins.

- Gene Portfolio: Through the usage of their own technology, Compugen was able to discover as much as 4000 new genes that were later tested *in-vivo* and patented by them. These genes are potential drug targets that pharmaceutical companies could use to develop new drugs.

Business Model:

- ASP Model: They offer access to their LEADS product through the website www.labonweb.com.

- Software licensing: The product is installed in the customer's equipment and they pay a fee for the right to use it.

- Drug targets: this is more like selling Intellectual Property. They have discovered genes that might be used as drug targets by pharmaceutical companies.

- Technical Service: They offer technical service as well as participation in joint projects.

Customers:

Alphagene, Avalon, Bayer, Lilly, Genome Therapeutics Corporation, Merck, Monsanto, National Institute of Health, San Diego Supercomputer Center, Serono, US Patent and Trademark Office, The European Molecular Biology Laboratory (EMBL), Jansen, Pharmacia & Upjohn, Weizmann Institute of Science, Amgen, Aventis Pharmaceuticals, Chugai Biopharmaceuticals, Human Genome Sciences, Incyte Pharmaceuticals, Motorola, Parke-Davis, Scios, SmithKlein Beecham, Wyeth Ayerst, The Eurpean Bioinformatics Institute, The Sangre Centre.

7.7.Spotfire

Webpage: www.spotfire.com

Locations:

Cambridge, MA and Goteborg, Sweden.

Background:

The company was founded in 1996 and maintains its Eurpean headquarters and development center in Sweden. They offer analytical tools that enable access to database and application of statistical analysis to a wide range of industries including the pharmaceutical and biotechnology industry.

Product Offering:

- Spotfire DecisionSite: A general-purpose tool to support decision-making based on statistical information. They offer a graphical interface that allows users to access data from different structures and databases in a consolidated way. They claim the tool is effective to assist product discovery, development, manufacturing and marketing.

- Spotfire Structure Visualizer: this component connects the DecisionSite platform to chemical structure information and functionality. Among other tasks it allows correlation analysis between chemical structures and other compound properties.

- SpotFire Array Explorer: implements the DecisionSite product to the analysis of gene expression and related information generated by microarrays. It unifies data, analytical tools and reporting capabilities in one integrated workspace.

- Spotfire Software Laboratory: this product is a development environment that incorporates functionality offered by their DecisionSite product in a way that can be used to create applications.
- SpotFire Computational Services: they offer this service through their website. It includes a set of data mining and statistical tools that can be accessed by authorized users.

Business Models

- Software License: Most of their products are offered under a software license agreement. The software is installed in the customer's computers.
- ASP Model: They offer some of their tools through their website and charge annual fees to allow users to access them.
- Professional Service: They offer to develop customized applications using their own development tool to specific customer's needs.

Customers:

Abbott, Acadia, American Home Products, Amgen, Ariad Pharmaceuticals, Arqule, AstraZeneca, Aurora Bioscience, Aventis, Bayer, Vertex, Biogen, Boehringer Ingelheim, Bristol-Myers Squibb, Chiron, Chugai Biopharmaceuticals, Ciphergen, Cytovia, Dainippon, DuPont Pharmaceuticals, Eli Lilly, EOS Biotechnology, FDA, Gene Logic, Genetics Institute, Genzyme, Glaxo, Human Genome Sciences, Incyte Pharmaceuticals, Janssen Pharmaceuticals, Lexicon Genetics Incorporated, Lundbeck, Medivir, Merck, Millenium, Motitix, Novartis, Novo Nordisk, OSI Pharmaceuticals, Pfizer, Pharmacia & Upjohn, Roche, Schering-Plough, Selectide, Sepracor, Serono, SIBIA, SIDDCO, SmithKline Beecham, Vertex, Warner Lambert, Yamanouchi.

They also have customers in the chemical process industry and manufacturing.

7.8.Genelogic

Webpage: www.genelogic.com

Locations:

Gaithersburg, MD and Berkeley CA.

Background:

The company was founded in 1994. They started by offering customized solutions in the areas of gene expression and management of microarrays information. In 1999 they started offering databases with information from human and animal tissues., leveraging the knowledge they acquired through their previous experience in managing information generated by microarrays.

Product Offering:

- GeneExpress Database: this database includes a reference library of gene expression information across a broad survey of organ types, diseases, and diseases stages in both human tissues and animal models. The data was generated by Gene Logic and is offered for a fee.

- Custom Discovery Research Databases: this is a service Gene Logic offers to pharmaceutical companies to assist them in the construction and analysis of gene expression databases. These databases are customized to specific needs of the customer and the resulted data are owned by the customer and can't be used by GeneLogic in their other product (GeneExpress).

Business Model:

- ASP Model: They ask for a fee in order to allow users to access GeneLogic database through their website.
- Professional Service: They offer their experience to assist pharmaceutical companies in building their own customized databases of gene expression.

Customers:

They don't disclose the name of their customers, however they claim to have six customers of their GeneExpress database for the year 2000.

7.9. Genomica

Webpage: www.genomica.com

Locations:

Boulder, Colorado – US and Hertfordshire - UK.

Background:

The company was founded in 1996 to commercialize bioinformatics tools. These tools include clinical, epidemiology and biochemistry applications. They claim to have raised $36 million in equity financing. Among the companies that funded Genomica we found Falcon Technology Partners, Boulder Ventures and The Kaufmann Fund. They went public through an IPO in September 2000.

