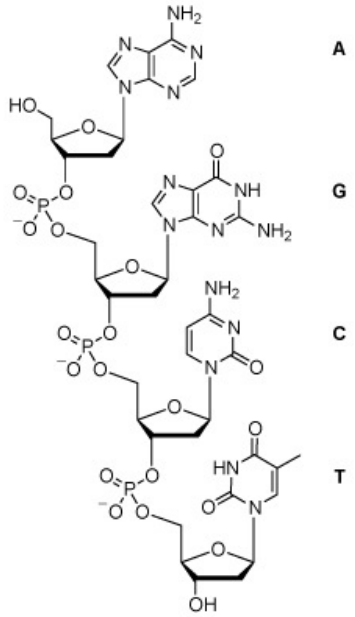
Digital computers store data and instructions as a string of two numbers (0,1) in a binary number (base-2) system. The on-off states of nanometer-size transistors in integrated circuits can represent the 0's and 1's, for example.

A DNA molecule stores data and instructions for a living organism as a string of four nanometer-size chemical groups (abbreviated A,T,G,C) along the backbone of a long DNA molecule.



We can think of the chemical groups as representing four numbers (0,1,2,3) in a quaternary number (base-4) system.

The sequence of the groups along the molecule - e.g., TGACCGATGTGC – encodes all the information necessary for life to live, grow and reproduce.

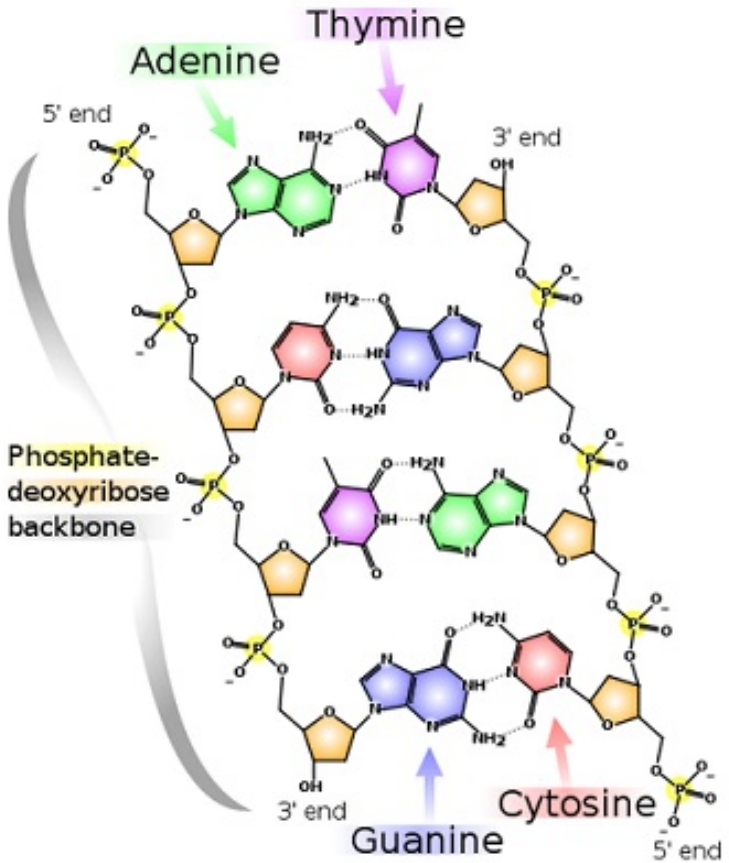
Can you think of any advantages to the quaternary system used in DNA over the binary system used by digital computers?

http://www.atdbio.com/content/2/Molecular-weight-and-mass

The figure shows a short section of two complementary DNA molecules. Each molecule consists of a long chain (phosphate deoxyribose backbone) with the ATGC groups or “bases” attached along the chain.

The groups along the molecule on the left are attracted only to complementary groups or “base pairs” along the molecule on the right: A-to-T and G-to-C.

