

COPYRIGHT FOR ENGINEERED DNA: AN IDEA WHOSE TIME HAS COME?

*Christopher M. Holman J.D., Ph.D.**

ABSTRACT

The rapidly emerging field of synthetic biology has tremendous potential to address some of the most compelling challenges facing our planet by providing clean renewable energy, nutritionally-enhanced and environmentally friendly agricultural products, and revolutionary new life-saving cures. However, leaders in the synthetic biology movement have voiced concern that biotechnology's current patent-centric approach to intellectual property is in many ways ill-suited to meet the challenge of synthetic biology, threatening to impede follow-on innovation and open access technology. For years, copyright and patent protection for computer software have existed side-by-side, the two forms of intellectual property complementing one another. Numerous academic commentators have rejected the notion of copyright for engineered genetic sequences (the primary output of synthetic biology), based in large part upon a misinterpretation of copyright law and/or a failure to appreciate the profound advances in synthetic biology that have occurred in recent years. This article makes a case for extending copyright protection to engineered DNA, based in large part upon the striking analogy between engineered genetic sequences and computer programs, made more compelling by the growing convergence of software engineering and synthetic biology. The major doctrinal leap occurred thirty years ago, when Congress and the courts sanctioned the use of copyright to protect computer programs. A further extension to engineered DNA would represent a comparatively modest incremental expansion, and would be comfortably supported by current copyright doctrine without any need to amend the Copyright Act (although certain clarifying and limiting amendments would be desirable). The Article

* Associate Professor of Law, University of Missouri-Kansas City School of Law. I wish to thank participants of the 2009 West Virginia Law Review Symposium on Digital Entrepreneurship and the 2010 Biolaw 4.0 Conference at the University of Kansas School of Law for thoughtful commentary.

critiques, and to a large extent refutes, a variety of doctrinal and policy arguments that have been raised by commentators against extending copyright protection to engineered DNA. It concludes by outlining a number of positive policy objectives that could be addressed by such an extension of copyright law.

I.	INTRODUCTION	701
II.	PREVIOUS PROPOSALS FOR EXTENDING COPYRIGHT PROTECTION TO ENGINEERED GENETIC SEQUENCES HAVE BEEN LARGELY DISMISSED	703
III.	HISTORICALLY, THE EXPANSION OF COPYRIGHT LAW HAS BEEN DRIVEN BY ADVANCES IN TECHNOLOGY	706
IV.	THE ANALOGY BETWEEN SOFTWARE AND SYNTHETIC GENETIC MATERIAL SUPPORTS AN EXTENSION OF COPYRIGHT TO ENGINEERED GENETIC SEQUENCES	711
V.	THE SYNTHETIC BIOLOGY REVOLUTION.....	716
VI.	SOME OF THE ARGUMENTS AGAINST COPYRIGHT PROTECTION FOR ENGINEERED DNA ARE BASED ON AN INCOMPLETE UNDERSTANDING OF THE STATE-OF-THE-ART IN GENETIC ENGINEERING.....	720
VII.	SOME OF THE ARGUMENTS AGAINST COPYRIGHT FOR DNA ARE BASED ON A MISREADING OF COPYRIGHT LAW.....	722
VIII.	THE U.S. CONSTITUTION AUTHORIZES COPYRIGHT PROTECTION FOR ENGINEERED GENETIC SEQUENCES.....	724
IX.	IT IS UNNECESSARY FOR THE COPYRIGHT ACT TO EXPLICITLY IDENTIFY ENGINEERED GENETIC SEQUENCES AS COPYRIGHTABLE SUBJECT MATTER.....	726
X.	THE DEVELOPMENT OF ENGINEERED GENETIC SEQUENCES ENTAILS SUFFICIENT EXPRESSIVE CHOICE TO WARRANT COPYRIGHT PROTECTION	729
XI.	COPYRIGHT IS NOT PRECLUDED BY THE PRACTICAL UTILITY OF GENETIC SEQUENCES	731
XII.	SOME LEVEL OF UNPREDICTABILITY DOES NOT PRECLUDE COPYRIGHT PROTECTION FOR ENGINEERED GENETIC SEQUENCES.....	732
XIII.	PROVIDING COPYRIGHT FOR DNA DOES NOT IMPLY THAT COPYRIGHT IS GENERALLY AVAILABLE FOR CHEMICAL COMPOUNDS CAPABLE OF CONVEYING INFORMATION	734
XIV.	CONCLUSION	737

I. INTRODUCTION

A team led by Craig Venter recently achieved a scientific milestone by “booting up” a viable bacterium using a synthetic bacterial genome composed entirely of DNA synthesized in the laboratory.¹ Their success has drawn heightened public attention to “synthetic biology,” the leading edge of the biotechnology revolution. For example, President Obama responded to the announcement by directing his bioethics commission to conduct a six-month study on the potential benefits and risks of synthetic biology, and asked the panel to recommend “any actions the Federal government should take to ensure that America reaps the benefits of this developing field of science while identifying appropriate ethical boundaries and minimizing identified risks.”²

Congress likewise wasted no time in responding to Dr. Venter’s announcement, promptly convening a hearing to explore the potential and risks of synthetic biology.³ Leading synthetic biologists provided compelling examples of the potential fruits of synthetic biology, including the production of clean water, new biofuels,⁴ and drugs for diseases that afflict people in the developing world.⁵ One of the scientists who testified at the hearing, Drew Endy of Stanford, identified an important role for intellectual property in the development of synthetic biology, but noted that for a variety of reasons patents might be ill-suited for the task.⁶ He proposed that the advance of synthetic biology might be better served by the development of alternative forms of intellectual property protection.⁷

Historically, patents have served as the primary form of intellectual property protection for biotechnology, but as we stand poised to move into a new era of synthetic biology, it is worth considering Professor Endy’s suggestion that other forms of intellectual property protection might be more suited to the task. As noted by Dr. Endy, patents are expensive and often take many years to issue, which is a problem because synthetic biology is characterized by rapid and profuse incremental innovations.⁸ Patents also have the potential to unduly impede follow-on research, a characteristic that has been the subject of much debate and concern in recent years, particularly with respect to patents

¹ Elizabeth Pennisi, *Genomics: Synthetic Genome Brings New Life to Bacterium*, 328 *SCIENCE* 958 (2010).

² Jeffery Mervis, *Obama Orders Review of Synthetic Biology*, *SCIENCE*, May 20, 2010, available at <http://news.sciencemag.org/scienceinsider/2010/05/obama-orders-review-of-synthetic.html>.

³ *Developments in Synthetic Genomics and Implications for Health and Energy: H. Comm. on Energy and Commerce*, 111th Cong. (2010).

⁴ *Id.* (statement of J. Craig Venter, Ph. D, President, J. Craig Venter Institute).

⁵ *Id.* (statement of Jay Keasling, Professor, University of California at Berkeley).

⁶ *Id.* (statement of Drew Endy, Professor, Stanford University).

⁷ *Id.* at 4.

⁸ *Id.*

directed towards genetic discoveries and life science research.⁹ The recent lawsuit brought by the ACLU challenging the validity and constitutionality of gene patents was motivated largely by a widely held perception that gene patents are hindering biomedical innovation,¹⁰ as have been proposals to limit the patentability of genetic inventions and research tools in general.¹¹

In short, there are many who believe that the current IP regime is suboptimal for promoting innovation in biotechnology, and the concern will continue to mount as synthetic biology advances. But perhaps history can help chart a better course. More than thirty years ago, a similar situation existed with respect to the then nascent software industry, with respect to which there was a similar concern that the IP status quo was inadequate to provide appropriate incentives for innovation and dissemination of technology.¹² At that time, it was an open question as to whether patent protection was available, or even appropriate, for computer programs, and copyright was enlisted to provide intellectual property protection for software.¹³

For many years copyright functioned as the primary mode of intellectual property protection for software. Eventually, the scope of subject matter recognized as eligible for patent protection expanded to encompass software, but to this day copyright continues to play an important, and in many ways complementary, role in the protection of software. This Article explores the feasibility of extending copyright protection further so as to encompass engineered synthetic genetic sequences and concludes that the extension is probably justified based in large part upon the analogy between software and engineered DNA.

My proposal to consider extending copyright protection to engineered genetic sequences is far from novel; in fact, it was persuasively made nearly thirty years ago in the early days of biotechnology.¹⁴ However, to date the proposal has not been seriously considered by Congress, the courts or the U.S. Copyright Office, and most commentators who have addressed the issue in recent years have rejected it.¹⁵ But in my view, recent advances in synthetic biology have changed the equation, and as noted by Dr. Endy, it is important that

⁹ COMM. ON INTELLECTUAL PROPERTY RIGHTS IN GENOMIC & PROTEIN RESEARCH & INNOVATION, NAT'L RESEARCH COUNCIL, REAPING THE BENEFITS OF GENOMIC AND PROTEOMIC RESEARCH: INTELLECTUAL PROPERTY RIGHTS, INNOVATION, AND PUBLIC HEALTH (2006).

¹⁰ *Ass'n for Molecular Pathology v. U. S. Patent & Trademark Office*, 702 F. Supp. 2d 181, 190 (S.D.N.Y. 2010).

¹¹ Joshua D. Sarnoff & Christopher Holman, *Recent Developments Affecting the Enforcement, Procurement, and Licensing of Research Tool Patents*, 23 BERKELEY TECH. L.J. 1299, 1342 (2008).

¹² *See infra* Part III.

¹³ *Id.*

¹⁴ Irving Kayton, *Copyright in Living Genetically Engineered Works*, 50 GEO. WASH. L. REV. 191, 191-92 (1982).

¹⁵ *See infra* Part II.

we think outside of the box when considering the optimal intellectual property regime for advancing synthetic biology.

In this Article, I describe the growing technological convergence of software and synthetic biology, and the close analogy between software and engineered genetic sequences, which to my mind justifies extension of copyright detection to engineered DNA. In a nutshell, my position is that the major doctrinal leap occurred thirty years ago when copyright protection was recognized for computer programs. In view of the close analogy between software and engineered DNA, the further extension to encompass engineered genetic sequences is a relatively modest incremental expansion. For the most part, the same considerations which led to the extension of copyright to software apply to engineered DNA.

II. PREVIOUS PROPOSALS FOR EXTENDING COPYRIGHT PROTECTION TO ENGINEERED GENETIC SEQUENCES HAVE BEEN LARGELY DISMISSED

In 1982, in the early days of recombinant DNA technology, Professor Irving Kayton published a prescient article explaining why recombinant genetic sequences are copyrightable subject matter under United States law.¹⁶ In his article he identifies a number of policy objectives that would be promoted by extending copyright protection to genetic information, and observes that “[u]nder certain circumstances, from a practical as well as legal viewpoint, copyright protection may be the only or the most effective way an ‘author’ can protect a valuable genetic ‘work.’”¹⁷

Professor Kayton’s article was published shortly after Congress amended the Copyright Statute to explicitly acknowledge copyright protection for computer programs.¹⁸ At the time, computer programming and genetic engineering were both generally recognized as nascent technologies poised to play an increasingly significant role in the nation’s technological future. He recognized the analogy between the two technologies and explained that a further expansion of copyrightable subject matter to include engineered genetic sequences would be generally consistent with copyright’s recent embrace of computer programs.¹⁹

Professor Kayton’s suggestion that copyright be extended to engineered DNA sequence has been seconded in a few subsequent articles. For example, a student comment set forth a similar proposal in 1987.²⁰ In 1988, shortly before

¹⁶ Kayton, *supra* note 14. As used in this article, the term “genetic engineering” refers to genetic modifications occurring through human intervention, rather than through natural processes.

¹⁷ *Id.* at 192.

¹⁸ *See infra* Part III.

¹⁹ Kayton, *supra* note 14, at 198.

²⁰ Donna Smith, *Copyright Protection for the Intellectual Property Rights to Recombinant Deoxyribonucleic Acid: A Proposal*, 19 ST. MARY’S L.J. 1083 (1988).

becoming a law professor, Dan Burk published an article essentially agreeing with Kayton that the logic behind extending copyright protection to software would also justify copyright protection for engineered genetic sequences, but ultimately concluding that public policy considerations weighed against providing copyright protection for DNA.²¹ In his view, the scope of copyrightable subject matter should be informed by public policy considerations, and for that reason he advised against extending copyright to DNA.

