Industry Collaboration and the Discipline of Academic Science:

The Case of *Arabidopsis* Research, 1974-2003

James A. Evans

The University of Chicago

January 2005

Prepared for the 2005 Meetings of the

American Sociological Association

As intellectual property protection has expanded and the distinction between basic life science research and applied biotechnology erodes, academic labs are increasingly collaborating with firms on research. Using citation analysis, content analysis, network analysis and interviews, I examine how industry collaboration shapes the practice of academic research. Focusing on the case of plant biotechnology, I map the relationships among scientists and the public, private, and nonprofit organizations involved in research on the genetic model organism *Arabidopsis thaliana*. My findings suggest that industrial partnerships make science less novel and influence scientists to be less persistent in their inquiry. These partnerships exert their effects unequally, however, by improving the academic quality of research performed by central, high status scientists and universities while degrading those qualities for peripheral ones. By engaging central scientists on the basis of their academic quality and peripheral scientists on the basis of their time and effort, industry creates a striking, new division of scientific labor.

Introduction

National systems of innovation involve three primary institutional actors: universities, government agencies and private companies. The past thirty years represent a shift in the division of labor between these actors, especially in the life sciences and related industries. Greater collaboration between universities and businesses has raised questions about the unique mission and identity of each (Press and Washburn 2000). Agricultural, chemical, and pharmaceutical industries have long built on basic science discoveries made in universities, but historically the downstream development of these ideas did not engage university researchers. More recently, however, a stream of fundamental discoveries in the fields of biochemistry, molecular biology, and genetics have proven relevant to both basic questions of bioscience and also applied problems of medicine (e.g., breast cancer gene) and agriculture (e.g., plant steroids).

As a result of this convergence, academic scientists have founded many of the new life science firms, and most firms maintain ongoing engagement with academic science and scientists (Powell 1996). Research contracts between commercial firms and university departments, which were once unusual (Kenney 1986: chapter 3), have become frequent and more consequential in terms of real investment and duration (Merrill and Cooper 1999; Mowery 1999). Consequently, scientists at firms, government agencies, and universities frequently converge around similar research questions using the same scientific materials, techniques, and analyses.

At the same time, intellectual property (IP) protection has increased. Traditionally, patents were unavailable for many discoveries in basic science. Currently it is possible to patent the discovery of biological organisms and basic scientific “research tools” (Heller and Eisenberg 1998). These changes have enticed industry to collaborate with academic and government organizations. As scientific careers have become more mobile across public, private and nonprofit sectors in several areas of research (National Research Council 1997), cross-sector collaborations have become more likely and more successful.

Industry-Academy Ties

This surge in academic relationships with industry has sparked a tinderbox of opinion about the fate of universities, public science and technology. Much academic response has been polarized, weighing in favor or opposition to such relationships. Both the positive and negative perspectives on industrial collaboration, however, assume that any particular working relationship between firms and universities will have little questionable impact on the content of science for opposing reasons. Enthusiasts for academic-industrial collaboration, including scholars of technology transfer (Reimers 1984; Deurtozus et al 1989), regional innovation (Rogers and Larsen 1984; Saxenian 1985, 1994; Lee et al 2000), and economic history (Rosenberg and Nelson 1994; Rosenberg 2000), emphasize the importance of such ties to technology transfer into the economy. They assume, however, that the goals of academic science remain distinct from industrial science. This, in turn, presumes that scientists are so socialized to a culture of public science as that proposed by Robert Merton that it will continue robust in the face of industrial collaborations and incentives (Merton 1942; 1968). Critics of collaboration including the "new" economists of science (Dasgupta and David 1994; Nelson 2003) and scholars interested in the "triple helix" model of tightly bound government, university and industry interests (Etzkowitz 1983; Etzkowitz and Leydesdorff 1995; Nowotny, Scott and Gibbons 1994, 2001), emphasize the frailty of open science and the pervasiveness of industrial influence. To these economists of science, the reliance of some academic scientists on industrial funds will alter the strategic behavior of all academics. To Etzkowitz, Nowotny and others, norms from industrial science suffuse the new culture of science to the point that they guide academic behavior independent of particular collaborations with industry: An individual partnership is a symptom, but not a cause of the triple helix.

I argue that each of these positions is untenable. The technology transfer position assumes scientific culture is too independent from industrial concerns. The new economics of science and triple helix positions assume scientific culture is too fused with them. The technology transfer and triple helix positions assume that scientists are over-socialized to the culture of science, but they disagree about the nature of that culture. The new economics stance assumes that scientists are uniformly under-socialized to the culture of open science.[1] In contrast, I argue that industry has and does influence academic science through partnerships between academy and industry. Although commercial norms increasingly color the consciousness of academic science, their incidence is uneven and corresponds with industry relationships.

[1] The new economics and triple helix assumption of commercial dominance make sense for situations like U.S. Cold War spending on physics and engineering when the U.S. Department of Defense sponsored over 60% of the total research in these fields and physicists and historians have argued that it shaped much of what was studied (Lowen 1997). The position is less germane for contemporary academic-industrial ties. NIH-funded surveys of academic life science labs indicate that in both 1984 and 1994 the percentage of funds from industry averaged about 12% (Blumenthal 1986, 1996).

Through partnerships with universities, industry exerts influence on academic science in a number of ways: Industry may offer a commercial logic for academic science to mimic; it may exert normative pressure over how academics should alter their practice in a collaborative project; and in the extreme it may attempt coercion (DiMaggio and Powell 1983). While partnerships rarely bring all of these pressures to bear on academic science, different types of affiliation will likely differ in their influence on science. In research on commercialization in the life sciences, studies note the distinction between partnerships that involve collaborative research relationships versus those that resemble market transactions (Powell 1990). I expect that relational ties, involving cooperation between industry and academic scientists, will expose researchers to the industrial approach and engage them in bonds of trust and obligation with industrial partners whereas transactional relations may exert a wider variety of outcomes. In the worst case, however, industry transactions will burden academic researchers with the explicit knowledge of their partner's expectations, and pressure them to deliver a research product that meets those expectations.

Industry will also likely differ in their influence on academic science if they ally with scientists or universities. Companies may partner directly with academic scientists, funding their labs or specific projects within them. Companies may also enter into commitments with departments, institutes or universities which involve many scientists and are typically longer in duration and larger in scope. Industry ties between companies and academic scientists have the potential to directly shape academic science. Such ties between departments and companies may also involve collaboration between some university and industry scientists, but the distribution of those collaborations is almost always uneven across the department, and may only involve the transaction of intellectual property once it has left academic labs. As a result, although university or department level ties may have long-term effects on the culture of research, individual ties between companies and scientists portend a greater immediate impact.

Academy-industry alliances also differ in the status which their affiliated scientists or universities maintain within the scientific community. Central scientists and universities have more funding opportunities and a better bargaining position relative to their industrial partners than scientists and universities more peripheral in the academic status system. As a result, central scientists will likely experience less pressure to conform to the industrial norms of their partners. Finally, partnerships differ in the extent to which the science of affiliates is performed in areas with more or less commercial promise. Scientists already performing research in topics relevant to industry will likely experience less influence to change their mode of research. These lines of reasoning lead me to the following hypotheses:[2]

Hypothesis A1: Partnerships between academic scientists and companies will have a larger, negative impact on the academic quality of their science than partnerships between universities and companies, ceteris paribus.

Hypothesis A2: If academic scientists or institutions have high-status within the scientific community, then partnerships with industry will have a less negative effect on the academic quality of their science than if they have low-status within the scientific community, ceteris paribus.

Hypothesis A3: If academic scientists or institutions work in less applied subfields, then collaborations with industry will have a more negative effect on the academic quality of their science than if they work in more applied subfields, ceteris paribus.

With this theory of industry influence over academic science, we now need a theory of difference between science in these two institutional contexts.

Differences Between Academic and Industrial Science.

Many characterizations have been made of academic and industrial science in isolation from one another, but their distinguishing qualities reveal themselves more clearly through comparison. Most early comparisons of the two, however, suggest that the overriding difference is quality: industrial scientists are failed academic scientists (Marcson 1960; Kornhauser 1962). In recent years, with the boom of high technology industries and the slowdown in government funding for science, industrial science has become a more desirable career option and academic science less so. At points in the last decade, Genentech, a biotechnology company, has claimed the honor of having the most academic citations per paper for any department of any organization, university or otherwise. Thus, any contemporary distinction between the academic and industrial science must appeal to differences in the logic between them, and not to the generalized quality. The literature on science, organizations and innovation suggests two broad characterizations of the comparative activities of universities

[2] A more thorough discussion of each of these hypotheses is available in the extended version of this paper.

and businesses, and two corresponding characterizations of the organization of those activities. The most frequently cited difference between the activities of academy and industry is the contrast between the academic focus on developing theories compared with an industry focus on creating products (Sorenson and Fleming 2001). These activities are often organized differently with theory-building coordinated through the public sharing of discoveries while product building is managed privately. In this paper, I will develop and evaluate a second, complementary contrast between the activity of academic exploration versus industry exploitation, and the degree to which these activities are integrated through persistent or efficient (but scattered) research efforts.

Exploration or Exploitation. James March's notion of exploration versus exploitation (1991) suggests the first difference between the activities of universities and firms. University research often *explores* novel scientific possibilities, while firm research seeks to *exploit* existing scientific knowledge for commercial applications. University exploration of science leads to high variance in the quality of its scientific output—seeking breakthroughs that often fail. Occasionally, however, such risky efforts succeed dramatically, changing the basic paradigms of science (Kuhn 1962). This is why university budgets are organized as cost centers—to support scientists that do not, on average, generate useful innovations.

In contrast, company exploitations of science in a quest for new technologies seek high average quality in their efforts to maintain viability and steadily increase profits. Figure 1 illustrates the difference between the university and company approaches to variance and mean quality. The red line corresponding to company quality illustrates how company efforts may, on average, succeed more than university efforts, but by involving less risk of failure, they also minimize the risk of dramatic success. Einstein's employer, the Swiss patent office in Bern where Einstein worked between 1902 and 1909, probably cared about his daily, average ideas, but the academic world only cares about his best ones.

In the literature on technical innovation, invention is frequently characterized as the process of combining existing components into new technologies or processes (von Hippel 1988; Fleming and Sorenson 2001a). Systems of intellectual property, beginning with Britain's patent system in 1640, deem technologies worthy of patent protection based on the novelty of their combinations of known, or "prior art" components. Scientific innovation can be understood in a similar way, as the process of combining old ideas, methods and results into novel experiments and theories (David, Foray and Steinmueller 1998). In this combinatorial sense, universities specialize in high risk combinations of many components, while firms exploit proven recipes with incremental amendment.

Persistent or Efficient. Universities typically organize their explorations by *not* organizing them. Academic scientists self-organize as each independently pursues lines of inquiry that seem most fruitful. Moreover, scientists' research records—their vitas—undergo peer review as they seek tenure at universities and grants from government. This encourages academic scientists to persist in their chosen areas of research to build name recognition as they seek new insight.

Industrial science, on the other hand, coordinates its science centrally and often hierarchically within companies. This allows companies to hedge their bets against bad ideas, and quickly shift from one project to another when it becomes clear that development will take longer or be riskier than the company can afford. Investor-based economies exaggerate this quick shifting of projects as publicly traded companies experience intense quarterly scrutiny from analysts and shareholders. Monsanto, a chemical and life-sciences company, is infamous among scientists for reorganizing its departments and abandoning half of its projects—and people—every five years (Charles 2001). This results in fragmented but efficient scientific careers which quickly harvest commercial value from science or move on.

In contrast, Harvard University's Judah Folkman discovered the role of angiogenesis in cancer development—the process by which cancer tumors recruit new blood vessels—through his own experiences with tumor removal surgery in the early 1960's. He persisted for nearly 30 years—until 1989—before widespread skepticism gave way to agreement that this process played a major role in the growth of tumors (Cooke 2001). His persistence has resulted in one of the most promising classes of anti-tumor treatments—efforts a company would never have supported.[3] Certainly one can persist with a problem for too long and the British sociologist Harry Collins has shown how the academic realm facilitates this luxury (2000). In one paper, Collins illustrates how decades after the physics community had rejected high visibility gravitational radiation, its persistent academic priests continued to research, publish and proselytize, though forced into more and more peripheral publishing venues.

[3] Genentech, a large biotechnology company, eventually contributed to Folkman's research in the late 1980's after the scientific community came to agree on the criticality of the role of angiogenesis in cancer development.

While novelty and persistence characterize academic science, they are themselves opposing forces in either's extreme. A researcher's work could be so incessantly original that it persists in nothing but novelty. Alternately, a perfectly persistent researcher is novel, at most, once. My characterization of academic science involves a delicate balance between persistence in a stream of research, while combining ideas, materials and methods in novel ways within that stream. As such, it captures that quality of academic inquiry best termed "depth."

Social studies of science and of business contain subtle characterizations of academic and industrial science which may appear, on the surface, to conspire against my broad contrast. Kuhn's philosophy of science, for example, portrays most university science as "normal" rather than "revolutionary" in character (1962): scientists build onto accepted theory in increments more frequently than they challenge it with a daring new one. The presence of "invisible colleges" (Crane 1969) and model organism communities facilitate this conservatism by organizing agreement around platforms of existing theory and method.

On the side of industrial science, James March developed the concepts of exploration and exploitation to characterize companies in different phases of a company lifecycle (1991). In periods of excess resources, companies explore and innovate. As resources dwindle, companies hunker down to exploit existing knowledge. Some small, science-based companies take exploration to seeming extremes by continuously producing novel combinations in search of products. Diversa, a genomics company in San Diego, routinely gathers microbes from brackish water around the world, "shuffles" and multiplies variations of their genes, then screens the dizzying number of gene products for pharmaceutical, agricultural, chemical and industrial use.

Moreover, the image of a hierarchical chain of corporate command seems passé in the context of modern high technology industries. Life science companies in recent years have encouraged their scientists to publish independent work and engineering companies like 3M and HP are known for sponsoring "skunkworks," or unmandated, curiosity-driven research initiated by their employees. In short, academic science often appears more conservative and industrial science more novel—academic activity more organized and industrial activity less—than my simple dichotomy suggests. These observations, however, consider the academy and industry independently. Bringing them together reinforces my distinction.

