Book Report

A Molecular Biologist Visits *Jurassic Park*

Mark S. Boguski
National Center for Biotechnology Information, National Library of Medicine,
National Institutes of Health

What do you get when you mix dinosaurs and DNA, chaos and computers, Frankenstein and fractals, vortex briefly and then divide into several dozen movie scene-sized aliquots? The answer is Michael Crichton's techno-thriller *Jurassic Park* (1990, Knopf, New York) that has recently been released in paperback. Crichton is a Harvard Medical School graduate who achieved popular success as a writer with his book *The Andromeda Strain* and went on to direct hit movies such as *Westworld* and *Coma.* Having greatly enjoyed his earlier work, I eagerly awaited an opportunity to read *Jurassic Park* which is based on the premise that molecular biotechnology will permit the resurrection of extinct animals from fossilized DNA.

In 1963 Linus Pauling and Emile Zuckerkandl published a paper entitled "Chemical Paleogenetics: Molecular 'Restoration Studies' of Extinct Forms of Life" (8). Despite their obviously provocative title, Pauling and Zuckerkandl were only attempting to infer the sequences of ancestral polypeptide chains. Little did they realize at the time, however, that it would one day actually be possible to rescue portions of extinct genomes using approaches pioneered by the late Allan Wilson and co-workers (7). DNA has been extracted from a variety of ancient remains and there are, for example, two GenBank® entries (accession Nos. M30380, M30383) for extinct DNA from the quagga—a relative of horses, zebras and asses—the last of which died about 140 years ago. Sufficient DNA has been recovered from 7500-year-old human remains to permit HLA typing (5). However, the current record, I believe, for the oldest piece of cloned DNA is from the leaf of an extinct magnolia tree that fell during an autumn wind 17–20 million years ago (2).

Crichton's plot assumes that full genome equivalents of DNA are recoverable from nucleated erythrocytes in the gastrointestinal tracts of insects (mosquitos?) that ingested a blood meal from a dinosaur just before becoming fossilized in amber. From here on in, it is easy to imagine nuclear transplantation technology [which dates back more than 30 years (3,4)] with unfertilized chicken eggs being used to grow dinosaurs. But to what end? Why for a millionaire's theme park, of course! With entrepreneurial spirit (and sufficient venture capital), a fictional biotechnology firm called InGen creates a biological Disneyland on an island off the coast of Costa Rica. Merchandizing possibilities rival or exceed those of the Teenage Mutant Ninja Turtles.

Early in the book, InGen's chief scientist, Dr. Henry Wu, is giving a tour of the laboratory facilities which contain three Cray® XMP supercomputers and automated gene sequencers which are called "hoods" (an amusing allusion to Lee Hood of Applied Biosystems that will be lost on the general reader). At one stop on the tour (p. 103), we are shown an actual sample of dinosaur DNA sequence

```
   1 gcgttgctgg cgtttttcca taggctccgc cccctgacg agcatcacaa
  51 aaatcgacgc ggtggcgaaa cccgacagga ctataaagat accaggcgtt
 101 tccccctgga agctccctcg tgttcgacc ctgccgctta ccggatacct
 151 gtccgccttt ctcccttcgg gaagcgtggc tgctcacgct gtaggtatct
 201 cagttcggtg taggtcgttc gctccaagct gggctgtgtg ccgttcagcc
 251 cgaccgctgc gccttatccg gtaactatcg tcttgagtcc aacccggtaa
 301 agtaggacag gtgccggcag cgctctgggt cattttcggc gaggaccgct
 351 ttcgctggag atcggcctgt cgcttgcggt attcggaatc ttgcacgccc
 401 tcgctcaagc cttcgtcact ccaaacgttt cggcgagaag caggccatta
 451 tcgccggcat ggcggccgac gcgctgggct ggcgttcgcg acgcgaggct
 501 ggatggcctt ccccattatg attcttctcg cttccggcgg cccgcgttgc
 551 aggccatgct gtccaggcag gtagatgacg accatcaggg acagcttcaa
 601 cggctcttac cagcctaact tcgatcactg gaccgctgat cgtcacggcg
 651 atttatgccg cacatggacg cgttgctggc gtttttccat aggctccgcc
 701 cccctgacga gcatcacaaa caagtcagag gtggcgaaac ccgacaggac
 751 tataaagata ccaggcgttt cccccctggaa gcgctctcct gttccgaccc
 801 tgccgcttac cggatacctg tccgcctttc tcccttcggg ctttctcaat
 851 gctcacgctg taggtatctc agttcggtgt aggtcgttcg ctccaagctg
 901 acgaaccccc cgttcagccc gaccgctgcg ccttatccgg taactatcgt
 951 cttgagtcca acacgactta acgggttggc atggattgta ggcgccgccc
1001 tataccttgt ctgcctcccc gcggtgcatg gagccgggcc acctcgacct
1051 gaatggaagc cggcggcacc tcgctaacgg ccaagaattg gagccaatca
1101 attcttgcgg agaactgtga atgcgcaaac caaccttgg ccatcgcgtc
1151 cgccatctcc agcagccgca cgcggcgcat ctcgggcagc gttgggtcct
1201 gcgcatgatc gtgctagcct gtcgttgagg acccggctag gctggcgggg
1251 ttgccttact atgaatcacc gatacgcgag cgaacgtgaa gcgactgctg
1301 ctgcaaaacg tctgcgacct atgaatggtc ttcggtttcc gtgtttcgta
1351 aagtctggaa acgcggaagt cagcgccctg
```

