THE DIVISION OF INNOVATIVE LABOR

AMONG UNIVERSITIES, ENTREPRENEURS, AND CORPORATIONS

IN AGRICULTURAL BIOTECHNOLOGY

Gregory D. Graff

and

David Zilberman

University of California

Summer 2005

JEL Classifications: O32 - Management of Technological Innovation and R&D; Q16 - Agricultural R&D, Technology; L33 - Comparison of Public and Private Enterprises; Privatization; Contracting Out

---

Gregory D. Graff is a Research Economist with the Public Intellectual Property Resource for Agriculture at the University of California, Davis, and David Zilberman is a Professor in Agricultural and Resource Economics at the University of California, Berkeley.

Special thanks to Brian Wright, Alain de Janvry, Bronwyn Hall, David Mowery, Nicholas Kalaitzandonakis, Robert Evenson, Wally Huffman, GianCarlo Moschini, and Brent Mishler for helpful comments. Gratitude to Zhihai Zheng for excellent research assistance. The authors are responsible for all omissions and errors.

Correspondence to: Gregory Graff, gdgraff@ucdavis.edu.

1. Introduction

The conventional economic paradigm of research and development as a linear process has, at least since the Second World War, applied a convenient shorthand of ‘basic’ research, ‘applied’ research, and product ‘development,’ assumed to follow causally in some sense one from the other (Bush 1945; National Science Board 2003). This intuitive framework has informed science and technology policy at all levels and has profoundly shaped the public’s image of research activities carried out at federal laboratories and universities at public expense. One of the fundamental economic justifications for the public financing of knowledge creation has hinged on the ‘public good’ nature of knowledge, distinguishing basic research as that which generates results with more public-good attributes and broader impacts on economic and social welfare from applied research as that which generates results with more private-good attributes that are more likely to benefit private agents positioned to capture the associated rents. As such, the linear paradigm represents more the terms of a political-economic compromise over public funding of research than it does a definitive description of real R&D processes.

The conventional linear paradigm has, as a result, been consistently challenged by observers of the complex and multidimensional nature of real R&D processes. Some have noted that both practical utility and greater understanding provide crucial motivation and justification to spending on even the most basic of research programs (Rosenberg and Nelson 1994; Stokes 1997). Others have pointed to the dynamics of feedback and convergence flowing from mere technological applications to inform and advance basic science (Kealey 1996; Mowery 1995). While, in theory, the need for public provision of pure public goods is uncontroversial, the results of publicly supported research have not always evidenced an efficient or equitably distributed public benefit, at least in the short term.

In one major area of publicly funded R&D, that supporting agricultural production, the ‘basic’ versus ‘applied’ distinction has never been quite as clear as that made in the post-World-

War linear paradigm. Regardless, since the 1860s significant large public financial commitments have been made through the Land Grant system of U.S. universities, in effect designed to deliver private economic gains and practical results for students educated within the system, for farmers aided by advances in farming technologies and genetic materials, and for consumers provided with an ever cheaper and more abundant food supply (Alston, Norton, and Pardey 1995; Sunding and Zilberman 2001; Williams, 1991). Recently, in the agricultural life sciences, what might be considered some of the most basic research, namely work in molecular biology and recombinant DNA, has often stood very close to the market place, blurring the distinction between basic and applied research and the public versus private aspects of the knowledge assets thereby produced (Rausser 2001; Wright 1998, 2001).

In the agricultural life sciences the question has intensified since biotechnology arose in the 1970s and 1980s out of both publicly and privately funded R&D. The new biologically based technological regime has shaken up a relatively stable *status quo* in the division of innovative labor in mechanically and chemically intensive agriculture, challenging intersectoral relationships and giving rise to new arrangements for generating and appropriating value from innovation, in the wake of which a number of fundamental economic and research policy questions have emerged and persisted. Have the roles of public and private sector researchers become indistinguishable, as corporations invest private funds in projects—such as sequencing the genomes of important crop species—that could, and perhaps should be deemed as scientific public goods, while at the same time university and government laboratories make commercially valuable discoveries, such as genes, which they then privatize via patents, to be developed and marketed by private firms? Are universities and government laboratories being subsidized by the public to provide mere substitutes—of knowledge and technology—in a market where industry is now, in fact, *not* underinvesting? Still, how prevalent have the private sector corporations been in the creation of the very fundamental breakthrough technologies? And how effective have venture funded biotechnology entrepreneurs been as vehicles for technology development? While much has been made of these questions in recent debate, little systematic empirical work has documented the differences between the various sources of biological innovation for agriculture to provide sound basis for policy in agricultural R&D and more generally concerning the role and context of university R&D in driving entrepreneurship and economic development.

We suggest that one of the sources of sectoral R&D role confusion lies in a failure to accurately reconcile evolutionary economic theories on how heterogeneous technologies tend to emerge over time with predictions of theory that organizations in different sectors of the economy should enjoy different comparative advantages in R&D activities and output. Existing theories on innovation suggest that basic exploratory research serves, with some probability, to create the new problem-solving paradigms that, if successful, initiate new 'technological trajectories', temporal sequences of technological developments within a narrowly defined problem solving paradigm that result in new commercial processes or products. What we are questioning, in essence, is whether the common 'linear hypothesis' of innovation suffices to explain the different roles of public and private agricultural R&D, given instances of public R&D yielding some private-goods-like innovations and private R&D yielding some public-goods-like innovations. A simple comparative advantage argument suggests that, as a result of different organizational endowments and characteristics of the sectors, publicly funded researchers are more likely to specialize in more uncertain exploratory research and privately funded researchers will specialize in more narrowly focused, certain, and appropriable research.

We test several hypotheses implied by this idea using a comprehensive U.S. patent data set on life science inventions with relevance for crop agriculture granted from 1973 to 2001. Detailed data about the technological characteristics of each patent in the dataset are compiled from front page information and used to estimate the structure of the field's evolution by arranging the inventions into a pylogenetic tree using methods borrowed from biological systematics. The tree is used to identify the most significant technological trajectories emerging during these formative years of the agricultural biotechnology industry. Assignee designations on the patents are used to identify the type of organization that generated each invention--whether a government laboratory, a university, a non-profit research organization, an individual inventor, an entrepreneurial startup firm, or an established corporate firm. We also calculate established citations-based indices that proxy for a patent's quality, value, originality, generality, and appropriability (Hall, Jaffe, and Trajtenberg, 2001). Finally, within the technological trajectories framework we develop multinomial regressions that allow for partial correlation analysis capable of testing the hypotheses of sectoral specialization.

Patents are particularly useful for this exercise as they are a common measure across sectors of commercially relevant R&D output in agriculture. Whereas in other industries government and university patenting make an almost insignificant contribution—less than 3 percent on average, according to USPTO summary data—in the field of agricultural biotechnology we have found that government and university R&D contributes upwards of 25 percent of the U.S. patents, meaning that systematic comparison across sectors at an industry level of analysis is possible. The greatest drawback of using patent data is, of course, that not all inventions are patented, and differences in institutional significance of patents result in different propensities to patent across sectors. For example, economically significant inventions made at universities often show up in published research papers, not in patents, while many inventions made within companies are kept secret altogether. It is also important to note that the use of patent data necessarily constrains the investigation to issues of the original inventorship and not the current ownership of the technologies claimed in the patents. This is because a U.S. patent documents list only the names of the organizations to whom the property rights are originally assigned when the patent is granted; neither the patent document nor the patent office keeps a running record of who currently holds title to the property rights. For this reason, the questions investigated in this study concern only the economics of the generation of the new technologies and not their subsequent redistribution.

The results show that the data are consistent, both before and after controlling for the patents' places within specific technological trajectories, with systematic differences in innovation across sectors as predicted by a broad interpretation of the linear hypothesis. At the same time, these preliminary results reemphasize warnings often made in the literature against assuming a simple one-to-one relationship between basic and applied innovations, and provide clues for a more realistic albeit a more nuanced model of the innovation process to aid in considering R&D policies and strategies for organizations in the different sectors.

2. Framework for analysis

2.1. The theory of micro patterns in innovation: technological trajectories

Micro patterns of innovation have long been implicated in empirical studies of the determinants of R&D output (reviewed in Cohen 1995). Efforts to explain field-specific discrepancies or patterns in the rates and characteristics of innovation (Jaffe 1986; Levin, Klevorick, Nelson, Winter, Gilbert, and Griliches 1987; Scherer 1965) have led to the concept of the natural technological trajectory. Rosenberg (1969; 1974) describes innovative efforts as focused on solving a finite set of closely or sequentially related problems which he terms *focusing devices* or *technological imperatives*—bottlenecks, weak spots, and clear targets for improvement—resulting in *compulsive sequences* of innovations over time. Somewhat more focused on final markets, Abernathy and Utterback (1978; Utterback 1979) describe a *technology life cycle* with four phases: (1) the early experimental pre-paradigmatic phase, (2) the emergence of a *dominant design*, (3) the mature phase of refinement in which incremental innovations decrease costs and exploit economies of scale, and finally (4) the phase of decline and obsolescence, until the dominant design is replaced by radically new technologies and a new cycle begins again. Nelson and Winter's (1977) notion of *technological regimes* growing over time along *natural technological trajectories* contains may of the same elements discussed by Rosenberg and others and borrows, in addition, from the notion of R&D as a search mechanism (Evenson and Kislev 1976) and induced innovation theory (Binswanger 1974; Hayami and Ruttan 1985). Concerned with the relative inflexibility built into induced innovation models by deterministic conceptions of decision making in the R&D process, Nelson and Winter seek in their heuristic natural trajectories model to balance the simultaneous influences of *demand pull* and *technology push* in an explanation of the patterns of R&D output. They observe that R&D strategies adjust to the incentives and constraints of changing demand and cost conditions faced by the commercialized outputs of R&D as well as the fact that initiation and success of a given R&D project is a function of the expected time, cost, and feasibility of the project, which in turn depend on the general state of science and the technological knowledge base of the researchers and engineers being employed.