Though Kuhn classified science as only rarely “revolutionary” and more often “normal,” the two concepts anchor ends of an empirical continuum. Several scholars have criticized Kuhn’s work for not crediting radical experimental innovation with anything more than “normal” status (Nowotny, Scott and Gibbons 2003). Like layers of an onion, scientific theories reside within theories, and an inquiry may challenge a great many layers without challenging them all. In this way, academic science can be more or less revolutionary even if it shares a theoretical base (Ben-David 1973). Moreover, the set of components involved in novel academic research overlaps those involved in industrial development, but often extends beyond them. For example, the process of creating new theory in the molecular biology of photosynthesis involves the persistent recombination of ideas but also usually involves new methods (e.g., cDNA micro-arrays) , new biological materials (e.g., new mutant plants with parts of the photosynthetic apparatus inoperative), and new findings. With so many variables, academic scientists must think carefully about the limited sample of combinations they can actually attempt.

Industry typically produces scientific combinations in a narrower range. Companies like Diversa take a single scientific frame or research tool, like the technology of high-throughput molecular screens, and then exhaustively sample the entire population of possibilities. Even though the percentage of Diversa’s molecules with industrial promise will always be low, the yield is reasonably assured. Diversa’s process recalls that of Edison, the founder of the first industrial laboratory in Menlo Park, New Jersey, when he sought to improve the filament for his new electric light. Edison and his associates traveled the world, gathering tens of thousands of mineral and plant materials to lengthen its burn-time (Josephson 1959). Such combinations are novel, but narrow. “Revolutionary” engineering, in this sense, often remains scientifically more conservative than “normal science.”

On the organization of academic science, invisible colleges organize and solidify agreement in the scientific community, but scientists self-select into these colleges. Moreover, academics often retain membership in many subfields at once, giving their contributions an even greater independence. One of the largest policy criticisms for government science funding—the massive duplication of effort caused by scientific races (Dasgupta and Maskin 1987; Dasgupta and David 1994)—highlights how academics use their independence to persist in focal areas. Companies, while they may allow their scientists some freedom to explore, almost always coordinate the majority of their scientific effort. There are exceptions, like Xerox's Palo Alto Research Center (PARC), which allowed its researchers almost free rein to work on curiosity-inspired projects in the 1970's and 80's. But this exception proves the rule: Xerox quibbled for years before spinning off PARC in January, 2004 for not transferring enough product leads to its parent. Most companies seek efficiency in the allocation of intellectual resources, and they frequently redeploy their scientists to achieve it.

Differences between academic and industrial science in novelty and persistence provide clear opportunities for collaboration. Companies can lower their costs of discovery by harvesting the few successful ideas produced by universities. With these ideas in hand, companies can justify the lower-risk efforts required to develop them into products. Universities can profit from the sale of scientific consultation and patent licenses. University science may also benefit, both from the scientific materials often provided by firms, but also from the results of company-funded research which have exploited a narrow set of experiments or scientific frames.

If partnerships influence academic scientists and administrators to drift toward their industrial partners, however, universities may engage in scientific projects which are less risky or novel. Academics may stray from their own research commitments and into a set of projects integrated by the companies that support them. Furthermore, the company-sponsored research that feeds back to university scientists may speed the path to normal science (1962) by validating early, suboptimal theories before a broader range of perspectives is explored. This leads me to anticipate the following:

Hypothesis B1: If academic scientists or institutions collaborate with industry, then their subsequent research will be less novel than if they had not collaborated with industry, ceteris paribus.

Hypothesis B2: If academic scientists or institutions collaborate with industry, then they will become less persistent in their research than if they had not collaborated with industry, ceteris paribus.

With a theory of institutional difference and influence, I am prepared to evaluate these hypotheses and extend them in an empirical context. I do so in the setting of all the research and researchers engaged in science with the model organism *Arabidopsis thaliana*.

Arabidopsis thaliana

Arabidopsis was the first higher plant with a sequenced genome and has become the dominant genetic model organism in plant biology and agricultural biotechnology (Walbot 2000), just as the mouse and *Drosophila* (fly) serve as animal model organisms. Most major universities, many research institutes and government agencies, and all major plant biotech companies perform or fund research on *Arabidopsis*. *Arabidopsis*'s short, haploid genome and abundant seed production make it easy to simultaneously study molecular and classical genetics—linking DNA sequences to plant functions. Fundamental discoveries in *Arabidopsis* have fed back into basic science, revealing previously unknown similarities between plant and animal function. This research is increasingly being transferred into crop plants with social and profitable implications, such as drought resistance crops, and the plant-manufacture of oils and other pharmaceutical and industrial substances. As such, studying *Arabidopsis* research constitutes more than a single case. *Arabidopsis* provides an elegant platform from which to study organizational influences on the creation of knowledge and technological innovation in the world of plant biotechnology.

Additional reasons makes *Arabidopsis* research an ideal setting in which to examine the effect of business collaboration on the novelty and persistence of academic science. First, scientific use of *Arabidopsis*, and plant molecular biology in general, are little more than three decades old and have grown up within the new system of intimate connection between firms and universities. As such, this setting provides an opportunity to examine the consequences of commercial influence within this system over time rather than forcing me to focus on the evolution of that system. Second, *Arabidopsis* biology is a subfield of plant science, but because it is used as a model for economically and socially meaningful organisms like crop plants and humans, the *Arabidopsis* community is one step removed from implementation. As a result, concern within the *Arabidopsis* community is likely skewed toward the most fundamental issues of plant science and biology. Moreover, in interviews, scientists repeatedly noted the openness and sharing in the *Arabidopsis* community relative to the subfields of their colleagues. Because of its relatively small size compared with other model organism communities, many *Arabidopsis* researchers know one another and can easily sanction each other for unscientific behavior. This encourages competitive researchers to share. These factors make *Arabidopsis* science a conservative test of corporate influence and capture. If we observe corporate influence in *Arabidopsis*, we can be confident it is amplified in other fields.

Observations and interviews with *Arabidopsis* scientists and industry collaborators underlined the contrast between exploration in academic settings and exploitation in industrial ones. In the course of a conversation with the head of a mid-sized biotechnology company's agricultural division, I asked why he had only been talking about the control of single genes in plants and not the more complex traits that new technologies were beginning to isolate. The veteran entrepreneur and scientist, involved with technology transfer on three continents for over thirty years gasped audibly. "The answer to that is—uh—I think it's about the human brain and about the ability for us to actually conceive of complex traits," then he chuckled at my naïveté and continued:

> Notwithstanding the phenomenal advances in gene-chip technology that enable us to look at patterns of expression and change, as far as I'm concerned, that's totally irrelevant from the standpoint of developing new products in agriculture in my lifetime. I mean, it's just—it doesn't even count!… The concept of genomics in agriculture, that you alter a plant's performance by changing expression of its own genes, is an incredibly pretentious notion—I mean, a wonderful one—but to alter the performance of an organism in a way that is noticeable to a farmer, that's a hell of a challenge. I only mention that because my objective, from the standpoint of running a business, is to develop products in my own lifetime and so I have a very simplistic view of what transgenic biology can do, that is you can take protein molecules that kill stuff and you can make 'em better. And you kill stuff better than other people can kill stuff. And that's sort of—that's what agricultural chemistry is about, you know, designing fungicides and herbicides and insecticides.

For his company, exploring multi-gene traits was far too risky: it would take too long to develop into products and face too many technical dead ends along the way. What's more, the economics of agriculture conspired against it. Farmers paid little to seed companies for advances in crop yield compared with simpler products that "kill stuff." The worldwide plant breeding industry, which provides seeds to farmers, does an annual business worth 10 billion dollars; agricultural chemicals generate over 30 billion. A noted scientist, this executive with a penchant for efficiency was no longer an academic. In his frequent partnerships with academic scientists and universities—several of them using *Arabidopsis*—his eye was trained to the low-hanging fruits of science. In this paper, I evaluate the effect of this type of engagement with business on the novelty and persistence of academic science using the complete corpus of science publications.

Methods

Publication data. To evaluate novelty and persistence within the *Arabidopsis* community, I collected all 18,359 published articles utilizing *Arabidopsis thaliana* between 1907 and 2002 by combining publication data from BIOSIS, PubMed, Medline, AGRICOLA, and SciSearch where "Arabidopsis" or "thaliana" was mentioned in the title, abstract or author-provided-keywords of an article. I added abstracts from *Arabidopsis* papers presented in plant molecular biology conferences over the past several years with the help of the U.S. Department of Agriculture and the Arabidopsis Information Resource (TAIR). I also gathered citation data on the 11,000 *Arabidopsis* articles which are present in the SciSearch database between 1945 and 2002. SciSearch is produced by Thompson Scientific's Institute for Scientific Information (ISI). Citations link the 11,000 ISI *Arabidopsis* articles, by citation, to *Arabidopsis* articles within that set and to an additional 45,577 non-*Arabidopsis* articles which cite them.[4] Using this information, Figure 2 shows the growth in *Arabidopsis* papers and citations to *Arabidopsis* papers worldwide, from 1975 until 2002. In the course of this time period, the number of papers grew exponentially, from 91 and 112 in 1980 and 1985, respectively, to 373 in 1990, 1070 in 1995 and 2192 in 2000.

From the titles and abstracts of *Arabidopsis* publications, I extracted an extensive list of scientific terms, corresponding to *Arabidopsis* genes, proteins, species, techniques, biological processes, molecular functions, cellular components, *Arabidopsis* developmental stages and anatomical locations. Curators from TAIR with Ph.D.s in molecular plant biology hand-coded all *Arabidopsis* genes in the publications and then coded those genes with terms corresponding to their molecular function, the biological process of which they are part, the cellular components they affect, and the stage of development and anatomical position at which they are expressed.[5] For example, their codes indicate that the gene LEAFY is related to flowering as inferred from a mutant phenotype (knocking out the gene and observing flowering mutation).

I contributed to this process by testing for and classifying frequent single- and multi-word concepts in the *Arabidopsis* abstracts,[6] and also by using a fuzzy search algorithm[7] to match all of the TAIR terms directly to the titles and abstracts of articles in the database. Furthermore, I used a probabilistic algorithm to annotate

---

[4] Thompson Scientific generously gave me use of the entire Science Citation Index for the purpose of this project.

[5] Codes for molecular function, biological process and cellular component (13,976) used to code the *Arabidopsis* genes were developed in a joint effort among curators of several model organisms to characterize genes independently of their host-organisms, entitled the Gene Ontology (GO) project. Development and anatomical annotations were developed by TAIR specifically for *Arabidopsis*.

[6] A multi-word concept is established if a word frequently co-occurs with another word or string of words (e.g., Northern Blot; Polymerase Chain Reaction).

[7] I used a BLAST (Basic Local Alignment Search Tool)-like algorithm that searches through windows of text-characters looking for substring matches.

additional gene and protein names in articles (Chang, Schutze and Altmann 2003), and worked with a biology student, Christian Anderson, to build a typology of methods[8] and species that we matched independently into abstracts and titles. This matching process resulted 28,350 unique terms in nearly 400,000 publication-term matches with the 18,359 articles described above. I associated these publications and the terms they contain, with the scientists that produce them and the academic research organizations where they reside over time. From the term data, I generate a novelty and persistence score for each year in which the scientist or research organization publishes. To address the hypotheses stated earlier, I link these PIs and ROs—along with their newly minted scores for novelty and persistence—to the companies with whom they collaborate. Using statistical models, I predict how industrial collaboration in prior years affects the novelty and persistence of research in the present one. In the following sections, I detail these steps regarding data organization, coding and analysis.

Units of analysis: Principal Investigators and Research Organizations. For the purpose of this study, I focus on those scientists who ran their own labs at any time within the period of study—principal investigators (PIs)—and the research organizations (ROs) that employ them. Because students, technicians and post docs rarely come to independently organize biology research of their own, including them in the study would artificially inflate the effects I examine. I established the existence and location of Arabidopsis principal investigators and research labs through two separate sources of information. The first is provided by TAIR, which maintains a database of all individuals and organizations in the Arabidopsis community who have ever published Arabidopsis papers or used their extensive genetic information databases. TAIR data indicate the location and institutional affiliations of researchers and research labs that use Arabidopsis. Between 1974 and 2003, 5,725 principal investigators used Arabidopsis in their research, a third of whom reside in the U.S.

Where labs weren't specified in TAIR, the bibliographical data from ISI and BIOSIS electronically lists the institutional locations for the authors of each article in the database. In the field of molecular plant biology, the last author on a research article is almost always (with 90% accuracy) a PI—the head of a research lab. Across articles, this rule allowed me to identify most of the remaining principal investigators, and then verify their status by searching department websites and examining professional association directories. I associated

[8] We drew on the methodological classification scheme used in the BIOSIS database.

articles with specific universities and research institutes in the same manner described above for PI's.

I associated articles with specific universities and research institutes (ROs) in the same manner described above for PI's. In total, 3,163 research organizations have sponsored science utilizing *Arabidopsis*. I associated these ROs with a number of university characteristics from the World Education Database such as the presence of a medical or agricultural school, number of faculty and students, etc. I also coded the organization's identity as an agricultural, technical, medical or general research university and added an additional number of more detailed characteristics for U.S. organizations such as the number of research dollars received from the federal government, the total amount spent on research, and the university's charter as a land grant university.

Industry Ties. Documenting partnerships between firms and academic science is difficult and requires some *inference*. As a result, I document industrial collaboration and funding in a number of ways. I *infer* association between firms and PI's from articles co-authored and patents co-invented by university and firm-based scientists. I have also scanned the acknowledgements of all *Arabidopsis* papers published between 1975 and 2000 in the SciSearch database (8,400 documents) and have parsed these acknowledgements for the company collaborations and funding they mention. Figure 3 provides an example for the *Nature* article "Shatterproof MADS-box genes control seed disbursal in *Arabidopsis*" coauthored by PIs John Bowman, Martin Yanofsky and Beth Savidge from UC San Diego and UC Davis. Because of the funding mention in the acknowledgement, I code all three as having a tie to Monsanto in the year 2000. Figure 3 also highlights the underlying question of the study: Is Monsanto transforming esoteric research funded by the government into valuable products, or is it poaching the biggest fish in the public pool of science and muddying the waters in the process?