In 2002, Willem “Pim” Stemmer, the scientist who invented “DNA shuffling,”²² wrote a short article in which he assumed that copyright protection was not available for DNA sequences per se.²³ However, as a pioneering scientist actively engaged in the engineering and commercialization of synthetic DNA sequences, Stemmer believed that some form of copyright protection would be beneficial. To circumvent what he perceived to be a prohibition against direct copyright protection for engineered DNA, he outlined a proposal whereby a DNA sequence is converted into music, and then copyrighted as a musical work.²⁴ Although this proposal by a non-lawyer scientist is probably unworkable, it clearly shows that Drew Endy is not the only pioneering synthetic biologists to seriously consider the potential value of copyright protection for engineered DNA sequences.

Overall, however, the consensus appears to have come down heavily against the copyrightability of engineered genetic sequences.²⁵ The U.S. Copyright Office does not appear to have ever officially addressed the question, but has unofficially taken the position that DNA, whether naturally occurring or synthetic, is not copyrightable subject matter and will not grant copyright regis-

²¹ Dan Burk, *Copyrightability of Recombinant DNA Sequences*, 29 JURIMETRICS J. 469, 531–32 (1988) (expressing the view that the logic behind extending copyright protection to computer programs could also be used to justify copyright of recombinant DNA sequences, but that policy considerations appeared to counsel against copyright for DNA).

²² Willem P. C. Stemmer, *DNA Shuffling by Random Fragmentation and Reassembly: In vitro recombination for Molecular Evolution*, 91 PROC. NAT'L. ACAD. SCI. USA 10747 (1994). “DNA shuffling is a way to rapidly propagate beneficial mutations in a directed evolution experiment. It is used to rapidly increase DNA library size.” *DNA Shuffling*, WIKIPEDIA, http://en.wikipedia.org/wiki/Dna_shuffling (last visited Mar. 7, 2011).

²³ Willem P.C. Stemmer, *How to Publish DNA Sequences with Copyright Protection*, 20 NATURE BIOTECHNOLOGY 217 (2002). “DNA shuffling” is a powerful tool for engineering synthetic genes and proteins, and was the basis for the foundation of the biotechnology company Maxygen, for whom the author was previously employed.

²⁴ *Id.*

²⁵ See Brian Gargano, *The Quagmire of DNA Patents: Are DNA Sequences More Than Chemical Compositions of Matter?* 2005 SYRACUSE SCI. & TECH. L. REP. 3 (2005); Sapna Kumar & Arti Rai, *Frontiers of Intellectual Property: Synthetic Biology: The Intellectual Property Puzzle*, 85 TEX. L. REV. 1745 (2007); M. Scott McBride, *Bioinformatics and Intellectual Property Protection*, 17 BERKELEY TECH. L.J. 1331 (2002); Joseph N. Michelotti, *Genes As Intellectual Property*, 11 MICH. ST. J. MED. & LAW 71 (2007); James G. Silva, *Copyright Protection of Biotechnology Works: Into the Dustbin of History?*, 2000 B.C. INTELL. PROP. & TECH. F. 012801 (2000).

tration to gene sequences or DNA molecules.²⁶ There are reports that attempts have been made to register DNA sequences with the copyright office, but no one appears to have been successful or to have appealed the denial of registration.²⁷ My research on the topic failed to identify any judicial decisions addressing the question, and Congress seems to have ignored the issue. Moreover, the biotechnology establishment has for the most part not actively pushed for extending copyright protection to DNA.²⁸

The notion that copyright could extend to genetic sequences does seem counterintuitive, and it is not surprising the initial response of most people has been to reject the notion. Historically, copyright has been the realm of expressive works of art and literature, while patents have been the primary form of intellectual property protection available to technological innovation. The propriety of even patent protection for DNA sequences has been challenged, in part due to the widely held belief that DNA and genes are the blueprint of life, and as such unsuited to individual property rights.²⁹ This antipathy is reflected in current attempts to limit or even bar the patenting of genetic material and information.³⁰ The resistance against property rights in DNA is visceral, due in large part, I believe, to the deeply personal, some would say spiritual, link between DNA and the essence of what it means to be human, and between an individual and his or her own unique DNA.

Professor Kayton reported that he too was initially “shocked and perplexed” when an attendee at a continuing legal education program on copyrights first sparked his interest in the topic by asking him to explain exactly why it is that a genetically engineered organism could not be copyrighted.³¹ However, after shaking off his initial skepticism and his intuition that genetic sequences

²⁶ Silva, *supra* note 25 (citing MICHAEL A. EPSTEIN, MODERN INTELLECTUAL PROPERTY, 458–59 (2d ed. 1992)). The author (Holman) called the U.S. Copyright Office helpdesk in 2009 to ask if engineered genetic sequence could be registered and was informed that genetic sequences cannot be registered because the office considers them not copyrightable. The individual on the phone was unable to identify anything in writing that would support this position, but to the author’s knowledge, she accurately expressed the current views of the Copyright Office.

²⁷ Heidi Williams, *Intellectual Property Rights and Innovation: Evidence from the Human Genome*, Job Market Paper, Nov. 8, 2009 (on file with author).

²⁸ *But see* Stemmer, *supra* note 23 (proposing that genetic sequences could be transposed into music and thus rendered eligible for some form of copyright protection); *Engineering Biology: A Talk with Drew Endy*, EDGE: THE THIRD CULTURE, available at http://www.edge.org/3rd_culture/andy08/andy08_index.html (last visited Mar. 7, 2011) (predicting that in view of advances in synthetic biology “an ownership sharing and innovation framework needs to be developed that moves beyond patent-based intellectual property and recognizes that the information defining the genetic material’s going to be more important than the stuff itself and so you might transition away from patents to copyright”).

²⁹ See Christopher M. Holman, *The Impact of Human Gene Patents on Innovation and Access: A Survey of Human Gene Patent Litigation*, 76 UMKC L. REV. 295, 296 (2007).

³⁰ *Id.*

³¹ See Kayton, *supra* note 14, at 218.

could not be eligible for copyright protection, he opened his mind and embarked upon a rigorous inquiry into the question.

In the midst of this scholarly exercise, Kayton confessed that “every intellectual and emotional prejudice, both sophisticated and primitive, to which [a man] is subject opposed coming to the conclusion [I] finally reached. Copyright protection for engineered DNA sequences seemed ludicrous.”³² Ultimately, however, he was forced to conclude that logic did in fact dictate that some engineered DNA sequences should be eligible for copyright protection, based to a large extent upon the analogy between engineered genetic sequences and computer programs.³³ He also identified a number of policy objectives that would be promoted by extending copyright protection to recombinant DNA.³⁴

I myself shared similar prejudices until I attended a presentation by Professor Andrew Torrance at the University of Kansas, in which he suggested the possibility of copyright protection for engineered genetic sequences.³⁵ This inspired me to explore the issue for myself, and ultimately, I came to agree with Professor Kayton’s conclusion. In fact, I think the arguments for the copyrightability of engineered genetic sequences have been strengthened substantially by dramatic technological developments in genetic engineering and software technology that have transpired over the intervening thirty years, resulting in an increasing convergence of the two historically distinct disciplines.

I think that many readers of this Article, if they maintain an open mind, will have a similar response, and will realize that the real doctrinal leap occurred thirty years ago when Congress and the courts categorized computer programs as copyrightable “literary works.”³⁶ In comparison, the extension of copyright doctrine from computer software to engineered genetic sequences is relatively modest and flows logically and naturally once one has accepted the premise that a set of encoded instructions directed towards a non-human audience is copyrightable.

III. HISTORICALLY, THE EXPANSION OF COPYRIGHT LAW HAS BEEN DRIVEN BY ADVANCES IN TECHNOLOGY

The extension of copyright to encompass engineered genetic sequences would be entirely consistent with the historical development of copyright law, which has been marked by a progressive expansion in the scope of protectable subject matter. The original U.S. copyright statute, enacted in 1790, limited

³² *Id.*

³³ *Id.* at 203–09.

³⁴ *Id.* at 213–14 (“[I]n many circumstances it may be the single most vibrant and flexible form of protection for man-made genetic sequences.”).

³⁵ Professor Andrew Torrance, Presentation at Biolaw 2008 at the University of Kansas School of Law (Nov. 2008).

³⁶ CRAIG JOYCE ET AL., COPYRIGHT LAW 77 (4th ed. 2010). It is “perhaps surprising” that computer programs are comprehended under the rubric of “literary works.” *Id.*

copyright protection to a short list of explicitly enumerated categories of subject matter: essentially, maps, charts and books. Over the next two centuries it expanded to encompass “[d]esigns, engravings, and etchings” (1802), “[m]usical compositions” (1831), “[d]ramatic compositions” (1856), “[p]hotographs” (1865), “[s]tatuary and models” (1870), “[a]ll the writings of an author” (1909), “[m]otion pictures” (1912), “[s]ound recordings” (1972), and ultimately all “[o]riginal works of authorship” (1976).³⁷

As a general rule, copyright has tended to expand in response to new technologies that render the cost of reproducing a work much lower than the cost of initially creating the work, giving rise to an environment where free riders threaten to chill the incentive for creation. In order to maintain an adequate reward for creativity, the scope of copyrightable subject matter is allowed to expand in order to provide creators with a legal barrier to copying.

For example, the development of technologies that facilitated the production of multiple copies of a graphical work, first by etchings, and later lithography and photography, led to the embrace of these works by copyright law. The development of machine-readable media, initially in formats such as piano player rolls, later in analog media such as film and phonorecords, and ultimately in digital recordings, eventually resulted in an expansion of copyright law to encompass works fixed in these media.³⁸

Another recurring theme is that the expansion of copyright often lags far behind the technological advance that prompted the expansion. Copyright law tends to react to new technologies, rather than addressing them proactively. We see this today, for example, in the difficulty copyright law is experiencing in adapting to the development of digital media and new forms of content distribution, but the problem is nothing new.

For example, in the not-too-distant past, copyright protection was only afforded to works expressed in a medium that could be directly interpreted by a human. In *White-Smith*, decided in 1908, the Supreme Court was presented with the question of whether content fixed in a machine-readable format, specifically, music transcribed as an arrangement of holes on a player piano roll, fell within the ambit of copyright protection.³⁹ The Court, somewhat reluctantly, answered no, and held that a player piano roll is not a “copy” of the musical composition recorded thereon, because unlike conventional musical notation a human could not directly read the music encoded in this manner.⁴⁰

White-Smith established that in order to be protectable a “copy” must be “in a form which others can see and read.”⁴¹ Later, Congress implicitly adopted

³⁷ National Commission on New Technological Uses of Copyrighted Works, Final Report at 15 (1979) [hereinafter CONTU Report].

³⁸ MELVILLE NIMMER, NIMMER ON COPYRIGHTS § 2.03(B)(1).

³⁹ *White-Smith Music Publ'g Co. v. Apollo Co.*, 209 U.S. 1 (1908).

⁴⁰ *Id.* at 18.

⁴¹ *Id.* at 17.

White-Smith's restrictive definition of "copy" when it defined the scope of copyrightable subject matter in the 1909 Copyright Act by restricting copyright protection to works that one could "see and read" with the naked eye.⁴² The effect was to deny copyright protection not only for piano rolls, but also for later developed technologies such as sound recordings.

This limitation on copyright protection became increasingly problematic as advances in technology resulted in dramatic reductions in the cost of reproduction in comparison with the cost of producing such creative works. If this rule had been maintained to the present, copyright protection would be unavailable for the fruits of much of today's most important creative activities, such as software, music and other digitally-fixed content.

Fortunately, Congress eventually responded to the disconnect between copyright law and advances in recording technology, overruling its previous exclusion of copies fixed in a machine-readable format, at least with respect to recorded music, by enacting the Sound Recording Amendment in 1971.⁴³ But this reform only came about after the widespread use of technologies that facilitated easy copying of sound recordings created a huge problem of unauthorized "bootleg" copies of musical works, for which a patchwork of state anti-bootlegging statutes provided only partial relief.⁴⁴ It was not until passage of the 1976 Copyright Act, which specifies that fixation can be any medium capable of perception by means of any machine or device now known or later developed, that Congress finally overruled *White-Smith* and expanded copyright protection to encompass machine-readable works in general.⁴⁵

During the 1960s and 1970s, as computer programming became an increasingly important form of technological innovation, it became apparent that some form of intellectual property protection would be necessary to incentivize the optimal level of software development and dissemination. But unlike previous revolutionary technologies, which were generally eligible for patent protection, in the early days of computer programming it was unclear whether software could be patented.⁴⁶ Indeed, two Supreme Court decisions from the 1970s held that the software claims at issue in those cases were ineligible for patent protection, based on the Court's determination that the computer pro-

⁴² NIMMER, *supra* note 38, § 2.03(B)(1).

