

Patent coupling analysis of primary organizations in genetic engineering research

SZU-CHIA LO

Graduate Institute of Library and Information Science, National Chung Hsing University,
Taichung (Taiwan)

The aim of this study is to reveal the possible linkage among the 40 primary organizations in Genetic Engineering Research by taking the Patent Coupling approach. The primary organizations were defined by the productivity and identified by the patent count and Bradford Law. The author analyzed the cited patents of the patents granted by United States Patent and Trademark Office (USPTO) from 1991 to 2002 to the 40 primary organizations (assignees) in Genetics Engineering Research to establish the correlation.

780 coupling pairs formed by the 40 primary organizations and Coupling Index and Coupling Strength were calculated for each pair and primary organization. Correlation Analysis and Multiple-Dimension Scaling were applied further based on Coupling Index. Technological clusters were found in the results of the analyses.

Introduction

“Biotech Era” started to take shape back in the 1950s when the methods of locating and identifying chromosomes and genes were found. With the development of genetic engineering research, information technologies and global commerce, not only a new economic framework is created, but the re-engineer of human society is undergoing [RIFKIN, 1999]. The science research and technology development of genetic

Received February 12, 2007

Address for correspondence:

SZU-CHIA LO

Graduate Institute of Library and Information Science, National Chung Hsing University

250 Kuo-kuang Rd., Taichung 402, Taiwan

E-mail: loszuchia@gmail.com

0138–9130/US \$ 20.00

Copyright © 2007 Akadémiai Kiadó, Budapest

All rights reserved

engineering continue to be one of the important drive force of bringing new resources and global commerce. Patenting is still an essential strategy of protecting the intellectual property and potential monetary value. Continuing the study done by the author [LO, 2004], the aim of this study is to reveal the research links among the primary organizations in Genetic Engineering Research by taking the Patent Coupling approach. The method of Patent Coupling adopted from the Bibliographic Coupling in Bibliometrics is used to analyze the patents granted by United States Patent and Trademark Office (USPTO) from 1991 to 2002 to the 40 primary organizations (assignees) in Genetics Engineering Research that are identified by the patent count and Bradford Law. The author tried to enclose the research link among the primary organizations (assignees) in Genetic Engineering Research by taking the coupling approach. The concept of Bibliographic Coupling was first mentioned by Fano and formally introduced by Kessler [KESSLER, 1963]. The concept of Bibliographic Coupling is to define the relevance between two literatures by the number of shared references. Based on the Bibliographic Coupling concept, this study uses cited patents as the coupling base and the Coupling Index to define the link between organizations.

4,475 patents granted to the 40 primary organizations and 20,203 cited patents listed on the front pages of the 4,475 patents were examined in this study. 780 coupling pairs were formed with the 40 primary organizations. Two indicators, *Coupling Index (CI)* and *Coupling Strength (CS)*, were calculated for each coupled pair and primary organization. Correlation Analysis and Multi-dimensional Scaling were run based on the Coupling Index to present the clusters.

Literature review

The work done by Narin in 1994 established the use of patent bibliometrics approach [NARIN, 1994]. The bibliometrics approach has been applied in studies of research productivity and impact. Patent counting is heavily used to identify the productive entities, which include countries, assignees and inventors [BANERJEE & AL., GARG, 2000; GARG & PADHI, 1998; NARIN, 1994; NARIN, 1995]. The results of citation counts were used as indicators to present the level of impact [ALBERT & PLAZA, 2004; JAFFE & AL., 1997; MOED, 2000] and the density of direct citing was applied to describe the technical associations between science research and technology development [CARPENTER & AL., 1980; COLLINS & WYATT, 1988; VIANNEN & AL., 1990].

To establish relationships among documents, bibliographic coupling, citing same source documents [KESSLER, 1963, KESSLER, 1965], and co-citation, links cited documents through later documents [SMALL, 1973; CAWKELL, 1976; MARSHAKOVA, 1979; BELLARDO, 1980] have been applied besides direct referencing to earlier documents. The concept of Patent Coupling used in this study is transferred from the bibliographic coupling method, which was proposed by Kessler from Fano's idea

[KESSLER, 1963], onto patent analysis. The hypothesis is that two articles relate to each other if they share the same cited references. The relevance intensifies as the number of shared references increases. Every two-article form a bibliographic coupling pair and the ones with more shared cited references are more relevant to each other. Kessler applied the coupling concept on comparing the results of bibliographic coupling and analytic subject indexing. The similarity was found among the results of these two methods used in information retrieval [KESSLER, 1965]. The article-network could be built upon this hypothesis and the correlation could be also tested [WEINBERG, 1974; EGGHE & ROUSSEAU, 1990; GARFIELD, 1998]. The studies done based on the concept of bibliographic coupling were limited. Huang, Chiang and Chen tried to construct a patent citation map of Taiwan's high-tech companies by using bibliographic coupling [HUANG & AL., 2003]. In this study, the author tried to apply the concept of bibliographic coupling to establish the correlation among the primary organizations of genetic engineering research. The patents granted to the primary assignees and the shared patents cited by the assignees were examined. The Coupling Index and Coupling Strength were used to demonstrate the correlation among the primary institutes of genetic engineering research.