Sahal (1981) and Dosi (1982; 1988) take the concept further and characterize the technological regime as the set of parameters of the meta-production function of Hayami and Ruttan, the set of potentially feasible yet costly technological capabilities traded-off by the technology user under physical or budget constraints. Dosi argues that this is equivalent to ascribing a set of hedonic attributes (Lancaster 1966) to technologies, locating a particular set of coordinates in *technology characteristics space* around which individual innovations cluster to define a technological regime, either in the form of quantitative performance-cost characteristics as emphasized by Sahal or more cognitive or conceptual characteristics emphasized by Nelson and Winter.

These theories suggest that new innovations arise as results from different points along a spectrum of R&D modes. The R&D mode at one end of the spectrum tends to be of a more original and exploratory nature, testing the limits of the possible and probing the frontiers of known technology characteristic space. Most of the outcomes of such original exploratory research are dead-ends. Occasionally, however, one of these exploratory searches may happen upon a particularly promising problem-solving paradigm—in both the sense of creating new technological opportunities and in the sense of showing new ways to meet market demand—and may initiate a move toward the other end of the R&D spectrum by enabling or inspiring follow-on innovations. The original work may, then, in hindsight come to be considered as having been a breakthrough discovery or a radical innovation.

As the new idea and the attendant technical information for a successful problem-solving paradigm diffuse (either directly or indirectly) to other investigators working in the same area, success can beget success. Competitors may notice the threat of a new approach in solving an old problem and attempt with new vigor to build upon or to work around the ideas of the initial innovators. The diffusion of the new paradigm continues with the making of numerous refinements and improvements clustered at those coordinates in hedonic technology characteristic space that were first pinpointed by the original breakthrough.

As this focused cluster of innovations accumulate over time, they form a *technological trajectory* along the time axis at that set of coordinates within technology characteristic space. The generation of successful and prominent technological trajectories continues to be driven both from the innovation supply side, by each new development in the trajectory and its associated cost reductions, and drawn from the demand side, with express customer demand manifest in the adoption both by midstream technology users and by final consumers of the products created with or embodying the new technology.

2.2. Theories of organizational research capabilities

It should be noted that the preceding discussion contains no mention of the organizational or institutional nature of the agents—be they firms, individuals, universities, or governments—expending resources and R&D efforts to generate innovations. Two lines of discussion tend to dominate in the economic literature exploring the differential capabilities of different kinds of organizations at generating innovations, one roughly traced to Schumpeter and the other roughly traced to Nelson and Arrow.

In the Schumpeterian tradition, discussion is largely focused on the private sector and the question of the relative advantages of established firms versus new entrants in innovation. Beginning with Scherer's (1965) treatment, empirical studies have considered the effect of a broad range of firm characteristics in addition to size and market power on innovation—often controlling for field effects usually defined in terms of technological opportunity—relating firm characteristics to (homogeneous) quantities rather than (heterogeneous) qualities of the innovative output. Suggestive exceptions include Henderson (1993), whose framework effectively relates empirically different qualities of innovation—radical versus incremental and technical versus organizational—to characteristics of the firm, demonstrating that larger incumbent firms are more likely to pursue incremental innovation and less likely to pursue (disruptive) radical innovations. Cohen and Klepper (1996) show firm size giving a comparative advantage in exploiting process innovations relative to product innovations. In this Shumpeterian tradition public sector research is typically regarded as merely an exogenous factor creating technological opportunity to be exploited by the private sector agents that populate the models.

The theoretical notions posed by Nelson (1959) and Arrow (1962) concern a different economic question—that of the socially optimal provision of innovation given the uncertainty, inappropriability, and public-goods nature of the knowledge created—and the discussions that follow their lead largely focus on the tradeoffs between public and private sector provision of

R&D. Dasgupta and David (1994) examine how the different institutional structures and social communities of publicly supported *'open science'* and privately financed *'commercial technology'* influence the efficiency and output of these respective R&D systems. However it is Trajtenberg, Henderson, and Jaffe (1992) who propose empirical measures, adapted from bibliometric analysis, able to get at some of the more qualitative economic notions of 'basicness' and appropriability of individual inventions, with which they are able to show systematic differences between the more basic output of university research versus the more applied results of corporate R&D.

These theories suggest that, regardless of the sector in which they work, researchers can be considered to face a universal optimization problem: given the opportunities and incentives posed by their organizational environment as well as the budget and policy constraints they face, to maximize their own individual utility in pursuit of some combination of three fundamental sources of utility: fame, fortune, and freedom (Graff, Heiman, and Zilberman 2002). The specific incentives and constraints of the organizational environment include hiring and promotion practices, publication and peer review, salary and tenure (or seniority) ladders, as well as royalty-sharing and conflict-of-interest policies. Given a choice of sectors for which she could work, a researcher, will self select into an organization with the system of incentives and constraints that she expects will allows her to pursue the kind of research that will maximize her individual utility, taking into account her own skill set and her preferences over the different kinds of incentives offered. Once employed, a researcher makes specific choices of research projects, given that system of incentives and constraints. At the same time, taking into account the fact that faculty or research staff are all the time pursuing their own individual preferences—over a combination of fame, fortune, and freedom—the management of the organization constructs an organizational environment consisting of incentives attractive enough and constraints reasonable enough to engage talented researchers and induce them to be as innovative as possible in those kinds of research outputs that will maximize the benefits to the organization, its shareholders, or it constituents.

2.3. Combining theories of organizational comparative advantage with theories of technological trajectories

The ultimate result is that different types of R&D organizations become endowed with different kinds of research talent and differently optimized strategies for maximizing their payoff. Combined with the heterogeneity of research opportunities, defined by where a given research project arises at a given point in time within the evolving structure of knowledge and technology, it follows that differently endowed sectors will specialize according to their comparative advantage in different types of research.

R&D is assumed to proceed within distinct research paradigms such that the resulting technologies are generated along naturally occurring trajectories, $k = 1 \ldots K$, each of which can be assumed to be captured as a major branch on an estimated phylogenetic tree. Not all technological trajectories are presently at the same point of maturity in their growth or evolution: some constitute new (and thus currently poorly defined) areas of research with little accumulated prior knowledge; others are mature areas with large stocks of existing knowledge already in place.

There are J R&D sectors, $j = 1 \ldots J$, in the economy, distinct from one another in terms of the financing structures, prevailing cultures of open science versus proprietary technological R&D, and corresponding systems of incentives and constraints:

$$j = \begin{cases} 1 \text{ for universities, governments, and non-profit research laboratories,} \\ 2 \text{ for individuals, entrepreneurs, startup firms, and small businesses, and} \\ 3 \text{ for corporations.} \end{cases}$$

The underlying behavioral model of any researcher, in any sector, *j*, for generating research results and ultimately patents consists of several distinct steps.

1. A joint decision is made by a researcher and the research administrator or funding source in sector *j* to undertake a project in a specific research sub-field corresponding to an existing or emerging technological trajectory, *k*. It is presumed, although not observed, that the expected (joint) returns from this research project exceed the expected returns from the next best deployment of the researcher's time and talents and the R&D organization's resources.
2. With a certain probability, a successful research result is produced that meets the standard criteria for patentability of being a novel, non-obvious, and useful.
3. Another joint decision is made by the researcher and their host organization as to whether the (novel, non-obvious, and useful) research result be patented, versus alternative strategies of being kept as a trade secret, being published in the public domain, etc. It is presumed, although again not observable, that the expected (joint) returns to taking a patent on this invention, subject to policy restrictions and transaction cost constraints, are greater than the expected returns to taking a patent on the next best invention. Thus, with a certain probability, or *patenting propensity*, a patent is applied for and granted on the research result.
4. The patent, *n*, the R&D sector, *j,* of the assignee, the technological trajectory, *k,* to which it contributes, and the qualitative attributes, $\mathbf{X}_n$, of the patent are all observed. Some of the X can be observed immediately after the patent issues, others only after some time has elapsed.

Thus, when considering the universe of new research opportunities, the researcher's choice set may include whether to explore uncharted territory or to pursue work within an already established technological trajectory, whether to attempt uncertain original experiments or to make more certain incremental advancements, all the while factoring the probabilities of success and the expected payoffs in terms of fame, fortune, and freedom. It follows that each sector should therefore generate results of correspondingly different qualitative characteristics and perhaps concentrated at different particular phases of technological trajectories.

2.4. Development of hypotheses

Dasgupta and David's two sectors of *open science* and *commercial technology* effectively describe the two most broadly general alignments of incentives and constraints for R&D. However, to effectively summarize the major organizational arrangements most active in biotechnology and to reflect important differences between new entrants and incumbents, commercial technology is further divided into two sectors: *entrepreneurs, new entrants,* or *startup R&D* and *corporate* or *incumbent* R&D. Particularly apropos to biotechnology, these may be thought to reflect the primary mechanisms for financing R&D expenditures. The former raising funds from venture sources versus the later rising capital from public financial markets.

The result is three R&D sectors. First, the public sector consists mostly of universities and some government laboratories, funded largely by public funding sources and characterized organizationally by the incentives, constraints, and culture of open science. Second, the entrepreneurial sector consists of individual inventors and small often privately held biotech firms, funded largely by private venture capital, and characterized organizationally by the incentives, constraints, and culture of commercial technology development. Third, the corporate sector consists of the R&D laboratories of larger publicly traded firms, funded through corporate R&D expenditures, and likewise characterized organizationally by the incentives, constraints, and culture of commercial technology development. A division of innovative labor among these three can be hypothesized in terms of the expected characteristics of their respective contributions to a given technology's evolution.