⁴³ Act of October 15, 1971, Pub. L. No. 92-140, 85 Stat. 391 (1971) [hereinafter Sound Recording Act of 1971].

⁴⁴ NIMMER, *supra* note 38, at § 2.10[A].

⁴⁵ 17 U.S.C. § 102 (a) (2006).

⁴⁶ See CONTU Report, *supra* note 37, at 17 ("It is still unclear whether a patent may ever be obtained for a computer program"); *Diamond v. Diehr*, 450 U.S. 175, 195 (Stevens, J., dissenting). Until recent times, "well-established principles of patent law probably would have prevented the issuance of a valid patent on almost any conceivable computer program." *Diamond*, 450 U.S. at 195 (Stevens, J., dissenting).

grams were mere algorithms or “abstract ideas.”⁴⁷ The U.S. Patent Office also had a policy at the time of generally rejecting patents directed towards computer programs.⁴⁸

With patent protection apparently off the table, some suggested copyright as an alternative form of protection for software. However, there was serious doubt as to whether software was amenable to copyright protection, at least absent explicit action by Congress extending copyright protection to computer programs. At the time, the Copyright Act was silent on the issue, making no reference to computer programs. When first confronted with the issue, the Copyright Office expressed the view that software probably was not copyrightable, but in 1964, it began issuing some provisional registrations of computer programs under its “rule of doubt.”⁴⁹

In the 1970s, as an offshoot of the legislative processes leading to the Copyright Act of 1976, Congress established a commission of experts to study and make recommendations as to how the copyright law should respond to various technological developments, most notably the increasing significance of computer programs.⁵⁰ The commission was specifically directed to consider and make recommendations with respect to the question of whether, and to what extent, computer programs could be protected under current copyright law, and whether copyright law should be amended to accommodate computer programs.⁵¹ CONTU issued its highly influential report in 1978, which concluded not only that copyright protection for computer programs was justified both in terms of legal doctrine and innovation policy, but that computer programs in fact already were copyrightable under both the 1976 and 1909 Copyright Acts.⁵² Because computer programs were already copyrightable, no amendment to the statute was thought necessary to extend copyright to software, although some refinements to the statute were suggested to address some unique concerns associated with applying copyright law to software.

In its report, the commission stated that the underlying principle of copyright is that “if the cost of duplicating information is small, then it is simple for a less than scrupulous person to duplicate it[,] . . . legal as well as physical protection for the information is a necessary incentive if such information is to be

⁴⁷ See *Gottschalk v. Benson*, 409 U.S. 63, 72 (1972); *Parker v. Flook*, 437 U.S. 584, 594 (1978).

⁴⁸ *Diamond*, 450 U.S. at 195 (Stevens, J., dissenting). However, the majority opinion clarified that software could be patentable if it achieved some practical outcome, and over the next ten to twenty years, lower courts and the PTO progressively expanded the ability of software inventors to patent their inventions. *Id.* at 191–92.

⁴⁹ Pamela Samuelson, *Contu Revisited: The Case Against Copyright Protection for Computer Programs in Machine-Readable Form*, 1984 DUKE L.J. 663, 692–93 (1984).

⁵⁰ *Id.* at 693–94.

⁵¹ CONTU Report, *supra* note 37, at 1–5.

⁵² CONTU Report, *supra* note 37, at 16.

created and disseminated.”⁵³ As an illustration, the commission pointed out that in the nineteenth century, when music was recorded on a brass wheel to be played on a music box, “[t]he cost of making the wheel was inseparable from the cost of producing the . . . final product.”⁵⁴ But with the development of easily copied magnetic tapes, legal protection for recorded music became essential, and Congress eventually responded by passing the Sound Recording Act of 1971.⁵⁵

Analogously, with respect to software, the commission found that while there was little reason to protect the wired circuit or plug boards used to communicate instructions to early computers, the ease with which modern software can be copied weighed heavily in favor of providing effective intellectual property protection for computer programs in the form of copyright.⁵⁶ The commission further opined that not only was copyright protection desirable for computer programs, it was already available, based on the explicit declaration in the copyright act that “literary works” are copyrightable and the commission’s characterization of software as a form of literary work.⁵⁷

In 1980, Congress adopted the commission’s recommendation and amended the Copyright Act to include a definition of “computer programs.”⁵⁸ Although Congress has never enacted legislation explicitly stating that computer programs are copyrightable subject matter, the legislative history of the 1980 amendment defining computer programs in the Copyright Act clearly indicates that Congress viewed computer programs as a form of “literary work” falling within the scope of “original works of authorship,” and courts have universally interpreted the amendment as signaling congressional approval for copyright protection of software.⁵⁹ Indeed, by not explicitly adding computer programs to the enumerated list of copyrightable subject matter, Congress implicitly endorsed the commission’s view that the statute already provided copyright protection for software.

Even though the 1980 amendment implicitly sanctioned copyright protection for software, uncertainty remained as to the scope of this protection. For example, did copyright protection extend to object code, which could only be read by a computer, or was it limited to human readable source code? What about operating system software, whose only intended audience is a machine?

⁵³ *Id.* at 10.

⁵⁴ *Id.*

⁵⁵ Sound Recording Act of 1971, *supra* note 43.

⁵⁶ CONTU Report, *supra* note 37, at 10. As developed more fully below, the same reasoning would apply to DNA sequences, which like computer software are inherently susceptible to easy duplication but can require a substantial investment to develop.

⁵⁷ *Id.* at 16.

⁵⁸ NIMMER, *supra* note 38, § 2.04(C)(1).

⁵⁹ *Id.* § 2.04(C)(2).

Or a computer program embodied in computer readable media, such as a CD-ROM? Ultimately, all these questions are answered in the affirmative.

For example, some prominent members of the CONTU commission expressed the view that while copyright protection is appropriate for some computer programs, it should not be generally available to all computer programs. While Professor Nimmer joined with the majority in recommending copyright protection for some software, he filed a concurring opinion to the commission's report arguing that copyright protection should be limited to "computer programs which produce works which themselves qualify for copyright protection," i.e., word processors, graphics programs and the like.⁶⁰ In contrast, he felt that more utilitarian software, such as computer programs "which control the heating and air-conditioning in a building, or which determine the flow of fuel in an engine, or which control traffic signals [should] not be eligible for copyright because their operations do not result in copyrightable works."⁶¹ It took a number of judicial decisions in the 1980s to arrive at today's consensus that copyright protection is generally available for software, regardless of its format, practical utility, or the media on which it is recorded.⁶²

As has so often been the case, the extension of copyright protection to software lagged many years behind the initial development of the technology. Indeed, one can speculate that if patent protection had been available for computer programs from the beginning the case for extending copyright protection to software would have been far less compelling, and it seems quite possible that Congress and the courts would never have made a doctrinal leap to categorize software as a copyrightable literary work.

Contrast this with the situation in genetic engineering, where patents have been available and functioned reasonably well since the earliest days of the technology. With the availability of patent protection, there has been no compelling need for any alternative form of intellectual property. This lack of urgency likely explains why the scope of copyrightable subject matter has yet to undergo the natural evolution from software to engineered genetic sequences predicted by Professor Kayton.

IV. THE ANALOGY BETWEEN SOFTWARE AND SYNTHETIC GENETIC MATERIAL SUPPORTS AN EXTENSION OF COPYRIGHT TO ENGINEERED GENETIC SEQUENCES

When faced with the challenge of conforming copyright law to accommodate a new technology, Congress and the courts have relied heavily on analogy. The initial decision to extend copyright protection to software, for example, was based in large part upon the perceived analogy between software and

⁶⁰ CONTU Report, *supra* note 37, at 27.

⁶¹ *Id.*

⁶² MARK A. LEMLEY, ET AL., SOFTWARE AND INTERNET LAW 35-92 (3d ed. 2006).

traditional literary works.⁶³ Analogy has also been used extensively by courts faced with questions regarding the scope and extent of copyright protection for software.⁶⁴ It is thus informative to explore the analogy between computer programs and engineered DNA.

A computer program is essentially a set of instructions, each directing a computer or other machine to perform some function.⁶⁵ The program is written by a human and can be expressed in a format that is directed towards a human reader. For example, a computer program can be printed on a piece of paper in the form of source code, which is designed to be interpretable by a human. Alternatively, the same program can be expressed as object code, which is intended for communication with the machine itself, although in principle even object code could be directly interpreted by human conversant in that language.

However, the whole point of computer software is to communicate with a computer, so the computer program is compiled into a machine-readable format and then transcribed upon a computer readable medium. Computers normally do not read print on a page, but they can read the same set of instructions if encoded in a tangible medium that can be deciphered by machine. For example, a computer program can be transcribed upon a CD-ROM disk, in a process which entails physically altering the medium, by introducing microscopic indentations into the aluminum coating on the disc.⁶⁶

Another important attribute of software is that all but the most simple of computer programs are modular in nature, comprising a structured arrangement of subroutines, parameter lists and the like, which for simplicity I refer to in this article as “modules.” When software engineers develop a new computer program, they do not undertake the task *de novo* and write all the code from scratch; instead, to the extent possible, they incorporate modules of existing code. If a desired function appears in multiple contexts throughout a program, a single module is enlisted to perform that function, rather than writing new code in each instance. Often new modules will need to be written, and existing modules modified, but the use of modules greatly facilitates software engineering.

Because of its modular nature, software can be represented at different levels of abstraction. At the lowest level of abstraction, software is described by the string of zeros and ones that ultimately provide direction to the computer, or at a slightly higher level of abstraction, as lines of object code or source code. At the highest level, software can be represented by flow diagrams representing

⁶³ See CONTU Report, *supra* note 37, at 16.

⁶⁴ See, e.g., *Apple Computer, Inc. v. Franklin Computer Corp.*, 714 F.2d 1240, 1246–49 (3d Cir. 1983).

⁶⁵ A “computer program” is defined in the Copyright Act as “a set of statements or instructions to be used directly or indirectly in a computer in order to bring about a certain result.” 17 U.S.C. § 101 (2006).

⁶⁶ A “CD-ROM is a pre-pressed compact disc that contains data accessible to, but not writable by, a computer for data storage and music playback.” *CD-ROM*, WIKIPEDIA, <http://en.wikipedia.org/wiki/CD-ROM> (last visited Mar. 8, 2011).

the structure of the program selected sequence encodes of the arrangement and interaction of the modules. And at an intermediate level, the modules can themselves often be broken down into smaller subunits, such as subroutines and other more fundamental units of code. Thinking of software at various levels of abstraction permits an engineer working at a higher level of abstraction to focus on the selection and arrangement of modules without concerning herself with the specific code underlying those modules.⁶⁷

Returning to DNA, the analogy between software code and genetic code is striking. A genetic sequence provides a series of instructions directing a living cell to perform functions dictated by the instructions. Genetic engineering permits a human to dictate these instructions. Like a computer program, a genetic sequence can be expressed in a format directly interpretable by a human, albeit instead of a series of zeros and ones, it is a sequence of A, T, C and G's, representing the four primary nucleotides that make up DNA.

Genetic sequences can also be represented at various levels of abstraction. A three nucleotide codon representing an amino acid can be symbolized by a single letter representing that amino acid. A string of codons representing a protein domain can be expressed as a single symbol representing the domain, and a combination of domains can be expressed as a single protein. A string of nucleotides constituting a regulatory element, such as a promoter or enhancer, can be represented by a single symbol. For a good example of an engineered genetic sequence represented at a very high level of abstraction, see the figure used by Venter and colleagues to represent the full-length sequence of the synthetic bacterial genome they created.⁶⁸ By means of abstraction, they are able to represent a genetic sequence comprising 582,970 nucleotides essentially as a notated circle. A recent article describes the importance of being able to represent genetic sequences at a high level of abstraction in order to facilitate the design of complex synthetic DNA molecules.⁶⁹

Like software, in order to be useful, engineered genetic sequences must be transcribed into a format that can be interpreted by the primary intended audience, the difference being the audience in this case is a cell rather than a computer. In either case, this involves physically transcribing instructions into the appropriate medium of communication at a "nano" level. In the case of a CD-ROM, the reflective properties of a thin layer of aluminum are altered by making microscopic indentations, while in DNA the ordering of molecular subunits (individual nucleotides) conveys the message to the appropriate audience.

Importantly, as noted above, copyright protection extends to copyrightable subject matter embodied in a format that is not directly decipherable by a

⁶⁷ Computer Assoc. Int'l v. Altai, 982 F.2d 693, 698 (2d Cir. 1992).