Methodology

The data source used in this study is USPTO Patent database, one of the most exhaustive patent sources. The patents included in this study were selected by the International Patent Classification (IPC) numbers and the issued date was added to the searching criteria to identify the patents granted in Genetics Engineering Research during the period of 1991 to 2002. The patents that are with main IPC numbers C12N 15/00 (Mutation or genetic engineering), C12P 21/00, C07H 21/00, C07K 14/00 (Preparation of peptides or proteins) and C12Q 1/68 (Measuring or testing processes involving nucleic acids) were the candidates for analysis in this study. The U.S. Patent Classification (USPC) numbers were also included in the strategies to insure the completeness of the dataset. The USPC numbers relate to the Genetic Engineering, such as subclasses 435/440 and 435/69.1 were added to search strategies. "Patent Count" was used for productivity analysis and with Bradford's Law primary organizations that hold significant amount of patents were identified and Patent Coupling Analysis was used to demonstrate the link among primary organizations.

The concept of Patent Coupling was applied to reveal the correlation among the primary organizations. The assumption was that the organizations own patents cited same patents were more technological related than those which did not share same citations. The higher of the number of shared citations presents the higher correlation of the organizations. Most of the primary organizations do not show the research impact on others. By the common cited patents, the correlations could be further built based on

the results. Two indicators were developed to demonstrate the correlations among the primary organizations, Coupling Index (CI) and Coupling Strength (CS) were calculated for each coupling pair and organization. CI is the ratio of the number of shared citations over the sum of citations of paired organizations. The higher coupling index presents closer link between two paired organizations. Each organization got the CS that is the sum of the coupling indexes of the organization's coupling pairs.

$$CI_{ij} = C_{ij} \div (C_i + C_j)$$

CI_{ij} is the coupling index of entities i and j , C_i is the number of patent citations of entity i , C_j is the number of patent citations of entity j , C_{ij} is the number of shared patent citations of entities i and j .

$$CS_i = \sum_{j=1}^n I_{ij}$$

CS_i is coupling strength of entity i , CI_{ij} is the coupling index of entities i and j , the number of entity j could be from 1 to n , n is the number of core assignees minus 1.

Based on the Coupling Index, Correlation Analysis and Multi-dimensional Scaling were done to identify the technological clusters.

Results

13,055 USPTO patents were retrieved from USPTO Patent database. 40 primary organizations were identified by the patent count and Bradford's distribution analysis. Including University of California, INCYTE, SmithKline, Department of Health and Human Services, the 40 primary organizations hold one third of the Genetics Engineering related patents issued from 1991 to 2002.

Bibliographic coupling analysis

The 40 primary organizations form 780 coupling pairs and it was found that coupled organizations of 724 pairs shared same cited patents. 151 (19.36%) pairs had less than 10 shared patents. 261 (33.46%) pairs shared 10 to 29 cited patents. Only 19 (2.44%) pairs had higher coupling strength, shared more than 200 cited patents. Coupling Indexes were calculated further for the 724 coupling pairs. Among the pairs, Institut National de la Sante et de la Recherche Medicale and Institut Pasteur had the highest Coupling Index. Both organizations are Paris, France based institution and shared faired amount cited patents in the patents that involved the technology of gene sequencing. The pairs of Genetics and University of Michigan, Merck and Salk Institution, Genentech and Genetics, Thomas Jefferson University and University of Pennsylvania

were the other top 4 pairs that had higher Coupling Indexes. Patents granted to Genetics and University of Michigan and involved with the techniques of DNA recombinant cited common patents. Patents relate to the preparation of protein and owned by Genentech and Genetics shared the cited patents deal with recombinant techniques. Thomas Jefferson University and University of Pennsylvania are technically related in gene detecting and gene therapy by sharing the cited patents. Table 1 lists the 20 coupling pairs which were with higher Coupling Index.