This pairing of molecules allows the information coded on one molecule to be copied on the other molecule, which allows multiple copies of the information to be made during cell division and other important processes.



http://eduflix.tv/ask/questions/422/which-is-a-single-stranded-dna-virus

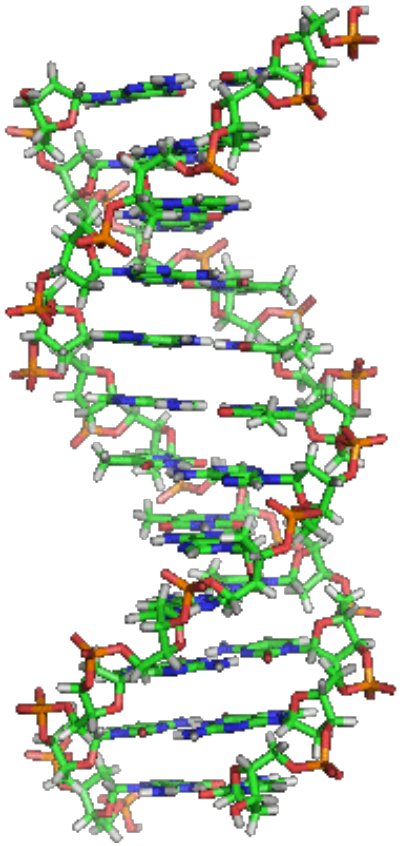
There are about six trillion base pairs in a human cell.

The DNA in one human chromosome, if stretched out, would be a few centimeters in length.

The DNA in one human cell nucleus, if stretched out and put end-to-end, would be a few meters in length.

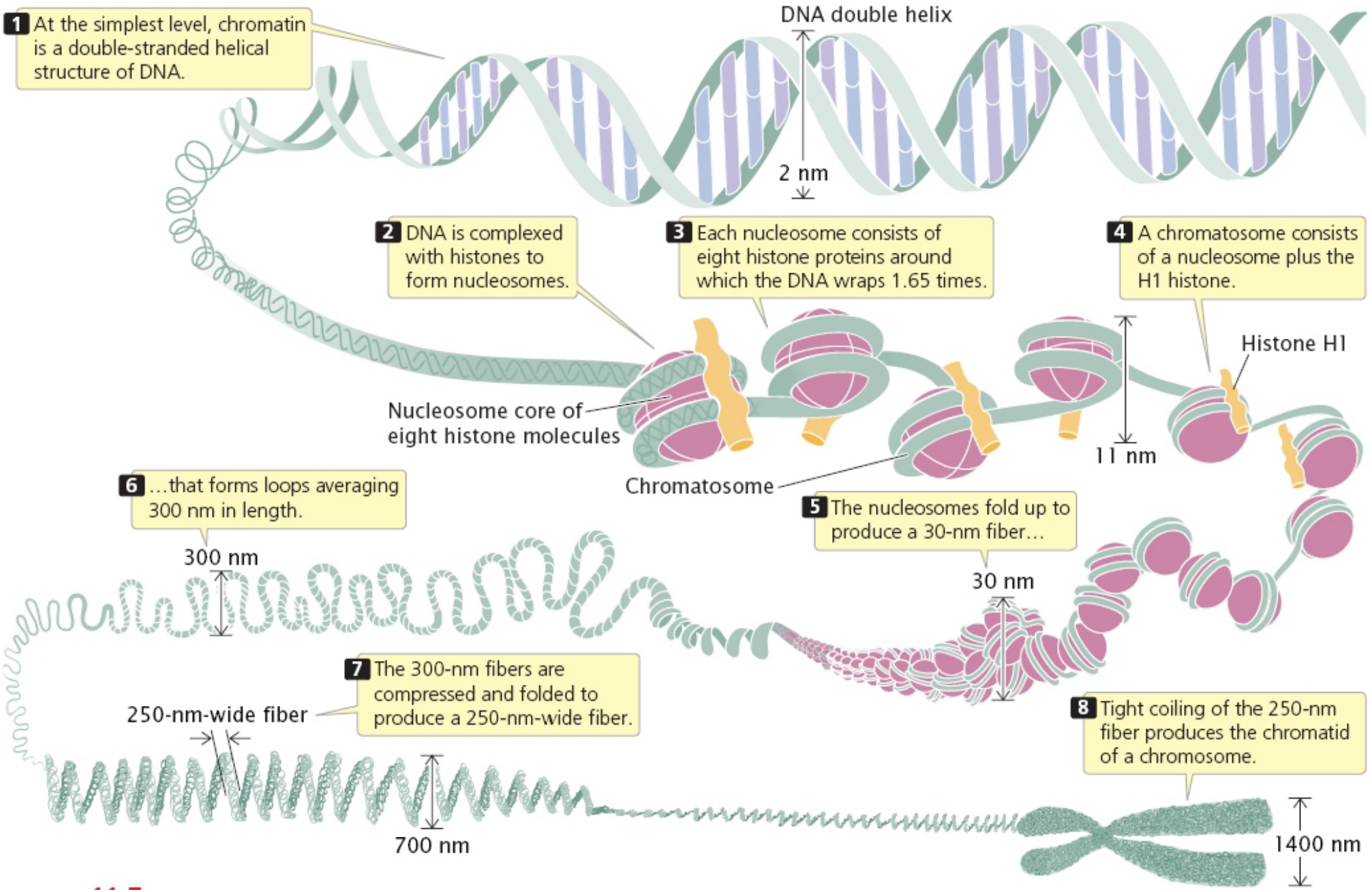
How is all this DNA packed into a tiny cell?