⁶⁸ J. Craig Venter et al., *Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome*, SCIENCE, July 2010, at 52, available at <http://www.sciencemag.org/cgi/content/full/329/5987/52>.

⁶⁹ Jonathan A. Goler et al., *Genetic Design: Rising Above the Sequence*, 26 TRENDS IN BIOTECHNOLOGY No.10 538 (2008).

human being. For that reason, copying of a CD-ROM can constitute copyright infringement, even though the copied medium could not be directly read by human being. In the same manner, there would appear to be no obstacle to finding copyright infringement based on the copying of a DNA molecule representing a copyrightable set of instructions.

Not only was the use of analogy instrumental in justifying the initial extension of copyright to software, courts have often resorted to analogy when faced with a novel application of copyright law to software. For example, in assessing the range of copyright protection available to functional and non-literal elements of a computer program, courts have repeatedly looked for guidance from *Baker v. Selden*, a nineteenth century Supreme Court decision involving a book that described a system for accounting and forms for use with that system.⁷⁰ In *Computer Associates International v. Altai*, when considering the similarity between two software programs, the Second Circuit adapted the theoretic framework for analyzing substantial similarity expounded by Judge Learned Hand in a 1930 decision involving a copyrighted play.⁷¹ And in *Lotus v. Borland*, the First Circuit found the menu command hierarchy of a computer spreadsheet program to be an unprotectable “method of operation” based on an analogy it saw between the command hierarchy and the control buttons on a VCR.⁷²

Similarly, the analogy between computer programs and engineered genetic sequences not only argues in favor of extending copyright to engineered genetic sequences, but will prove invaluable in establishing the contours of copyright protection for engineered genetic sequence. This will substantially facilitate adaptation of copyright law to accommodate the unique technical and policy issues which will arise if copyright is extended to DNA, since the courts have already grappled with, and to a large extent resolved, these issues with respect to computer programs.

The ability to conceptualize software at different levels of abstraction often plays an important role in determining the extent to which copyright can be enforced to protect various elements of a computer program from copying.⁷³ At the lowest level of abstraction, expressed as lines of code, it is well-established that copyright affords protection against unauthorized literal copying of a computer program.⁷⁴ But courts will also enforce copyright in cases that involve no literal copying of code, but where the infringer copies more abstract elements of a computer program, such as the structural arrangement of modules and functional elements.⁷⁵ This is often referred to as “non-literal copying,”

⁷⁰ *Baker v. Selden*, 101 U.S. 99, 99–100 (1879).

⁷¹ *Computer Assoc. Int'l*, 982 F.2d at 704.

⁷² *Lotus Dev. Corp. v. Borland Int'l, Inc.*, 49 F.3d 807, 815–16 (1st Cir. 1995).

⁷³ *See Computer Assoc. Int'l v. Altai*, 982 F.2d 693, 698 (2d Cir. 1992).

⁷⁴ *Id.* at 702.

⁷⁵ *Id.*

based on analogy to cases finding copyright infringement of more conventional literary works where the copying occurs at a more abstract level, such as plot or storyline.⁷⁶ However, in the case of both conventional literary works and computer programs, the courts will not provide copyright protection for software elements expressed at the highest levels of abstraction, characterizing these elements as the unprotectable “idea” of the work as opposed to a protectable expression of that idea.⁷⁷ For that reason, subsequent innovators are free to copy higher order concepts from existing copyrighted computer programs of others so long as they implement those concepts in an original matter.

Like software, engineered DNA can be expressed at different levels of abstraction, and thus should be amenable to copyright protection even in the absence of literal copying at the level of nucleotide sequence. For example, copyright protection could be available to protect against the copying of an engineered arrangement of genetic modules, i.e., protein coding domains and regulatory elements, even in the absence of copying at the genetic sequence level. However, analogous to the case with software, higher level concepts embodied in an engineered genetic sequence would be unprotectable and hence available to enrich the public domain. Case law pertaining to copyright infringement of software at various levels of abstraction will provide guidance for courts confronted with analogous facts relating to engineered DNA.

Another important characteristic shared by software and engineered genetic sequences is that, for both, the cost of development greatly exceeds the cost of duplication owing to the fact that both can serve as the template for their own reproduction.⁷⁸ A computer program encoded in digital form can be easily and repeatedly copied, resulting in a virtually unlimited number of essentially perfect copies. In the same way, DNA serves as its own template for the reproduction of the exact copies by biological DNA replication processes, both *in vitro* and *in vivo*.

Significantly, because the copies reproduced are identical to the original, and therefore can function as another template for further copying, software and DNA are both susceptible to viral replication. Each time a computer program is copied, it creates a new template for copying, and the iterative copying of copies can result in the production of copies at an exponential rate, mimicking the spread of a virus. In the same manner, a DNA sequence serves as the template for production of an exact copy, and each copy likewise can serve as a template for subsequent copies. Indeed, the terms “viral replication” and “computer virus,” widely used in connection with digital files and computer software, arise out of the remarkable propensity of viral DNA to self replicate.

⁷⁶ *Id.* at 696. Infringement of software at a higher level of abstraction is sometimes referred to as non-literal infringement, as opposed to copying of the code itself, which is referred to as literal infringement.

⁷⁷ *Id.* at 703.

⁷⁸ *See supra* Part III.

V. THE SYNTHETIC BIOLOGY REVOLUTION

Earlier arguments for extending copyright to engineered DNA, as exemplified by Professor Kayton's article, were proposed at a time when the capabilities of genetic engineers were still quite limited and engineered genetic sequences tended to deviate only slightly from naturally occurring sequences. Subsequent advances have greatly expanded the molecular tool chests of genetic engineers, permitting the design and deployment of complex engineered genetic sequences bearing even less resemblance to naturally occurring sequences. These advances, which are often referred to under the trendy moniker of "synthetic biology,"⁷⁹ are bringing genetic engineering and software engineering into ever closer alignment and are prompting legal academics like myself and Professor Torrance, along with synthetic biologists such as Stemmer and Endy, to raise the issue anew in light of these developments.

In the early days of biotechnology, genetic engineering primarily entailed recombining naturally occurring genetic sequences.⁸⁰ The objective was often simply to use recombinant technology to produce a microbial cell capable of producing large amounts of a naturally occurring human protein in culture.⁸¹ This would generally entail cloning the genetic sequence encoding the desired human protein, slightly modifying the sequence as necessary to conform with the requirements of the microbe, and then linking the gene sequence to the appropriate regulatory elements in the orientation necessary to direct expression of the protein in the microbial host.⁸²

This classic example of engineered DNA is characterized by minimal deviation from naturally occurring sequences and deviations dictated by relatively strict and straightforward functional constraints. To use the language of copyright law, relatively minimal "expressive choice" is exercised in designing the deviations from the naturally occurring starting material.

The limited number of reasonable alternate designs available for some of the most straightforward genetic engineering projects could have some bearing on the copyrightability of the resulting genetic sequence. Under the merger doctrine, the expression of an idea can be uncopyrightable if there are only a limited number of alternate ways of effectively expressing that idea.⁸³ The policy behind the merger doctrine is that under such circumstances, protection of expression can be tantamount to protection of the underlying idea.⁸⁴ For exam-

⁷⁹ Synthetic biology has been defined as "the design and construction of new biological functions and systems not found in nature." *Synthetic Biology*, WIKIPEDIA, http://en.wikipedia.org/wiki/Synthetic_biology (last visited Mar. 8, 2011).

⁸⁰ Goler, *supra* note 69, at 538.

⁸¹ Holman, *supra* note 29, at 323-40.

⁸² *Id.*

⁸³ *Morrissey v. Procter & Gamble*, 379 F.2d 675, 678-79 (1st Cir. 1967).

⁸⁴ *Id.* at 678.

ple, in *Morrissey v. Procter & Gamble*, the court held that a set of instructions for entering a sweepstakes could not be copyrighted because there were only a limited number of ways of expressing the substance of the contest.⁸⁵ The merger doctrine could arguably apply to the some of the simplest and most straightforward engineered genetic sequences in view of the relatively limited number of reasonable alternate designs available at the time the sequence was engineered. But as discussed below, contemporary synthetic biology provides genetic engineers with a much wider range of design choices that, in my view, should overcome objections to copyrightability based on the merger doctrine.

As stated by the Supreme Court in *Feist*, at least some “modicum of creativity” is necessary in order for a work to receive copyright protection.⁸⁶ But the creativity bar set by *Feist* is quite low, and even software programs of minimal creativity are assumed to satisfy it. The complex creative works resulting from modern synthetic biology should easily reach the *Feist* threshold.

Today, the ascent of synthetic biology is transforming genetic engineering in fundamental ways, enabling an entirely new level of control and precision.⁸⁷ Synthetic biologists are increasingly able to design and synthesize genetic sequences that deviate substantially from anything occurring naturally and capable of performing novel and often highly useful functions.⁸⁸ Techniques that rely upon naturally occurring DNA sequences as starting material, such as DNA shuffling and other modes of directed molecular evolution, have been successfully deployed to create synthetic gene sequences deviating substantially from anything found in nature.⁸⁹ Work is progressing on methods for de novo genetic design, which results in genetic sequences bearing even less resemblance to any natural counterpart.⁹⁰

Genetic engineering has historically relied heavily upon the physical manipulation of DNA molecules and empirical testing to identify engineered sequences having desired functional characteristics. For example, classic DNA shuffling involves generating huge libraries of variant DNA sequences, and physically screening those sequences to identify variants having improvements and desired function.⁹¹

However, the heavy lifting in synthetic biology is increasingly being conducted in computers, where genetic sequences are designed and manipulated

⁸⁵ *Id.*

⁸⁶ *Feist Publ'n, Inc. v. Rural Tel. Serv. Co., Inc.*, 499 U.S. 340, 362 (1991).

⁸⁷ Alok Jha, *From the Cells Up*, THE GUARDIAN, Mar. 2005, available at www.guardian.co.uk/science/2005/mar/10/science.research/print. (“Do not be fooled into thinking that [synthetic biology] is just a new era of biotechnology. . . . In comparison, [biotechnology], which relies heavily on trial and error and is difficult to control, is a shot in the dark.”).

⁸⁸ Arti Rai and James Boyle, *Synthetic Biology: Caught Between Property Rights, the Public Domain, and the Commons*, PLOS BIOLOGY, Mar. 2007, at 0389.

⁸⁹ SHELDON J. PARK & JENNIFER R. COCHRAN, PROTEIN ENGINEERING AND DESIGN (2010).

⁹⁰ *Id.*

⁹¹ Jon Cohen, *How DNA Shuffling Works*, 293 SCIENCE 237 (2001).

at the level of information—these techniques have been referred to as *in silico* genetic engineering.⁹² Thanks to advances in DNA synthesis technology, companies have sprung up that can quickly and efficiently synthesize virtually any gene sequence that a synthetic biologist has designed *in silico*.⁹³ *In silico* genetic engineering illustrates the increasing analogy between the processes used to develop engineered genetic sequences and software.

The development of the glyphosate N-acetyltransferase (GAT) gene provides a good example of synthetic biologists successfully applying classic gene shuffling augmented by computer design to create an entirely novel protein with unique and useful functional attributes. Glyphosate, more commonly known by its tradename ROUNDUP, is the most important herbicide in the world, widely used because of its efficacy and relatively benign environmental and safety profile.⁹⁴ Using computer-assisted DNA shuffling, synthetic biologists were able to create GAT, a novel gene that, when introduced into a plant, allows the plant to breakdown and detoxify glyphosate.⁹⁵ For example, genetically modified corn bearing the GAT is resistant to glyphosate, which permits a farmer to spray a cornfield with glyphosate to kill weeds without harming the corn.⁹⁶

The development and commercialization of the GAT gene exemplifies the progress that has been made in genetic engineering at the lowest level of abstraction, *i.e.*, the gene sequence itself. Similar advances are occurring at higher levels of abstraction, involving the design and synthesis of complex rearrangements of DNA modules. For example, undergraduates and even high school students are now able to design genetic constructs by rearranging DNA modules in creative and often ingenious ways.⁹⁷ The BioBricks Foundation is assembling a set of DNA modules, which it refers to as “standard biological parts,” for use in this sort of higher level genetic engineering.⁹⁸

And while genetic engineering has clearly experienced dramatic progress in the last thirty years, the current state-of-the-art pales in comparison with advances likely in the not-too-distant future. Drew Endy, a thought leader in the synthetic biology movement, recently acknowledged that despite tre-

⁹² For example, DNA 2.0, a gene synthesis company, has developed an *in silico* protein engineering system, “ProteinGPS™,” that it claims “quickly and efficiently designs proteins with improved characteristics.” *Protein Engineering, DNA 2.0*, <https://www.dna20.com/index.php?pageID=65> (last visited Mar. 8, 2011).