Further examining was done on the coupling density of each organization with other primary organization. University of California coupled with other 39 primary organizations and the value of Coupling Strength is the highest, but none of the coupling pairs had significant coupling index. Chiron, Baylor College, General Hospital (Massachusetts, U.S.) and Gen-Probe also had stronger Coupling Strength, Different from the organizations listed above; Eli Lilly, Human Genome, Nordisk and Ludwig Institute demonstrated weaker Coupling Strength. Table 2 lists more details.

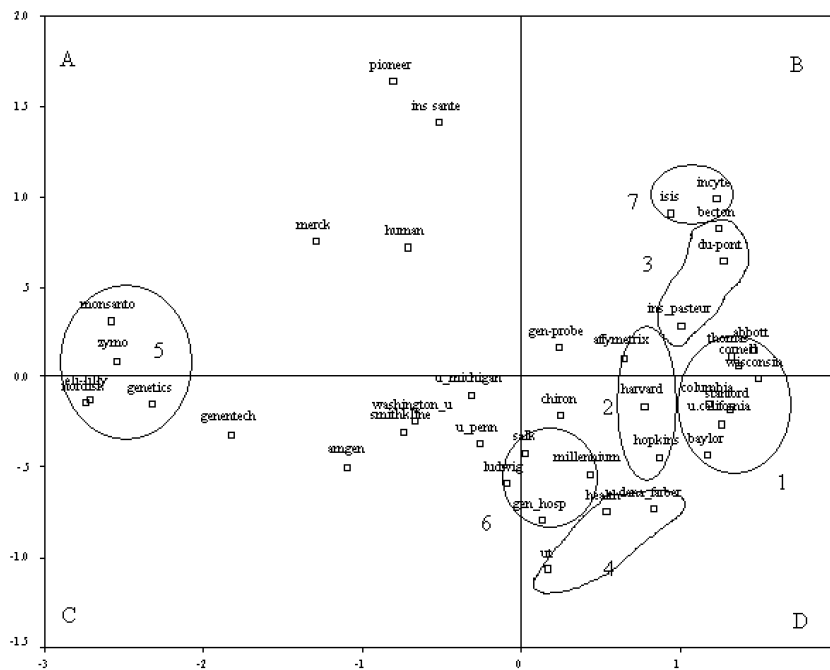


Figure 1. Technological Clusters of Primary Organizations
 (A: Modify forms of DNA or RNA Segments; B: Introduction of Genetic Materials/Preparation of Vectors;
 C: Preparation of Peptide or Protein; D: DNA Recombinant Technology)

Table 1. The first 20 coupling pairs – by coupling index

Rank	Coupling pair	Coupling Index
1	INSRM – Pasteur	0.339
2	Genetics – U. Michigan	0.202
3	Merck – Salk Inst.	0.195
4	Genentech – Genetics	0.194
5	Thomas Jefferson U. – U. Penn	0.193
6	Chiron – Washington U	0.189
7	Abbott – Gen Probe	0.157
8	Columbia – Gen Probe	0.154
9	Chiron – Gen Probe	0.153
10	General Hosp. – Harvard U	0.151
11	Chiron – Genetics	0.150
12	Becton – Gen Probe	0.147
12	Cornell – U Penn	0.147
14	Gen Probe – Pasteur	0.140
15	Columbia – Salk Inst.	0.135
16	Affymetrix – U Calif.	0.134
17	Chiron – U. Calif.	0.131
17	Human Genome – SmithKline	0.131
19	Amgen – Genentech	0.128
20	Monsanto – Pioneer	0.127

Correlation Analysis and Multi-dimension Scaling Analysis were further done based on the Coupling Indexes. 7 clusters were found in the results. Reviewing the patents granted to the organizations in each cluster and the clusters could be named as technological clusters. Figure 1 presents a visual of the clusters and Table 3 lists the organizations of the clusters. Four zones of Figure 1 present four groups of technologies, A: Modify forms of DNA or RNA Segments; B: Introduction of Genetic Materials/Preparation of Vectors; C: Preparation of Peptide or Protein; D: DNA Recombinant Technology. The patents granted to the organizations positioned close to the center of A and B zones involved the technologies of modifying DNA or RNA segments by introducing genetic materials and the ones located between zone C and zone D relate to the preparation of peptide by the recombinant technology. The center is the Gene Sequencing zone. The cluster that included University of California, Stanford University, Wisconsin University, Columbia University, Abbott, Thomas Jefferson University, Baylor College and Cornell University was identified as the “Modifying DNA or RNA fragments” Group. The cluster of John Hopkins University, Harvard and Affymetrix focus more on the “Measuring or Testing the Nucleic Acids”. Institute Pasteur, Du Pont and Becton belong to the cluster concentrated on the “Preparation of Vectors”. The Department of Health and Human Services (US), University of Texas and Dana-Farber specialized in “Applying the techniques of DNA recombinant in modifying DNA or RNA segments”. Genetics Institution, Eli-Lilly, Nordisk, Monsanto and ZymoGenetics were in the technological cluster of the “Preparation of Peptide and