In the public sector, operating in the mode of open science, the most important criteria by which success is judged and rewarded are (1) the significance of the problem addressed, (2) the originality and creativity of the solution posed, and (3) being the first to make the contribution. This triad of criteria—*significance*, *creativity*, and *priority*—drive university researchers to differentiate and diversify their projects, their careers, and themselves, ultimately creating a broad portfolio of research results that span the conventional continuum from the most purely basic of scientific discoveries to readily applied product and process inventions. Inevitably, however, a system that emphasizes creativity and originality tends to produce more discoveries that are less likely to be of immediate commercial use. In commercial R&D, by contrast, the output must contribute within a fairly tight time horizon to a firm's ability to create value and perpetuate itself by either increasing market share and revenues or reducing costs. Commercial R&D is therefore more directed and sequential, often charted out in routine phases leading up to the introductions of new production processes or the launches of new products, where success is ultimately judged by the market.

Age*:* Priority in discovery is important for researchers working under all three regimes. It is hypothesized that earlier inventions within a given technological trajectory are more likely to arise from researchers in universities and government labs. These researchers have greater incentives to do initial work in exploratory and unestablished areas, given the driving criteria they face for creativity and self-differentiation. Then, the middle phases of a trajectory are more likely to arise from entrepreneurial firms, as the common business model in biotechnology involves startups backed by venture capital to explore technically proven but still uncertain commercial opportunities. Finally, corporations are hypothesized to be more likely to innovate in the later phases of more established trajectories, refining and scaling up technologies for market. Thus, considering age alone, this hypothesis describes the classic *linear model* of R&D. (See this and following hypotheses summarized in Table 1.)

Value*:* The distribution of values of inventions is highly skew (Scherer, Harhoff, and Kukies 2000). The few high value successes tend to be paradigm-setting patents that dominate in large areas of follow-on innovation and product development. Such patents are often the unexpected results of exploratory research or research directed toward other questions. Since the breadth of sampling in open science is greater and because such researchers may be more on the lookout for new ideas and applications, it is expected that the probability of occurrence of top value inventions would be higher in universities. Yet, the stakes and uncertainties of investing in such research and the desire to capture the value of such top value inventions is so great that the entrepreneurial sector may in fact be more likely to actually pursue and patent such inventions. This might be called a *'value filter'* hypothesis: venture capitalists and biotech startups will bet their investments only on the cream of the crop, and the entrepreneurial biotechnology sector may show the most valuable patents, followed by the open science sector. The incentives and dynamics of the corporate R&D sector seem less likely to consistently generate top value inventions.

Generality*:* Technologies characterized by greater generality are those that enable and drive follow-on innovation among a wider diversity of subsequent technological trajectories, affecting a wider diversity of markets. Proposing a measure of the diversity of technology fields from which a patent is cited, Trajtenberg, Henderson, and Jaffe (1992) find randomly selected university patents to be somewhat more general than corporate patents. The potential value of a more general invention is likely greater, but the transaction costs of licensing to other firms or fixed costs of vertically integrating into multiple markets makes that value more difficult to appropriate. Given the diversity of research programs found in open science organizations, their interest in broad social impact of results, and their lower regard for the appropriation of returns, it is hypothesized that organizations in the open science sector should be more likely to generate such measurably general inventions. In addition, many among the current generation of startups in biotechnology are created around expertise in general technology platforms, such as micro-array or genetic sequencing technologies, and they essentially sell the services of that platform. In contrast, corporations, which are more focused on final markets, are hypothesized to be least likely to innovate in general technologies.

***Originality**:* The breadth of prior knowledge on which an invention draws is proposed by Trajtenberg, Henderson, and Jaffe to define its *originality.* Based upon the assumption that new ideas are influenced by existing knowledge, drawing on a broad base versus a narrow base indicates more original or synthetic thinking[1]. Since originality or breadth of thinking is a key criterion of success in open science it is hypothesized that original inventions are more likely to be observed coming from universities and other public sector sources. Since breadth of inspiration is not as important a criterion in the incentives of corporate R&D, original patents are least likely to come from that source. Startups are presumed to be somewhere in between.

***Pace of Innovation**:* The pace of innovation, measured as the average lag between the grant date of the observed patent and the grant date of the patents it cites, will presumably be slower earlier in a technological trajectory, as larger conceptual and technological feasibility issues are being worked out. The pace likely quickens as innovation in that trajectory becomes more routine and as competition intensifies to get products to market. University and government researchers are less likely to be working in an area with short lag times, while startups or corporations will have an advantage in fast paced innovation, although, between these two it is not clear which would have the upper hand in terms of pace.

***Appropriability**:* The appropriability of a technology is the proportion of its value that can be captured as private rents. A common bibliometric proxy for appropriability is 'self citation,' defined as the proportion of citations that were made to patents invented by the same organization

---

[1] Arguably, an invention that draws on no prior knowledge at all is original. The definition based on breadth of influence conforms to the measure that is available to test this hypothesis. See Table 2.

as the citing patent, reflecting the degree to which a patent builds upon the existing technologies already owned by the same organization. This measure of appropriability reflects a technology's dependence upon the internalized transfer of knowledge: a more highly appropriable technology does not transmit as readily through external spillovers. Conversely, the greater the degree of external spillovers from a technology, the smaller the proportion of its value that is left to be appropriated by the inventing organization. The building of protective 'patent fences' by filing extensively around a valuable patent position is a commonly discussed strategy in intellectual asset management. It is thus hypothesized that both types of private sector R&D organization will strongly emphasize innovation with high appropriability, while researchers in open science will be much less concerned with appropriation or building upon their organization's own prior patents.

Table 1. Summary of hypotheses

Characteristic of invention:	Sector of inventor:		
	University/Public	**Entrepreneurial**	**Corporate**
Age or priority of the invention	+++	++	+
Scope of the invention	++	++	+
Value of the invention	++	+++	+
Generality of application	+++	++	+
Originality of idea	+++	++	+
Pace of innovation	+	++	++
Appropriability	+	+++	+++

+++ is most likely, ++ less likely, and + least likely sector to specialize in each characteristic.

3. The data

The data for this study seek to encompass all inventions made in the life sciences over the last 30 years with relevance to crop agriculture, combined from two previously published collections of U.S. utility patents. The first set, assembled in 1999 from MicroPatent data, is 2477 U.S.

patents granted from 1973 to 1998 (De Janvry, Graff, Sadoulet, and Zilberman 1999; Graff, Rausser, and Small 2003). The second data set, assembled in late 2001, consists of 4303 U.S. patents granted between 1982 and 2001 (Graff, Cullen, Bradford, Zilberman, and Bennett 2003). The intersect of the two sets is 1678 patents, yielding a combined collection for this study of 5102 U.S. patents granted between 1973 and 2001. The patents in both of the collections were selected using complex iterative data base searches over patent classification numbers, technology terms, and patent citation links, and both were thoroughly screened by experts in biotechnology, in an effort to include all patents with subject matter pertinent to the industry but to exclude any patents with non-pertinent subject matter.

Table 2. Independent variable definitions and sources

Variable	Definition	Indicates
Age	Calculated from the application date of the patent	***Priority*** of the invention.
Number of citations made	Count of how many prior patents the observed patent cites as relevant *prior art* in its References Cited section.	***Quality*** of the invention (Lanjouw and Schankerman, 1999), or ***scope*** of the invention (various authors).
Originality index	Ranges between 0 if the prior art patents cited by the observed patent are concentrated in a single technology class and 1 if the cited patents are spread out.	***Originality*** of the invention: the ***breadth of knowledge drawn upon*** by the invention (Trajtenberg et al, 1992, 1997; Hall et al, 2001)
Average backward citation lag	The average age of the prior art patents cited by the observed patent, at the time it was granted.	The ***pace or rate of innovation*** in the area of the invention (Trajtenberg et al, 1992, 1997; Hall et al, 2001)
(within sample) Self-citation ratio	The percentage of cited prior art patents that are assigned to the same assignee (aggregated under parent organization) as the observed patent.	***Appropriability*** of the invention: transmission of the knowledge is easier internally than via external spillovers (Trajtenberg et al, 1992; Hall et al, 2001)
Number of citations received	Count of how many subsequently granted patents had cited the observed patent as relevant prior art, by the end of 2002.	***Value, importance*** (Trajtenberg, 1990; Hall et al, 2000; Hall et al, 2001), or ***quality*** (Lanjouw and Schankerman, 1999) of the invention.
Generality index	Ranges between 0 if the subsequent patents citing the observed patent are concentrated in a single technology class and 1 if they are spread out across separate classes.	***Generality*** of the invention: the ***breadth of technological impact*** resulting from the invention (Trajtenberg et al, 1992, 1997; Hall et al, 2001)

3.1. Patent quality indicators: independent variables

Not all inventions are created equal. Bibliometric indicators introduced in Trajtenberg (1990) and in Trajtenberg, Henderson, and Jaffe (1992) and several other studies have been linked to important aspects of economic heterogeneity in the technologies underlying patents (Hall, Trajtenberg, and Jaffe 2000; Harhoff, Narin, Scherer, and Vopel 1999; Lanjouw and Schankerman 1999). The characteristics driving the hypotheses developed in the preceding section and listed in Table 1 have all been developed in this literature. Table 2 lists the set of indicators employed in this study, defines each briefly, and describes the economic qualities of the invention measured. Most are adapted directly from the NBER Patent Citations Data File, and detailed definitions are available in the reference paper that accompanies the data file (Hall, Jaffe, and Trajtenberg 2001). Summary statistics are provided in Table 3.