⁹³ *See, e.g., id.*

⁹⁴ *Glyphosate*, WIKIPEDIA, <http://en.wikipedia.org/wiki/Glyphosate> (last visited Mar. 8, 2011).

⁹⁵ Linda A. Castle et al., *Discovery and Directed Evolution of a Glyphosate Tolerance Gene*, 304 SCIENCE 1151 (2004).

⁹⁶ *Optimum GAT Trait*, DUPONT (Mar. 8, 2011, 10:52 AM), http://www2.dupont.com/Production_Agriculture/en_US/science_of/Optimum_GAT.html (website describing DuPont Pioneer Optimum GAT product).

⁹⁷ *See, e.g., iGEM 2009* (Mar. 8, 2011), http://2009.igem.org/Main_Page.

⁹⁸ BIOBRICKS FOUNDATION, <http://bbf.openwetware.org/> (last visited Mar. 8, 2011).

mendous individual successes, genetic engineering and biotechnology are still shackled by an unacceptably high degree of unpredictability that renders each new genetic engineering project an “expensive, unreliable and ad hoc research process[.]”⁹⁹ He proposes addressing this unpredictability by a concerted effort to invent and implement foundational technologies that will facilitate and engender the efficient and predictable engineering of complex genetic systems, using principles borrowed from conventional engineering, such as standardization, decoupling, and abstraction.¹⁰⁰

To encourage the creation of the foundational tools and methodologies that will be necessary to facilitate advances in synthetic biology, synthetic biologists have joined forces to create the BioBricks Foundation. The primary mission of the BioBricks Foundation is to promote the development and availability of “standard biological parts” for use in higher order genetic engineering, based on design principles taken from traditional engineering practice.¹⁰¹ To the extent these efforts succeed, they will render the engineering of DNA much more analogous to conventional engineering, and particularly software engineering, than was the case in the early days of genetic engineering.

Another example of the growing convergence of engineered DNA and software can be seen in the ongoing development of DNA-based computers¹⁰² and logic gates.¹⁰³ The blurring of the line between genetic engineering and software engineering can also be seen in the language used by modern synthetic biologists. For example, the BioBricks Foundation states on its website that “[u]sing BioBrick standard biological parts, a synthetic biologist or biological engineer can already, to some extent, program living organisms in the same way a computer scientist can program the computer.”¹⁰⁴ Similarly, Amyris Biotechnologies of Emeryville, California, states on its website that its

technology makes it possible to alter the metabolic pathway of microorganisms such as these, creating living factories that produce molecules to practical applications. While reading, writing, and analyzing the DNA of microbes once took years, Amy-

⁹⁹ Drew Endy, *Foundations for Engineering Biology*, 438 NATURE 449, 449 (2005).

¹⁰⁰ *Id.* at 450–52.

¹⁰¹ BIOBRICKS FOUNDATION, *supra* note 98. The development of the tools and methodologies for higher order design-oriented genetic engineering has been described in a number of published reviews available at <http://syntheticbiology.org/Documents.html>.

¹⁰² Yaakov Benenson et al., *An Autonomous Molecular Computer for Logical Control of Gene Expression*, 429 NATURE 423, 423–29 (2004).

¹⁰³ Kate McAlpine, *DNA Logic Gates Herald Injectable Computers*, 206 NEW SCI. 9 (2010), available at <http://www.newscientist.com/article/dn18989-dna-logic-gates-herald-injectable-computers.html> (the development of DNA logic gates “brings the prospect of injectable biocomputers programmed to target diseases as they arise”).

¹⁰⁴ BIOBRICKS FOUNDATION, *supra* note 98.

ris can now reprogram microorganisms and test our ability to produce desired molecules in days to weeks.¹⁰⁵

The convergence of the technologies is also evident in the incorporation of biological principles in software design. Examples include artificial intelligence and the use of neural networks in computing. One manifestation of artificial intelligence, genetic programming, applies evolutionary algorithms, inspired by biological evolution, to create new computer programs.¹⁰⁶ In genetic programming, functions are represented as “chromosomes,” and the main operators used in the evolutionary algorithms are “crossover” and “mutation,” concepts taken from genetics.¹⁰⁷

Meta-Genetic Programming is a proposed technique of evolving a genetic programming system using genetic programming itself. It is based upon the premise that software chromosomes, crossovers, and mutations should, like their real-life counterparts, be allowed to change on their own rather than being determined by a human programmer.¹⁰⁸ As these trends continue, the justification for maintaining copyright protection for software while denying it for engineered DNA becomes increasingly questionable.

VI. SOME OF THE ARGUMENTS AGAINST COPYRIGHT PROTECTION FOR ENGINEERED DNA ARE BASED ON AN INCOMPLETE UNDERSTANDING OF THE STATE-OF-THE-ART IN GENETIC ENGINEERING

Some commentators have rejected copyright for engineered DNA based on an incomplete or outmoded understanding of the underlying technology and a failure to recognize the increasing analogy between software and DNA. For example, a recently published article by Michelotti states that of the “thousands [of genes] for which patent applications have been filed, the number authored by humans, that is, the number whose base pair sequence is the product of human intellect, is exactly zero.”¹⁰⁹ He goes on to assert that “[t]o date, no person has actually composed an original base pair sequence for even a relatively short segment of DNA with the expectation that it would encode predictable and useful instructions in a living organism.”¹¹⁰

Michelotti’s assessment of the state-of-the-art in genetic engineering is inaccurate. As explained above, synthetic biologists have made great strides in

¹⁰⁵ Robert Rapier, Consumer Energy Report, R Squared Robert Rapier Energy Blog, *Amyris is Looking Promising* (Nov. 13, 2008), <http://www.consumerenergyreport.com/2008/11/13/amyris-is-looking-promising/> (quoting AMYRIS, <http://www.amyris.com/>).

¹⁰⁶ *Genetic programming*, WIKIPEDIA, http://en.wikipedia.org/wiki/Genetic_programming (last visited Mar. 9, 2011).

¹⁰⁷ *Id.*

¹⁰⁸ *Id.*

¹⁰⁹ Michelotti, *supra* note 25, at 87.

¹¹⁰ *Id.*

rationally engineering complex and original genetic sequences, as illustrated by examples such as the development of the GAT gene and the many creative genetic constructs designed by undergraduate students in conjunction with the IGEM competition.¹¹¹

Ironically, Michelotti goes on to posit that “[w]hen an individual understands biochemical systems well enough to conceive an original design for a non-obvious protein and compose genetic instructions for synthesizing that protein in a living organism in a manner that expresses a useful trait, then copyright protection . . . will be well deserved.”¹¹² In other words, he agrees that copyright protection for engineered DNA will be “well deserved” once the capabilities of genetic engineers have reached a certain level. In fact, however, that competency has arguably already been attained, and the level of control he describes is the essence of modern synthetic biology. There are many examples of patented (and hence presumably nonobvious) genetic sequences that encode rationally engineered proteins that confer a useful trait an organism expressing the gene. Engineered genes that confer herbicide resistance on plants, such as the GAT gene, are but a few of the host of examples provided by biotechnology.¹¹³ Michelotti’s conclusion that copyright protection is not available for engineered DNA is evidently rooted in unfamiliarity with the current state-of-the-art in biotechnology.

In the same article, Michelotti further asserts that engineered genetic sequences are insufficiently original for copyright protection because, in his view, genetic engineering has been limited to trivial modifications of genetic sequences.¹¹⁴ To the contrary, the market is full of important biotechnology products that have been produced by modifying naturally occurring genetic sequences to achieve substantially different function than any naturally occurring sequence. For example, there are engineered variations of naturally occurring proteins that are used as drugs and have very different pharmaceutical characteristics than the naturally occurring protein.¹¹⁵ The GAT gene described above varies dramatically from anything that exists in nature. The current success of the biotechnology industry is based in large part upon genetic modifications that are far from trivial, and the future of synthetic biology promises even more dramatic departure from naturally occurring genetic sequences.

Another recent article, this one by Gargano, argues that DNA sequences are not copyrightable because they are “obtained from nature and are probably not original.”¹¹⁶ His arguments for the most part echo Michelotti’s and likewise

¹¹¹ See *supra* Part V.

¹¹² Michelotti, *supra* note 25, at 87.

¹¹³ See *supra* Part V.

¹¹⁴ Michelotti, *supra* note 25, at 87.

¹¹⁵ Christopher M. Holman, *Learning from Litigation: What Can Lawsuits Teach Us About the Role of Human Gene Patents in Research and Innovation?*, 18 KAN. J.L. & PUB. POL’Y 215, 221–33 (2009) (describing engineered variants of erythropoietin and human growth hormone).

¹¹⁶ Gargano, *supra* note 25, at 29–33.

fail to appreciate the significant differences between engineered genetic sequences and naturally occurring sequences, and the power of modern synthetic biology.

VII. SOME OF THE ARGUMENTS AGAINST COPYRIGHT FOR DNA ARE BASED ON A MISREADING OF COPYRIGHT LAW

In some cases, the arguments raised against copyright protection for genetic sequences are based on a misinterpretation of copyright law. For example, Michelotti contends that “trivial” modifications of a naturally occurring genetic sequence would be uncopyrightable “derivative works,” apparently based on a misunderstanding as to the meaning of the term “derivative work.”¹¹⁷

The Copyright Act defines a derivative work as a “work based upon one or more *pre-existing works*.”¹¹⁸ While not explicitly defined, the term “work” is used extensively throughout the Act to refer to works of authorship, i.e., expressive works amenable to copyright protection. It necessarily follows that, in order for subject matter to satisfy the definition of a “derivative work,” it must be based on another copyrightable work. Thus, the only way engineered DNA could be a derivative work would be if the naturally occurring DNA upon which it is based is itself copyrightable, which is clearly not what Michelotti is suggesting.

Moreover, the very suggestion that derivative works are uncopyrightable is itself completely wrong; true, the statute provides authors with the exclusive right to make derivative works, but the derivative works are themselves copyrightable, at least to the extent they were made with the consent of the owner of the copyright in the original work.¹¹⁹ An understanding of copyright law reveals that the argument that engineered DNA is uncopyrightable because it is a derivative work is entirely baseless.

Kumar and Rai also base their conclusion that copyright is doubtful for genetic sequences in part on a misinterpretation of copyright law. For example, they assert that computer “source code may become unprotectable if it represents a method of operation” and infer that genetic sequences might likewise be treated by the courts as uncopyrightable “methods of operation.”¹²⁰ In support of this assertion, they point to a single reported decision, *Lotus v. Borland*, which they characterize as having held that the computer source code at issue in the case was “a method of operation and therefore unprotectable.”¹²¹

¹¹⁷ Michelotti, *supra* note 25, at 87.

¹¹⁸ 17 U.S.C. § 101 (2006) (emphasis added).

¹¹⁹ *Anderson v. Stallone*, 11 U.S.P.Q.2d (BNA) 1161, 1165 (C.D. Cal. 1989).

¹²⁰ Kumar & Rai, *supra* note 25, at 1763–64 (citing 17 U.S.C. § 102(b) (2000)) (“excluding from copyright protection any ‘idea, procedure, process, system, [or] method of operation’”).

¹²¹ *Id.* at 1764.

But Kumar and Rai misread the *Borland* decision—the case is not even on point. In *Borland*, the copyrightability of software code was not even a question before the court. Instead, the question was whether copyright protection extended to the computer menu command hierarchy, which is something very different from software code.¹²² In fact, the defendant explicitly conceded that the plaintiff had a valid copyright in the underlying software code and never suggested that the code itself was a method of operation.¹²³ To my knowledge, there is nothing in the case law that would support Kumar and Rai’s contention that a court would characterize an engineered genetic sequence as an uncopyrightable method of operation.