Protein”. Salk Institution, General Hospital, Millennium and Ludwig Institution owned the techniques covered the basics of “DNA Sequences Coding for Fusing Protein”. INCYTE and ISIS focus on the methods of “Introduction the Genetic Materials”.

Table 2. Coupling counts of primary organizations

	Primary organization (assignees)	Number of coupling	Times of coupling	Average coupling	Coupling strength	Strong pairs*
1	U California	39	5,329	14.06	2.787(01)	27
2	INCYTE	31	649	1.76	0.816(36)	5
3	SmithKline	37	598	1.93	0.842(33)	3
4	Dept of Health & Human (US Gov.)	39	2,181	10.39	2.248(06)	26
5	Genentech	39	3,231	18.57	2.104(10)	23
6	Chiron	39	3,906	24.26	2.351(02)	22
7	U Texas	39	1,863	13.03	1.790(17)	19
8	Johns Hopkins U	38	1,737	12.41	2.104(10)	25
9	Human Genome	34	497	3.79	0.676(39)	2
10	General Hospital	38	2,052	16.03	2.316(04)	27
11	Institut Pasteur	38	1,750	14.23	2.089(12)	24
12	ISIS	35	1,185	11.07	0.874(31)	7
13	Harvard U	38	1,593	15.17	2.032(14)	24
14	Stanford U	39	1,912	19.31	2.204(09)	24
15	Genetics Institute	36	1,485	15.47	1.292(27)	6
16	Millennium	38	1,152	12.52	1.627(30)	16
17	Affymetrix	36	4,969	59.87	1.297(26)	14
17	Gen-Probe	36	3,316	39.95	2.282(05)	22
19	Eli Lilly	22	240	2.96	0.449(40)	3
20	Becton	33	1,106	13.83	1.258(28)	10
21	Du Pont	38	1,165	14.94	1.501(22)	16
21	Salk Institute	38	1,267	16.24	1.956(15)	20
23	Pioneer Hi-Bred	32	564	7.32	0.823(35)	5
24	U Pennsylvania	36	1,350	17.76	1.818(16)	18
25	Wisconsin U	37	1,145	15.27	1.407(24)	17
26	Columbia U	38	1,486	20.93	2.077(13)	24
26	Merck	34	488	6.87	0.869(32)	4
28	Abbott	39	1,970	28.14	2.248(06)	24
28	Nordisk	32	467	6.67	0.694(37)	6
30	Ludwig Institute	37	414	6.00	0.694(37)	3
30	Thomas Jefferson U	38	1,750	25.36	2.207(08)	26
32	Baylor College	38	1,830	27.73	2.322(03)	30
32	Dana-Farber	37	732	11.09	1.301(25)	15
32	Monsanto	29	618	9.36	0.836(34)	6
32	Washington U	37	1,074	16.27	1.413(23)	14
36	U Michigan	36	1,609	24.75	1.248(29)	12
37	Amgen	38	1,489	23.27	1.656(18)	16
38	Cornell U	38	1,752	28.26	1.643(19)	18
39	INSRM	38	806	13.21	1.605(21)	14
39	ZymoGenetics	34	637	10.44	1.118(30)	11

* Coupling Index > 1

Table 3. Technological clusters of primary organizations – listing

Cluster	Organizations (assignees)
1	U. of California, Stanford U., Wisconsin U., Columbia U., Abbott, Thomas Jefferson U., Baylor College, Cornell U.
2	John Hopkins Univ., Harvard Univ., Affymetrix
3	Institut Pasteur, Du Pont, Becton
4	Dept. of Health and Human Services, Univ. of Texas, Dana-Farber
5	Genetics Inst., Eli-Lilly, Nordisk, Monsanto, ZymoGenetics
6	Salk Inst, General Hospital, Millennium, Ludwig Inst.
7	INCYTE, ISIS

1=“Modifying DNA or RNA fragments”; 2=“Measuring or Testing the Nucleic Acids”; 3=“Preparation of Vectors”; 4=“Applying the techniques of DNA recombinant in modifying DNA or RNA segments”; 5=“Preparation of Peptide and Protein”; 6=“DNA Sequences Coding for Fusing Protein”; 7=“Introduction the Genetic Materials”.