Table 3. Summary statistics

Variable	type	Obs.	Mean	Std. Dev.	Min.	Max.
Age	years	5102	6.22	4.35	1.32	27.98
Number of citations made	count	5102	3.25	6.33	0	125
Originality index	(0,1)	5102	0.15	0.25	0	1
Average backward citation lag	years	3511	7.47	5.62	0	102
(within sample) Self-citation ratio	ratio (0,1)	5102	0.17	0.35	0	1
Number of citations received	count	5102	3.97	10.89	0	180
Generality index	(0,1)	5102	0.14	0.24	0	1

3.2. Determination of inventing sector: dependent variables

The question of *what sector generated each new invention* was determined for each patent in the collection by examining the organization to which it was assigned when issued by the patent office. The names of such 'assignees-at-issue'[2] are recorded in the patent data. However, several issues complicate the usefulness of the names thus obtained. First, something as simple as the consistent identification of an individual organization is complicated in a data set of 5,103 documents by the fact that an assignee may be listed under different names or under different spellings (and misspellings) of those names on different patents. Second, different business units or subsidiaries of a single larger organization may each receive patents in the name of the business unit or subsidiary rather than in the name of the larger organization. And, third, a small fraction, about 6 percent, of the patents have more than one assignee, and some of those involve collaboration across different sectors.

The approach taken to solve the first complication was to clean the names of the assignees, uniformly giving all patents for each assignee the single most-common spelling. In response to the second challenge, all documents assigned to a smaller entity known to have been majority owned by a larger entity at the time the patent was filed were reassigned in the data set to the parent entity. In response to the third complication, co-assigned patents were simply attributed to the first assignee listed on the patent, since priority listing often indicates the lead institution in a collaborative relationship.

Each assignee was then identified as a university, government agency, non-profit organization, individual, small entrepreneurial firm, or large corporate firm. The most difficult differentiation was between the last two, both because some medium sized firms defy easy classification as either 'entrepreneurial' or 'corporate', and because several of the most active small biotech firms in the industry were acquired by the large corporations in the industry during the timeframe considered. Since the fundamental research question seeks to relate organizational

---

[2] Inventors are the original owners of intellectual property rights, but they usually assign the rights to their employer. In the case that a patent's inventor is independent there may not be any assignee, and the patent simply remains the property of the independent inventor. However, only 1.7 percent of the documents in this data set went to independent inventors. (See Table 4.)

comparative advantages to innovative outcomes, the rules of thumb used to determine between these two categories considered issues of size, age, the nature of ownership and financing (privately held versus publicly listed), and the publicly projected culture of the firm. In the cases of acquired firms, patents assigned in the name of a small firm were tabulated as 'entrepreneurial' if filed before the date that the firm was acquired by its corporate parent. Applications made after that date were then considered 'corporate'.

Table 4. Three R&D sectors generating inventions in the agricultural life sciences

Type of R&D Organization	Patent Count	Percent of Total
University / Public Sector		
Universities	957	18.8%
Government laboratories, agencies	291	5.7%
Non-profit research centers, foundations	37	0.7%
Total Public Sector	***1,285***	***25.2%***
Entrepreneurial Sector		
Individual inventors	89	1.7%
Entrepreneurial firms: biotech startups, small private firms	916	18.0%
Total Entrepreneurial Sector	***1,005***	***19.7%***
Corporate Sector		
Corporate firms: large, diversified, publicly listed	2,812	55.1%
Total Corporate Sector	***2,812***	***55.1%***
Grand Total:	**5,102**	**100.0%**

3.3. Determination of technological trajectories by phylogenetic analysis

A technological trajectory may be considered, in its essence, as an evolutionary lineage of conceptual constructs. This paper approaches the problem of empirically identifying the main technological trajectories in an industry-defined collection of patents by borrowing from the methodologies of phylogenetic systematics (Hennig, 1966; Swofford, Olsen, Waddell, and Hillis

1996) to infer the evolutionary relationships among individual observations, based upon similarities and differences in shared characteristics and under explicit assumptions about the inheritance and evolutionary change of those characteristics. The result of a phylogenetic analysis is a network or tree describing a hypothesis of the historical order and pattern by which the 'observed taxanomic units' arose from common ancestors. Generally two taxa are placed more closely together in a phylogeny if they have more similar heritable anatomical, physiological, or genetic characteristics. Analogously two patents, treated as observed taxa, may be inferred to be evolutionarily related (in an economic sense) given similarity in conceptual or technological characteristics[3]—indicating perhaps that one built sequentially upon ideas contained in the other or that both arose simultaneously in response to similar demand conditions, technological opportunities and spillovers. In short, they may be placed within a common technological trajectory. Phylogenetic analogies have already been applied to the conceptual development of science (Hull 1988; Mishler 1990), and similar clustering and mapping approaches, albeit lacking explicit evolutionary assumptions, have been applied to scientific documents, including patents (Callon, Law, and Rip 1986; Noyons, Moed, and Luwel 1999).

The conceptual and technological characteristics of the patents on which this preliminary phylogenetic analysis was based included (1) whether a patent fit any of 136 possible categories of technological application determined and coded by an expert in the field[4], (2) whether a patent was given any of 162 relevant International Patent Classifications (IPCs) indices, and (3) whether a patent cited any of 667 commonly-cited older patents in the collection (akin to the bibliometric method of tracking 'co-citations': Small and Griffith 1974; Small 1973; Zitt and Bassecoulard

---

[3] An analogy between a 'species' and a 'patented technology' holds, as both are defined by a discrete and novel step setting them apart from their closest neighbors.

[4] This expert analysis was carried out on just the earlier, 1973-1998 data set. Data for technological application characteristics were coded as "missing" for patents unique to the later 1982-2001 data set.

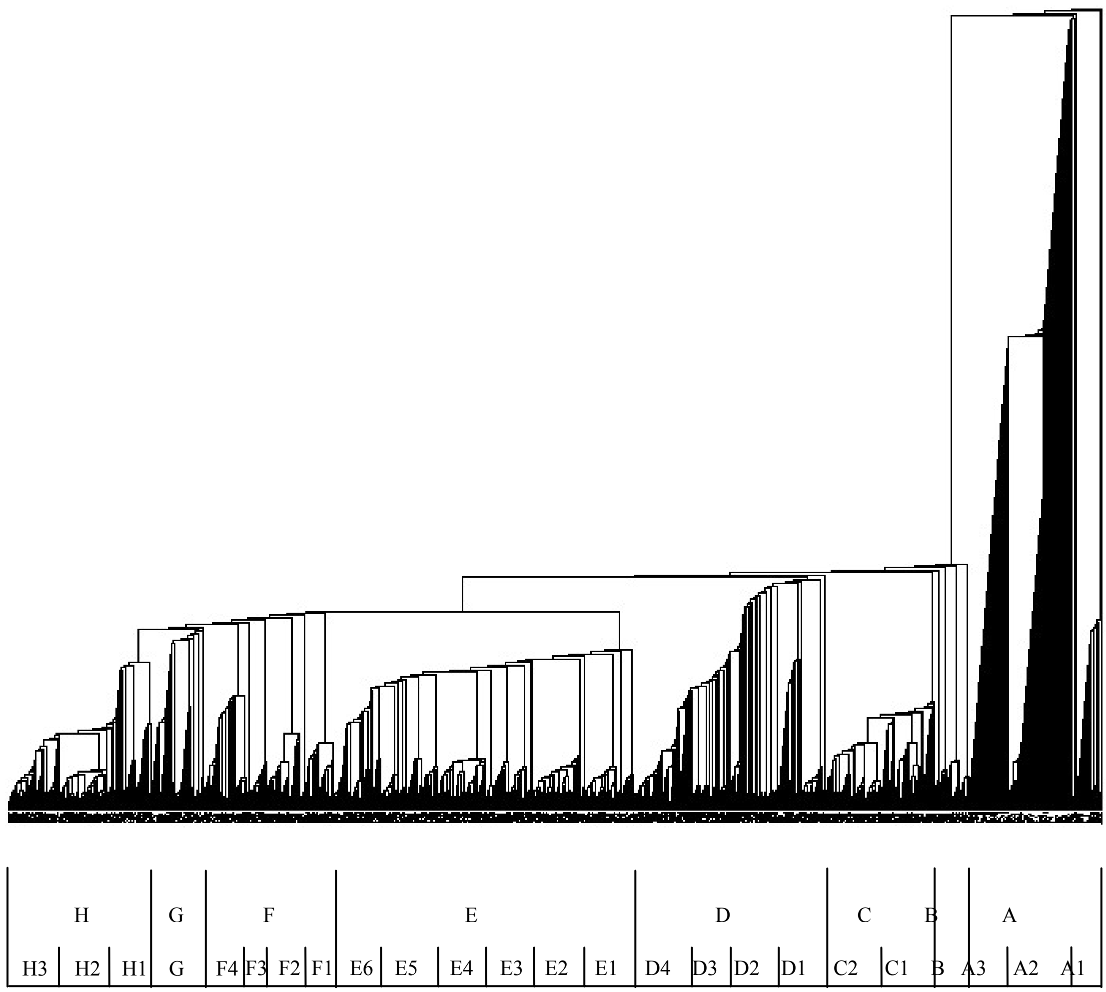
1996). In each case a character was coded as ‘present’ (=1), ‘absent’ (=0) or ‘missing’ (=?), resulting in a character matrix of 5102 patent ‘taxa’ described by 875 characters.

Model choice for such a large data set is limited. Global optimization methods, based on maximum parsimony and maximum likelihood optimality criteria, are feasible only for small data sets given their computational intensity. Heuristic searches for local parsimony or likelihood optima are feasible on larger data sets. However, based on congruence of initial results with previous clustering analysis of this patent collection (see Graff 2003), an additive tree building algorithm called the ‘neighbor-joining method’ (Saitou and Nei 1987) was chosen for this preliminary exploration into applications of phylogenetic methods.