After my initial coding of PI-company ties, I recoded them to indicate whether they were relational or transactional in nature. I coded coauthorship, collaboration and the sharing of scientific materials with industry as relational ties. Alternately, I coded funding and employment as transactional. In addition, I coded transactional relationships between a PI and a company if the PI served on the board of directors or scientific advisory board of the company. Such relationships are common in the life sciences, they often provide an additional source of revenue for the scientist, and they rarely involve joint research (Audretsch and Stephan 1996; Zucker, Darby and Armstrong 1998). I also coded a transactional tie if the PI was listed as an inventor on a patent assigned partially or exclusively to that company.[9] Where collaboration was not specifically suggested by the data—where companies were acknowledged for nothing, in particular—I assumed that the relationship was a transactional one. This assumption is grounded in the insight I received from interviews with executives at plant biotechnology companies. These executives frequently described seeking public acknowledgement for their academic contributions, and getting it.[10]

If the dates of ties between firms and scientists or their research organizations were not specified, I coded them as occurring in both the year when they are published and in the two prior years. Through interviews with scientists, I learned that three years is the modal duration for a firm to support an academic lab. Moreover, for coauthorship or collaboration, three years roughly corresponds to the time required to run a series of experiments leading to a published paper.

With the complete list of co-authoring, inventing and acknowledged companies just described, I organized a group of students to search the news databases of Lexis-Nexis and related databases for instances where each company name co-occurred in the same article with "institute" or "university." We then recorded the collaborative and funding relationships linking these companies with the universities, departments and researchers in my sample. Figure 4 gives an example of a tie between Novartis and UC Berkeley's Department of Plant and Microbial Biology from 1998 until 2003. Relationship described as materials sharing or joint efforts in research I coded as relational, while funding, licensing and marketing I treated as transactional. Because some relational ties centered on projects outside the usual province of academic science, such as the joint development or marketing of products, I also coded these relationships according to an alternate scheme which indicates the degree to which both parties are collaborating on academic science versus commercial projects. I used this more restrictive coding scheme in my subsequent analysis.

In total, these data include 3054 relational ties and 736 transactional partnerships between companies and *principal investigators*. These data also include 1584 relational partnerships and 693 transactional ties between companies and academic *research organizations*. For the purposes of the analysis in this paper, I gave

[9] I purchased acknowledgement data for all of the publications cited by patents after 1992 from CHI Research, Inc. with funds from NSF grant #0242971.

[10] Where partnerships were coded as multiple types, I considered them transactional only if they were not also coded as relational, consistent with interviews which suggested that relational ties had to do with scientific relationships independent of compensation.

academic labs and organizations a dichotomous score of “0” or “1” corresponding with whether or not they maintained relational ties with one or more firms, and contractual ties with one or more firms.

Measuring the research approach of a PI or RO is a delicate issue, and yet this project depends upon my ability to reliably make this assessment. In this section, I briefly sketch the metrics I use to indicate: 1) the novelty of a research program, 2) the persistence of PIs and ROs in particular streams of research over time, 3) the status of a PI or RO within the scientific community over time, and 4) the application-orientation of a PI or RO’s scientific subfield.

Novelty. I derive my measure of novelty from the combinatorial novelty of the biological terms linked to *Arabidopsis* articles (i.e., genes, proteins, species, methods, biological processes, molecular functions, cellular components, *Arabidopsis* developmental stages and anatomical locations). For the purposes of this analysis, I consider novelty a function of the number of times biological terms have ever been used together before and the number of times they have ever been used before at all. Specifically, I measure novelty within each publication as:

$$\frac{\sum_{i=1}^{n}\sum_{j=i+1}^{n}\left[1-\left(\frac{\sum_{ij=1}^{N}\left(\chi_i \cap \chi_j\right)}{\min\left(\sum_{i=1}^{N}\chi_i \wedge \sum_{j=1}^{N}\chi_j\right)}\right)\right]}{\left[\frac{n}{2}\right]}$$

where *i* and *j* are terms in article *x*, *n* is the total number of terms in *x*, and *N* is the total number of articles within the entire field of *Arabidopsis* published prior to article *x*. Hence, novelty is composed of one minus the frequency with which any two terms have been used together in the previous literature divided by the number of times they could have been combined—the frequency of the least frequent term. This measure is averaged for the pairwise combinations of all terms within an article *x*. If a term, such as a gene, is mentioned for the first time in an article, it is given the weight of a completely novel combination even if no other relevant terms are mentioned.

For example, consider the abstract for the “Shatterproof” article illustrated in Figure 3 by Yanofsky et al. Only two genes, SHP1 and SHP2, are mentioned in the article. If these genes were mentioned together in five previous articles, and the least frequent (SHP2) was mentioned in seven total articles, then the gene novelty score for the article would equal .287. I used this measure to develop novelty scores for individual classes of scientific terms (e.g., gene-novelty, method-novelty, biological process-novelty, etc.), theoretically meaningful combinations of terms (e.g., method-by-gene novelty), and the entire "kitchen sink" of all scientific terms. I generated a score for each publication and then associated the score with the PIs and ROs that produced it.

I tested the validity of my novelty index against a database of expert rankings, entitled "Faculty of 1000." In this database, prominent scientists nominate the best articles in the field and classify them as a "new finding," "confirmation," "methodological innovation," "hypothesis" or some combination of these. Three hundred and ninety-one articles in Faculty of 1000 use *Arabidopsis* and are located in my database. I correlated the presence of each of the classifications of Faculty of 1000 articles against my measures of novelty. One-tailed t-tests indicate that articles proposed as "new finding" and "hypothesis" are significantly (p<.05) more novel, according to my index (using all terms), than those which are merely a "confirmation" of former assumptions.[11]

In the tables of model coefficients, I list only the results for gene novelty which were representative of most other novelty scores. The linear models I use in this analysis rely on the assumption that the dependent variable is normally distributed. In order to make these novelty indices conform to a normal distribution, I transformed them by taking the exponential for gene novelty, and the double exponential for all term novelty.

Persistence. I measure an entity's persistence in research at any given time as a function of the stretch of time across which a PI or RO has used a certain scientific term in their research, and the length of time since they first used the term. Specifically, I measure persistence as

$$\frac{\sum_{i=1}^{N} \frac{d(x_{in}) - d(x_{i1})}{d(x_n) - d(x_{i1})}}{N}$$

where N is the number of terms (out of the 28,350 unique described earlier) that a PI or RO has ever used in any article x, $d(x_{i1})$ and $d(x_{in})$ are the dates of its first and most recent articles containing term i, and $d(x_n)$ is the publication date of its most recent article of any type. For example, again consider again the "Shatterproof"

[11] Methodological novelties are almost always "less novel" according to our index because their abstracts rarely specify the nature of the subtle improvements that make them different from older methods.

article in Figure 3. According to the measure described above, Martin Yanofsky's persistence in using mutant analysis is composed of the difference, in days, between the publication of his first and most recent paper using mutant analysis, divided by the difference, in days, between the publication of his first mutant analysis article and his most recent article of whatever type. Because Yanofsky has used mutant analysis from his first paper (published on April 5, 1984) until his most recent paper of any type (published on April 13, 2000), he is perfectly persistent in his use of mutant analysis. Yanofsky's persistence score for the year 2000 would average his persistence in mutant analysis (i.e.,1) with his persistence in using every other coded scientific term he has ever used in an *Arabidopsis* publication (i.e., 332 distinct terms).

As with novelty, I can compute the persistence of a PI or RO within a specific class of scientific terms (e.g., method persistence, gene persistence, biological process persistence) or within the set of all scientific terms. Because many PIs and ROs were not at all persistent with an unvarying score of 0, I estimated separate models to predict whether PIs and ROs were persistent, and then, if they were persistent, to predict how persistent they were. I took the natural log of gene persistence to normalize its distribution.

Status. I measured the social status of PIs and ROs within the community of *Arabidopsis* science in two ways. I measured it first with the centrality of a PI or RO's articles within the network of scientific references inscribed by the bibliographies of all *Arabidopsis* articles listed in ISI's SciSearch. Secondly, I measured social status as a function of centrality within the social network defined by PI article coauthorship, accumulating over time.[12] Both measures can be easily aggregated to the RO-level of analysis. The first captures the position of PIs and ROs in the reputational structure of science, and the second captures their position within the social structure. Coauthorship centrality represents a PI or RO's position within a network of collaborations, and it also represents their scientific "lineage" as students and post docs coauthor with faculty mentors before managing their own labs and students. I measured centrality according to Sabidussi's straightforward formula:

$$C_{closeness}(v_i) = (n-1)\left(\sum_{j=1}^{n} d(v_i, v_j)\right)^{-1}$$

[12] Because the great majority of *Arabidopsis* papers have been produced so recently—after 1985—I did not allow citations or coauthorship ties to atrophy in their contribution to the centrality score over time.

where v_i represents the PI or RO of interest and v_j represents all of the other PIs or ROs in the network (1966).[13] Collaborations within the *Arabidopsis* community represent opportunities to work with other scientists, receiving special access to their materials and methods.

Application-Oriented Subfields. In order to establish the application-oriented subfields of *Arabidopsis* science, I first grouped all of the *Arabidopsis* publications into a set of seventeen subfields by clustering all 18,563 articles on the 28,350 unique scientific terms that linked them together.[14] I used the principal component of the matrix of articles by terms to divisively partition the data into substantively meaningful subfields (Boley 1997). With a team of biologists, I examined these clusters to select the appropriate number, reunited related clusters separated by gross initial divisions, and characterized their scientific content. In order to visualize the development of these clusters over time, I created lay-outs with the graph drawing software Pajek (Batagelj and Mrvar 1998). Figure 5 shows a layout of the publications in each cluster, networked by the citations between them.[15] The figure reveals the variety in terms of size, centrality and cohesiveness of the various subfields, ranging from the large and central organogenesis cluster, the scientific area concerned with the differentiation of cells within the tissues of complex higher organisms, to the smaller and more peripheral cluster of commercial disease resistance.

I then identified all of the patent citations to publications within each cluster and flagged those clusters with a higher-than-median patent citation per publication as potentially commercial. Note that less commercial fields like “photosynthesis” examine one of the most biologically significant and distinctive aspects of plant metabolism; while more commercial fields like “terpines” examine a class of protein substances that are biologically inessential, but have historically been very useful as the basis for industrial substances such as turpentine and rubber.

Controls. I controlled for the length of time that a PI and RO had ever worked with *Arabidopsis*, which

[13] Only in-degrees were used in the calculation of the citation centrality, while both in- and out-degrees were used in the calculation of coauthorship centrality.

[14] I tried an alternate process of “finding” subfields in the data by assigning a series of random variables to each publication which I conceptualized as “latent concepts”, and then simulating an influence process by which the latent concepts in each article influenced those they cited and those that cited them (Moody 2001). This process created subfields which were more integrated by author and institution, but were much less distinguishable by their content.

[15] Recall that the citations were *not* used to develop the clusters and represent a form of structure related but not bound to the organization of the clusters themselves. An author can publish in multiple subfields, and citations tend to reference friends and coauthors over strangers if both have produced relevant work.

I call "tenure" in the tables and subsequent discussion. I also controlled for the square of tenure, and the measures for citation and coauthorship closeness-centrality described earlier. These measures correlated so highly with measures of cumulative citations and cumulative coauthoring relationships that only the centrality measures were used. I also controlled for the degree to which PIs and ROs brokered relationships between other PIs and ROs by using Freeman's measure of betweenness centrality within coauthorship networks.[16] In addition, I included dummy variables corresponding to the type of department housing the PI's lab. These included departments of agriculture, plant biology, medicine, chemistry, food science, and other departments (e.g., math, statistics…sociology), compared with the reference category of general biology departments. I also included the type of RO such as science and technology universities (e.g., Rensselaer Polytechnic, MIT), agricultural universities, research institutes, government agencies, hospitals, etc., compared with the reference category of general universities (e.g., Harvard University). I included dummy variables for all 63 countries and the 17 scientific subfields listed in Figure 5. The addition of these controls, especially those for scientific content which reference PIs involvement in specific, heterogeneous research communities, represents an important advance in the research on science and innovation studies which too often make loose cross-sector comparisons. Different subfields represent not only varying scientific content, but also trace different scientific practices, sources of funding, and research audiences.

Modeling Strategy. I use panel models to take advantage of the dynamic nature of my data; I associate past relational and transactional ties with subsequent scientific content and activity. One can imagine using a hierarchical modeling strategy to simultaneously capture the effect of PI and RO relationships on PI novelty and persistence. My data, however, reflect what is becoming increasingly commonplace in academic science: scientists move through research organizations, just as research organizations "move through" scientists, transgressing any simple notion of hierarchy. Instead, I estimate the models separately at the PI and RO levels of analysis, but include ties between companies

---

[16] $C_{between}(v_i) = \sum_{k=1}^{n} \sum_{j=1}^{n} \frac{g_{jk}(v_i)}{g_{jk}}$. I could also have used Burt's measure of structural constraint (1992) which correlated inversely with the Freeman betweenness measure.

and both PIs and ROs in all models. In other words, I measure the effect of ties between a PI and companies on that PI's subsequent novelty and persistence, as well as the effect of independent ties between that PI's RO and companies. I measure the effect of ties between an RO and companies on subsequent novelty and persistence, as well as the aggregate effect of all ties between that RO's PIs and companies. Because PIs and ROs are distinct strategic actors, and because they vary in their distribution across each other, sometimes corresponding variables (e.g., PI relational ties at the PI-level versus aggregated PI relational ties at the RO-level) will have different effects on the same dependent variable at different levels of analysis.

The specific unit of analysis in all models is PI-publication-years and RO-publication-years. For example, if Martin Yanofsky from UC San Diego published one article in 1984, three articles in 1986 and two articles in 1987, he would receive three separate entries in the regression corresponding to 1984, 1986 and 1987. For 1986 and 1987, when he published more than one article, I take the mean values of Yanofsky's novelty, persistence, and independent variables across all articles published that year. This yields unbalanced panels with some gaps corresponding to years that PIs and ROs do not publish any *Arabidopsis* articles.

The basic models, specifying random effects,[17] are:

$$y_{it} = \alpha_{it} + x_{it}\beta + v_i + \varepsilon_{it} \quad , \quad y \sim N(\;)$$

$$\Phi^{-1}(y_{it}) = \alpha_{it} + x_{it}\beta + v_i + \varepsilon_{it} \quad , \quad y \sim Bernoulli$$

for normally distributed and dichotomous dependent variables, respectively, where time *t* is measured in years, and each *i* is a PI or research organization. I estimate the models using a General Estimation Equations strategy, allowing me to use the Huber/White/sandwich estimator of variance to produce semi-robust standard error estimates from which all of the significance levels are derived (Huber 1967; White 1980, 1982). By allowing the errors to vary in size, I take into account the fact that larger ROs and PIs with larger labs will likely produce correspondingly larger errors than smaller ROs and PI labs.