To the contrary, courts that have addressed the issue have rejected the argument that computer code is an uncopyrightable method of operation. For example, in *Apple Computer v. Franklin Computer*,¹²⁴ the court rejected the argument that the Apple operating system software is uncopyrightable because it is a method of operating a computer.¹²⁵ This decision is entirely consistent with the CONTU Report, which concluded that while the bar against copyright for methods of operation would prohibit the protection of the electrical or mechanical functioning of a computer, or the processes performed by the computer, it in no way implicates protection for the series of instructions controlling these processes, as embodied in computer software.¹²⁶ It is also consistent with the House and Senate reports on the 1976 Copyright Act, which state that “Section 102(b) is intended, among other things, to make clear that the expression adopted by the programmer is the copyrightable element in a computer program, and that the actual processes or methods embodied in the program are not within the scope of copyright law.”¹²⁷

Silva and Gargano have independently argued that genetic sequences are uncopyrightable based on their assertion that the utilitarian aspects of genetic sequences are not conceptually or physically separable from any “aesthetic” copyrightable expression.¹²⁸ In referring to conceptual and physical separability, these authors are clearly implying that engineered genetic sequences are barred from copyright by the useful article doctrine, but again this argument is based upon a clear misunderstanding of the scope and nature of this doctrine.

Under the useful article doctrine, the courts have held that copyright does not extend to the utilitarian aspects of a useful article, but only to aesthetic

¹²² *Lotus Dev. Corp. v. Borland Int’l, Inc.*, 49 F.3d 807, 814–15 (1st Cir. 1995). The computer menu command hierarchy is the arrangement of command terms such as Quit, Copy, Print, etc. *Id.*

¹²³ *Id.*

¹²⁴ *Apple Computer, Inc. v. Franklin Computer Corp.*, 714 F.2d 1240 (3d Cir. 1983).

¹²⁵ *Id.* at 1250–51.

¹²⁶ CONTU Report, *supra* note 37, at 18–20.

¹²⁷ S. REP. NO. 94-473, at 54 (1975); H.R. REP. NO. 94-473, at 57 (1976).

¹²⁸ See Gargano, *supra* note 25, at 29–33; Silva, *supra* note 25.

design features that are either physically or conceptually separable from the utilitarian features of the article.¹²⁹ Importantly, this doctrine has only been applied to “pictorial, graphic and sculptural” works. The Copyright Act explicitly defines a “useful article” as “an article having an intrinsic utilitarian function that is not . . . merely to convey information.”¹³⁰ Courts have explicitly noted that computer programs are copyrightable as “literary works,” and thus “not subject to a useful-article exception.”¹³¹

A genetic sequence, which is essentially coded information providing a set of instructions to a biological machine, is more analogous to software than a picture, graphic or sculpture. The concepts of physical and conceptual separability, while pertinent in the context of assessing the availability of copyright protection for aesthetic elements of a useful article of manufacture, such as a bicycle rack or toaster, clearly have no place in determining the availability of copyright protection for engineered genetic sequences.

VIII. THE U.S. CONSTITUTION AUTHORIZES COPYRIGHT PROTECTION FOR ENGINEERED GENETIC SEQUENCES

The Copyright Clause of the U.S. Constitution confers upon Congress the power to grant “Authors” exclusive rights in their “Writings,” in order to promote the progress of science and the useful arts.¹³² In the past, when copyright expanded to accommodate a new technology, some have argued that the extension exceeds the scope of congressional authority under the Constitution. But the courts have consistently interpreted the Copyright Clause in a flexible manner that accommodates new technologies.

For example, although the term “writing” could be construed literally, the Supreme Court has repeatedly interpreted the term expansively to reflect new developments in technology that would have been difficult if not impossible to foresee at the time the Constitution was drafted. For example, faced with the question of whether photographs qualify as “writings” in the constitutional sense, the Supreme Court pointed out that the first U.S. copyright statute enacted in 1790 explicitly afforded copyright protection to maps and charts, and in 1802 Congress amended the Act to include engravings, etchings and prints, none of which are “writings” in the literal sense.¹³³ Many of those in Congress who voted for the statute in 1790 and its 1802 amendment were members of the convention which framed the Constitution, from which the Court inferred that the drafters of the Constitution did not intend the term “writings” to be inter-

¹²⁹ See *Brandir Int'l v. Cascade Pac. Lumber Co.*, 834 F.2d 1142, 1146–47 (2d Cir. 1987).

¹³⁰ 17 U.S.C. § 101 (2006).

¹³¹ *Apple Computer, Inc. v. Franklin Computer Corp.*, 714 F.2d 1240, 1247–49 (3d Cir. 1983).

¹³² U.S. CONST. art. 1, § 8, cl. 8.

¹³³ *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53, 55 (1884).

preted in a narrow, literal sense.¹³⁴ The Court concluded that “we entertain no doubt that the Constitution is broad enough to cover an act authorizing copyright of photographs, so far as they are representative of original intellectual conceptions of the author.”¹³⁵ Subsequent Supreme Court decisions have held, consistent with *Burrow-Giles*, that commercial art, motion pictures, and sound recordings are all “writings.”¹³⁶

In a judicial opinion that Professor Nimmer, the foremost commentator on U.S. copyright law, has characterized as the “touchstone” for interpreting the constitutional “writing” requirement, Judge Learned Hand wrote that “our Constitution [does not] embalm inflexibly the habits of 1789 . . . its grant of power to Congress comprise, not only what was then known, but what the ingenuity of man should devise thereafter.”¹³⁷ Based on this judicial precedent, the CONTU commission concluded that computer programs are writings in the constitutional sense, and thus that the Constitution posed no obstacle to copyright protection for software.¹³⁸ For similar reasons, the constitutional “writing” requirement should not pose a bar to copyright for engineered DNA.

In *Feist*, the Supreme Court affirmed that there is a constitutional requirement of originality in copyrighted works.¹³⁹ Thus, one might envision an argument that engineered genetic sequences lack the requisite originality. However, in *Feist*, the Supreme Court set a very low bar for originality, requiring a mere “modicum of creativity.”¹⁴⁰ As described above, the ingenuity and creativity of any reasonably complex engineered genetic sequences should easily satisfy the constitutional “modicum of creativity” standard.

Some commentators have expressed a concern that copyright on genetic sequences would likely impede subsequent research and innovation. This could conceivably form the basis for an argument that copyright protection for genetic sequence is unconstitutional because it violates the preambular language of the IP Clause, which arguably directs Congress to enact copyright laws that promote progress in science and the useful arts.¹⁴¹

Interestingly, the analogous argument has been raised in *Association for Molecular Pathology v. U.S. Patent and Trademark Office*,¹⁴² the case brought

¹³⁴ *Id.* at 58.

¹³⁵ *Id.*

¹³⁶ See *Burrow-Giles*, 111 U.S. 53; *Bleistein v. Donaldson Lithographing Co.*, 188 U.S. 239 (1903); *Kalem Co. v. Harper Bros.*, 222 U.S. 55 (1911); *Goldstein v. California*, 412 U.S. 546 (1973).

¹³⁷ *Reiss v. Nat'l Quotation Bureau, Inc.*, 276 F. 717, 719 (S.D.N.Y. 1921).

¹³⁸ CONTU Report, *supra* note 37, at 14–15.

¹³⁹ *Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., Inc.*, 499 U.S. 340, 362 (1991).

¹⁴⁰ *Id.*

¹⁴¹ U.S. CONST. art. 1, § 8, cl. 8.

¹⁴² 702 F. Supp. 2d 181 (S.D.N.Y. 2010).

by the ACLU challenging the validity and constitutionality gene patents.¹⁴³ One of the plaintiffs' constitutional arguments was based on an assertion that gene patents tend to impede rather than promote innovation, which they argued violates the preambular directive to promote science and the useful arts.¹⁴⁴

But in the context of copyright, this argument has already been raised and rejected by the Supreme Court in the 2003 *Eldred v. Ashcroft* decision.¹⁴⁵ In that case, the Court denied a constitutional challenge to the 1998 Copyright Term Extension Act, holding that Congress has a great deal of discretion in deciding how best to implement its constitutional mandate to "promote the progress of science and useful arts."¹⁴⁶ In *Eldred*, the Court "stressed [that] it is generally for Congress, not the courts, to decide how best to pursue the Copyright Clause's objectives."¹⁴⁷

Thus, so long as there is a rational basis for concluding that copyright protection for genetic sequences would promote innovation, the preambular language of the Copyright Clause should not stand in the way of this extension of copyright law. In fact, there are a number of important policy considerations that could be advanced favoring some form of copyright protection on genetic sequences, thereby promoting the advance of science and technology.¹⁴⁸

IX. IT IS UNNECESSARY FOR THE COPYRIGHT ACT TO EXPLICITLY IDENTIFY ENGINEERED GENETIC SEQUENCES AS COPYRIGHTABLE SUBJECT MATTER

In their 2007 article, Kumar and Rai correctly noted that "[u]nlike software . . . the products of synthetic biology are not discussed as copyrightable subject matter in the statute[, and thus] a court that wished to find that material copyrightable would have to do so by analogy."¹⁴⁹ While this statement is correct as far as it goes, in context those authors seemed to be implying that the lack of explicit reference to DNA in the copyright statute weakens the case for extending copyright to engineered DNA. In fact, however, the expansion of copyright to cover software clearly illustrates that specific mention in the copyright statute is not a prerequisite for copyright protection, and that courts routinely and justifiably rely on analogy to adapt copyright to newly arising technologies.

It is important to remember that prior to 1980, Congress, the courts, and the CONTU commission all concluded that the Copyright Act encompassed software, notwithstanding the fact that the Copyright Act made absolutely no

¹⁴³ *Id.* at 208.

¹⁴⁴ *Id.*

¹⁴⁵ 537 U.S. 186 (2003).

¹⁴⁶ *Id.* at 208.

¹⁴⁷ *Id.* at 212.

¹⁴⁸ *See infra* Part XIV.

¹⁴⁹ Kumar & Rai, *supra* note 25, at 1763.

mention of computer programs until the 1980 amendment.¹⁵⁰ For example, the legislative history of the 1976 Copyright Act explicitly indicates that both houses of Congress viewed computer programs as falling within the scope of copyrightable subject matter under both the 1909 and 1976 Copyright Acts.¹⁵¹

Prior to enactment of the 1976 Act, some proponents of software copyright suggested that computer programs be included in the nonexclusive list of exemplary categories of copyrightable subject matter enumerated in § 102 of the 1976 Act.¹⁵² However, Congress decided not to include computer programs in the list because, in its view, the Copyright Act already encompassed software, so no explicit language to that effect was necessary. Both the House and Senate reports on the bill, using identical language, explained why explicit identification of software as copyrightable would be unnecessary:

[T]he history of copyright law has been one of gradual expansion in the types of works accorded protection, and the subject matter affected by this expression has fallen into two general categories. In the first, scientific discoveries and technological developments have made possible new forms of creative expression that never existed before. In some of these cases the new expressive forms-- electronic music, film strips, and *computer programs*, for example-- could be regarded as an extension of copyrightable subject matter Congress had already intended to protect, and were thus *considered copyrightable from the outset without the need of new legislation*. In other cases, such as photographs, sound recordings, and motion pictures, statutory enactment was deemed necessary to give them full recognition as copyrightable works.¹⁵³

Clearly, Congress did not see any need to explicitly identify all categories of copyrightable subject matter as such in the Copyright Act. To the contrary, the legislative history of the Copyright Act of 1976 evidences that Congress intended an expansive reading of works of authorship.¹⁵⁴ In 1980, the Copyright Statute was amended to include certain restrictions on the scope of copyright protection for computer programs, along with a definition of “computer programs,” but since Congress considered the Act to already cover software, there was no need to explicitly add software to the nonexclusive list of

¹⁵⁰ See *supra* Part III.

¹⁵¹ CONTU Report, *supra* note 37, at 16.

¹⁵² Transcript, CONTU Meeting No. 6, at 13.

¹⁵³ S. REP. NO. 94-473, at 50–51 (1975); H.R. REP. NO. 94-1476, at 51 (1976) (emphasis added).

¹⁵⁴ See Doreen M. Hogle, *Copyright for Innovative Biotechnological Research: An Attractive Alternative to Patent or Trade Secret Protection*, 5 HIGH TECH. L.J. 75, 89–90 (1990) (discussing H.R. 94-1476).

copyrightable subject matter provided by the statute.¹⁵⁵ In fact, if Congress had added computer programs to the list of copyrightable subject matter by amendment in 1980, this could have been misconstrued as implying that computer programs were not copyrightable prior to 1980, which clearly was not the intent or understanding of Congress.