The neighbor-joining algorithm begins by calculating an (N x N) matrix of pairwise distances[5] among the N taxa from the character matrix. It then normalizes the distance matrix and constructs the first branch of the hypothesized evolutionary tree by joining the closest pair of taxa in the normalized distance matrix. It then removes the joined pair of neighbors from the sample, recalculates a new (N-1 x N-1) distance matrix, normalizes it, and constructs the next branch of the tree by joining the closest pair of neighbors in the new normalized distance matrix. Each subsequent cycle generates an additional internal node supporting another branch, resulting in a final tree constructed out of shortest normalized branch lengths.

The phylogenetic tree estimated for this patent dataset is displayed in Figure 1. The patents are listed across the bottom of the figure. (Because of reduction in size they are not legible in this reproduction of the tree.) The tree emanates from a putatively most ancestral patent, which roots the tree on the right hand side. The vertical lengths of branches indicate the computed distance between the characteristics of the patents in one branch and the characteristics of the patents in the closest neighboring branches.

Figure 1. Hypothesized evolution of patented inventions in the agricultural life sciences, calculated from descriptive characteristics of 5103 U.S. patents using a neighbor-joining phylogenetic algorithm in PAUP* 4.O (Swofford 2002).



The primary branches and sub-branches represent significant lines of technological development in crop genetics and serve our purpose of identifying separate technological trajectories. The most 'ancestral' branch—designated 'A' in the phylogram—consists, not surprisingly, of crop variety germplasm breeding lines. A large leap is then made to the rest of the tree, which consists of a variety of genetic traits and process technologies. For example, the

---

[5] Pairwise distance may be defined alternately as least squares, weighted least squares, or "minimum evolution".

trajectory of genetically engineered herbicide resistance technologies is contained in sub-branch 'D1'. However, insect control technologies based on Bt (*Bacillus thuringiensis*) genes and proteins are scattered across four branches: two that represent the development of spray-on Bt biopesticides (C2 and E1) and two that represent the development of transgenic Bt crops (F3 and H2). Interestingly, branch E2, occurring relatively late in the evolutionary hierarchy, is the branch that contains most of the genetic transformation process technologies. This is a logical result of the algorithm since it does not take any time variable into account: these molecular biology patents contain little reference to specific crop varieties, making them distantly removed from the ancestral 'outgroup' to which the breeding lines we closely related. This illustrates some of the limitations of the neighbor-joining algorithm, and further work to refine this empirical evolutionary model of technological development will help to reduce this and other such inconsistencies.

4. The econometric model

Econometric models of discrete random outcomes, such as multinomial probit and logit analysis, have been adapted and employed by economists to estimate latent variable models of choice behavior (McFadden 1974; Ruud 2000) in which each outcome is interpreted as the choice of an individual economic agent whose unobserved or 'latent' utility, construed as a random variable, is assumed to have been maximized by the observed choice, also a random variable, made relative to all other available options. McCullagh and Nelder (1983) argue that the statistical model employed to analyze joint sample distributions of polytomous data and the underlying behavioral model used to describe the unobserved latent variable are, however, indeed separate models, and in most cases the latent variable, while useful for the internal consistency of the behavioral hypothesis, is often unverifiable in practice. Given the data limitations and the behavioral complexity of the innovation phenomena addressed in this study, it is not possible to identify a single, behaviorally meaningful latent variable. Instead we simplify the complex decisions effected by the many unobservable parameters and latent behavioral variables at play in the data generating process into a single 'black box' probability index that relates the qualitative characteristics of a patent with the probability that it is observed to arise from research conducted in a particular sector of the economy.

In essence, this exercise is the same as the classical statistical problem of drawing a randomly distributed sample, pulling n colored chips from j barrels. For each observed patent, n, in each technological trajectory, k, the probability index that the technology is found to be invented and patent by the jth organizational type is denoted by

$$\mathbf{y}^*_{nj} = \mathbf{X}_n\mathbf{B}_j + \boldsymbol{\varepsilon}_{nj},$$

where $\mathbf{X}_n$ is a vector of attributes of the n^{th} patent and the $\mathbf{B}$ are unknown coefficients. The ε_{nj} are the unobserved differences in the probability of that patent arising in the j^{th} type of R&D organization, resulting from unobserved features of the behavioral model including the institutional features of the organization, and are assumed to be i.i.d. random variables with a Weibull probability distribution.

When the j^{th} organizational type actually undertakes the research and receives the n^{th} patent, the observed outcome is described with the J dummy variables where

$$y_{nj} = \begin{cases} 1 \text{ if the } n^{\text{th}} \text{ patent is issued to the } j^{\text{th}} \text{ organizational type} \\ 0 \text{ otherwise.} \end{cases}$$

From the probability index equation the probability of the n^{th} patent coming from the j^{th} organizational type is

$$\begin{aligned} P_{nj} &= \Pr[y_{nj} = 1 \mid \mathbf{X}_n] \\ &= \Pr[y^*_{ni} \leq y^*_{nj},\ \forall i \neq j \mid \mathbf{X}_n] \\ &= \Pr[\varepsilon_{ni} - \varepsilon_{nj} \leq (\mathbf{X}_{nj} - \mathbf{X}_{ni})`\mathbf{B},\ \forall i \neq j \mid \mathbf{X}_n] \end{aligned}$$

Which is equivalent, given the assumptions made about the distribution of the εs, to

$$\mathrm{P}_{nj} \;=\; \mathrm{P}(\mathrm{y}_{nj} = 1) \;=\; \frac{e^{X_n B_j}}{\sum_{i=1}^{J} e^{X_n B_i}}$$

This can be normalized and written as the multinomial logit:

$$\mathrm{P}_{nj} \;=\; \frac{e^{X_{nj} B}}{1 + \sum_{i=1}^{J-1} e^{X_{ni} B}}, \qquad \forall i \neq j$$

where the values of the J different 'P's are conditional probabilities of a patent's occurrence in the J different sectors given the independent variables describing the patent's attributes.

Because they do not enter the probabilities linearly, the coefficients on these patent attributes, the magnitude of the **B**s, cannot be interpreted directly. However, an interpretation is possible from the definition

$$\ln\left(\frac{\Pr_{nj}}{\Pr_{no}}\right) = X_n' B_{jo} \text{ where } o \neq j = 1, 2, 3$$

or more conveniently for interpretation

$$\frac{\Pr_{nj}}{\Pr_{no}} = \exp\left(X_n' B_{jo}\right)$$

which is the probability ratio (also known as the relative risk ratio) of a given type of patent arising from a research organization of type *j* relative to a research organization of type *o*. The *q* parameters in the vector **B** are the marginal effects of the *q*th regressor in X_n on the probability ratio. Finally, since the multinomial logit system is solved by maximum likelihood, testing hypotheses about coefficients follows standard methods based on the covariance matrix from the maximum likelihood estimation.

5. Results and Analysis

5.1. Evidence of R&D specialization in agricultural biotechnology

Multinomial regressions on the entire data set explore the general significance of the patent indicators as predictors of R&D sector of invention and test the hypotheses of technological evolution at the industry level (Table 5.) This, in essence, treats all patented innovation in agricultural biotechnology and crop science according to our framework as a single technological trajectory.

By the construction of the multinomial logit, one of the possible values of the explained variable is treated as the baseline comparison group. We designated the corporate sector, which consists of 55 percent of the patents, as the comparison group. Seperate equations are then estimated together to compare the likelihood of observing each of the other possible values of the explained variable (in this case 'university/public sector' and 'entrepreneurial sector') relative to observing the comparison group, given the values of the explanatory variables. Thus, the sign of a coefficient obtained in the university/public sector equation indicates whether patents strong in the corresponding independent variable are more or less likely to be from the university or public sector, relative to the corporate sector, controlling for all of the other explanatory patent quality variables. Similarly, the sign of a coefficient obtained in the entrepreneurial sector equation only explains the relative likelihood of a patent having come from a new entrant relative to an established corporate firm. The likelihood of university versus entrepreneurial invention of patents with a particular quality can be imputed by comparing the coefficients of the corresponding variable in both equations, since they are both computed relative to the corporate sector comparison group.

The four different equation specifications alternately include and exclude a variable on citation lag and controls for technological trajectories respectively. The variable of average backward citation lags is missing for all patents that do not cite prior art, thus decreasing the number of observations entered and weakening the estimations. Dummies for the technology trajectories are adapted from the structure of sub-branches estimated in the phylogenetic analysis in Section 3.3. The explanatory power of these full industry regressions is limited, given the cross-sectoral and dynamic complexity summarized in each. Even so, several significant results are obtained.

Coefficients on 'citations made' are negative and significant in all four specifications of the university equation. Thus, the less a patent cites prior art, the more likely it is to be from a university than from a corporation, rejecting our hypothesis of broader patent scope from universities. In addition, citations are only slightly more likely from biotech entrepreneurs than from corporations, also challenging our hypothesis. However, significant positive coefficients on 'originality' in all specifications of the university equation indicate the more broadly a patent cites, the more likely it is to be from a university than a corporation. Biotech entrepreneurs are also, although to a lesser degree, a more likely source of original patents than are corporations. These two results together support our hypotheses of greatest originality from universities, followed by entrepreneurs, and then corporations.

Coefficients on the self-citation ratio in the university equations are significantly negative. Inventions that cite—and thus presumably build upon and further develop—patents from the same organization, are not as likely to be found amongst universities or government labs. Considering the size and significance of corporate portfolios this result may be considered somewhat surprising: it is much easier to cite and build on one of your own patents when there are so many more of them on average. The insignificance of the result thus concurs with our hypothesis: appropriability will not, all else being equal, differentiate patents from the entrepreneurial and corporate sectors.

We have attempted to separate the hypothesis of appropriability from the hypothesis of value creation. The significant positive coefficient on 'citations received' in the entrepreneurial sector equation supports what we call the value filter hypothesis. And, accordingly, the number of citations received is not a significant predictor of university patents relative to corporate patents.