Product Offering:

- Genomica Discovery Manager: this software provides integration of information from different database sources. This information can be from clinical data, pharmacogenomics, SNP or sequence analysis.
- Genomica Reference Database: this database integrates mapping data from different public genome centers. The Whitehead Institute SNP maps and radiation hybrid maps as well as the SNP Consortium markers maps are some of the sources that are integrated through Genomica Reference Database.
- LinkMapper: this software complements Applied Biosystems Genemapper. It provides quality checking and genetic data management and analysis in an integrated

way. Together with GeneMapper provides connectivity from the instrument to the analysis tool

Business Model:

- Software Licensing: Each of the products mentioned above are offered under a license agreement.
- Professional service: they also offer installation, training, upgrade and related services for a fee.

Customers:

Parke-Davis, Oxagen, Glaxo Wellcome, AstraZeneca, Biognosis, University of Oxford, National Cancer Institute, National Institutes of Health, Genome Therapeutics, Aventis.

7.10. NetGenics

Webpage: www.netgenics.com

Locations:

Cleveland, OH. Columbus, OH. San Diego, CA. Boston, MA. London, UK.

Background:

Netgenics was founded in 1966. Since then they have developed integrated informatics solutions for pharmaceutical companies. Dr. Walter Gilbert, a Nobel Laureate and co-founder of Diogen, serves as chairman of the board. Their sources of funding include Oxford Bioscience Partners, Incyte Pharmaceuticals and Edgewater Private Equity, among others.

Product Offering:

- NetGenics Gene Expression DataMart: this product allows storing, managing and analyzing gene expression data derived from a variety of experiments with microarrays. It also allows users to include annotations to describe each set of results as well as the specific characteristics of the test (date, times, biological sources, etc.)
- Synergy Discovery Environment: this component provides an integrated interface to access the Gene Expression Datamart and Sequence Analysis.
- Synergy Sequence Analysis: the goal of this product is to provide an analysis platform for DNA and Protein sequence data. The product includes implementations of bioinformatics algorithms and it enables access to public and proprietary sequence databases.

- Synergy Plug-in Service: a set of modules that can be incorporated to the rest of their products in order to extend functionality. It uses XML as a data interchange mechanism.

- NetGenics ChemSymphony: uses a set of components for the creation of Web-based applications. It includes 2D and 3D annotations functions for chemical structures. They use RasMol (a commonly used tool for 3D drawing of molecules) interpreters as well as SQL database access and templates.

- NetGenics MetaSymphony: integrated interface that enables access to a variety of research databases through the Internet.

Business Models:

- Software Licensing: they offer licenses of their products for a fee.

- Professional service: they offer professional service solutions development and consulting.

Customers:

Abbott Laboratories, Aventis and Pfizer, among others.

7.11. Structural Bioinformatics

Webpage: www.strubix.com

Locations:

San Diego, CA.

Background:

The company was founded in 1996 in San Diego, Ca. They have a subsidiary in Denmark called SBI Advanced Technologies and totally owned by them. They offer an information protein structure platform to assist drug development processes. Part of their technology was originally developed at Immuno Pharmaceutics.

Product Offering:

- Drug Discovery Collaboration: joint projects with pharmaceutical companies oriented to drug discovery using genomic and proteomic information.
- SbdBase: database of protein structures derived from genomic information. Including screening capabilities over chemical and combinatorial libraries.

Business Models:

- ASP model: they offer access to their database through subscriptions.
- Professional Services: joint development projects with pharmaceutical companies

Customers:

Biochem Pharma, Cyberchemics, Yamanouchi Pharmaceutical, De Novo Pharmaceuticals, Arqule.

7.12. Double Twist

Webpage: www.doubletwist.com

Locations:

Oakland, CA. Muenchen, Germany. Basel, Switzerland. Paris, France.

Background:

Double Twist was originally founded as Pangea Systems, Inc. in 1993. In 1997 they got their first round of VC funding and they launched the company as Double Twist in December 1999 They offer bioinformatics analysis tools and since 1999 they also offer a bioinformatics portal powered by their own tools.

Product Offering:

- DoubleTwist.com: web-based application that allows users to retrieve and analyze genomic information to support research activities. They offer comparisons of DNA and protein sequences, gene discovery, patent monitoring and annotation functionality.
- Prophecy: this product includes proprietary annotated genomic data and bioinformatics tools to analyze these data. They offer access to this resource through their DoubleTwist.com portal service.
- GeneForest: this product includes an index of annotated human genes and associated software to analyze this information. It can be licensed with Prophecy as an integrated product.

- Clustering and Alignment: this product includes a set of tools that allows comparison and analysis of multiple sequences simultaneously. It uses sequence alignment algorithms to do these comparisons.
- SNPTwist: consists of a set of tools that allows detection of single nucleotide polymorphisms (SNPs) within public sequence databases.
- Prophecy Toolkit: a Unix-based development platform that allows incorporating customer's proprietary data into DoubleTwist annotated genome database. They offer it as an optional enhancement to the Prophecy product.

This was the only company that offered pricing information in their website. They offer the following subscription options:

Subscription Level	Price per Month
Level 0	Free
Level 1	$ 50
Level 2	$ 150
Level 3	$ 900

Each additional level allows access to a major set of tools and proprietary data.

Business Model:

- ASP model: they offer subscriptions to their portal service DoubleTwist.com.
- Software Licensing: they offer some of their tools under software licensing agreements. It seems they are trying to gradually migrate all their service to the ASP model.