[17] Hausman's specification test (1978) indicated that the coefficients of random and fixed effects models did not sufficiently differ (p<.05), and so I used a random effects model to calculate the effects of time-invariant characteristics.

Descriptive Statistics

Tables 1 and 2 list descriptive statistics for the variables described in previous sections. Each table includes the mean variable value for all PI-publication-years or RO-publication-years, the standard deviation for the entire sample and the standard deviation of variables within individual PIs and ROs. The final three columns of the table correspond to the number of PI- or RO-publication-years, the number of distinct PIs and ROs which they represent, and the average number of publication-years per PI or RO.

The dependent variable section of Tables 1 and 2 show the reduced number of cases available for analysis where scientists and research organizations are persistent in the study of genes (956 PIs; 421 ROs) or terms of any kind (1945 PIs). This is why I also model whether or not PIs and ROs are at all persistent with the dichotomous persistence variables listed. The negative mean values for the natural logarithm of PI-level gene and all term persistence (-1.854 and -1.831) correspond to actual mean persistence scores of .157 and .160, respectively. Differences between the same novelty and persistence scores at the PI and RO levels of analysis suggest that novel and persistent *Arabidopsis* research is not randomly distributed across ROs. Rather, high quality academic work is clustered within a smaller set of organizations, reducing all the novel or persistent research produced by each of them to a single score. This highlights that although related, the PI and RO level results provide distinct insights into the effect of industry on science. By equally weighting the research performed by organizations specializing in *Arabidopsis* with those that use it only occasionally, the RO level of analysis effectively gives additional influence to more peripheral researchers who are more institutionally isolated in their *Arabidopsis* research.

Most authors are relatively new to the use of *Arabidopsis*, with an average of 2.8 years experience publishing about the plant. Organizations, on the other hand, have had some scientist working with *Arabidopsis* for over four years, on average. Forty-one percent of the PI-publication-years are spent in some type of general biology department (e.g., biology, biochemistry, cell biology, molecular biology), with plant biology departments (e.g., botany, plant molecular biology) a close second at 32%, and departments of agriculture (e.g., agronomy, plant pathology, soil science) in third with 23%. Table 1 also suggests that more than three-fourths of all scientists stay within the same type of department and research organization over the course of their careers researching *Arabidopsis*.

While Table 1 states that over 75% of PI-publication-years were spent in universities and over 15% in government agencies, Table 2 reveals that these same institutions account for 68% and 13% of the RO-publication-years, suggesting that these are the two institutions with the largest clusters of scientists. While only two percent of the universities are explicitly agricultural, a full 22% have some agricultural influence, such as the presence of agriculture-related schools or departments. Thirty-one percent of the universities have a medical school, and 3.5% are explicitly science and technology institutions (e.g., MIT, Georgia Tech). After universities, research institutes like Cold Spring Harbor and the John Innes Centre in the UK make up the next most populous set of academic research institutions (26%).

Thirty-four percent and 27% of the PI- and RO-publication-years, respectively, took place in the U.S., while a similar proportion (30% and 34% of PI- and RO-pub-years) took place in Western Europe (not including Great Britain). A substantial number of *Arabidopsis* researchers are in Asia (16.3% and 18.3% of PI- and RO-pub-years) and England, Australia and Canada (15.2% and 13.6% of PI- and RO-pub-years). The remaining cases, in order from most to least, come from Eastern Europe, the Middle East, South America, and Africa. Differences between the PI and RO averages suggest that the densest communities of *Arabidopsis* researchers are clustered in the U.S., Canada, Britain, Australia, and New Zealand.

Descriptive statistics for scientific subfields in Table 1 reveal that the standard deviations within PIs are smaller, but in every case not less than half of the overall deviations. In other words, while the research careers of PIs are more focused within a subfield than the population of research papers, individual research continues to cross subfield boundaries over time. The size of subfields range from organogenesis, representing 14% of the PI- and RO-pub-years to targeting / splicing and photosynthesis, each with about 1% of the PI- and RO-pub-years.

Twenty percent of the PI-pub-years are coded as having relational ties with firms and 9.9% are coded as members of research organizations that have such ties. A smaller proportion of PI-pub-years, 7.6% and 14.3%, are coded as having direct transactional ties with firms or of being part of an RO with such ties, respectively. The distribution of ties shifts in the RO-level data, where a larger proportion of RO-pub-years are coded as having transactional ties: 11.2% for those with distinct, organization-level ties to firms and 20.4% for those with aggregated PI-firm ties. Thus, firms are much more likely to collaborate with academics working in dense communities of *Arabidopsis* scientists, but will transact with them wherever they are located.

Results

Table 3 lists estimates of covariate coefficients as regressed on the exponential of gene novelty and the double-exponential of all term novelty at the PI level of analysis. Table 4 presents the same analysis at the organization-level. The all-term novelty models (4-6 in each table) were run on more cases than the gene novelty models because not all *Arabidopsis* publications discussed genes. All models have x^2 statistics that indicate a statistically significant relationship between novelty and the complete set of covariates. Tables 3 and 4 indicate that tenure has a significant negative effect on novelty while tenure-squared has a small but significant positive effect. This suggests a slight curvilinear relationship between tenure and novelty: as PIs progress in their academic careers, their likelihood to combine genes, techniques, and theories in novel ways drops the most in the first few years and then levels off. Interestingly, the effect of tenure on novelty appears to be even initially more negative for research organizations, suggesting that organizations producing the most *Arabidopsis* research are the most likely to maintain novelty in their pool of research over time.

Citation centrality, coauthorship centrality and coauthorship betweenness centrality each have negative effects on gene and all term novelty. This could suggest that as scientific actors become more socially and scientifically central, they become more conservative in their combinations of distinct scientific terms, exploiting what they have learned in earlier, more risky studies.

At the PI level, departments of medicine inspire more novelty of both genes and all terms than general biology departments. Similarly, agriculture departments also inspire more all term novelty. Different types of institutions also had heterogeneous effects on scientific novelty. PIs at science and technology universities, like MIT or Caltech, generated more novel combinations of genes than PIs at general purpose universities. Hospitals generated more all term novelty, while medical schools generated less than research produced by general purpose universities. At the level of research organizations, the institutional effects were somewhat different. In the RO-level models, agriculturally influenced universities and those with medical schools had a negative effect on novelty. Because the RO novelty models average the effect of all PIs within each RO, an organizational effect which is more negative in the RO models than in the PI ones suggests that PIs with the most novelty are clustered within a few of that type of organization, diluting the impact of their scores on the total effect for that organization-type. In this way, novel researchers are clustered within a few agriculturally influenced universities and those with medical schools, lowering the average impact of those institutions on novelty.

The addition of dummy variables indicating whether PIs and ROs had relational and transactional ties to firms significantly improved the x^2 statistic associated with each model (p<.001). Furthermore, adding the interaction between those ties and the centrality and application-orientation of the subfield significantly improved the x^2 statistic associated with both gene novelty models (p<.001). Table 3 reveals that both relational and transactional ties between PIs and companies have a significant negative effect on gene novelty (model 3: -.195, -.162, p<.001), while transactional ties between PI's ROs and companies have a negative effect on all-term novelty (model 5: -.167, p<.01; model 6: -.362, p<.1), providing support for Hypothesis B1. Accordingly, if a scientist has a gene novelty score of .5, engaging in a corporate collaboration will decrease her willingness to explore novel combinations of genes by 35% in the subsequent year.[18]

Consistent with Hypothesis A1, PI-level ties do exert a stronger effect on PI-level gene persistence than RO-level ties, but the relationships are not significantly different between the two types of ties in the models predicting all-term persistence. Table 3 also suggests that the effects of relational and transactional ties exert a similar effect on scientists. Relational ties appear to be no less corrosive of novel scientific production than transactional ties, and in some cases more so.

Table 4 shows a similar relationship between industrial ties and the exponential of gene novelty and double exponential of all term novelty at the organization level. Transactional ties between research organizations and companies have a negative effect on gene novelty. Aggregated PI-level relational and transactional ties have a negative effect on both gene novelty (models 2 and 3, p<.001) and all term novelty (models 5 and 6, p<.05). When industrial relationship dummies are interacted with the aggregate social centrality of its *Arabidopsis* scientists, however, those relationships exert a positive effect on organizations' willingness to combine genes and scientific terms in novel ways. Only a few organizations receive this novelty premium—for the centrality interaction to counterbalance the dampening effect of industrial collaboration,

---

[18] This calculation assumes that the PI has an average centrality score (.156). In this way, .156 (average centrality) *.238 (coefficient estimate for centrality-by-industry collaboration) -.211 (estimate for industry collaboration) = -.326 (predicted novelty score) which is 35% lower than a novelty score of .5.

universities must have a centrality score more than one and a half standard deviations above the mean. While a similar effect is not significant at the PI-level in models 2 and 3 of table 2, the signs of the coefficient estimates are consistent with it, providing partial support for Hypothesis A2. When scientific actors have status and centrality within the *Arabidopsis* community, they are able to avoid the negative effects of firm ties on their tendency to combine genes in novel ways. A slightly different interpretation would be that central PIs may simply attract a different kind of tie—one in which the company is willing to take a greater risk by exerting no influence.

At the PI level, interactions between application-orientation of PI subfields and ties with firms showed the positive effect on novelty predicted by Hypothesis A3. The negative impact of ties between academic scientists and companies on novelty is partially ameliorated where researchers work in the areas most consistently cited by patents. This effect is not significant at the RO-level.

Tables 5, 6 and 7 indicate that the effect of university-industry partnerships on academic persistence is less equivocal. Each table describes two nested models: the first estimating the effect of a vector of covariates on whether or not a PI or RO is persistent at all; the second measuring the effect of those same covariates on the degree to which the PI or RO is persistent. As such, the sample size for the second models are smaller. The same slight curvilinear relationship occurs between tenure and the log of persistence that occurred between tenure and the exponential of novelty. These models predict that only after an average of fifteen years researching with *Arabidopsis* would tenure's effect reach its nadir and begin to correlate positively with persistence. The first several years of a research career involve an expansion of the set of scientific terms used, after which scientists refocus and persist within these areas. Closeness and betweenness centrality in coauthoring networks both exert modest positive effects on gene persistence, while centrality in citation networks exerts a negative effect. It should be noted that because the mean of betweenness centrality is so small (.001), the effect is actually smaller than for the other centrality measure. PIs in chemistry departments persist in using the same scientific terms more than those in biology departments, while those in research institutes have lower persistence than their peers at general universities.

Table 5 shows that relational and transactional ties to industry exert a negative effect on academic persistence. Consistent with expectations, relational ties with firms exert a negative effect on whether or not a

PI persists in using the same genes (model 2) and this effect is contingent on the PI's scientific status for all four types of industry ties (model 3). For high status PIs with centrality scores one standard deviation above the mean (.237), ties to firms have a small positive effect on whether or not they subsequently persist in a research area. For the majority of less central scientists, however, ties to firms exert a negative influence on their persistence. A smaller centrality effect, but in the same direction, is true for the log of gene persistence. This effect is only significant when all of the tie and tie-by-centrality interactions are considered together and is not strong enough to reverse the effect of industrial collaboration on gene persistence (Table 5, model 6: x^2=7.49, p <.05).

The effects of industrial collaboration on scientists' persistence within the complete set of scientific terms are somewhat different, but complement the log gene persistence effects. While industrial partnerships are better at predicting whether PIs are persistent in their use of genes, industrial partnerships are better at predicting the degree to which scientists persist in their use of all scientific terms. This confirms that my collection of scientific terms improves the resolution of the picture of a PI's research program when compared to analyzing only the genes they have studied. As more detail is available regarding a PI's scientific activity, industrial collaboration improves in its ability to predict the actual degree and not just the existence of scientific persistence.

Table 7 provides the coefficient estimates for models predicting whether or not research organizations were persistent in their research, and the degree of their persistence. Regional variables, which I included in the probit model in lieu of specific country dummies, suggest that the academic quality of persistence is not simply a U.S. phenomenon—researchers in the UK are significantly more persistent in working with genes than their U.S. counterparts.

Organization-level models partially corroborate the negative influence of industrial collaboration on scientific persistence. Model 3 shows that company ties to an organization's PIs actually exert a slightly positive effect on the likelihood that organizations with an average aggregate PI centrality will persist in studying their genes of choice. More peripheral scientists, however, are influenced to become less persistent when they maintain industrial ties. This dampening effect of industrial ties on persistence appears to become even stronger in model 6, which predicts organizations' persistence in studying genes. In this model, universities with an average aggregate PI centrality are less persistent in their study of genes in the subsequent year if they have industrial ties. The difference between the PI and organization-level effects of industry ties on persistence—that the PI level effects appear stronger—suggests that collaborations with the worst effects on PI centrality are clustered within a subset of the research organizations.

Discussion

Consistent with Hypotheses B1 and B2, PI-company and RO-company ties exert a relatively consistent, negative effect on the novelty and persistence that characterizes the academic research of principal investigators and research organizations. In the opening section of this paper, I described how genes in *Arabidopsis* articles were hand-coded by scientists at TAIR, but that I used a computational approach for matching scientific terms within papers. The consistency of the effects of industrial collaboration on novelty and persistence across these very different specifications underscores the robustness of the effects.

Hypothesis A1, predicting that PI-level ties would have a greater effect on scientists than organization-level ties, received modest support. Its lack of stronger support, however, is notable. Many of the firms maintaining the most ties to universities have businesses aligned with pharmaceutical applications, such as those of Bristol Myers Squibb and Rosetta Inpharmatics. Their relationships, as described in the business press, probably involved researchers studying mammalian models more frequently than researchers studying plant models. That these ties had any effect on *Arabidopsis* science at either the PI or organization level, and often a strong, significant effect at both, suggests that organization-level ties alter the environment in which academic science is performed in ways similar to those of direct, interpersonal ties.

Relational and transactional ties to firms had complementary effects on academic persistence and novelty. When a transactional tie failed to show the predicted effect, the relational tie within the same model would exert the influence. Most often, the two types of relationships had a simultaneous effect. In general, relational ties were more harmful to academic novelty and persistence across models. This suggests that the influence of mimetic and normative pressures, as channeled through relational ties with firms, may have more weight on academic persistence and novelty than the legal or coercive pressures brought to bear in more formal relationships. Relational ties, because they were often indicated by industrial coauthorship on academic papers, were also the most reliable ties to measure. Another interpretation of the large PI-level relational tie effect is that it is simply the most accurate of the four measured relationships.