The Copyright Office likewise recognized long before the 1980 amendment that a computer program can comprise copyrightable subject matter. In fact, in 1964 the Register of Copyrights announced that computer programs would be accepted for registration under the “rule of doubt,” indicating that the Copyright Office viewed software as potentially copyrightable even under the 1909 Act.¹⁵⁶ Since the 1980 amendments, courts have consistently found computer programs eligible for copyright protection, even in cases where the program was written before the 1980 amendments, thus implying that programs were copyrightable prior to the amendment.¹⁵⁷

Although the absence of specific reference to genetic sequences in the copyright statute should not render genetic sequences uncopyrightable any more than it excluded software from copyright protection prior to the 1980 amendments, Kumar and Rai are probably correct that courts will be less likely to recognize copyright protection for genetic sequences in the absence of some amendment of the Copyright Act acknowledging the copyrightability of engineered DNA. This occurred in the case of computer programs; prior to the 1980 amendments, some courts found that computer programs are not copyrightable.¹⁵⁸ The 1980 amendments resolved any ambiguity, and since that time the courts have consistently recognized the availability of copyright protection for software. In the same manner, amendment of the Copyright Act to explicitly mention engineered DNA would signal Congressional assent, and probably greatly facilitate acceptance of the idea by the courts.

Furthermore, Kumar and Rai are correct that analogy to other copyrighted works will be significant in extending copyright to engineered DNA. The analogy between computer software and literary works was important in the expansion of copyright to software, and the much closer analogy between engineered DNA and computer programs will greatly facilitate the incorporation of genetic sequences into the realm of copyrightable subject matter.

¹⁵⁵ 21 U.S.C. §§ 101, 116 (2006) (providing definitions).

¹⁵⁶ See CONTU Report, *supra* note 36, at 15.

¹⁵⁷ See *Apple Computer, Inc. v. Franklin Computer Corp.*, 714 F.2d 1240, 1251–52 (3d Cir. 1983).

¹⁵⁸ See *Data Cash Sys. Inc. v. JS&A Group, Inc.*, 480 F. Supp. 1063, 1066–67 (N.D. Ill. 1979).

X. THE DEVELOPMENT OF ENGINEERED GENETIC SEQUENCES ENTAILS SUFFICIENT EXPRESSIVE CHOICE TO WARRANT COPYRIGHT PROTECTION

Perhaps the most fundamental precept of copyright law is the idea/expression dichotomy which holds that copyright protection does not extend to ideas but is limited to the expression of ideas.¹⁵⁹ As a corollary, under the merger doctrine, copyright protection will be denied for the expression of an idea if there is only one or a relatively small number of ways of expressing the idea.¹⁶⁰ The policy behind the merger doctrine is that if there are insufficient alternative ways of expressing an idea, then extending copyright protection to the expression would unduly restrict access to the underlying protected idea.¹⁶¹

Some commentators have invoked the idea/expression dichotomy and merger doctrine against copyright for genetic sequences, arguing that there is insufficient expressive choice available for genetic engineers to justify copyright protection. However, this argument is based on a flawed premise; in fact, there are generally an astronomical number of alternate coding sequences that would direct essentially equivalent function in a cell in much the same way that multiple redundant software codes can be used to direct essentially the same function in a computer.

Commentators who have raised this argument against copyright for genetic sequences have often based it on the notion that there is a single “genetic code” which they allege substantially limits expressive choice in the design of genetic sequences. For example, Kumar and Rai argue that copyright protection for genetic sequences is “less likely” because the genetic code constrains the expressive choices available to synthetic biologists. They go on to suggest that if a synthetic biologist were to use an alternate genetic code, he or she would be more likely to incorporate sufficient expression to warrant copyright protection.¹⁶² Gargano argues along the same lines, finding that while in computer programs a single instruction can be expressed in numerous ways using different program languages, for “genetic programs” there is only one program language (by which he presumably means the genetic code).¹⁶³

Similarly, Silva argues

that the most important and stark difference [between a genetic sequence and a computer program] is that there is only one way to express a “genetic program.” This is by various combinations of the four nucleotide bases. The DNA instructions for producing proteins can only exist in the form of nucleotide se-

¹⁵⁹ See *Apple Computer*, 714 F.2d at 1252–53; *Mazer v. Stein*, 347 U.S. 201, 217 (1954).

¹⁶⁰ *Id.* at 1253.

¹⁶¹ See *Kregos v. Associated Press*, 937 F.2d 700, 705 (2d Cir. 1991).

¹⁶² Kumar & Rai, *supra* note 25, at 1764.

¹⁶³ Gargano, *supra* note 25.

quences. In essence, there is only one “program language” to express the method of producing proteins in cells. Computer programs, on the other hand, can a single instruction expressed in numerous ways via different program languages.¹⁶⁴

But these arguments cannot survive serious scrutiny. The genetic code is merely the “set of rules by which information encoded in genetic material (DNA or mRNA sequences) is translated into proteins (amino acid sequences) by living cells.”¹⁶⁵ To argue that alternate genetic codes are required to render an engineered genetic sequence copyrightable is equivalent to arguing that alternate software languages are necessary to render software copyrightable, or for that matter, alternate alphabets are necessary to render a novel copyrightable. Alphabets, software languages, and the genetic code are merely the raw materials used by authors to convey expressive content, and all provide ample opportunity for highly creative expression.

There are only twenty-six letters in the English alphabet, but clearly this has not impeded the ability of authors to create a wide range of diverse expressive works using the same small collection of letters. Although there are multiple computer languages in existence, a wide range of expressive computer programs can be written using a single language. And by the same token, a single genetic code is sufficient to allow for an astronomical range of diverse expression. One need merely look around at the genetic diversity of humans and other living organisms, the vast majority of which are using the same standard genetic code, to appreciate the huge expressive potential of the standard genetic code. To argue that alternate genetic codes are somehow necessary to provide synthetic biologists with sufficient expressive choices to create copyrightable works reflects a misunderstanding of biology and/or copyright law.

At one point in his article, Silva asserts that “[a] particular sequence is scientifically required to produce a protein. Any significant variation will result in no protein or production of useless protein.”¹⁶⁶ This is clearly incorrect. As noted above, for any given protein there are an astronomical number of different genetic sequences that will encode the identical protein.¹⁶⁷ Furthermore, it is well-established that substantial alterations can be introduced into a protein sequence without destroying function.¹⁶⁸ It is routine to find variants of a given protein structure differing by up to fifty percent that retain substantially equiva-

¹⁶⁴ Silva, *supra* note 25.

¹⁶⁵ “The code defines a mapping between tri-nucleotide sequences, called codons, and amino acids.” *Genetic code*, WIKIPEDIA, http://en.wikipedia.org/wiki/Genetic_code (last visited Mar. 9, 2011).

¹⁶⁶ Silva, *supra* note 25.

¹⁶⁷ Brief for Law Professor Christopher M. Holman, as Amicus Curiae Supporting Neither Party, *Ariad Pharms., Inc. v. Eli Lilly and Co.*, 2009 WL 3711551 (Fed. Cir. 2009).

¹⁶⁸ *Id.*

lent function.¹⁶⁹ Indeed, much of the recent progress in synthetic biology has been based on the ability to create huge libraries of protein variants sharing some degree of structural similarity and screening to identify proteins having different or improved function relative to the starting protein.¹⁷⁰

Gargano has similarly argued that gene sequences are barred from copyright protection by the merger doctrine, based on his assertion that “the idea of combining promoters, plasmids, genes and bacteria can only be expressed in limited ways, and therefore is not protectable.”¹⁷¹ While the earliest feats of genetic engineering were quite straightforward, at least when considered retrospectively, today’s synthetic biologists have at their disposal a host of different plasmids, genes, promoters and other genetic regulatory elements that can be combined in a virtually unlimited number of combinations, and introduced into any of a astronomical number of bacterial hosts. The diversity of available starting materials is expanding exponentially, with the discovery of previously unidentified naturally occurring bacteria and genetic elements, and increasingly through the creation of new genes and other genetic elements that do not exist in nature.

In assessing whether there are sufficient alternate ways for expressing an idea to permit copyright of those expressions, the definition of the idea is critical. If the idea is expressed in very abstract terms, such as Gargano’s “idea of combining promoters, plasmids, genes and bacteria,” then clearly there are a virtually unlimited number of ways of expressing it. But even if the idea behind a genetic sequence is expressed very specifically, such as coding a specific amino acid sequence, it is still the case that for virtually any real protein there exists an astronomical number of redundant genetic sequences coding for it.

XI. COPYRIGHT IS NOT PRECLUDED BY THE PRACTICAL UTILITY OF GENETIC SEQUENCES

Some have argued that the utilitarian nature of engineered genetic sequences renders them ineligible for copyright protection.¹⁷² This same argument was forcefully raised with respect to software in the 1970s and early 1980s, but resoundingly rejected by the courts.

For example, some prominent members of the CONTU commission voiced their opposition to extending copyright protection to computer programs having a highly utilitarian nature. One member of the commission, Hersey, dissented from the majority’s conclusion that software should be copyrightable at all, based in large part on his opinion that computer programs are too utilita-

¹⁶⁹ *Id.*

¹⁷⁰ *Id.*

¹⁷¹ Gargano, *supra* note 25.

¹⁷² *See* Gargano, *supra* note 25; Kumar & Rai, *supra* note 25; Silva, *supra* note 25.

rian to be eligible for copyright protection.¹⁷³ Professor Nimmer, another member of the commission, joined with the majority in recommending recognition of copyright for some computer programs, but in a concurring opinion he argued forcefully that utilitarian computer programs should not be eligible for copyright.¹⁷⁴ In a law review article published in 1984, years after the eligibility of computer programs for copyright protection would seem to have been firmly established, Professor Pamela Samuelson continued to argue forcefully that the utilitarian nature of software rendered copyright protection highly inappropriate.¹⁷⁵

Nonetheless, during the 1980s, courts extended copyright protection to even the most utilitarian software, such as operating system programs, and since then the copyrightability of highly utilitarian software has become firmly entrenched.¹⁷⁶ Today, there is ample authority firmly establishing that the utilitarian nature of software is no impediment to copyrightability, and there is no reason to think that the same standard should not apply to engineered genetic sequences.

XII. SOME LEVEL OF UNPREDICTABILITY DOES NOT PRECLUDE COPYRIGHT PROTECTION FOR ENGINEERED GENETIC SEQUENCES

Some of the arguments against copyright for genetic sequences appear to be based on a perception that genetic engineering is unpredictable and that copyright protection is only available for predictable creative works.¹⁷⁷ As explained above, synthetic biologists have made huge strides in the rational design and engineering of synthetic DNA sequences encoding useful and predictable biological function.¹⁷⁸ But more to the point, it is well-established that copyright protection is not precluded simply because it is impossible to predict the exact nature of the creative work prior to its creation. So while it is true that the specific functional attributes of an engineered genetic sequence are often not entirely predictable prior to its synthesis and empirical testing, this should not stand as a bar to copyright protection.

An artist creating a painting or sculpture, a photographer taking a picture, or a musician recording an improvised musical performance often cannot exactly predict the final form of the work, but that does not render the work uncopyrightable. Much highly regarded modern art is created by means of semi-random processes, but it is eligible for copyright protection nonetheless.

¹⁷³ CONTU Report, *supra* note 37, at 27.

¹⁷⁴ *See id.* at 26.

¹⁷⁵ Samuelson, *supra* note 49.

¹⁷⁶ *See* Apple Computer, Inc. v. Franklin Computer Corp., 714 F.2d 1240 (3d Cir. 1983).

¹⁷⁷ *See* S. REP. NO. 94-473, at 54 (1975); H.R. REP. NO. 94-1476, at 57 (1976).

¹⁷⁸ *See supra* Part V.

As noted by Professor Nimmer, the “independent effort that constitutes originality may be inadvertent and still satisfy the requirements of copyright.”¹⁷⁹

In the realm of computer programming, software engineers (being human) lack the cognitive ability to accurately predict the performance of a complex modern computer program without actually empirically testing it, as reflected in the widespread practice of beta testing and the common occurrence of unanticipated problems with computer software products. In principle, of course, a hypothetical engineer of unlimited cognition could predict all possible outcomes of even the most complex software, but the same might be said with respect to engineered DNA because genes and proteins follow established rules of chemistry and physics. In practice, both must be tested empirically to confirm function and identify flaws, but this should not stand as an impediment to copyright protection.

In fact, software engineering is developing in a manner in which the final form of the computer program is less and less predictable based on the original work of the software engineer. Many computer programs now include self-modifying code that alters its own instructions while executing.¹⁸⁰ Evolutionary computing systems such as genetic programming are designed to permit a computer program to learn from experience and optimize a computer program’s ability to perform a given computational task.¹⁸¹ Meta-Genetic Programming is a proposed technique for evolving a genetic programming system using genetic programming itself.¹⁸² The unpredictability associated with the development of these computer programs does not detract from their copyrightability nor should it detract from the copyrightability of engineered genetic sequences.