Customers:

Berlex, Rigel, Ontogeny, EOS Technology, Axys Pharmaceuticals, GeneCor International, Biosource, Genetics Institute, Monsanto, Bristol-Myers Squibb, Lilly, Isis Pharmaceuticals, among others.

7.13. Entigen

Webpage: www.entigen.com

Locations:

Eveleigh, Australia.

Background:

Entigen was originally called eBionformatics. The company was founded in May 1998 in Australia with the main purpose of commercializing web-based bioinformatics tools.

Product Offering:

- Bionavigator: This web-based service allows users to execute bioinformatics algorithms for sequence analysis. It also enables integrated access to public sequence databases. End users can upload and download their own data as well and the product enables sharing results among colleagues in a collaborative way.

Business Model:

- ASP model: The pricing scheme is somehow complicated. Users can log on for free but then they have certain amount of units that allows them to execute specific operations. As they finish their units credit they have to buy units for a fee. Tools in the service consume more or less units depending on their complexity.

Customers:

They don't disclose any information about their customers in their Website.

7.14. Silicon Genetics

Webpage: www.sigenetics.com

Locations:

Redwood City, CA.

Background:

The company was created in 1998 based on expression analysis techniques developed by one of his founders. The company specializes in delivering software solutions for analyzing genomic expression information.

Product Offering:

- GeneSpring: pc-based software that enables researchers to analyze and visualize data from gene expression experiments.
- GeNet: web database that allows scientists to share results of gene expression analysis. It includes an easy-to-use interface that allows users to publish results of these analysis and include text and images as well.
- Metamine: this data mining tool works as a background agent that continuously searches gene expression databases and tries to discover new genes or correlations among data sets. Any finding is reported through automatic emails to the end user.

Business Model:

- Software Licensing: the software is installed on customer's networks and they are charged for a license agreement.

- Professional services: They offer training in their products as well as consulting in gene expression experiments and related activities.

Customers:

They claim they have 22 pharmaceutical customers but they don't disclose their names.

7.15. Molecular Mining

Webpage: www.molecularmining.com

Locations:

Ontario, Canada.

Background:

The company was founded in 1999. They obtained their initial funding from Cardinal Health Partners and S.R. One Limited. Both US venture capital firms. They don't have their products in the market yet.

Product Offering:

- MMC GeneLinker Products: This product is not yet released, however they claim it will consist of three packages. One for data exploration of gene expression results, another one for biomedical inference based on gene expression and a third one for gene and protein discovery and modeling.

Customers:

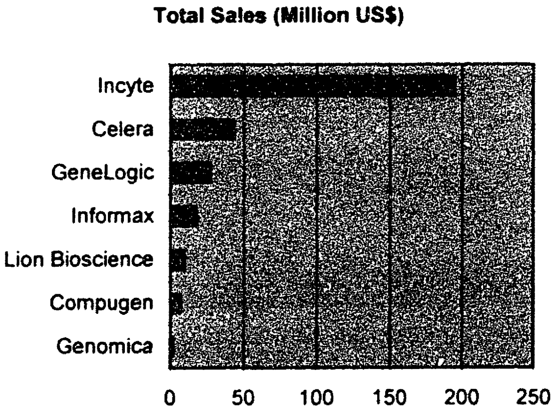
They are in an early stage. They don't have any product in the market yet.

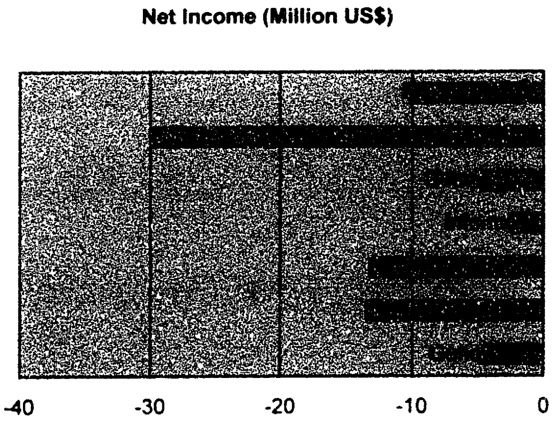
8. Comparisons

8.1. Relative Size

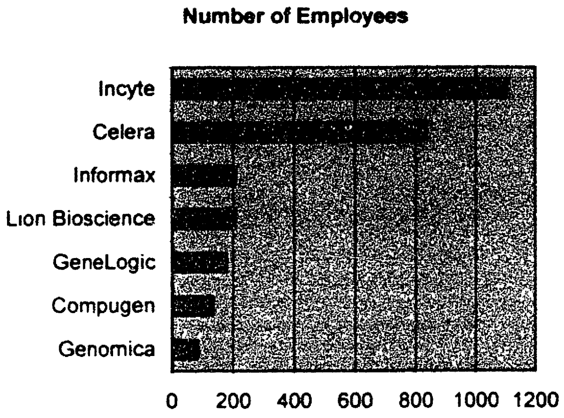
In order to determine the relative size of each of the companies analyzed, I retrieved some financial information on the seven firms that were public and ranked each company according to this information (see appendix 1 for data table and sources).