The mediating influence of centrality was substantially confirmed in both PI and organization-level models. Central, high status PIs and research organizations use industry ties to support their science with no harmful effects to the academic quality of their science, while the novelty and persistence of research produced by researchers in less central positions erodes with industrial collaboration. In some cases, the most central PIs were actually able to generate more novel and persistent research subsequent to industry partnerships. Central actors have greater bargaining power to “cut better deals” in their negotiation of industry collaborations. Similarly, firms may think differently about these relationships. In an interview with one prominent *Arabidopsis* biologist, the academic scientist explained that during one period in the early 1990s, despite his unwillingness to directly accept money from companies after a bad experience with DuPont, “Monsanto was very anxious to put money in my lab.” He eventually accommodated them by letting them fund some additional post docs to work in his lab.

Following this pattern, many central scientists succeeded in harnessing industry relationships to do better science, which subsequently generated greater industry recognition and resources. In this way, the science of central principal investigators and research organizations takes on a multi-vocal quality (Padgett and Ansell 1993; Powell, White, Koput and Owen-Smith 2005), signaling different competencies to different audiences. Central scientists use resources from the academy to achieve greater rewards from industry, and resources from industry to achieve greater rewards in the academy (Owen-Smith 2003).

Due to many of the processes just described, companies may have come to use central scientists and organizations for different purposes than those which are more peripheral. For the chance to have access to central academic labs, which produce the highest quantity and quality of science (Cole and Cole 1973), firms may be willing to take smaller and less defined shares in the output of those labs. As with Monsanto and the *Arabidopsis* researcher cited above, companies may simply want to place a bet on high quality scientists in hopes that they will have a share in the use of discoveries.

The consequence of these patterns is illustrated in Figure 6. By boosting the novelty and persistence of central scientists and their institutions, industry collaborations may increase the variance of their scientific success while lowering the average success of their research. This would mean that industry ties help central scientists and institutions produce occasional, breakthrough successes. Collaborations between industry and peripheral academic scientists and institutions may decrease the variance of their scientific output and increase the average success of their work. This brings the risk profile of peripheral academic science into closer conformity with an industrial approach. In short, companies put less central academic scientists “to work.” Industry ties thus increase the inequality among academic scientists: they render the academic science of more central scientists more novel and focused and that of less central scientists less so. The combined effect of industrial ties on academic novelty and persistence, however, is negative.

There are limitations to my analyses which point to a series of next steps that would extend the questions I pose and strengthen the answers I provide. I could make refinements to the measures of novelty and persistence to broaden the reach of my results. For example, a refinement to the persistence measure might take into account the different ways in which scientific careers can be persistent, such as within a method or within a topic area.[19] Future research may also seek to develop a direct measure of the balance between novelty and persistence, rather than individual measures of each quality. This approach might help to better distinguish, *a priori*, between hyperactive experimentation and productive novelty and courageous, insightful persistence from the single-mindedness of a plodding, unimaginative, or compulsive scientific mind.

This research also indicates other levels of analysis and settings at which researchers might profitably explore the relations between the academy and industry. For example, while this study has focused on the impact of industry collaboration on academic scientists, research organizations and subfields, significant differences between the region effects in my models suggest that examining industrial embeddedness at the national level might reveal similar dynamics. Germany and France, for example, have recently structured their largest plant biotechnology grant programs to facilitate university-company collaboration and the spill over of research into the R&D of homeland agricultural and life-science firms. Other governments, like Britain, Australia and Spain, do not have a strong agricultural private sector and sponsor research in very different and

[19] An improved persistence metric would measure a scientist’s or research organization’s persistence in whatever area their career is most integrated, rather than focusing on a single area (i.e., genes) or averaging across all areas (i.e., all terms).

less commercial ways. Subsequent research could examine how industrial ties at the level of nation states shape the funding, channeling and approach of national science.

Furthermore, this research suggests the importance of exploring the influence of industry ties on novelty and persistence in other scientific environments. Research in areas with even greater commercial potential, like many areas of mammalian biology and crop research, will likely be much more susceptible to industrial influences. Moreover, to understand the ultimate impact of industry on academic science, future investigations will need to reach further through the intellectual property and product markets where companies transform research ideas into economic growth.

In the past thirty years, government support for academic science has ebbed and academic collaboration with industry has flowed. Partnerships between companies and academic scientists have channeled the stream of ideas out of the ivory tower: Significantly more published ideas from scientists partnering with industry enter patents and products. Private sector contributions have, in return, contributed to the limited funds and materials dedicated to academic research. Though these contributions are modest by many measures, they have not left academic science unchanged. By engaging central scientists on the basis of their academic quality and peripheral scientists on the basis of their time and effort, industry creates a striking, new division of scientific labor.

Table 1. Descriptive Statistics for Principal Investigator by Publication-Year Data

Variable	Mean	S.D. overall	S.D. within PI	*N* overall	*n* of PIs	*N / n*
Dependent variables						
exp(gene novelty)	1.945	0.581	0.366	6,073	3,152	1.926
gene persistence (0/1)	0.774	0.419	0.324	12,920	5,768	2.240
ln(gene persistence)	-1.854	0.925	0.563	3,099	956	3.241
PI qualities						
tenure in *Arabidopsis* research	2.479	3.374	2.279	12,920	5,768	2.240
tenure-squared	17.523	42.984	29.870	12,920	5,768	2.240
citation centrality	0.257	0.171	0.105	12,907	5,755	2.242
coauthorship centrality (closeness)	0.156	0.081	0.051	12,907	5,755	2.242
coauthorship centrality (betweenness)	0.001	0.003	0.002	12,907	5,755	2.242
Departments						
biology department[a]	0.407	0.485	0.471	12,882	5,755	2.238
plant biology department	0.320	0.460	0.438	12,882	5,755	2.238
agricultural department	0.231	0.417	0.416	12,882	5,755	2.238
medical department	0.030	0.170	0.183	12,882	5,755	2.238
other department	0.010	0.091	0.103	12,882	5,755	2.238
Organizations [a]						
University	0.751	0.427	0.416	12,920	5,768	2.240
agricultural university	0.023	0.146	0.140	12,920	5,768	2.240
agricultural influence	0.338	0.468	0.448	12,920	5,768	2.240
medical school	0.423	0.487	0.470	12,920	5,768	2.240
science and technology school	0.030	0.168	0.166	12,920	5,768	2.240
research institute	0.250	0.427	0.417	12,920	5,768	2.240
government agency	0.156	0.359	0.349	12,920	5,768	2.240
hospital	0.014	0.117	0.113	12,920	5,768	2.240
Regions [a]						
U.S.	0.347	0.473	0.459	12,920	5,768	2.240
U.K., Canada, Australia	0.152	0.357	0.356	12,920	5,768	2.240
Western Europe	0.302	0.457	0.454	12,920	5,768	2.240
Eastern Europe	0.020	0.138	0.161	12,920	5,768	2.240
Asia	0.163	0.367	0.365	12,920	5,768	2.240
Cluster application orientation	0.352	0.438	0.293	12,920	5,768	2.240
Relationships [a]						
relational	0.204	0.403	0.209	11,002	5,146	2.138
contractual	0.076	0.266	0.154	11,002	5,146	2.138
relational (org-level)	0.099	0.299	0.170	11,002	5,146	2.138
contractual (org-level)	0.143	0.350	0.189	11,002	5,146	2.138

[a] Dichotomous variables whose means can be interpreted as proportions or percentages of PI-publication-years with those characteristics.

Table 2. Descriptive Statistics for Research Organization by Publication-Year Data

Variable	Mean	S.D. overall	S.D. within RO	*N* overall	*n* within RO	*N / n*
Dependent variables						
exp(gene novelty)	1.734	0.548	0.361	3,187	1,092	2.919
exp(exp(all novelty)	8.296	2.118	1.670	5,728	1,905	3.007
gene persistence (0/1)	0.746	0.435	0.351	6,048	1,988	3.042
ln(gene persistence)	-2.338	0.928	0.605	2,029	421	4.819
all term persistence (0/1)	1.000	0.000	0.000	6,048	1,988	3.042
PI qualities						
time since first *Arabidopsis* article	4.276	5.059	3.221	6,048	1,988	3.042
time-squared	43.872	93.827	57.816	6,048	1,988	3.042
citation centrality	0.170	0.140	0.109	6,048	1,988	3.042
coauthorship centrality	0.125	0.086	0.069	6,048	1,988	3.042
coauthorship centrality (between)	0.001	0.002	0.002	6,048	1,988	3.042
Organization [a]						
university	0.682			6,048	1,988	3.042
agricultural university	0.017			6,048	1,988	3.042
agricultural influence	0.219			6,048	1,988	3.042
medical school	0.316			6,048	1,988	3.042
science and technology school	0.035			6,048	1,988	3.042
research institute	0.260			6,048	1,988	3.042
government agency	0.132			6,048	1,988	3.042
hospital	0.018			6,048	1,988	3.042
company	0.063			6,048	1,988	3.042
Regions [a]						
U.S.	0.271			6,048	1,988	3.042
U.K., Canada, Australia	0.136			6,048	1,988	3.042
Western Europe	0.339			6,048	1,988	3.042
Eastern Europe	0.041			6,048	1,988	3.042
Africa	0.003			6,048	1,988	3.042
South America	0.015			6,048	1,988	3.042
Asia	0.183			6,048	1,988	3.042
Middle East	0.011			6,048	1,988	3.042
Relationships [a]						
relational	0.094	0.284	0.139	5,315	1,801	2.951
contractual	0.118	0.315	0.158	5,315	1,801	2.951
relational (pi-level)	0.141	0.350	0.173	5,315	1,801	2.951
contractual (pi-level)	0.206	0.403	0.222	5,315	1,801	2.951

[a] Dichotomous variables whose means can be interpreted as proportions or percentages of PI-publication-years with those characteristics.

Table 3. MLE Estimates for PI-level Scientific Novelty

	exp (gene novelty)			exp(exp(all term novelty))		
Model	1	2	3	4	5	6
Constant	2.121***	2.133***	2.169***	8.473***	8.490***	8.488***
Principal Investigator						
tenure	-0.035***	-0.032***	-0.031***	-0.037*	-0.035*	-0.033**
tenure^2	0.002***	0.002***	0.002**	0.002	0.002	0.002*
citation centrality	-0.253**	-0.241**	-0.237**	-1.050***	-1.027***	-1.028***
coauthor centrality	-0.360*	-0.351*	-0.439*	-0.877*	-0.880*	-0.863
coauthor betweenness centrality	-14.782***	-10.226***	-10.914***	-11.937	-8.527	-7.253
Department						
(reference: general biology)						
plant biology	0.004	0.003	0.005	0.144**	0.144**	0.139*
agriculture	0.007	0.008	0.010	0.251***	0.249***	0.254***
medical	0.107*	0.107*	0.106*	0.373**	0.373**	0.377**
chemistry	-0.169†	-0.152	-0.151	-0.083	0.082	0.084
food science	-0.169	-0.183	-0.210	0.059	0.060	0.081
other	0.827***	0.788***	0.794***	0.151	-0.158	-0.155
School						
(reference: general university)						
agricultural university	0.046	0.050	0.062	0.262	0.247	0.247
agriculturally influenced univ.	-0.009	-0.011	-0.012	0.017	0.016	-0.008
medical school	-0.029	-0.034	-0.032	-0.056	-0.055†	-0.059†
science & technology university	0.076	0.083†	0.081†	0.068	0.064	0.056
research institute	-0.023	-0.018	-0.015	0.023	-0.022	-0.025
government agency	-0.049†	-0.051†	-0.054†	-0.119	-0.119*	-0.123*
hospital	-0.061	-0.066	-0.071	0.504**	0.519*	0.539*
Industry ties and interactions						
relational tie		-0.083***	-0.211**		-0.026	0.040
transactional tie		-0.114***	-0.154†		-0.108	-0.037
relational tie (organization-level)		0.031	0.084		0.085	0.357
transactional tie (org)		-0.004	-0.105		-0.122	-0.398†
relational tie x coauthor centrality			0.238			-1.017
transactional tie x c.			0.252			0.279
relational tie (org) x c.			-0.123			-1.956
transactional tie (org) x c.			0.156			2.126*
relational tie x applied subfield			0.229***			0.249†
transactional tie x a.			-0.012			-0.269
relational tie (org) x a.			-0.096			-0.047
transactional tie (org) x a.			0.197**			0.111
x^2	54,177.65***	58,365.33***	58,822.65***	13,740.22***	9,993.19***	10,248.43***
Number	4,933	4,933	4,933	10,051	10,051	10,051
Number of PI's	2,659	2,659	2,659	4,786	4,786	4,786
Range of PI-pub-years	1:13	1:13	1:13	1:19	1:19	1:19

† $p < .1$ * $p < .05$ ** $p < .01$ *** $p < .001$

Standard errors (coefficient significances) are determined using the Huber-White sandwich estimator of variance

All models include fixed country effects

Table 4. MLE estimates for Organization-level Scientific Novelty

	exp(gene novelty)			exp(exp(all term novelty))		
Model	1	2	3	4	5	6
Constant	1.922***	1.959***	1.992***	8.006***	8.052***	8.113***
Research Organization						
tenure	-0.054***	-0.046***	-0.042***	-0.069**	-0.055**	-0.053*
tenure^2	0.002***	0.002***	0.001***	0.002*	0.002*	0.002*
citation centrality	-0.559*	-0.556*	-0.614**	-0.959†	-0.911†	-0.957*
coauthor centrality	0.471	0.313	0.093	1.105	0.834	0.578
coauthor betweenness centrality	-6.874*	-1.826	-7.973*	-20.609*	-4.470	-18.575†
School						
(reference: general university)						
agricultural university	0.031	-0.041	-0.038	0.189	0.205	0.225
agriculturally influenced univ.	-0.093**	-0.079*	-0.076*	-0.122	-0.112	-0.109
medical school	-0.061*	-0.056*	-0.053†	-0.070	-0.067	-0.061
science & technology university	0.049	0.059	0.060	0.038	0.064	-0.062
research institute	-0.019	-0.012	-0.013*	0.007	0.033	0.037
government agency	-0.023	-0.030	-0.026	0.064	0.042	0.049
hospital	-0.006	0.013	0.019	0.410*	0.441*	0.405*
Industry ties and interactions						
relational tie (pi-aggregated)		-0.094***	-0.243***		-0.265**	-0.698**
transactional tie (pi)		-0.102***	-0.116**		-0.210*	-0.339
relational tie		0.045	0.144		0.039	-0.168
transactional tie		-0.082†	-0.230*		-0.010	0.145
relational tie x coauthor centrality			0.827*			2.459*
transactional tie x c.			0.262			0.165
relational tie (org) x c.			-0.752†			0.649
transactional tie (org) x c.			1.204*			-0.259
relational tie x applied subfield			0.064			0.434
transactional tie x a.			-0.036			0.297
relational tie (org) x a.			-0.078			0.362
transactional tie (org) x a.			0.026			-0.313
x^2	9,923.82***	13,484.91***	4,134.14***	55,758.08***	25,612.57***	2,840,000.00***
Number	2,628	2,628	2,628	4,685	4,685	4,685
Number of organizations	935	935	935	1,572	1,572	1,572
Range of organization-pub-years	1:14	1:14	1:14	1:19	1:19	1:19