Discussion and conclusions

4,475 patents and 20,203 cited USPTO patents were examined in this study. The author took Patent Coupling approach to reveal the technological links among the primary organizations. 7 technological clusters were identified based on the result of the correlation analysis and multiple dimensional scaling on Coupling Index. The author tried to show the feasibility of utilization of “Patent Coupling” on correlation analysis. The results demonstrated the technique relations among various clusters that did not appeal in the results of direct Citing-Cited Analysis [LO & HUANG, 2005]. The outputs showed a different viewpoint of cluster linking which is worth further investigation for strategic planning in technology development.

References

- ALBERT, A., PLAZA, L. M. (2004), The transfer of knowledge from the Spanish public R & D system to the productive sectors in the field of biotechnology. *Scientometric*, 59 (1) : 3–14.
- BANERJEE, P., GUPTA, B. M., GARG, K. C. (2000), Patent statistics as indicators of competition: an analysis of patenting in biotechnology. *Scientometrics*, 47 (1) : 95–116.
- BELLARDO, T. (1980), The use of co-citations to study science. *Library Research*, 2 (3) : 231–237.
- CARPENTER, M. P., COOPER, M., NARIN, F. (1980), Linkage between basic research literature and patents. *Research Management*, 23 : 30-35.
- CAWKELL, A. E. (1976), Understanding science by analyzing its literature. *The Information Scientist*, 10 (1) : 3–10.
- COLLINS, P., WYATT, S. (1988), Citations in patents to the basic research literatures. *Research Policy*, 17 : 65–74.
- EGGHE, L., ROUSSEAU, R. (1990), *Introduction to informetrics: quantitative methods in library, documentation and information science*. Elsevier Science Publishers, Amsterdam, Netherlands.
- GARFIELD, F. (1998), From citation indexes to informetrics: is the tail now wagging the dog? *Libri*, 48 : 67–80.
- GARG, K. C., PADHI, P. (1998), Scientometric study of laser patent literature. *Scientometrics*, 43 (3) : 443–454.

- HUANG, MU-HSUAN, CHIANG, LI-YUN, CHEN, DAR-ZEN (2003), Constructing a patent citation map using bibliographic coupling: a study of Taiwan's high-tech companies. *Scientometrics*, 58 (3) : 489–506.
- JAFFE, A. B., FOGARTY, M. S., BANKS, B. A. (1997), *Evidence from patents and patent citations on the impact of NASA and other federal labs on commercial innovation* (NBER Working Paper 6044). NBER, Cambridge, MA.
- KESSLER, M. (1963), Bibliographic coupling between scientific papers. *American Documentation*, 14 : 10–25.
- KESSLER, M. M. (1965), Comparison of the results of bibliographic coupling and analytic subject indexing. *American Documentation*, 16 (3) : 223–233.
- LO, SZU-CHIA (2004), *A Study of the Productivity of Genetic Engineering Research from 1991 to 2002 Using the Patentometrics Approach*. Paper presented at International Workshop on Webometrics, Informetrics and Scientometrics & Fifth COLLNET Meeting, Roorkee, India.
- LO, SZU-CHIA, HUANG, MU-SHUAN (2005), *Research impact on genetic engineering research: patent citation analysis*. Poster presented at 10th International Conference of the International Society for Scientometrics and Informetrics, Post Session, Stockholm, Sweden.
- MARSHAKOVA, I. V. (1979), Citation networks in information science. *Scientometrics*, 3 (1) : 13–26.
- MOED, H. F. (2000), Bibliometric indicators reflect publication and management strategies. *Scientometrics*, 47 (2) : 323–346.
- NARIN, F. (1994), Patent bibliometrics. *Scientometrics*, 30 (1) : 147–155.
- NARIN, F. (1995), Patents as indicators for the evaluation of industrial research output. *Scientometrics*, 34 (3) : 489–496.
- RIFKIN, J. (1999), *The Biotech Century: how genetic commerce will change the world*. Phoenix, London.
- SMALL, H. (1973), Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24 (3) :265–269.
- VIANNEN, B. G. VAN, MOED, H. F., RAAN, A. F. J. VAN (1990), An exploration of the science base of recent technology. *Science Policy*, 19 : 61–81.
- WEINBERG, B. H. (1974), Bibliographic coupling: a review. *Information Storage and Retrieval*, 10 (5/6) : 189–196.