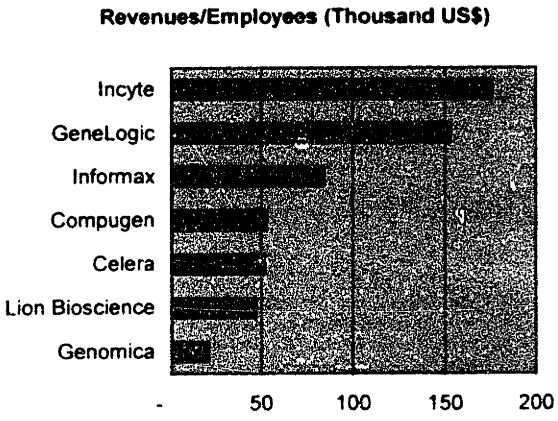
Ranking by Revenues and Net Income



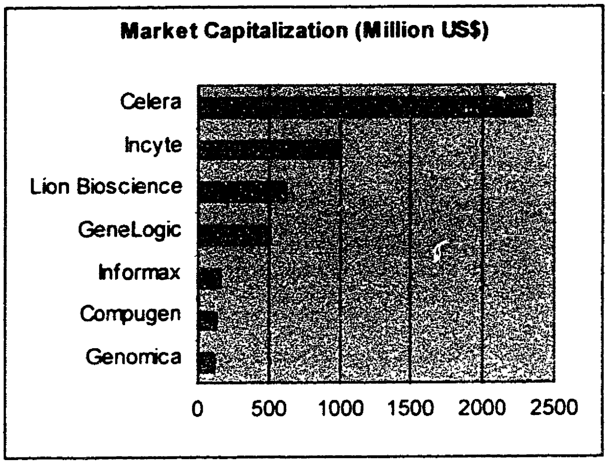


Ranking by Number of Employees and Revenue per Employee:





Ranking by Market Capitalization:



8.2. Business Models

From the companies analyzed the following business models have been used so far and this is the way companies are making money:

- Software Licensing: Software is delivered to the customer and installed in their computing installations. Customers pay a fee per user.
- In-House Development: Bioinformatics companies offer consultants to help developing bioinformatics tools inside the customer's corporations.
- ASP (Application Service Provider): Companies allow users to log in and pay by what they use. Usually a contract signed previously gives the user access to the website. But then the customer is billed based on the tools he used and the computer power consumed.
- Website – Annual Contract: Customers sign an annual contract that allows a pre-established number of users to have access to the analysis tools.

- Hardware/Software solutions: Timelogic offered an integrated solution of software and hardware, and Compugen used to offer it in the past. This model seems to be disappearing as computer capacities grow and software adapts to parallel computing.

8.3. Competing offerings

I analyzed each of the products offered by the companies analyzed, and estimated the following table.

	TimeLogic	Informax	Incyte	Compugen	Genelogic	Genomica	NetGenics	Spotfire	StrucBioinf	DoubleTw.	Lion Biosc.	Entigen	Silicon Gen.	Celera	Molec.Min.
Timelogic	■			■		■									
Informax		■		■		■				■	■	■			
Incyte			■			■				■				■	
Compugen	■	■		■	■	■	■			■	■	■	■		■
GeneLogic				■	■		■	■					■		■
Genomica	■	■	■	■		■				■	■	■		■	
Netgenics				■	■		■	■		■			■		■
Spotfire					■		■	■			■		■		
Str.Bioinf.									■						
Doubletwist		■	■	■		■	■			■	■	■	■	■	
Lion Biosc.		■		■		■		■		■	■	■			■
Entigen		■		■		■				■	■	■			
Silicon Gen				■	■		■	■		■			■		■
Celera			■			■				■				■	
Molec.Min.				■	■		■				■		■		■

Dark gray cells indicate that products of the company indicated in the column and products of the company indicated in the row, are direct competitors. In other words one or some of their products have the same function. Light gray cells on the other hand indicate that products compete but at the same time offer complementarities that the other competitor doesn't offer.

Clear examples of this case are companies offering access to their own databases as Incyte and Celera. The service is similar; they offer access to common data sets but at the same time include other datasets unique to their offering.

Lion Bioscience is the company with more direct competitors according to my estimations. This is due to their horizontal integration.

On the other hand, I couldn't find a direct competitor for Structural Bioinformatics. This company offers access to a database of proteins deducted from genetic information using Structural Bioinformatics' technology.