† $p < .1$ * $p < .05$ ** $p < .01$ *** $p < .001$

Standard errors (coefficient significances) are determined using the Huber-White sandwich estimator of variance

All models include fixed country and scientific subfield effects

Table 5. MLE Estimates for PI-level Gene Persistence

	gene persistence (dichotomous) [a]			Ln(gene persistence)		
Model	1	2	3	4	5	6
Constant	1.018***	1.777***	1.957***	-0.742***	-0.759***	-0.712***
Principal Investigator						
tenure	-0.345***	-0.343***	-0.343***	-0.061**	-0.062**	-0.059**
tenure^2	0.023***	0.023***	0.023***	0.003*	0.003*	0.002*
citation centrality	-2.773***	-2.735***	-2.699***	-3.154***	-3.167***	-3.184***
coauthor centrality	3.396***	3.410***	2.370***	2.760***	2.754***	2.460**
coauthor betweenness centrality	89.152***	100.506***	90.080***	-3.626	-3.846	-3.441
Department						
(reference: general biology)						
plant biology	-0.039	-0.040	-0.032	0.021	-0.026	-0.023
agriculture	-0.007	-0.006	0.014	-0.070	-0.068	-0.066
medical	0.082	0.088	0.086	-0.028	0.028	0.016
chemistry	0.012	0.010	0.013	0.617**	0.634**	0.654**
food science	-0.081	-0.112	-0.090	0.346	0.320	0.403
other	0.381	0.414	.418			
School						
(reference: general university)						
agricultural university	-0.113	-0.110	-0.128	0.110	0.135	0.134
agriculturally influenced univ.	0.046	0.046	0.048	-0.075	-0.078†	-0.075†
medical school	0.037	0.037	0.040	0.018	0.008	0.011
science & technology university	0.057	0.057	0.089	0.017	0.024	0.025
research institute	-0.032	-0.027	-0.030	-0.126†	-0.126†	-0.126†
government agency	-0.096†	-0.097†	-0.097†	-0.072	-0.079	-0.074
hospital	-0.208	-0.202	-0.159	0.029	0.046	0.052
Industry ties and interactions						
relational tie		-0.162**	-0.604***		0.022	0.063
transactional tie		-0.083	-0.565**		0.025	-0.049
relational tie (organization-level)		-0.006	-0.352†		-0.053	0.005
transactional tie (org)		-0.097	-0.425*		0.124	-0.188
relational tie x coauthor centrality			2.821***			-0.513
transactional tie x c.			2.653**			0.375
relational tie (org) x c.			1.725***			-0.521
transactional tie (org) x c.			1.733†			1.860
relational tie x applied subfield			-0.116			0.007
transactional tie x a.			0.047			-0.014
relational tie (org) x a.			0.243			0.091
transactional tie (org) x a.			0.128†			-0.073
x^2	1,484.62		1,569.890***	899.86***	911.56***	937.00***
Number	10,611	10,611	10,611	2,459	2,459	2,459
Number of PI's	4,989	4,989	4,989	821	821	821
Range of PI-pub-years	1:19	1:19	1:19	1:13	1:13	1:13

† $p < .1$ * $p < .05$ ** $p < .01$ *** $p < .001$

Standard errors (coefficient significances) are determined using the Huber-White sandwich estimator of variance

All models include fixed country and scientific subgroup effects

[a] Probit models. In each (models 1-3), significant coefficient estimates for Europe, Asia and Britain and its former colonies reveal that they are less persistent in studying genes than their U.S. peers.

Table 6. MLE Estimates for PI-level All Term Persistence

	All term persistence (dichotomous)[a]			Ln(all term persistence)		
Model	1	2	3	4	5	6
Constant	1.627***	1.082***	1.618***	1.567***	1.572***	1.513***
Principal Investigator						
tenure	-0.373***	-0.372***	-0.373***	0.044***	0.044***	0.045***
tenure^2	0.035***	0.035***	0.035***	-0.001*	-0.001*	-0.001*
citation centrality	-0.269	-0.295	-0.295	-0.428	-0.438	-0.424
coauthor centrality	1.210**	1.221**	1.161*	-0.403	-0.392	-0.763
coauthor betweenness centrality	245.827***	234.818***	233.356***	17.065***	16.049***	13.770***
Department						
(reference: general biology)						
plant biology	-0.019	-0.018	-0.017	-0.065†	-0.064†	-0.062†
agriculture	-0.040	-0.037	-0.034	-0.062	-0.062	-0.060
medical	-0.026	-0.031	-0.030	0.012	0.006	0.017
chemistry	0.059	0.059	0.057	-0.256	-0.263	-0.271
food science	-0.211	-0.208	-0.220	0.186	0.205†	0.212†
other	-0.550	-0.567	-.677	0.097	0.091	0.081
School						
(reference: general university)						
agricultural university	-0.112	-0.099	-0.113	-0.037	-0.045	-0.051
agriculturally influenced univ.	-0.050	-0.050	-0.051	-0.045	-0.045	-0.042
medical school	0.073†	0.075†	0.075	-0.018	-0.016	-0.017
science & technology university	0.055	0.050	0.060	-0.125	-0.128	-0.119
research institute	-0.016	-0.015	-0.018	-0.101*	-0.102*	-0.103*
government agency	-0.016	-0.013	-0.014	-0.071	-0.069	-0.065
hospital	-0.146	-0.163	-0.167	-0.107	-0.116	-0.098
Industry ties and interactions						
relational tie		0.053	0.024		0.011	-0.146†
transactional tie		0.170†	-0.087		0.049*	-0.015
relational tie (organization-level)		-0.116	-0.034		0.044	0.154
transactional tie (org)		0.109	0.012		-0.034	-0.225†
relational tie x coauthor centrality			0.728			0.808*
transactional tie x c.			-0.391			0.497
relational tie (org) x c.			-0.216			-0.699
transactional tie (org) x c.			0.246			0.978†
relational tie x applied subfield			-0.215†			0.026
transactional tie x a.			0.095			-0.080
relational tie (org) x a.			-0.141			0.025
transactional tie (org) x a.			0.172			0.059
x^2	572.47***	580.21***	584.92***	34,072.29***	51,509.59***	50,020.43***
Number	10,611	10,611	10,611	4,613	4,613	4,613
Number of PI's	4,989	4,989	4,989	1,613	1,613	1,613
Range of PI-pub-years	1:19			1:17	1:17	1:17

† $p < .1$ * $p < .05$ ** $p < .01$ *** $p < .001$

Standard errors (coefficient significances) are determined using the Huber-White sandwich estimator of variance

All models include fixed country and scientific subgroup effects

[a] Probit models. The regional coefficient estimates of these models (1-3) indicate that science produced in Asia is less persistent than that produced in the U.S.

Table 7. MLE Estimates for RO-level Gene Persistence

	gene persistence (dichotomous)[a]			Ln(gene persistence)		
Model	1	2	3	4	5	6
Constant	1.516***	1.503***	1.650***	-1.734***	-1.723***	-1.593***
Research Organization						
tenure	-0.237***	-0.236***	-0.234***	-0.045*	-0.041†	-0.030
tenure^2	0.010***	0.010***	0.010***	0.002**	0.002*	0.001
citation centrality	-5.020***	-5.075***	-5.248***	-0.193	-0.188	-.462
coauthor centrality	5.732***	5.725***	4.756***	-0.882	1.00	-1.78
coauthor betweenness centrality	943.998***	945.450***	859.143***	41.620***	42.115***	27.164***
School						
(reference: general university)						
agricultural university	-0.344	-0.330	-0.284	0.296	0.276	0.255
agriculturally influenced univ.	-0.053	-0.061	-0.049	-0.044	-0.029	-0.026
medical school	-0.222**	-0.231**	-0.207*	-0.085	-0.091	-0.090
science & technology university	0.122	-0.114	-0.134	-0.256	-0.267	0.255
research institute	-0.037	-0.040	0.039	-0.106	-0.109	-0.117
government agency	-0.146†	-0.145	-0.145†	-0.198	-0.207	-0.197
Hospital	-0.126	-0.164	-0.115	0.095	0.085	0.095
Region						
(reference: U.S.)						
Britain & former colonies: AU, NZ, CA	0.271*	-0.261*	-0.291	n/a[b]	n/a	n/a
Europe	-0.183*	0.158†	-0.192			
Africa	0.154	0.182	0.136†			
South America	-0.376	-0.343	-0.354			
Asia	-0.238*	-0.216*	0.213			
Middle East	0.083	0.112	0.022			
Industry ties and interactions						
relational tie		-0.174*	-0.477***		-0.109†	-0.409***
transactional tie		0.109	-0.517***		0.057	0.034
relational tie (organization-level)		0.260†	-0.207		0.103	0.213
transactional tie (org)		-0.064	-0.241		0.111	-0.465*
relational tie x coauthor centrality			4.088***			2.423***
transactional tie x c.			6.199***			0.555
relational tie (org) x c.			4.466			0.386
transactional tie (org) x c.			0.533			1.356
relational tie x applied subfield			-0.078			-0.126
transactional tie x a.			-0.062			-0.212
relational tie (org) x a.			0.028			0.483
transactional tie (org) x a.			0.488			-0.486
x^2	798.69***	807.98***	888.96***	3247.20***	2,873.77***	2,262.21***
Number	4,974	4,974	4,974	1,633	1,633	1,633
Number of PI's	1,645	1,645	1,645	354	354	354
Range of PI-pub-years	1:21	1:21	1:21	1:13	1:13	1:13

† p < .1 * p < .05 ** p < .01 *** p < .001

Standard errors (coefficient significances) are determined using the Huber-White sandwich estimator of variance

[a] Probit models

[b] Included individual country and scientific subfield dummies

Figure 1. University Exploration versus Industry Exploitation Strategies in Science

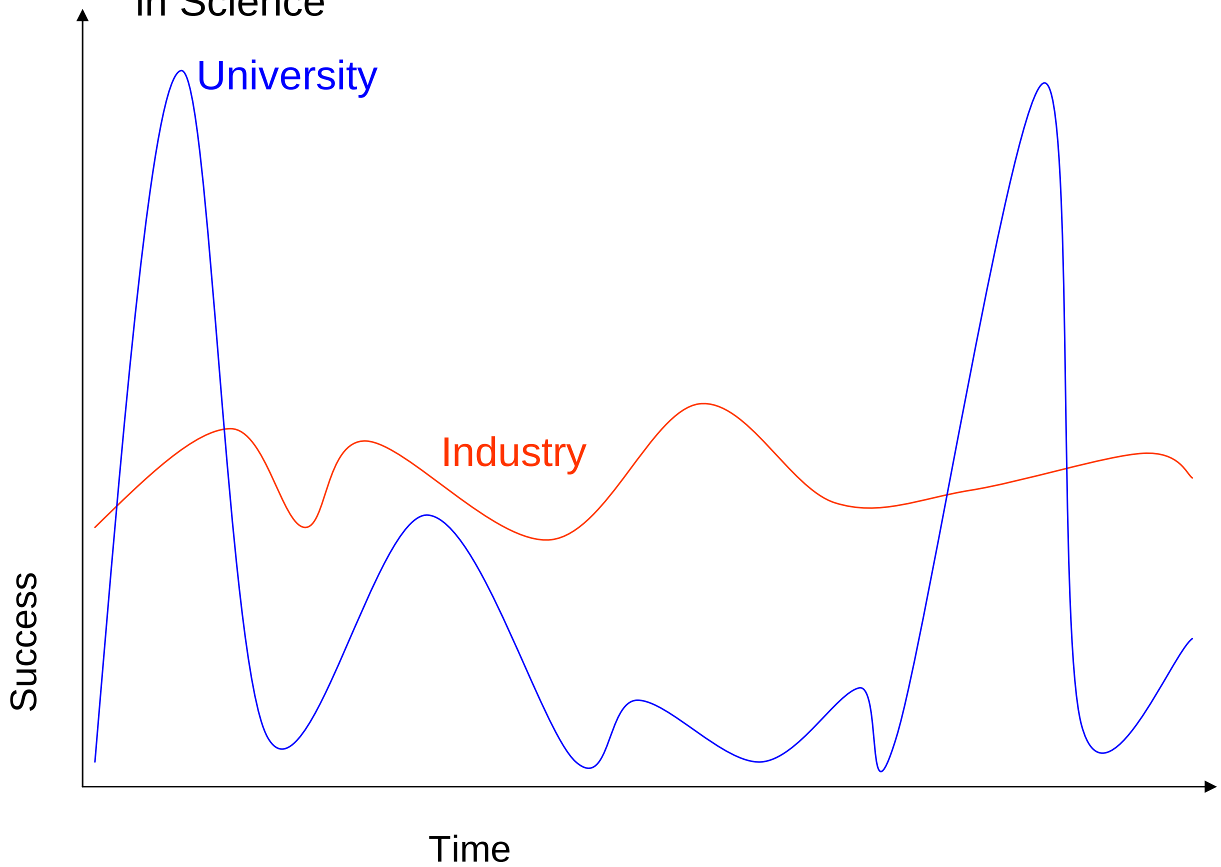


Figure 2.