that Dr. Wu indicates "probably contains instructions to make a single protein — say, a hormone or an enzyme." Unable to stifle my scientific curiosity, I simply had to find out what this DNA *really* was. My guess was that it was either random sequence, or if Crichton and his technical consultants were really clever, it could be a chicken gene artfully "mutated" to take into account 270 million years of molecular evolution.

```
National Center for Biotechnology Information (NCBI)

Experimental GENINFO(R) BLAST Network Service]

Query= "Dinosaur DNA" from p. 103 of JURASSIC PARK
       (1380 nucleotides, both strands)

Database:  GenBank(R) Release 71.0, March 15, 1992
           65,100 sequences; 83,894,652 total residues.

                                                                    Smallest
                                                                     Poisson
                                                              High  Probability
Sequences producing High-scoring Segment Pairs:               Score  P(N)      N

SYNPBR322  Cloning vector plasmid pBR322, complete genome.     328  0.0       17
PLMCG      Plasmid pAA3.7X from E.coli, complete genome.       328  0.0       17
SYNAA113M  Sequencing vector pAA113M DNA.                      328  0.0       17
SYNPKK223  pKK223-3 - Cloning Vector from PL-Pharmacia.        328  0.0       17

>SYNPBR322 Cloning vector plasmid pBR322, complete genome.
           Length = 4361

  Plus Strand HSPs:

 Score = 328 (90.6 bits), Expect = 1.9e-17, P = 1.9e-17
 Identities = 68/71 (95%), Positives = 68/71 (95%), Strand = Plus

Query:   721 CAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAA 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2581 CAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAA 2640

Query:   781 GCGCTCTCCTG 791
             || | ||| ||
Sbjct:  2641 GCTCCCTCGTG 2651


 Score = 320 (88.4 bits), Expect = 8.8e-17, Poisson P(2) = 1.0e-37
 Identities = 64/64 (100%), Positives = 64/64 (100%), Strand = Plus

Query:   478 GCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGG 537
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   964 GCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGG 1023

Query:   538 CGGC 541
             ||||
Sbjct:  1024 CGGC 1027


 Score = 320 (88.4 bits), Expect = 8.8e-17, Poisson P(3) = 7.8e-59
 Identities = 68/73 (93%), Positives = 68/73 (93%), Strand = Plus

Query:   530 GCTTCCGGCGGCCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGG 589
             || || ||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1026 GCATCGGGATGCCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGG 1085

Query:   590 GACAGCTTCAACG 602
             ||||||||||| |
Sbjct:  1086 GACAGCTTCAAGG 1098

 Total cpu time:  5.58 sec
```

Figure 1. Database search results using "Dinosaur DNA" as a query to search GenBank (release 71.0) with the BLASTN program (1). Multiple segments of the query were identical in sequence to various regions of pBR322; for space considerations, only the first three local alignments are shown. There is not a simple, colinear relationship between the query sequence and pBR322. In fact it seemed that the query was composed of approximately 60-nucleotide segments of 100% identity separated by 10-nucleotide, non-matching sequences of undetermined origin. Dot matrix analysis (Figure 2) clarified the relationship.