Not only do more highly cited patents come more often from the entrepreneurial sector, but more widely cited or 'general' patents are also found to be much more likely from the entrepreneurial sector than the corporate sector. Greater generality is also a significant qualitative indicator of university patents relative to corporate patents, although the effect is erased when technology controls are included in the equation. Contrary to our stated hypothesis, the difference in generality between entrepreneurial and corporate appears to be greater than it does between university and corporate: we had expected greater generality in the university equation.

Results on the timing of innovation are less clear in the full industry equations. A greater average backward citation lag appears to be more likely among patents in the entrepreneurial sector than the corporate sector. In other words the pace of innovation is slower in entrepreneurial patents than in corporate patents. The more interesting result is that a greater backward citation lag makes it less likely that the patent is from a university, implying that the pace of innovation is faster in university patents than in corporate patents. This may be an artifact of the fact that university patents tend to cite less on average: they also tend to cite more recent work on average.

Finally, we do not expect when considering the full industry to explain the sector of invention from the age of a patent. Interestingly, however, the age coefficient is significant in the first entrepreneurial sector equation, an artifact perhaps of the fact that most of the prolific biotech startups were acquired by the mid to late 1990s (see Graff, Rausser, & Small, 2003) and recent patenting in the industry has seen much less input from such firms.

Table 5. Patent qualities related to sector of invention: Full industry mulinomial logit regression results

Sector 3- Corporate sector is the comparison group — Regression coefficients displayed; Standard errors in parenthesis

	(1)		(2)		(3)		(4)	
	university	**entrep.**	**university**	**entrep.**	**university**	**entrep.**	**university**	**entrep.**
Age	0.012	0.031	0.016	0.017	0.011	0.015	0.018	0.000
	(0.010)	(0.010)**	(0.012)	(0.012)	(0.011)	(0.011)	(0.013)	(0.013)
Citations made	-0.054	-0.012	-0.050	-0.016	-0.049	-0.007	-0.044	-0.008
	(0.010)**	(0.007)	(0.010)**	(0.007)*	(0.011)**	(0.007)	(0.011)**	(0.008)
no citations made	0.017	0.065	0.239	-0.201	0.196	0.133	0.135	-0.385
	(0.088)	(0.101)	(1.105)	(1.110)	(0.093)*	(0.105)	(1.121)	(1.128)
Originality index	1.166	0.974	1.171	1.000	0.771	0.470	0.782	0.490
	(0.180)**	(0.183)**	(0.181)**	(0.185)**	(0.185)**	(0.192)*	(0.187)**	(0.196)*
Self citation ratio	-1.200	0.018	-1.235	0.093	-0.977	0.126	-0.992	0.213
	(0.131)**	(0.112)	(0.134)**	(0.115)	(0.137)**	(0.120)	(0.141)**	(0.126)
Citations received	-0.003	0.006	-0.004	0.009	0.001	0.010	0.000	0.013
	(0.004)	(0.003)	(0.005)	(0.004)*	(0.004)	(0.003)**	(0.005)	(0.004)**
no citations received	-0.108	-0.332	-0.093	-0.428	-0.017	-0.289	-0.096	-0.430
	(0.087)	(0.099)**	(0.108)	(0.119)**	(0.090)	(0.102)**	(0.112)	(0.123)**
Generality index	0.653	1.226	0.587	1.096	0.250	0.818	0.262	0.660
	(0.184)**	(0.184)**	(0.217)**	(0.212)**	(0.192)	(0.193)**	(0.226)	(0.224)**
Backward citation lag			-0.016	0.020			-0.008	0.026
			(0.009)*	(0.007)**			(0.009)	(0.008)**
Sub-branch technology dummy variables					sub-branch tech coeff's	sub-branch tech coeff's	sub-branch tech coeff's	sub-branch tech coeff's
Constant	-0.743	-1.456	-0.659	-1.497				
	(0.110)**	(0.120)**	(0.135)**	(0.141)**				
Observations	5102		3511		5102		3511	
Pseudo R2	0.04		0.05		0.18		0.18	

* is significant at 5%, ** is significant at 1%

Table 6. Germplasm patent qualities related to sector of invention: Trajectory specific mulinomial logit regression results

Sector 3- Corporate sector is the comparison group — Regression coefficients displayed; Standard errors in parenthesis

	A1		A2		B		G		H1	
	soy varieties		**maize parental lines/ hybrid maize**		**disease resistant hybrid maize**		**modified amino acid/ protein content maize varieties**		**modified fatty acid/ oil content soy varieties**	
	university	**entrep.**	**university**	**entrep.**	**university**	**entrep.**	**university**	**entrep.**	**university**	**entrep.**
Age	-0.004	0.173	-0.667	-0.333	0.021	0.066	0.238	0.358	0.025	0.084
	(0.198)	(0.081)*	(0.254)**	(0.183)	(0.066)	(0.049)	(0.106)*	(0.115)**	(0.090)	(0.092)
Citations made	0.411	0.341	-0.038	-0.001	0.050	0.211	-0.260	0.025	-0.451	0.059
	(0.262)	(0.233)	(0.058)	(0.161)	(0.066)	(0.090)*	(0.233)	(0 .076)	(0.192)**	(0.098)
no citations made	-1.366	-1.033	-0.885	-0.865	0.218	0.776	-0.113	-0.390	0.722	0.353
	(1.320)	(0.955)	(0.647)	(0.641)	(0.716)	(0.622)	(0.697)	(0.862)	(0.673)	(0.802)
Originality index	1.307		2.162	-0.459	3.475	-3.629	0.235	1.546	4.730	0.354
	(4.124)		(1.702)	(2.502)	(1.630)*	(2.054)	(1.781)	(1.578)	(1.316)**	(1.716)
Self citation ratio	-3.193	-2.081	-1.251	-1.961	-5.968	-0.017	0.460	0.097	-0.239	-1.398
	(1.601)*	(1.029)*	(0.994)	(1.129)	(3.532)*	(0.945)	(0.927)	(1.059)	(0.606)	(1.007)
Citations received	-18.741	-0.333	-0.081	-21.927	-0.124	-0.011	-0.387	0.072	0.006	0.011
	(1.864)**	(0.251)	(0.211)	(1.428)**	(0.129)	(0.011)	(0.224)*	(0.097)	(0.017)	(0.014)
no citations received	-18.723	-0.200	-0.909	-22.923	-1.089	-1.170	-0.339	1.612	0.367	0.219
	(1.587)**	(0.735)	(0.651)	(1.702)**	(0.705)	(0.628)	(0.657)	(0.942)	(0.561)	(0.674)
Generality index		-0.944	2.928		-1.672	0.043	3.165	2.012	1.691	0.879
		(3.349)	(2.444)		(1.519)	(1.127)	(1.782)*	(1.649)	(1.130)	(1.455)
Constant	17.389	-1.101	1.370	22.365	-1.148	-1.807	-1.673	-5.252	-1.643	-2.803
	(0.000)**	(1.296)	(1.265)	(1.973)**	(0.801)	(0.739)*	(0.979)*	(1.267)**	(0.776)*	(0.755)**
Observations	132		342		134		152		200	
Pseudo R2	0.2163		0.1458		0.1823		0.2365		0.1244	

* is significant at 5%, ** is significant at 1%

Table 7: Enabling technology patent qualities related to sector of invention:
Trajectory specific mulinomial logit regression results

Sector 3- Corporate sector is the comparison group
Regression coefficients displayed; Standard errors in parenthesis

	E2		**F1**		**F4**	
	plant genetic transformation tools		**cell/tissue culture and plant fertility**		**cell/tissue culture**	
	university	**entrep.**	**university**	**entrep.**	**university**	**entrep.**
Age	-0.003	0.069	-0.041	0.028	0.051	0.050
	(0.045)	(0.050)	(0.053)	(0.041)	(0.055)	(0.048)
Citations made	-0.016	-0.054	-0.072	0.101	-0.140	-0.263
	(0.072)	(0.112)	(0.110)	(0.048)*	(0.058)**	(0.080)**
no citations made	0.202	-0.503	-0.094	0.352	0.644	-0.132
	(0.379)	(0.466)	(0.750)	(0.649)	(0.547)	(0.505)
Originality index	-0.324	-1.316	-0.476	-0.945	2.727	2.159
	(1.016)	(1.319)	(1.601)	(1.201)	(1.127)**	(0.995)*
Self citation ratio	-1.051	-0.616	-3.669	-2.213	-1.514	-0.627
	(0.717)	(0.745)	(2.158)*	(1.201)	(1.093)	(0.631)
Citations received	0.012	0.015	0.023	-0.009	-0.076	-0.001
	(0.014)	(0.015)	(0.012)*	(0.017)	(0.043)*	(0.024)
no citations received	-0.166	-0.236	-1.039	-0.733	0.238	0.768
	(0.388)	(0.492)	(0.721)	(0.622)	(0.605)	(0.515)
Generality index	-0.577	-0.848	0.055	0.234	2.246	1.411
	(0.760)	(0.897)	(1.322)	(1.085)	(1.093)*	(1.013)
Constant	-0.379	-1.054	-0.248	-1.014	-1.707	-1.026
	(0.492)	(0.604)	(0.747)	(0.653)	(0.698)**	(0.606)
Observations	235		138		216	
Pseudo R2	0.0305		0.1583		0.1695	

* is significant at 5%, ** is significant at 1%

Table 8. Plant genetic trait patent qualities related to sector of invention: Trajectory specific mulinomial logit regression results

Sector 3- Corporate sector is the comparison group — Regression coefficients displayed; Standard errors in parenthesis