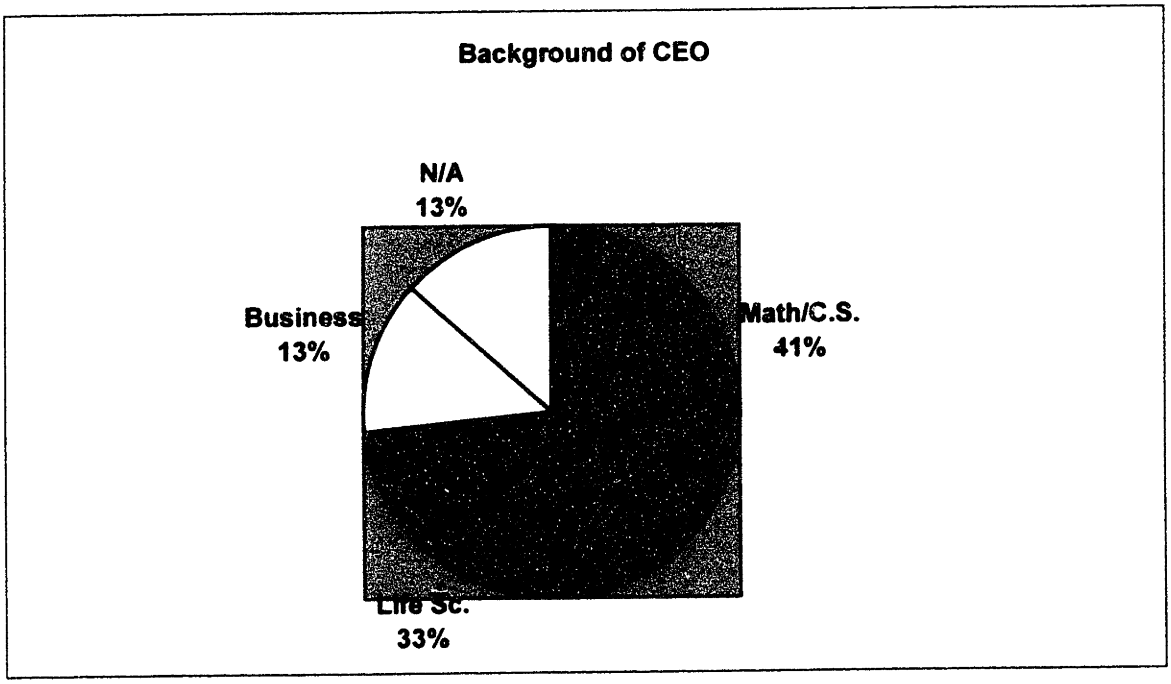
8.4. Who is running the Industry

I looked for the CEO's background of each of the companies analyzed. Most of the CEOs are also the founders and according to my findings, most of them have scientific background.

Company	CEO	Founder (Y/N)	Background
Timeline		N/A	N/A
Informax	Dr. Alex Titomirov	Yes	Life Sciences
Incyte	Roy Whitfield	Yes	Mathematics + MBA
Compugen	Mor Amitai	No	Mathematics
Gene Logic	Mark D. Gessler	No	Business
Genomica	Daniel R. Hudspeth	No	Business
Netgenics	Manuel J. Glynias	Yes	Life Sciences
Spotfire	Christopher Ahlberg	Yes	Computer Science
S.Bioinformatics	Edward T. Maggio	N/A	N/A

Company	CEO	Founder (Y/N)	Background
Double Twist	John Couch	No	Computer Science
Lion Bioscience	Friedrich von Bohlen	Yes	Life Science
Entigen	Howard D. Goldstein	No	Life Science + MBA
Silicon Genetics	Andrew Conway	Yes	Mathematics
Celera	J. Craig Venter	Yes	Life Science
Molecular Mining	Evan Steeg	Yes	Mathematics

Taking the above information I obtained the following proportions. I added Mathematics and Computer Science under the same category.



9. Bioinformatics Market and Strategies

9.1. Market Size

With only three in 10 drugs generating after tax revenues exceeding development costs, pharmaceutical companies are pressured to both improve the efficiency of research and development dollars and expand the global market for drugs to cover R&D investments[17]. Pharmaceutical profitability is also threatened by the fact that drugs with sales of approximately US $25 billion in revenues will come off patent by the year 2002[18].

Below are the phases of development of a new drug as well as some average estimation of time and cost for each phase.

	Preclinical	**Phase I**	**Phase II**	**Phase III**	**Phase IV**
Purpose	Assesses safety and biological activity	Determine human safety and dosage	Evaluate efficacy and adverse effects	Confirm benefits and monitor side effects on broad population	Assess cost-effectiveness and safety
Years	6.5	1.5	2	3.5	1+
Test Population	Laboratory and animal studies	20 to 80 healthy volunteers	100 to 300 patient volunteers	1,000 to 3,000 patient volunteers	100 to 5,000+ patient volunteers
Success Rate	5,000 compounds evaluated		5 enter trials		1 approved
Cost	$25-30 million	$30-35 million	$80-90 million	$200 million	

Source: Asher Dewhurst (2000). Pharmaceutical Informatics. Equity Research, 10.

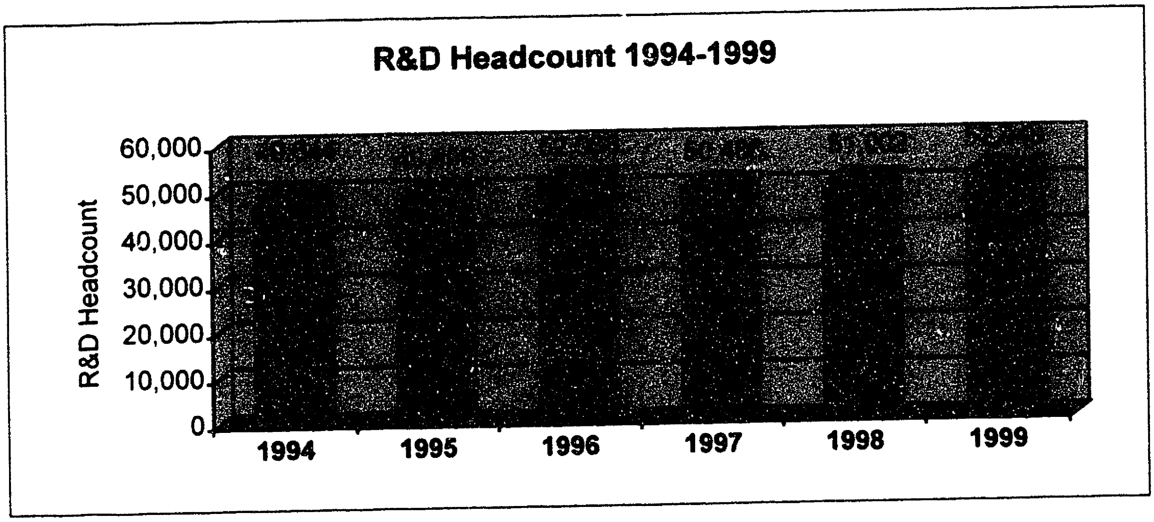
---

[17] Asher Dewhurst (2000). Pharmaceutical Informatics. Equity Research, 4-5.