¹⁷⁹ NIMMER, *supra* note 38, § 2.01[B]; *see, e.g.*, Alfred Bell & Co. v. Catalda Fine Arts Inc., 191 F.2d 99, 105 (2d Cir. 1951) (“A copyist’s bad eyesight or defective musculature, or a shock caused by a clap of thunder, may yield sufficiently distinguishable variations. Having hit upon such a variation unintentionally, the ‘author’ may adopt it as his and copyright it.”).

¹⁸⁰ “[C]ode that alters its own instructions while it is executing - usually to reduce the instruction path length and improve performance or simply to reduce otherwise repetitively similar code, thus simplifying maintenance.” *Self-modifying code*, WIKIPEDIA, http://en.wikipedia.org/wiki/Self_modifying_code (last visited Mar. 9, 2011).

¹⁸¹ *Genetic programming*, WIKIPEDIA, http://en.wikipedia.org/wiki/Genetic_programming (last visited Mar. 9, 2011) (“[A]n evolutionary algorithm-based methodology inspired by biological evolution to find computer programs that perform a user-defined task. It is a specialization of genetic algorithms (GA) where each individual is a computer program. It is a machine learning technique used to optimize a population of computer programs according to a fitness landscape determined by a program’s ability to perform a given computational task.”).

¹⁸² *Id.*

XIII. PROVIDING COPYRIGHT FOR DNA DOES NOT IMPLY THAT COPYRIGHT IS GENERALLY AVAILABLE FOR CHEMICAL COMPOUNDS CAPABLE OF CONVEYING INFORMATION

It has been pointed out that DNA is far from unique in its ability to convey information. Hormones, neurotransmitters, and transcription regulators, for example, convey biological information. Some have expressed a concern that if copyright is extended to genetic sequences there will be no principled basis for excluding other molecules capable of conveying biological information, potentially opening the floodgates to a dramatic and unwise expansion of copyrightable subject matter. While such concerns should not be dismissed out of hand, it is important to recognize that similar concerns were expressed when copyright was extended to computer programs.

For example, Professor Nimmer opined in a concurring opinion to the CONTU report that “[w]hat is most troubling about the Commission’s recommendation of open-ended copyright protection for all computer software is its failure to articulate any rationale which would not equally justify copyright protection for the tangible expression of any and all original ideas (whether or not computer technology, business, or otherwise).”¹⁸³ Nimmer believed that the extension of copyright to software threatened to transform copyright into a general misappropriation law. To avoid this slippery slope, he suggested that copyright protection should be limited only to “computer programs which produce works which themselves qualify for copyright protection,”¹⁸⁴ i.e., word processors, graphics programs and the like. In contrast, he felt that more utilitarian software, such as computer programs “which control the heating and air-conditioning in a building, or which determine the flow of fuel in an engine, or which control traffic signals would not be eligible for copyright because their operations do not result in copyrightable works.”¹⁸⁵

Of course, with the benefit of hindsight, it is apparent that Nimmer’s fear has never come to pass. The courts have extended copyright protection to highly functional software that does not itself produce copyrightable works, and this has not led to a further expansion of copyrightable subject matter, nor has copyright law been converted into a general misappropriation law. In fact, fundamental differences exist between software and other tangible expressions of ideas thus avoiding the slippery slope problem identified by Professor Nimmer. In the same manner, fundamental differences between engineered genetic sequences and other molecules capable of conveying information provided a principled basis for limiting copyright to DNA.

For one thing, DNA is imbued with informational characteristics that distinguish it from other molecules in ways that are both fundamental and qua-

¹⁸³ CONTU Report, *supra* note 37, at 26.

¹⁸⁴ *Id.* at 27.

¹⁸⁵ *Id.*

litative. In the related context of gene and DNA patents, academic commentators have repeatedly argued that DNA is unique in this regard. For example, Arti Rai and Eileen Kane have opined that, in determining the applicability of intellectual property rights, DNA is better conceptualized as a carrier of information than a molecule.¹⁸⁶ Similarly, Rebecca Eisenberg has suggested that in view of the primarily informational role played by genetic sequences, such sequences might constitute “non-functional descriptive material,” and as such be unpatentable.¹⁸⁷ Eileen Kane has argued that “a DNA gene sequence possesses the duality of being both a static chemical and a dynamic template executed by the genetic code.”¹⁸⁸ These sorts of concerns have not been raised with respect to other molecules capable of conveying information.

The unique informational properties of DNA have been noted outside the legal academy. In a lawsuit challenging the patentability of naturally occurring gene sequences and genetic information, a number of prominent scientists and non-legal scholars have submitted declarations strongly advocating for a differential treatment of DNA under the intellectual property laws based on the unique informational characteristics of DNA.¹⁸⁹

For example, Myles Jackson, a professor of the history of science and technology at New York University, stated in his declaration that DNA and genes are not just chemicals and should not be analyzed for patent purposes in the same manner as other chemicals.¹⁹⁰ John Sulston, a noted genomic researcher stated in his declaration that a genetic sequence “is the *biological information itself*.”¹⁹¹ These and a host of other expert declarants in the case repeatedly stressed that DNA conveys complex information in a manner fundamentally different than other chemicals. The informational role of genetic sequences even led the plaintiffs in the lawsuit to challenge the constitutionality of gene patents under the First Amendment based on a theory that restrictions on the use of genetic sequences are tantamount to restrictions on speech.¹⁹²

Christopher Mason, a postdoctoral associate in the program on neurogenetics at Yale University and a visiting fellow at Yale Law School, pointed out in his declaration that the information content of a gene is identical, whether represented as a string of letters in a human readable format, or as a DNA mole-

¹⁸⁶ Eileen Kane, *Splitting the Gene*, 71 TENN. L. REV. 707, 743 (2004).

¹⁸⁷ Rebecca Eisenberg, *How Can You Patent Genes?*, 2 AM. J. BIOETHICS 3, 7 (2002).

¹⁸⁸ Kane, *supra* note 186, at 744.

¹⁸⁹ Ass’n for Molecular Pathology v. U.S. Patent and Trademark Office, 702 F. Supp. 2d 181 (S.D.N.Y. 2010).

¹⁹⁰ Declaration of Myles W. Jackson, Ass’n for Molecular Pathology v. U.S. Patent and Trademark Office, No. 09-4515, 2009 WL 6634217, ¶¶ 12, 14, 15, 49.

¹⁹¹ Declaration of Sir John E. Sulston, Ass’n for Molecular Pathology v. U.S. Patent and Trademark Office, No. 09-4515, 2009 WL 6634213, ¶ 16.

¹⁹² Ass’n for Molecular Pathology, 702 F. Supp. 2d at 181.

cule in a test tube or in the body.¹⁹³ He went on to explain that a cell interprets the set of instructions provided in a genetic sequence by means of processes referred to as transcription and translation.¹⁹⁴ It is telling that the very words coined by biologists to describe these natural processes, transcription and translation reflect the unique informational role played by DNA and the close analogy to a written set of instructions.

The biological transcription and translation machinery present in cells, composed primarily of proteins and RNA molecules, is analogous to the components of a computer, such as the compiler, that process the instructions encoded in a computer program to cause the computer and any associated machinery to perform useful functions. While other molecules can convey information in a biological system, DNA is unique in that cells contain dedicated processes and molecular machinery for decoding and implementing the information coded by the genetic sequence. DNA is much more analogous to a computer program than other molecules, in that both can be expressed equivalently either as human-directed instructions printed on a piece of paper or instructions embodied in an arrangement of matter directed towards machine (e.g., a DNA molecule or CD-ROM).

Another distinction between DNA and other molecules capable of conveying information is the complexity of the information that can be conveyed by a genetic sequence. The human body, whose structure and function is encoded by DNA, provides compelling testimony of the level of complexity possible in DNA encoded instructions. This complexity is possible because a genetic sequence can convey a long, ordered set of instructions in the form of amino acid-encoding codons, transcription regulators, genes, etc., which is analogous to the complex set of instructions provided by a computer program. The set of genetic instructions can be modified, recombined, and expanded in much the same way as a computer program. As is the case for computer programs, the range of expressive content inherent in DNA is limited only by the imagination.

It is significant that the Copyright Act's definition of a computer program specifies a *set* of instructions.¹⁹⁵ The fact that a computer program comprises a plurality of instructions that can be strung together in creative ways is probably essential to their copyrightability. A single instruction, like a short phrase, slogan or unit of information, would generally not be copyrightable. Similarly, a biological molecule capable of conveying a single or limited number of instructions, should not be copyrightable. An engineered genetic sequence, on the other hand, that encodes a complex set of instructions that can be recombined in creative ways to achieve useful functions is more analogous to a computer program than other molecules capable of conveying a much more limited range of information.

¹⁹³ Declaration of Christopher E. Mason, *Ass'n for Molecular Pathology v. U.S. Patent and Trademark Office*, No. 09-4515, 2009 WL 6634216, ¶ 32.

¹⁹⁴ *Id.* ¶¶ 11–12.

¹⁹⁵ 17 U.S.C. § 101 (2006).

Like software, a genetic sequence can be modified in predictable ways to alter the information content. The same cannot be said for most molecules. There is no generally predictable way to modify a steroid hormone to cause it to convey a different message in a living organism. As synthetic biology matures, the power and predictability with which genetic sequences will be amenable to modification, recombination and rational design will continue to expand, bringing software and engineered genetic sequences into even closer alignment.

XIV. CONCLUSION

The decision to treat software as a copyrightable literary work was a major conceptual leap in copyright doctrine. In this Article, I have argued that further expansion to encompass engineered genetic sequences would be modest by comparison and would appear to flow naturally from the logic supporting the initial decision to open up copyright to software. However, the question remains whether such an expansion of copyrightable subject matter would further public policy, and if so, how copyright could best be implemented in a manner that fosters innovation and maximizes access. In 1988, Dan Burk likewise concluded that copyright doctrine could readily accommodate engineered DNA, but at the time he felt that public policy considerations weighed against doing so. But much has changed in the intervening years, and developments in synthetic biology call for a serious reconsideration of the potential policy benefits of copyright.

Copyright protection would not entirely supplant the role of patents but in many instances might be preferable for protecting some aspects of innovation in synthetic biology. As noted by Drew Endy, patents can be prohibitively expensive to obtain and enforce, and the long delay between the filing of a patent application and issuance of a patent is inconsistent with the rapid pace of innovation in synthetic biology. Copyright is particularly suited for protecting against infringement by direct copying, which is a highly useful characteristic for technologies such as software and DNA that are extremely vulnerable to pirating due to the ease with which an unlimited number of identical copies can be produced. Criminal penalties and provisions for blocking importation of infringing materials at the border substantially augment the practical enforceability of copyrights, as do the availability of statutory damages and litigation costs for prevailing copyright owners.

Copyright for engineered DNA could also facilitate the development of true open source biology. Open source has been successfully implemented in software development due in large part to the important role copyright place in protecting software. In contrast, attempts at open source biotechnology have been much less successful, due in large part to the use of patents rather than copyright in biotechnology. If synthetic biologists could rely on copyright protection as an alternate to patent, at least for certain aspects of their innovation, it might be possible to fashion some sort of working open source biotechnology based on principles successfully developed in the realm of software.

In many ways, the protection afforded by copyright is more limited than patent, which could be a boon to subsequent innovation. A fundamental precept of copyright law is that only the expression of an idea can be protected, not the idea itself. Thus, subsequent innovators could examine and learn from an engineered genetic sequence and even incorporate the ideas of that sequence in their own work so long as they do not simply copy the original innovator's choice of expression.

Copyright law has a well-developed doctrine known as fair use, which allows for certain limited uses of a copyrighted work without any liability for copyright infringement, particularly when the use is scholarly in nature or results in some significant transformation of the original work or does not unduly harm the economic interests of the copyright owner. In cases where fair use is not available, Congress has created a number of compulsory licensing provisions which allow for the use of copyrighted material upon the payment of some fee to the copyright owner.

There is ample precedent in copyright law for providing only limited protection to certain works. For example, copyright in sound recordings does not include the right to public performance which is afforded to the creators of most works. The copyright statute contains a number of restrictions aimed specifically at software. In the same manner, Congress could provide some limited form of copyright protection to engineered DNA in a manner that furthers public policy, without creating problems that might occur if unlimited copyright protection is provided to DNA.