	D1		C2		E1		F3		H2	
	engineered herbicide tolerance genes and traits		**biopesticides and biocontrol (including Bt)**		**engineered insect resistance (including Bt) and product quality genes and traits**		**engineered Bt insect resistance genes and traits**		**engineered Bt insect resistance genes and traits**	
	university	**entrep.**	**university**	**entrep.**	**university**	**entrep.**	**university**	**entrep.**	**university**	**entrep.**
Age	0.076	0.047	0.155	0.054	0.033	-0.091	0.134	0.044	0.020	0.007
	(0.061)	(0.082)	(0.051)**	(0.052)	(0.033)	(0.047)	(0.093)	(0.075)	(0.043)	(0.048)
Citations made	-0.061	-0.207	-0.051	-0.005	0.162	0.204	0.085	-0.037	-0.065	0.025
	(0.027)*	(0.083)*	(0.045)	(0.037)	(0.059)**	(0.059)**	(0.143)	(0.095)	(0.088)	(0.087)
no citations made	-0.322	-0.819	0.640	1.101	0.579	0.367	1.026	0.655	-0.031	0.377
	(0.596)	(0.809)	(0.569)	(0.540)*	(0.471)	(0.605)	(0.852)	(0.095)	(0.439)	(0.526)
Originality index	0.392	-0.665	1.246	0.509	-1.151	-1.104	-1.032	1.478	0.341	1.580
	(0.726)	(1.139)	(0.754)*	(0.745)	(0.909)	(0.978)	(2.199)	(1.296)	(0.938)	(0.935)
Self citation ratio	-1.409	-0.410	0.951	2.104	-0.406	1.979	-0.158	0.976	-0.178	1.469
	(0.624)*	(0.687)	(0.588)	(0.534)**	(0.652)	(0.572)**	(1.369)	(0.704)	(0.673)	(0.573)**
Citations received	0.006	0.028	-0.024	0.006	-0.035	-0.005	-1.130	-0.014	-0.010	0.015
	(0.015)	(0.016)	(0.027)	(0.022)	(0.027)	(0.018)	(0.744)	(0.017)	(0.019)	(0.017)
no citations received	-0.487	0.297	0.264	-0.091	0.643	-1.160	-2.695	-1.591	0.331	-0.276
	(0.464)	(0.701)	(0.508)	(0.550)	(0.460)	(0.572)*	(1.219)*	(0.598)**	(0.462)	(0.542)
Generality index	0.633	-0.113	-1.838	-0.241	1.877	3.289	2.678	0.452	1.072	1.012
	(0.827)	(1.278)	(0.797)*	(0.776)	(0.916)*	(0.924)**	(2.614)	(1.089)	(0.755)	(0.795)
Constant	-0.999	-1.364	-0.735	-0.603	-1.387	-1.254	0.195	-0.059	-0.659	-1.551
	(0.611)	(0.867)	(0.611)	(0.619)	(0.534)**	(0.611)*	(1.421)	(0.805)	(0.621)	(0.696)*
Observations	244		278		230		104		221	
Pseudo R2	0.1577		0.0910		0.1978		0.1783		0.0593	

* is significant at 5%, ** is significant at 1%

5.2. Patterns of R&D specialization in the technological trajectories

A series of multinomial regressions on thirteen seperate technological trajectories identified from the phylogenetic tree are intended to test hypotheses of sectoral specialization within each technology's own evolution. The thirteen are divided into germplasm (in Table 6), enabling technologies (in Table 7), and gene/trait technologies in (Table 8). By reducing sample size and by controlling explicitly for technological homology (and thus reducing technological heterogeneity within these sub-samples), the explanatory power of the individual estimation procedures is improved: pseudo R-squares are in some cases several fold greater than in the industry wide regressions.

Additional complications are also introduced, deriving both from smaller sample sizes and inconsistencies carrying over from the phylogenetic estimation of technological trajectories. Specifically, issues arise in trying to capture trajectories of technologies that are so mature that they have only observed the later phases, dominated by the corporate sector, or technologies of such ubiquitous general-purpose nature that they are diffused across the various branches of the phylogenetic tree. Examples of mature technologies are soy and maize germplasm (A1 and A2), where involvement by university or entrepreneurial inventors is limited to just a few dozen patents compared to the corporate sector's hundreds of patents. As a result, the ability of the multinomial procedure to compare between sectors becomes compromised. An example of a general purpose technology is plant genetic transformation tools (E2), where no significant differences result from the multinomial estimation, possibly because this branch of the tree was a catch all for nondescript incremental modifications to transformation techniques, while the most significant patents in this general purpose technology clustered within the branches of application that they enabled.

Some results obtained for the full industry in the previous section are recapitulated here at the level of individual technological trajectories. Greater originality is a significant indicator of university inventions in several technologies, including disease resistant maize (B), improved oil soybeans (H1), cell and tissue cultures (F4), and Bt biopesticides (C2). Self citations ratio, proxying for appropriability, is significantly smaller for university inventions in several technologies, including soy varieties (A1), disease resistant maize (B), cell culture and plant fertility (F1), and herbicide tolerance genetic traits (D1). Greater generality is associated with university patents in several, including high protein maize (G), cell and tissue culture (F4), and engineered Bt traits (E1). University/public sector patents receive significantly fewer citations, indicating lower average value of inventions, in several technologies including soy varieties (A1), high protein maize (G), cell and tissue culture (F4), yet received significantly more in the cell tissue and plant fertility trajectory (F1).

Unique results are also found looking within the technological trajectories. The time-dependent linear hypothesis is meaningfully tested in this context. In the university/public sector equation the coefficient point estimate on age is positive in nine out of thirteen trajectories, it is not significantly different from zero. However, in only two of these technologies does older age significantly identify university/public sector patenting. These are high protein maize (G) and Bt biopesticides (C2). Point estimates indicate patents from the entrepreneurial sector are at least as old as corporate patents in eleven technologies, although significantly so only in two as well, those being soy varieties (A2) and high protein maize (G). In three of the four Bt trajectories high self citation ratios are more likely in entrepreneurial biotech patents than corporate patents, interesting rejection of our hypothesis that corporate patents would be more likely to cite their own previous work.

6. Conclusions

So, have the roles of public and private sector agricultural scientists become indistinguishable in the era of biotechnology? Indeed they have not. Systematic differences are found in the attributes of the patents filed by university and public sector inventors, biotech entrepreneurs, and corporations. And those differences do indicate specialization in the qualitative nature of knowledge production under the different systems and policies of each sector. Four specific conclusions can be advanced based on the results.

The first conclusion is one of weak support for the simple age-defined linear hypothesis. As defined, this hypothesis can only be tested within the trajectory regressions. However in only two trajectories There is strong indication that older Bt patents are more likely to be from public sector sources and weak indication that older herbicide resistance patents are more likely to be from entrepreneurial sources. In most cases, however, patent age does not appear consistent with the scenario of first public sector invention and then entrepreneurial invention before corporate invention. More tests on more technological trajectories are needed. Until then, acceptance or rejection of this formulation of the linear hypothesis cannot be definitive.

The second is the acceptance of the ‘value filter’ hypothesis. The entrepreneurial sector is clearly the most likely source of high value inventions (as well as the least likely source of low value inventions), both within the specific trajectories and in the industry at large. While the public sector is less likely than the corporate sector to produce low value patents, it is indistinguishable from the corporate sector in high value patents. Biotech companies and their venture capital financiers appear to have filtered out the highest value talent and succeeded in creating more of the leading technologies in agriculture.

The third conclusion is the originality of publicly funded R&D. The results on the originality index for the public sector are the most significant and persistent throughout the study. The case for originality is further strengthened by the fact that the number of citations made is a negative predictor of public sector patents. Since the originality index is constructed from citations made, it has a slight positive correlation with that variable: if more citations are made, there are more of them to be spread out over more technology classes. As a result, the already high originality related with public sector patents is probably biased downward.

The fourth conclusion is the acceptance of the appropriability hypothesis, but with some qualification. University, government, and non-profit organizations are, as hypothesized, significantly predicted by low self-citation ratios and thus by low-appropriability technologies. The twist comes in the result that startups are a much more likely source than corporations of high self-citation ratios and, by correlation, of high appropriability technologies. Moreover, corporations have on average much larger internal portfolios on which to build and from which to draw self-citations. Thus, we would expect self-citations to be biased higher for corporations. This resonates with the value filter hypothesis: biotech startups and entrepreneurs are looking for technologies that are not only more valuable but technologies upon which they are able to build and from which they are able to appropriate the value created.

These results, however preliminary, show a world of commercially useful agricultural R&D in which public sector researchers generate the most original and often most general work, and do it a bit earlier. Entrepreneurs make their entry in the private sector around higher value tool or trait technologies that promise to be highly appropriable, and they build upon them. Corporations undertake the most innovation, in terms of generating sheer numbers of patents, but those patents tends to be less original or general, and in some cases of more moderate appropriability. While this is not the picture of the conventional linear paradigm, it does support the notion of a productive division of innovative labor within the complex and multidimensional nature of real R&D processes.

References

Abernathy, William J. and James M. Utterback. 1978. "Patterns of Industrial Innovation." *Technology Review*, pp. 41-47.

Alston, Julian M., George W. Norton, and Philip G. Pardey. 1995. *Science under Scarcity: Principles and Practice for Agricultural Research Evaluation and Priority Setting*. Ithaca, NY: Cornell University Press.

Arrow, Kenneth J. 1962. "Economic Welfare and the Allocation of Resources for Invention." in *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton University Press: National Bureau of Economic Research.

Binswanger, Hans. 1974. "A Microeconomic Approach to Induced Innovation." *Economic Journal*, 84, pp. 940-58.

Bush, Vannevar. 1945. "Science, the Endless Frontier." United States Government Printing Office: Washington D.C.

Callon, M, J. Law, and A. Rip. 1986. *Mapping the Dynamics of Science and Technology*. London: MacMillan Press.

Cohen, Wesley M. 1995. "Empirical Studies of Innovative Activity," in *Handbook of the Economics of Innovation and Technological Change*. Paul Stoneman ed. Cambridge: Blackwell, pp. 182-264.