[18] Purcell, D. J. (1998). Navigating biotechnology's new fiscal opportunities. Nature Biotechnology, 16:51-53

Bioinformatics companies that can alleviate the pressures of drug discovery and development and accelerate the time from discovery to peak sales will be the long-term strategic partners to the cash-rich global drug companies.

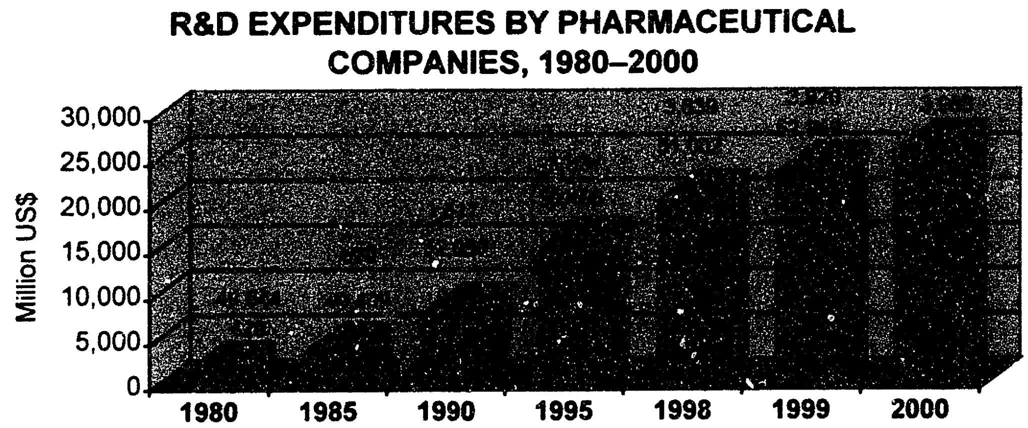
Considering that most of the products analyzed are offered under software type of agreements (licensing, ASP), one way of quantifying the market is using the number of potential end-users for these products. The total number of drug researchers employed by pharmaceutical companies in the US increased only 1.3% annually from 1994 to 1999.



Source: Asher Dewhurst (2000). Pharmaceutical Informatics. Equity Research, 8-9.

For 1999 there were 52,949 researchers working for pharmaceutical companies in the US. Actual bioinformatics tools are only assisting early drug discovery phases and late clinical trials. However, even assuming that an unrealistic 100% of the population of researchers will be assisted by these tools we end up with a very small market. If we consider Doulbetwist's price for their wider set of tools of $900 per month per user, we end up with a maximum hypothetical market of 52,949 researchers x $900 months x 12 months = $571,849,200. On the other hand expenditures in R&D during 1999 outside the US was 16% of the amount invested in the US. Adding an extra 16% to our previous calculation results in a maximum estimated market for bioinformatics tools worldwide of *$663 Million.*

A different approach would be to calculate the approximate amount of savings that bioinformatics tools will bring to pharmaceutical companies.



Source: PhRMAAnnual Survey, 2000.

The previous chart shows total expenditures in R&D by pharmaceutical companies. Savings through the application of bioinformatics technologies are going to be derived from efficiencies in the process and not from reduction in investments. In other words, pharmaceutical companies will be able to do more and better drugs with the same allocation of financial resources. This illustrates the complexity of this approach in calculating potential market size. It only requires an estimation of the number of additional drugs that pharmaceutical companies will be able to produce by using bioinformatics tools, but also an estimation of how more efficient these drugs are going to be by the only application of bioinformatics. But even a hypothetical increase of efficiency in the total drug development process of 10% will translate into a total market (calculated as the size of the savings) of $2,2 Billion. However, there is no business model being used that could generate profits as a portion of the savings generated, and in my opinion it is going to be very difficult to implement one. In the following section I explore one possible strategy for this approach.

9.2.Strategies

In this section I will include some of the strategies that are being used or that could be used in order to take advantage of the current situation in the bioinformatics industry. Although there are still a relatively low number of competitors and I was not able to identify a clear standard product, the potential market size looks too small. The adoption of a right strategy early in the life of a bioinformatics company might be critical in order to grow and succeed in this industry.

9.2.1. Business Model

As explained previously in chapter 8, the business models that are being used by the competitors analyzed in this study are similar to the traditional software/hardware models. Software licensing, ASP and hardware/software solutions have all been used in the computer industry for years and seemed to work for other segments. However, my estimations of the market indicate that the potential number of end-users is very small in order to succeed with this kind of model. The association of this business model with traditional software ones will limit pricing strategies and will probably require prices to have some proportional relation with other software products acquired previously by the organization.

One alternative model would be for bioinformatics companies to have a stake in the final developed drug. If the bioinformatics technology is as good as the providers claim, and the early drug candidates will have better chances to become a new drug in the market, then bioinformatics companies could propose a scheme where they share both the risk and the benefit of each of these drugs. The implementation could be negotiated in terms of royalties or percentage of profits. This model implicates that bioinformatics companies will need to package the product and the service in a consolidated offer.

