

How to publish DNA sequences with copyright protection

Willem P.C. Stemmer

I propose here an approach to encoding that will allow genomics companies to make their sequences available to the public while retaining some intellectual property (IP) protection. Using this approach, such companies would transform the DNA sequences in their databases into music files (e.g., MP3 format). An external database user would copy this music file, transfer the copy to himself, and re-convert the music file into DNA sequence using a back-translation program. This back-translated DNA sequence itself would not be covered by copyright. However, IP protection may exist because the external user can access a DNA sequence only by copying of a copyright-protected music file. One hopes that the ability to obtain limited IP protection for DNA sequences will encourage institutions with genomic data to make their sequence databases publicly available.

Original works of authorship, such as music, literature and sculpture, are covered by copyright¹. Nonnatural DNA sequences, to the extent they constitute works of authorship, may (N. Boorstyn, personal communication) or may not² be protected by copyright. An example is the DNA sequence of an improved variant of a gene, generated by rational design or directed evolution. However, natural DNA sequences are generally considered not to be covered by copyright, as they are not works of authorship (N. Boorstyn, personal communication; ref. 2). They were not created by scientists, but simply uncovered.

If significant functional information and a utility are described, isolated natural and nonnatural DNA sequences can be protected by patents. However, sufficient information is typically only obtainable for a small fraction of the genes in a genome. The great majority of genomic DNA sequences can therefore not readily be protected by patents.

In contrast to patent protection, which lasts for only 17 years, copyright protection continues for up to 100 years, depending on the identity of the author¹. To some extent, copyright also protects portions and derivative works.

Willem P.C. Stemmer is vice president of R&D, Maxygen, Inc., 515 Galveston Drive, Redwood City, CA 94063 (pim.stemmer@maxygen.com).

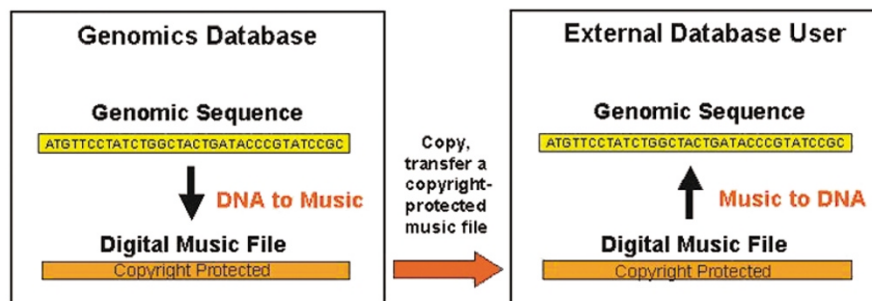


Figure 1. Harmonious sequence access. An external database user accesses a genomics database and copies a DNA sequence file that is encoded as a music file (e.g. MP3 format). The external user then converts the MP3 file back to the DNA sequence using the accompanying back-translation program. Both conversions can be transparent to the user.

When a genomics company publishes its DNA sequences, unless patent applications have been filed previously, they lose the ability to file for patents on these sequences. For this reason, most genomics sequences remain unpublished and inac-

Using the approach suggested here, genomics companies would be able to publish their genomics sequences and still retain some intellectual property in these data as a reward for their investment.

cessible to the scientific community. Using the approach suggested here, genomics companies would be able to publish their genomics sequences and still retain some intellectual property in these data as a reward for their investment.

As the Internet provider Napster recently discovered, music files are clearly protected by copyright¹, and unauthorized copying of a digital music file from a music database clearly constitutes copyright infringement. Music is just an example, as most types of encoding can result in copyright protection.

Implementation of the “musical DNA” approach could proceed as follows (see Fig.1): DNA sequences in the genomics

database are encoded as music files using an existing program, such as Bio2MIDI (ref. 3). The external user accesses the genomics database, copies a music file corresponding to the DNA segment desired, and then back-translates this music file to the DNA sequence. The back-translation program could be sent together with the music file. Both conversion steps, from DNA to music and from music to DNA, can be transparent to the user as they do not require any user input.

In the end, the DNA sequences thus obtained by the database user are still not copyrightable. However, only by copying a copyright-protected music file from the database could the external user obtain the DNA sequences. Such duplication may be copyright infringement, and this encoding approach therefore may provide a form of IP protection.

Clearly, there is a certain degree of whimsy in this proposal. But there are also more than a few quanta of genuine intent that this approach, or a similar one, may help to set the industry back on a (pardon the pun) less dissonant course, especially with regard to an increasingly negative public perception.

Acknowledgments

Thanks to Neil Boorstyn and Sharon Fujita for their comments.

- Boorstyn, N. *Boorstyn on copyright*, Vol. 1/2, 2d edn. (Boorstyn, N., 1994).
- Cooper, I.P. *Biotechnology and the law*, Vol. 1–2. (Clark Boardman Callaghan, 1982).
- <http://geneticmusic.com/>; <http://www.algoart.com/>