Cohen, Wesley M. and Steven Klepper. 1996. "Firm Size and the Nature of Innovation within Industries: The Case of Process and Product R&D." *Review of Economics & Statistics*, 78:2, pp. 232-43.

Dasgupta, Partha and Paul A. David. 1994. "Toward a New Economics of Science." *Research Policy*, 23:5, pp. 487-521.

De Janvry, Alain, Gregory Graff, Elizabeth Sadoulet, and David Zilberman. 1999. "Agriculture Biotechnology and Poverty: Can the Potential Be Made a Reality?" *3rd annual International Consortium on Agricultural Biotechnology Research (ICABR) conference "The Shape of the Coming Agricultural Biotechnology Transformation: Strategic Investment and Policy Approaches from an Economic Perspective"*: University of Rome at Tor Vergata.

Dosi, Giovanni. 1982. "Technological Paradigms and Technological Trajectories: A Suggested Interpretation of the Determinants and Directions of Technical Change." *Research Policy*, 11:3, pp. 147-62.

Dosi, Giovanni. 1988. "Sources, Procedures, and Microeconomic Effects of Innovation." *Journal of Economic Literature*, 26:3, pp. 1120-71.

Evenson, Robert E. and Y. Kislev. 1976. "A Stochastic Model of Applied Research." *Journal of Political Economy*, 84:2, pp. 265-81.

Graff, Gregory D., Amir Heiman, and David Zilberman. 2002. "University Research and Offices of Technology Transfer." *California Management Review*, 45:1, pp. 88-115.

Graff, Gregory D. 2003. "Observing Technological Trajectories in Patent Data: Empirical Methods to Study the Emergence and Growth of New Technologies." *American Journal of Agricultural Economics*, 85:5, pp. 1266-74.

Graff, Gregory D., Susan E. Cullen, Kent J. Bradford, David Zilberman, and Alan B. Bennett. 2003. "The Public-Private Structure of Intellectual Property Ownership in Agricultural Biotechnology." *Nature Biotechnology*, 21:9, pp. 989-95.

Graff, Gregory D., Gordon C. Rausser, and Arthur A. Small. 2003. "Agricultural Biotechnology's Complementary Intellectual Assets." *Review of Economics & Statistics*, 85:2, pp. 349-63.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg. 2001. "The NBER Patent Citations Data File: Lessons, Insights, and Methodological Tools." NBER Working Paper No. 8498. National Bureau of Economic Research: Cambridge.

Hall, Bronwyn H., Manuel Trajtenberg, and Adam B. Jaffe. 2000. "Market Value and Patent Citations : A First Look." NBER Working Paper No. 7741. National Bureau of Economic Research: Cambridge, MA.

Harhoff, Dietmar, Francis Narin, F. M. Scherer, and Katrin Vopel. 1999. "Citation Frequency and the Value of Patented Inventions." *Review of Economics & Statistics*, 81:3, pp. 511-15.

Hayami, Yujiro, and Vernon Ruttan. 1985. *Agricultural Development: An International Perspective*. Baltimore: Johns Hopkins University Press.

Henderson, Rebecca. 1993. "Underinvestment and Incompetence as Responses to Radical Innovation: Evidence from the Photolithographic Alignment Equipment Industry." *RAND Journal of Economics*, 24:2, pp. 248-70.

Hennig, Willi. 1966. *Phylogenetic Systematics*. Urbana, IL: University of Illinois Press.

Hull, D.L. 1988. *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*. Chicago: University of Chicago Press.

Jaffe, Adam B. 1986. "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value." *American Economic Review*, 76:5, pp. 984-1001.

Kealey, T. 1996. *The Economic Laws of Basic Research*. London: Macmillan Publishers.

Lancaster, Kelvin J. 1966. "A New Approach to Consumer Theory." *Journal of Political Economy*, 74:2, pp. 132-57.

Lanjouw, Jean O., and Mark Schankerman. 1999. "The Quality of Ideas: Measuring Innovation with Multiple Indicators." NBER Working Paper No. 7345. National Bureau of Economic Research: Cambridge, MA.

Levin, Richard C., Alvin K. Klevorick, Richard R. Nelson, Sidney G. Winter, Richard Gilbert, and Zvi Griliches. 1987. "Appropriating the Returns from Industrial Research and Development; Comments and Discussion." *Brookings Papers on Economic Activity*:3, pp. 783-831.

McCullagh, P., and J.A. Nelder. 1983. *Generalized Linear Models*. New York: Chapman and Hall.

McFadden, Daniel. 1974. "The Measurement of Urban Travel Demand." *Journal of Public Economics*, 3:4, pp. 303-28.

Mishler, Brent D. 1990. "Phylogenetic Analogies in the Conceptual Development of Science." *Proceedings of the Philosophy of Science Association*, 2, pp. 225-35.

Mowery, David C. 1995. "The Practice of Technology Policy," in *Handbook of the Economics of Innovation and Technological Change*. Paul Stoneman ed. Oxford: Blackwell Publishers.

National Science Board. 2003. "Science and Engineering Indicators." National Science Foundation: Arlington, VA.

Nelson, Richard R., and United States Air Force. 1959. *The Economics of Parallel R&D Efforts: A Sequential-Decision Analysis*. Santa Monica, Calif.: Rand Corporation.

Nelson, Richard R., and Sidney G. Winter. 1977. "In Search of a Useful Theory of Innovation." *Research Policy*, 6, pp. 36-76.

Noyons, E. C. M., H. F. Moed, and M. Luwel. 1999. "Combining Mapping and Citation Analysis for Evaluative Bibliometric Purposes: A Bibliometric Study." *J. Am. Soc. Inf. Sci.* (J. Am. Soc. Inf. Sci.), 50:2, pp. 115-31.

Rausser, Gordon. 2001. "Public Research/ Private Alignments," in *Knowledge Generation and Technical Change: Institutional Innovation in Agriculture*. Steven Wolf and David Zilberman eds. Boston: Kluwer Academic Publishers, pp. 55-62.

Rosenberg, Nathan. 1969. "The Direction of Technological Change: Inducement Mechanisms and Focusing Devices." *Economic Development and Cultural Change*, 18, pp. 1-24.

Rosenberg, Nathan. 1974. "Science, Invention, and Economic Growth." *Economic Journal*, 84, pp. 90-108.

Rosenberg, Nathan, and Richard R. Nelson. 1994. "American Universities and Technical Advance in Industry." *Research Policy*, 23:3, pp. 323-48.
Ruud, Paul. 2000. *An Introduction to Classical Econometric Theory*. New York: Oxford University Press.
Sahal, Devendra. 1981. *Patterns of Technological Innovation*. Reading, MA: Addison-Wesley.
Saitou, N., and M. Nei. 1987. "The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees." *Molecular Biology and Evolution*, 4, pp. 406-25.
Scherer, F. M. 1965. "Firm Size, Market Structure, Opportunity, and the Output of Patented Inventions." *American Economic Review*, 55, pp. 1097-125.
Scherer, F. M., Dietmar Harhoff, and Jorg Kukies. 2000. "Uncertainty and the Size Distribution of Rewards from Innovation." *Journal of Evolutionary Economics*, 10, pp. 175-200.
Small, H., and B. C. Griffith. 1974. "Structure of Scientific Literatures .1. Identifying and Graphing Specialties." *Science Studies*, 4:1, pp. 17-40.
Small, H.G. 1973. "Co-Citation in the Scientific Literature: A New Measure of the Relationship between Publications." *J. Am. Soc. Inf. Sci.*, 24, pp. 265-69.
Stokes, Donald E. 1997. *Pasteur's Quadrant: Basic Science and Technological Innovation*. Washington D.C.: Brookings Institution Press.
Sunding, David, and David Zilberman. 2001. "The Agricultural Innovation Process: Understanding Research and Technology Adoption in a Changing Agricultural Sector," in *Handbook of Agricultural Economics*. Bruce L. Gardner and Gordon C. Rausser eds: Elsevier Science, pp. 207-61.
Swofford, David L. 2002. "PAUP* 4.0: Phylogenetic Analysis Using Parsimony (and Other Methods)." Sunderland, MA: Sinauer Associates.
Swofford, David L., Gary J. Olsen, Peter J. Waddell, and David M. Hillis. 1996. "Phylogenetic Inference," in *Molecular Systematics*. David M. Hillis, Craig Moritz and Barbara K. Mable eds. Sunderland, MA: Sinauer Associates, pp. 407-514.
Trajtenberg, Manuel. 1990. "A Penny for Your Quotes: Patent Citations and the Value of Innovations." *RAND Journal of Economics*, 21:1, pp. 172-87.
Trajtenberg, Manuel, Rebecca Henderson, and Adam Jaffe. 1992. "Ivory Tower Versus Corporate Lab: An Empirical Study of Basic Research and Appropriability." NBER Working Paper No. 4146. National Bureau of Economic Research: Cambridge, MA.
Utterback, James M. 1979. "The Dynamics of Product and Process Innovation in Industry," in *Technological Innovation for a Dynamic Economy*. Christopher T. Hill and James M. Utterback eds. New York: Pergamon Press.
Williams, Roger L. 1991. *The Origins of Federal Support for Higher Education*, University Park, PA: The Pennsylvania State University Press.
Wright, Brian D. 1998. "Public Germplasm Development at a Crossroads: Biotechnology and Intellectual Property." *California Agriculture*, 52:6, pp. 8-13.
Wright, Brian D. 2001. "Challenges for Public Agricultural Research and Extension in a World of Proprietary Science and Technology," in *Knowledge Generation and Technical Change: Institutional Innovation in Agriculture*. Steven Wolf and David Zilberman Eds. Boston: Kluwer Academic Publishers, pp. 63-78.
Zitt, M., and E. Bassecoulard. 1996. "Reassessment of Co-Citation Methods for Science Indicators: Effect of Methods Improving Recall Rates." *Scientometrics*, 37:2, pp. 223-44.