9.2.2. Horizontal Integration

Only Lion Bioscience has the clear strategy of expanding horizontally in the drug development process. Most of the other competitors are supplying tools for specific stages of the process, while Lion is trying to provide an information technology platform that will support pharmaceutical companies from drug discovery phases to clinical trials and market. In my opinion this is the right approach to try to create a standard in this market. Integrating products horizontally makes the offering much more attractive to the customer and can eventually contribute to create a standard de-facto in the market. However, not even Lion is able yet to provide this platform. They bought Trega recently in order to integrate it with Lion's previous line of products but they still have not adjusted both products to implement this integration. On the other hand they still lack tools to support clinical trials, the longer and more expensive phase of drug developments.

9.2.3. Vertical Integration

Bioinformatics knowledge is becoming a key factor in revolutionizing the way drugs are being developed. Although the real change comes from genomics discoveries, bioinformatics is the way these discoveries are practically implemented and applied. Because of the size and nature of traditional pharmaceutical companies, they have not been able to adopt these technologies rapidly enough. Bioinformatics companies have the advantage that they possess this key knowledge and technology. So one possible strategy for bioinformatics companies is to integrate horizontally and use their own technologies to develop new drugs. This means they have to go from drug discovery (which is familiar to them) to drug development, market and distribution (which is totally new for them). This is a risky approach and requires a considerable investment in infrastructure and technology. On the other hand the revenue generation occurs approximately 10 years latter, and only if the selected target works its way through pre-clinical and clinical trials. However I do think it is feasible either by partnering with actual pharmaceutical companies or by obtaining the required financial resources to do it in-house or out-sourced. Even pharmaceutical companies are outsourcing most of clinical trials today.

9.2.4. Patenting

Companies as Celera, Compugen and Incyte are actively patenting gene discoveries made with their own bioinformatics technologies. The strategy behind this approach is to expect that these genes will be targeted sometime in the future for some drug application and therefore they will be able to charge royalties or other type of fees for the usage of these genes.

It is still not clear if this kind of patent will be defendable in court. As of today, there is no case of usage of these patents in new drugs but this might change in the future. In my opinion it is very risky to rely only on this strategy as a way of generating revenues, but it does work if it is coordinated with other related strategies for the short-term.

9.2.5. Partnering

Vertical integration can be achieved by partnering with pharmaceutical companies as mentioned above, while horizontal integration can be achieved by partnering with other bioinformatics providers. Another approach, however, is to try to partner with solution providers of other type of tools for pharmaceutical companies. One of these partners could be the Enterprise Resource Planning (ERP) software provider. Companies such as SAP already have a customer base in the pharmaceutical industry and have experience in developing and implementing large-scale software applications in pharmaceutical organizations. Integrating bioinformatics tools in their solutions might be a way of extending their actual product offering. On the other hand, for a bioinformatics company it could be an opportunity to supply bioinformatics technology while obtaining a customer base and distribution channel.

9.3. The Future

Assuming the business models being used today and the estimated market size, there is no potential multi-billion dollar player in this industry. I mentioned some strategies in the sections above where a company might integrate vertically or horizontally to capture more market share.

Horizontal integration means that the company is still a bioinformatics player but vertical integration means that the company becomes a pharmaceutical company. I estimate that there are two possible future scenarios where the bioinformatics companies can boost its revenues per-year:

- B2C: Today the customer for bioinformatics tools are either pharmaceutical or biotechnology companies. But if the regular individual becomes a customer of bioinformatics tools then this market can grow exponentially. The way this can happen is by offering services to end users where they can check what probabilities they have of suffering specific diseases based on the correlation of their genes with the ones of controlled populations. The major impediment to achieve this application is the cost of sequencing or expressing the genes of each individual. However improvements in sequencing and DNA chips might make this application feasible in the future.

- Growth in Pharmaceutical Companies: until today the pharmaceutical industry has been mostly controlled by few major players. Its almost impossible to enter this market as a new competitor: First, because of the financial resources required to create an infrastructure to develop new drugs, and secondly because of the difficulties in taking new drugs into the market. If genomics is able to simplify the process and reduce the costs and times required to develop new drug this might generate an emerging number of new pharmaceutical companies that will be potential users of bioinformatics products.

10. Summary

- After analyzing the technology and the market information on the bioinformatics industry, the industry seems to be in a very early stage. The number of competitors has been constantly increasing in the last 10 years but still there is no major technology standard in the market. Most of the competitors analyzed have revenues below $20M, less than 200 employees and all of them are generating losses.

- The business models being used are very similar to the ones used traditionally in the hardware and software industry: licensing, ASP (Application Service Provider), joint developments and hardware/software solutions.

- The bioinformatics field merges two disciplines, computer science and biology. Almost 41% of the companies' CEOs had backgrounds in mathematics or computer science, while 33% had life science backgrounds and only 13% had business educational background.