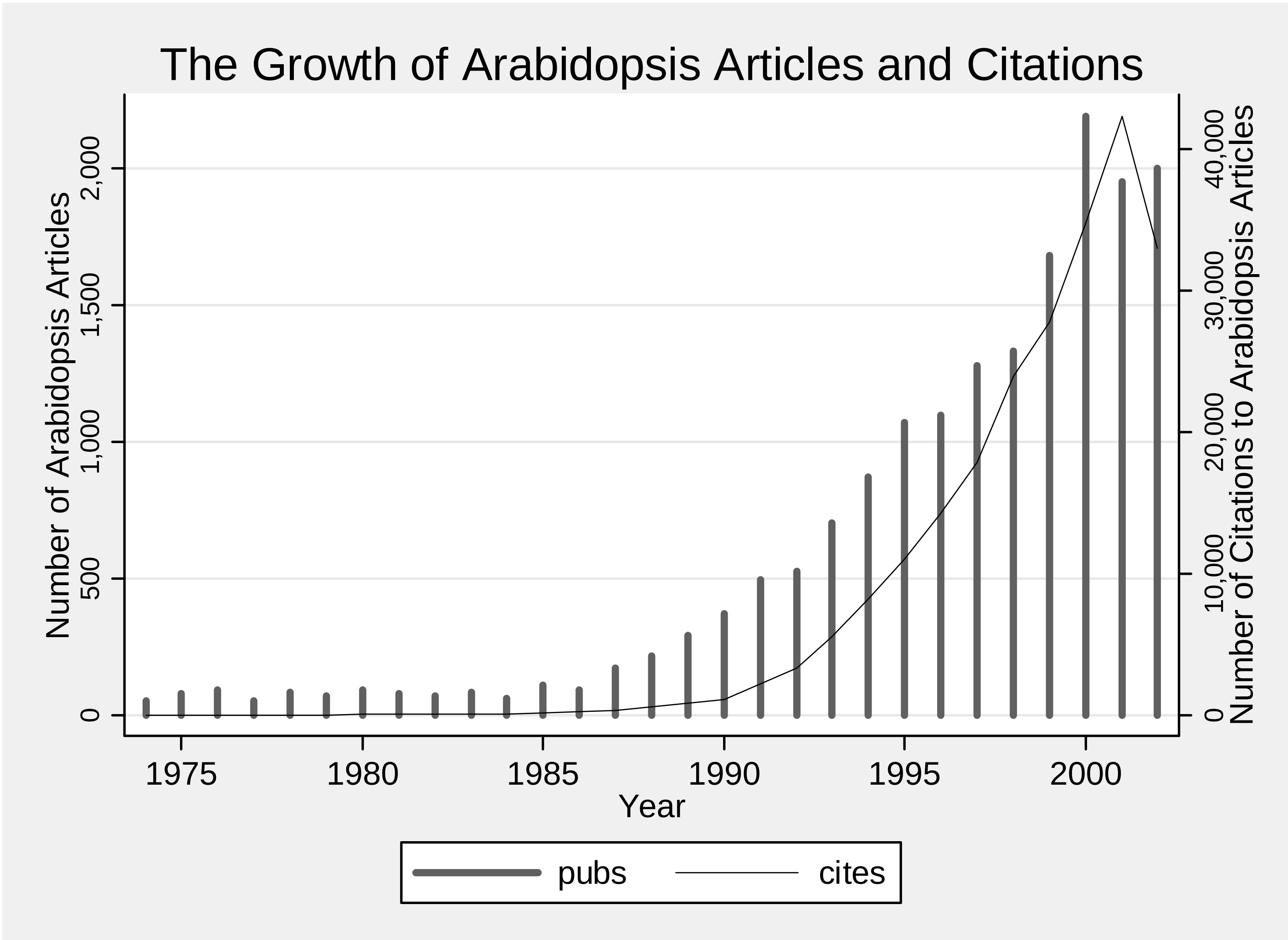

The Growth of Arabidopsis Articles and Citations
Number of Arabidopsis Articles
0
500
1,000
1,500
2,000
Number of Citations to Arabidopsis Articles
0
10,000
20,000
30,000
40,000
1975
1980
1985
1990
1995
2000
Year
pubs
cites

Figure 3. Coding Principle Investigator (PI) Ties to Companies

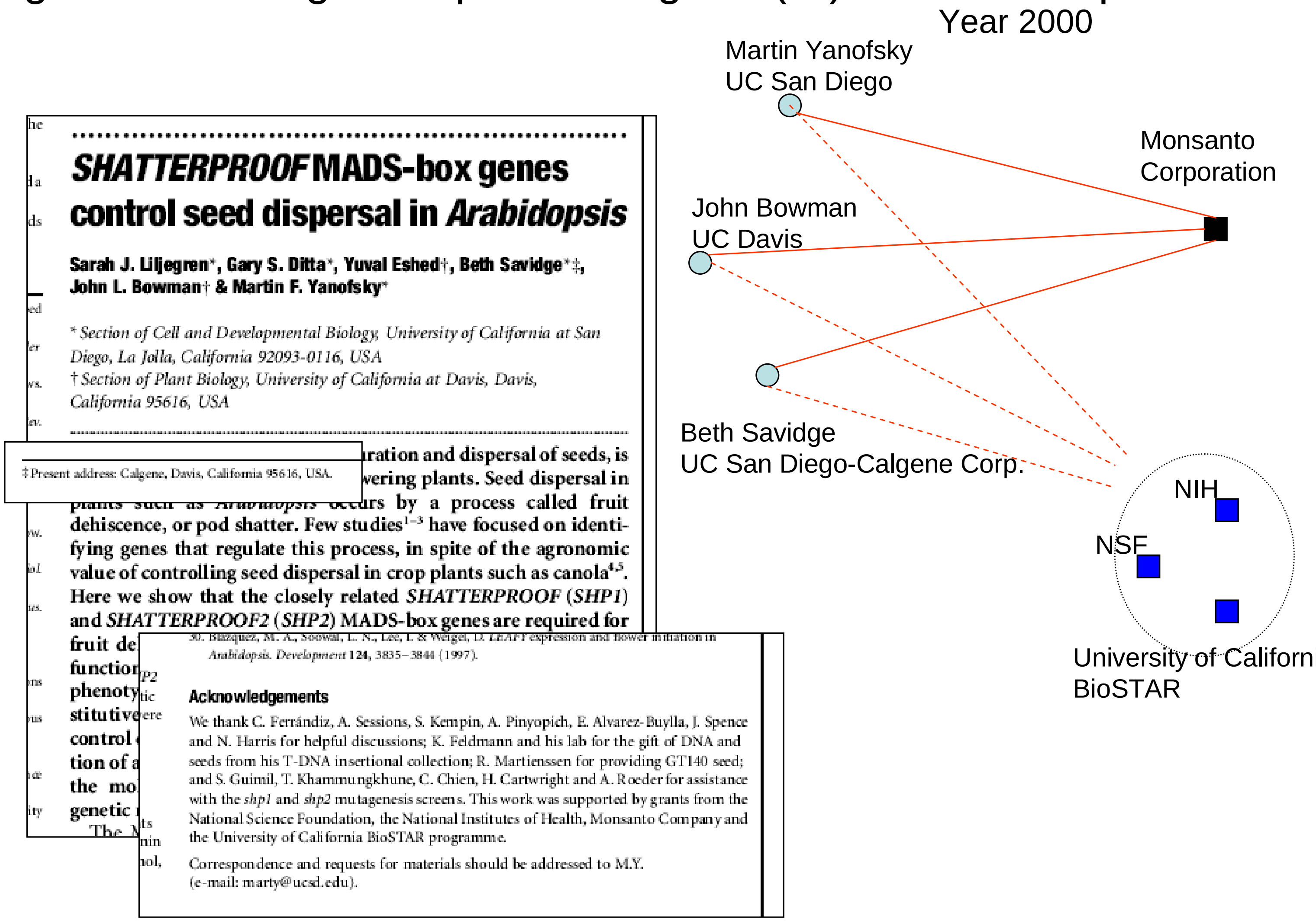
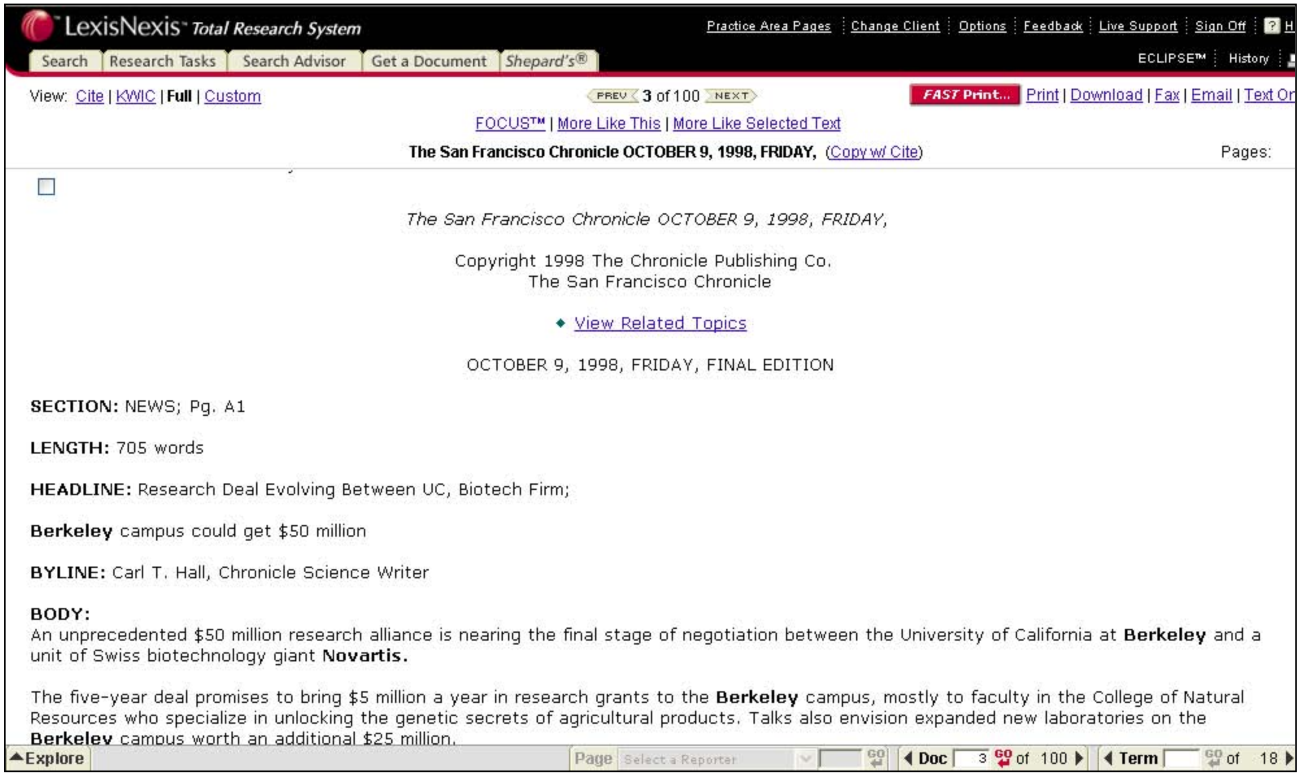


Figure 4. Coding Research Organization (RO) Ties to Companies


LexisNexis Total Research System

Practice Area Pages | Change Client | Options | Feedback | Live Support | Sign Off

Search | Research Tasks | Search Advisor | Get a Document | Shepard's®

ECLIPSE™ | History

View: Cite | KWIC | **Full** | Custom

PREV 3 of 100 NEXT

FAST Print... Print | Download | Fax | Email | Text On

FOCUS™ | More Like This | More Like Selected Text

The San Francisco Chronicle OCTOBER 9, 1998, FRIDAY, (Copy w/ Cite) Pages:

The San Francisco Chronicle OCTOBER 9, 1998, FRIDAY,

Copyright 1998 The Chronicle Publishing Co.
The San Francisco Chronicle

View Related Topics

OCTOBER 9, 1998, FRIDAY, FINAL EDITION

SECTION: NEWS; Pg. A1

LENGTH: 705 words

HEADLINE: Research Deal Evolving Between UC, Biotech Firm;

Berkeley campus could get $50 million

BYLINE: Carl T. Hall, Chronicle Science Writer

BODY:
An unprecedented $50 million research alliance is nearing the final stage of negotiation between the University of California at **Berkeley** and a unit of Swiss biotechnology giant **Novartis.**

The five-year deal promises to bring $5 million a year in research grants to the **Berkeley** campus, mostly to faculty in the College of Natural Resources who specialize in unlocking the genetic secrets of agricultural products. Talks also envision expanded new laboratories on the **Berkeley** campus worth an additional $25 million.