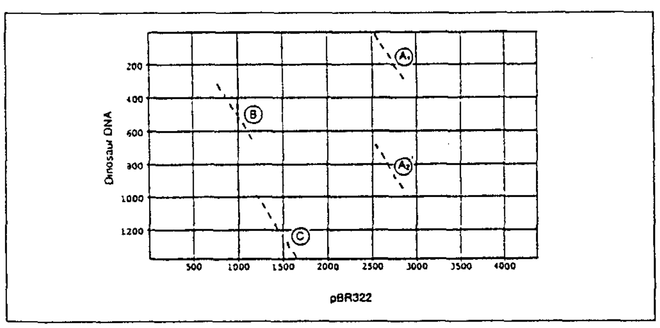

Figure 2. Comparison matrix analysis of "Dinosaur DNA" vs. pBR322. IBI MacVector™/Pustell DNA matrix option (International Biotechnologies, New Haven, CT) was used with "window size" parameter set to 35 and "min. % score" set to 100. It can be seen that the putative dinosaur DNA is a quadripartite sequence with each of the four subsequences constructed of short, identical segments of pBR322 strung together with shorter, non-pBR spacers. One of the subsequences is used twice (A1, A2) with an intervening sequence (B) and a 3′ distal sequence (C) drawn from other regions of pBR322. Crichton has obviously taken some pains to disguise the modern, prokaryotic origins of his "dinosaur DNA." He should have gone to the extra trouble of at least starting with a eukaryotic gene.

You can imagine my disappointment when this exotic sequence actually turned out to consist of pieces of lowly old pBR322, a man-made cloning vector and one of the most common pieces of DNA in the (contemporary) world (Figures 1 and 2)! My respect for Dr. Wu's scientific ability vanished. After all, he was unable to determine with three Crays what it took me two minutes on a Macintosh® to discover, that he was either dealing with contamination by vector sequences (6) or that dinosaurs were contemporaneous with early molecular biologists.

From this point on, the story was not the same for me, but I persevered only to learn, once again, that scientists and businessmen are evil (or at least amoral), that you shouldn't fool with Mother Nature and that writing science fiction probably pays better than writing science fact. Maybe I can get a part in the movie.

Note added in proof: The potential public relations impact of *Jurassic Park* on the biotechnology industry has recently been discussed in Genetic Engineering News *12*:22–24, 1992.

REFERENCES

1. **Altschul, S.F., W. Gish, W. Miller, E.W. Myers and D.J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. *215*:403-410.
2. **Golenberg, E.M., D.E. Giannasi, M.T. Clegg, C.J. Smiley, M. Durbin, D. Henderson and G. Zurawski.** 1990. Chloroplast DNA sequence from a miocene Magnolia species. Nature *344*:656-658.
3. **Gurdon, J.B.** 1962. The developmental capacity of nuclei taken from intestinal epithelial cells of feeding tadpoles. J. Embryol. Exp. Morphol. *10*:622-640.
4. **King, T.J. and R. Briggs.** 1956. Serial transplantation of embryonic nuclei. CSHSQB *21*:271-289.
5. **Lawlor, D.A., C.D. Dickel, W.W. Hauswirth and P. Parham.** 1991. Ancient HLA genes from 7,500-year-old archaeological remains. Nature *349*:785-788.
6. **Lopez, R., T. Kristensen and H. Prydz.** 1992. Database contamination. Nature *355*:211.
7. **Paabo, S., R.G. Higuchi and A.C. Wilson.** 1989. Ancient DNA and the polymerase chain reaction: The emerging field of molecular archaeology. J. Biol. Chem. *264*:9709-9712.
8. **Pauling, L. and E. Zuckerkandl.** 1963. Chemical paleogenetics: molecular "restoration studies" of extinct forms of life. Acta Chem. Scand. *17*:S9-S16.

Address correspondence to Mark S. Boguski, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894. Internet: boguski@ncbi.nlm.nih.gov