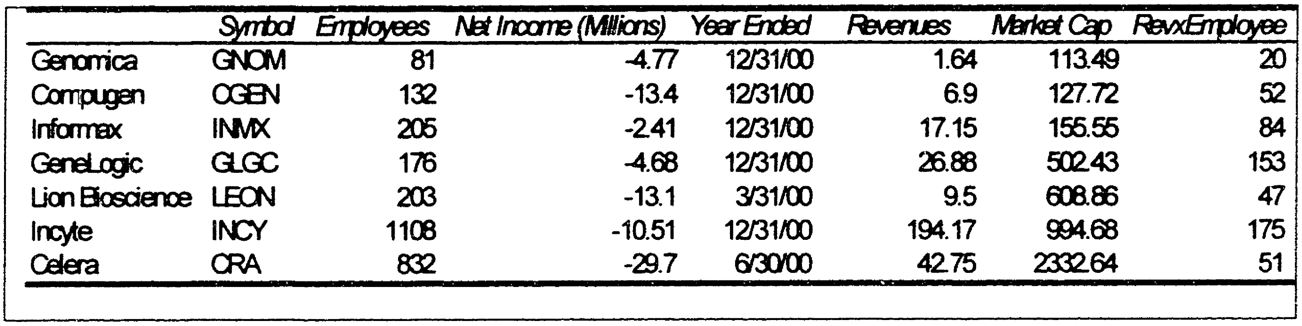
- The market potential is relatively small. My estimation is that the maximum market size considering the potential user population will be below *$663 Million*. Some authors claim that the actual market is $300M and will grow to $2B by 2005 (Jason, 2000).

- Implementing strategies of horizontal or vertical integration might boost revenues in the long term for bioinformatics players.

- In the future revenues might boost by offering bioinformatics services to individuals, allowing them to know whether they have high probabilities of suffering particular diseases based on their genetic information. However the required technology for sequencing or expressing gene information is not yet accessible to the regular individual.

11. Appendix 1

Financial information. Source: The Wall Street Journal.



	Symbol	Employees	Net Income (Millions)	Year Ended	Revenues	Market Cap	RevxEmployee
Genomica	GNOM	81	-4.77	12/31/00	1.64	113.49	20
Compugen	CGEN	132	-13.4	12/31/00	6.9	127.72	52
Informax	INMX	205	-2.41	12/31/00	17.15	155.55	84
GeneLogic	GLGC	176	-4.68	12/31/00	26.88	502.43	153
Lion Bioscience	LEON	203	-13.1	3/31/00	9.5	608.86	47
Incyte	INCY	1108	-10.51	12/31/00	194.17	994.68	175
Celera	CRA	832	-29.7	6/30/00	42.75	2332.64	51

12. Bibliography

- [1] Lyn Larrys. Machines and the Human Genome Project. Spring 1992.
- [2] The Human Project website. http://www.ornl.gov/hgmis/project/timeline.html
- [3] Elizabeth Pennisi. The Human Genome. Science Magazine. Volume 291, Number 5507, February 16, 2001.
- [4] Jason Reed. Trends in Commercial Bioinformatics. Oscar Gruss Biotechnology Review. March 13, 2000.
- [5] Business Week Online. http://www.businessweek.com/bwdaily/dnflash/dec2000/nf2000128_877.htm. December 8, 2000. By Amy Tsao.
- [6] Compugen Homepage. www.compugen.co.il
- [7] Akja Agrawal. A Software Model That Fathoms the Human Heart. Business Week. November 21, 2000.
- [8] Bellavance, L. L., Donlan. Bioinformatics Primer. Genetic Engineering News, 2000. 19:32-3
- [9] Gardner, S. (1999). The Evolution of Bioinformatics. BITS Journal. http://www.bitsjournal.com
- [10] Philip B.C. Jones. The Commercialization of Bioinformatics. EJB Electronic Journal of Biotechnology. Vol. 3 No. 2, August 15, 2000.
- [11] Lisa Althoff. DNA Chip - Genetic Testing of the Future, 1999.
- [12] Alice Park, Dan Cray, Cathy Booth and Dick Thompson. Brave New Pharmacy. Time Magazine, January 15, 2001. Pg.64-65.
- [13] Stanley T. Crooke. 1998. Optimizing the Impact of Genomics on Drug Discovery and Development. Nature Biotechnology Magazine. Volume 16. Supplement. Pg 29-30.
- [14] Robert Stevens, Carole Globe, Patricia Baker, Andy Brass. A Classification of Tasks in Bioinformatics. Bioinformatics Magazine. February 2001. Pg.180-188.
- [15] Tom Headrik. Celera's Algorithmic Edge. COMMUNITY PERSPECTIVE. February 6, 2001
- [16] Utterback, James M., Mastering the Dynamics of Innovation, (MA: Harvard Business School Press, 1994), pp.23-55.
- [17] Asher Dewhurst (2000). Pharmaceutical Informatics. Equity Research, 4-5.
- [18] Purcell, D. J. (1998). Navigating biotechnology's new fiscal opportunities. Nature Biotechnology, 16:51-53

THESIS PROCESSING SLIP

FIXED FIELD: ill. ______________ name ______________

index ______________ biblio ______________

► COPIES: [✓ Archives] Aero [✓ Dewey] Barker Hum

Lindgren Music Rotch Science Sche-Plough

TITLE VARIES: ► ☐ ______________

NAME VARIES: ► ☑ [illegible]

IMPRINT: (COPYRIGHT) ______________

► COLLATION: ______________

► ADD: DEGREE: ______ ► DEPT.: ______________

► ADD: DEGREE: ______ ► DEPT.: ______________

SUPERVISORS: ______________

NOTES:

cat'r: date:

page:

► 7709

► DEPT: Mgt

► YEAR: 2001 ► DEGREE: S.M. M.C.I.

► NAME: TALEON, Carlos Benjami