Explore | Page | Select a Reporter | Doc 3 of 100 | Term of 18

Years 1998-2003

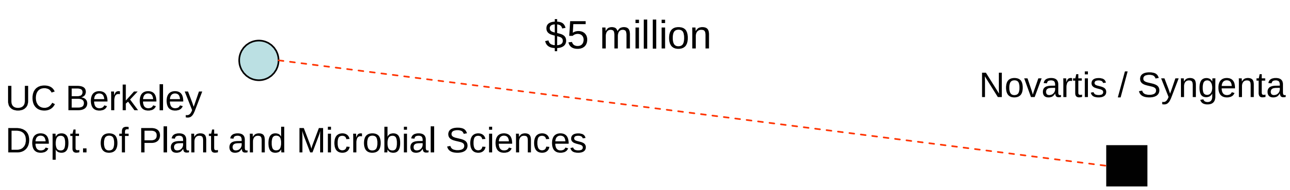


Figure 5. Citation Network of *Arabidopsis* Articles

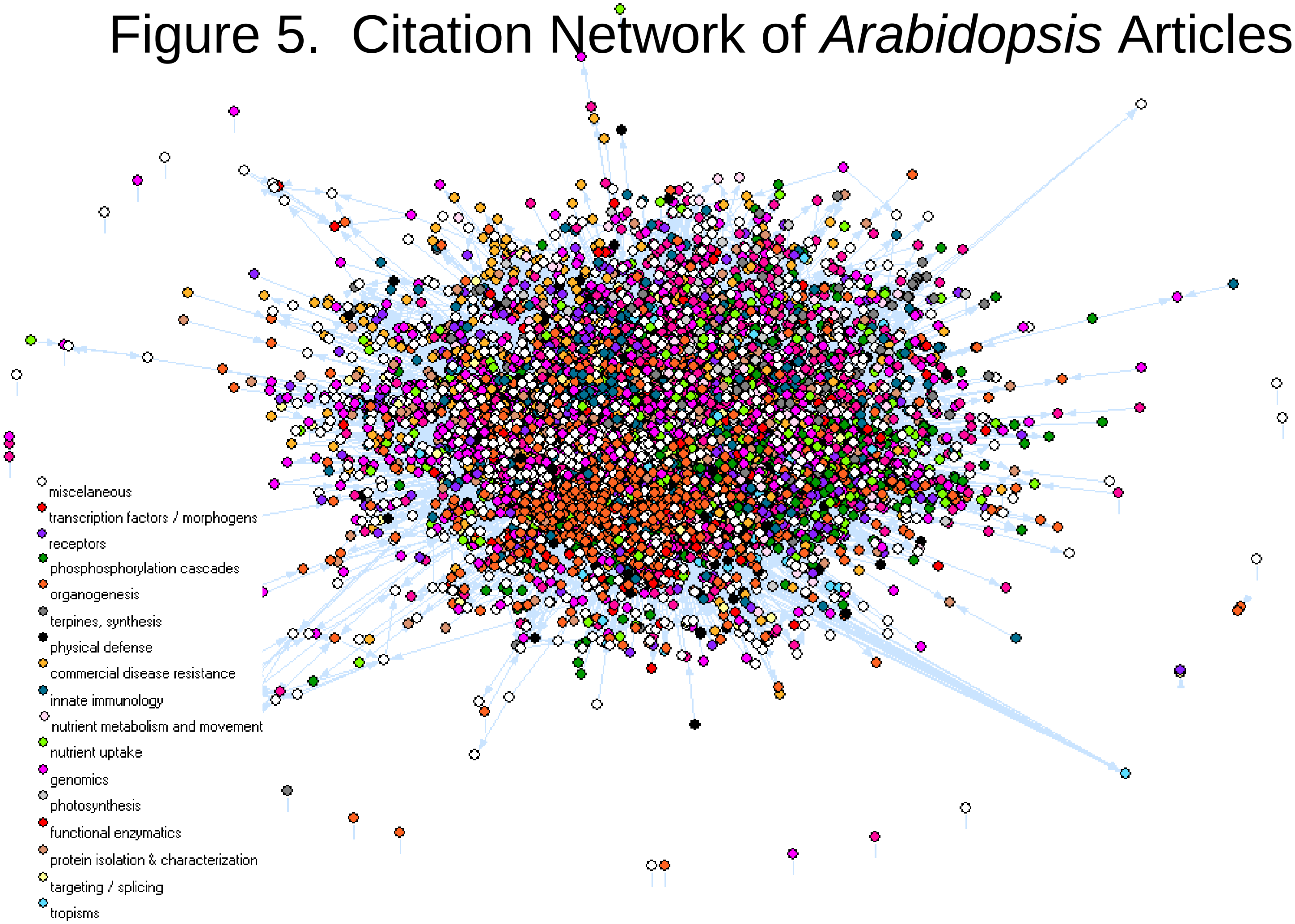
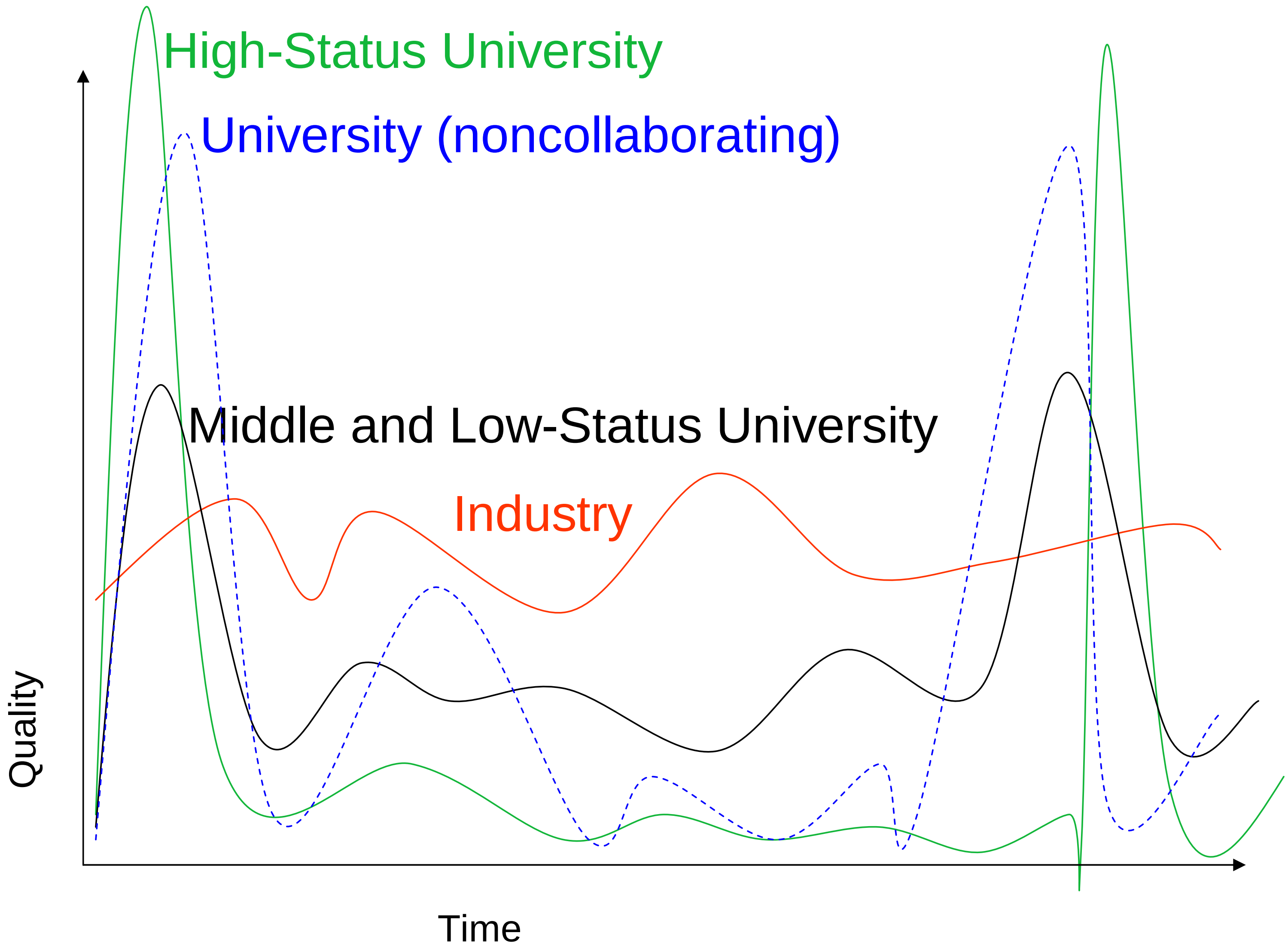


Figure 6. Universities Collaborating with Industry



List of References

Audretsch, David B., and Paula E. Stephan. 1996. “Company-Scientist Locational Links: The case of biotechnology.” *American Economic Review* 86:641-652.

Batelj, Vladimir, and A. Mvar. 1998. “Pajek - Program for Large Network Analysis.” *Connections* 212:47-57.

Ben-David, Joseph. 1991. “Scientific Growth” in *Scientific Growth*. Berkeley: University of California Press.

Boley, Daniel. 1997. “Principle Direction Divisive Partitioning.” : http://www-users.cs.umn.edu/~boley/Distribution/PDDP2.html.

Chang, Jeffrey, Hinrich Schutz, and Russ Altman. 2003. “GAPSCORE: Finding gene and protein names one word at a time.” *Stanford Working Paper* .

Charles, Daniel. 2001. *Lords of the Harvest: Biotech, big money, and the future of food*. Cambridge, MA: Perseus publishing.

Cole, Jonathan, and Stephen Cole. 1973. *Social Stratification in Science*. Chicago: University of Chicago Press.

Collins, Harry M. 2000. “Surviving Closure: Post-rejection adaptation and plurality in science.” *American Sociological Review* 65:824-845.

Cooke, Robert. 2001. *Dr. Folkman's War: Angiogenesis and the struggle to defeat cancer*: Random House.

Crane, Diana. 1969. “Social Structure in a Group of Scientists: A test of the "invisible college" hypothesis.” *American Sociological Review* 34:335-352.

Dasgupta, Partha, and E. Maskin. 1987. “The Simple Economics of Research Portfolios.” *Economic Journal* 97.

Dasgupta, Partha, and Paul David. 1994. “Towards a New Economics of Science.” *Research Policy* 23:487-521.

David, Paul A., Dominique Foray, and W. Edward Steinmueller. 1998. “The Research Network and the New Economics of Science: From metaphors to organizational behaviors.” in *The Organization of Innovative Activities in Europe*, edited by A. Gambardella and F. Malerba. Cambridge, UK: Cambridge University Press.

Dertouzos, Michael L., R. K. Lester, Robert Solow, and MIT Commission on Industrial Productivity. 1989. *Made in America: Regaining the productive edge*. Cambridge, MA: MIT Press.

DiMaggio, Paul J., and Walter W. Powell. 1983. “The Iron Cage Revisited: Institutional isomorphism and collective rationality in organizational fields.” *American Journal of Sociology* 48:147-160.

Etzkowitz, Henry. 1983. “Entrepreneurial Scientists and Entrepreneurial Universities in American Academic Science.” *Minerva* 21:198-233.

Etzkowitz, Henry, and Loet Leydesdorff (Eds.). 1995. *Universities and the Global Knowledge Economy: A triple helix of university-industry-government relations.* Amsterdam: University of Amsterday.

Fleming, Lee, and Olav Sorenson. 2001. “Technology as a Complex Adaptive System: Evidence from patent data.” *Research Policy* 30:1019-1039.

Heller, Michael A., and Rebecca S. Eisenberg. 1998. “Can Patents Deter Innovation? The anticommons in biomedical research.” *Science* 280:698.

Josephson, Matthew. 1959. *Edison*. New York: McGraw-Hill.

Kenney, Martin. 1986. *Biotechnology: The university-industrial complex*. New Haven: Yale University Press.

Kornhauser, William. 1962. *Scientists in Industry: Conflict and accommodation*. Berkeley: University of California Press.

Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lee, Chong-Moon, William F. Miller, Marguerite Gong Hancock, and Henry Rowen (Eds.). 2000. *The Silicon Valley Edge: A habitat for innovation and entrepreneurship*. Palo Alto: Stanford University Press.

March, James G. 1991. “Exploration and Exploitation in Organizational Learning.” *Organization Science* 2:71-87.

Marcson, Simon. 1960. *The Scientist in American Industry*. New York: Harper.

Merrill, S. A., and R. S. Cooper. 1999. “Trends in Industrial Research and Development: Evidence from national data sources.” Pp. 99-116 in *Securing America's Industrial Strength, National Research Council Board on Science, Technology and Economic Policy*. Washington, D. C.: National Academy Press.

Merton, Robert K. 1942. “Science and Technology in a Democratic Order.” *Journal of Legal and Political Sociology* 1:115-126.

Merton, Robert K. 1968. “The Matthew Effect in Science: The Reward and communication system of science.” *Science* 199:55-63.

Moody, James. 2001. “Peer Influence Groups: Identifying dense clusters in large networks.” *Social Networks* 23:261-283.

Mowery, David C. 1999. “America's Industrial Resurgence?: An Overview.” Pp. 1-16 in *U.S. Industry in 2000: Studies in Competitive Performance*. Washington, D.C.: National Academy Press.

National Research Council. 1997. *Intellectual Property Rights and Research Tools in Molecular Biology*. Washington, D.C.: National Academy Press.

Nelson, Richard R. 2003. “The Market Economy and the Scientific Commons.” *Research Policy* 33:455-471.

Nowotny, Helga, Peter Scott, and Michael Gibbons. 1994. *The New Production of Knowledge: The production of science and research in contemporary societies*. London: Sage Publications.

Nowotny, Helga, Peter Scott, and Michael Gibbons. 2001. *Re-Thinking Science: Knowledge and the public in the age of uncertainty*. Cambridge: Polity Press.

Nowotny, Helga, Peter Scott, and Michael Gibbons. 2003. “Mode 2 Revisited: The new production of knowledge.” *Minerva* forthcoming.

Owen-Smith, Jason. 2003. “From Separate Systems to a Hybrid Order: Accumulative advantage across public and private science.” *Research Policy* 32:1081-1104.

Padgett, John F., Ansell, Christopher K. (1993) "Robust Action and the Rise of the Medici 1400-1434." *American Journal of Sociology* 98:1259-1319.

Powell, Walter W. 1996. “Interorganizational Collaboration in the Biotechnology Industry.” *Journal of*

Institutional and Theoretical Economics 120:197-215.

Powell, Walter W., Douglas R. White, Kenneth W. Koput, and Jason Owen-Smith. 2005. “Network Dynamics and Field Evolution: The Growth of Inter-organizational Collaboration in the Life Sciences.” *American Journal of Sociology* forthcoming.

Press, Eyal, and Jennifer Washburn. 2000. “The Kept University.” *Atlantic Monthly* 285:39-54.

Reimers, Niels. 1984. *The Government-Industry-University Interface: Improving the Innovation Process*. Englewood Cliffs, NJ: Prentice-Hall.

Rogers, E., and L. Larsen. 1984. *Silicon Valley Fever*. New York: Basic Books.

Rosenberg, Nathan, and Richard R. Nelson. 1994. “American Universities and Technical Advance in Industry.” *Research Policy* 23:323-348.

Rosenberg, Nathan. 2000. “America's University/Industry Interfaces, 1945-2000.” *Manuscript, Department of Economics, Stanford University* .

Saxenian, AnnaLee. 1985. “Silicon Valley and Route 128: Regional prototypes or historic exceptions?” Pp. 91-105 in *High Technology, Space, and Society*, edited by Manuel Castells. Beverly Hills: Sage Publications.

Saxenian, AnnaLee. 1994. *Regional Advantage: Culture and competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.

Sorenson, Olav, and Lee Fleming. 2001a. “Science as a Map in Technological Search.” *Working Paper* .

von Hippel, Eric. 1988. *The Sources of Innovation*. New York: Oxford University Press.

Walbot, Virginia. 2000. “A Green Chapter in the Book of Life.” *Nature* 408:794-795.

Zucker, Lynne G., Michael R. Darby, and Jeff S. Armstrong. 1998. “Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises.” *American Economic Review* 88:290-306.