By lots of twisting and winding and wrapping!



http://www.karelianbeardog.us/kbd_science.html

http://hypertextbook.com/facts/1998/StevenChen.shtml



http://www.nature.com/scitable/resource?action=showFullImageForTopic&imgSrc=/scitable/content/18847/

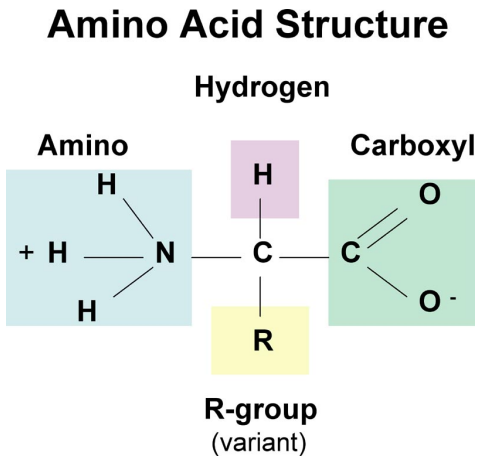
Richard K. Herz, ReactorLab.net

We asked earlier: “Can you think of any advantages to the quaternary system used in DNA over the binary system used by digital computers?”

There are 20 types of amino acid molecules in human bodies. Amino acid molecules are used to build enzymes (biological catalysts) and other protein molecules.

How many bits of information (0’s and 1’s) are required to encode a type of amino acid in the binary (base-2) number system?

How many bits of information (ATGC groups) are required to encode a type of amino acid in DNA’s quaternary (base-4) system?



Human glyoxalase I enzyme

http://www.whitetigernaturalmedicine.com/nutrition/proteins-amino-acids

http://en.wikipedia.org/wiki/Enzyme

In the binary (base-2) system,

00000 represents decimal 0

01111 represents decimal 15 = $0 + 2^3 + 2^2 + 2^1 + 2^0$

10000 represents decimal 16 = $2^4 + 0 + 0 + 0 + 0$

11111 represents decimal 31

You need 5 bits to code for 20 amino acids

In the quaternary (base-4) system,

000 represents decimal 0

003 represents decimal 3 = $3*4^0 = 3*1$

033 represents decimal 15 = $3*4^1 + 3*4^0$

300 represents decimal 48 = $3*4^2 = 3*16$

You need 3 bits to code for 20 amino acids

So you need 5 bits in a digital computer's code but only 3 bits - 40% fewer - in DNA's code.

This reason makes it significantly easier to pack all the information required into a cell nucleus. There may be chemical reasons for the need for more than two bases.

Quick story. Many years ago I stopped my car at a stop light at the top of Mt. Soledad. I glanced to my right and saw a car with the California license plate “ATGC.” I looked up to see that the driver was Francis Crick, the co-discover of the structure of DNA! Crick was then doing research at the Salk Institute.

The sequence of the ATGC bases in DNA stores both the data and the instructions – the program – for life.

The Nobel Prize in Medicine in 2012 was shared by Professor Shinya Yamanaka of Kyoto University in Japan. Professor Yamanaka discovered how to “reprogram” a cell. He was able to take a cell that had differentiated into one type of tissue cell and reprogram the cell such that the cell became a stem cell, able to grow any kind of tissue. This reprogramming was done by changing the “transcription” of the DNA in the cell - how the ATGC sequence was read.

http://www.nytimes.com/2012/10/09/health/research/cloning-and-stem-cell-discoveries-earn-nobel-prize-in-medicine.html

http://www.cc.com/video-clips/fkt99i/the-colbert-report-george-church

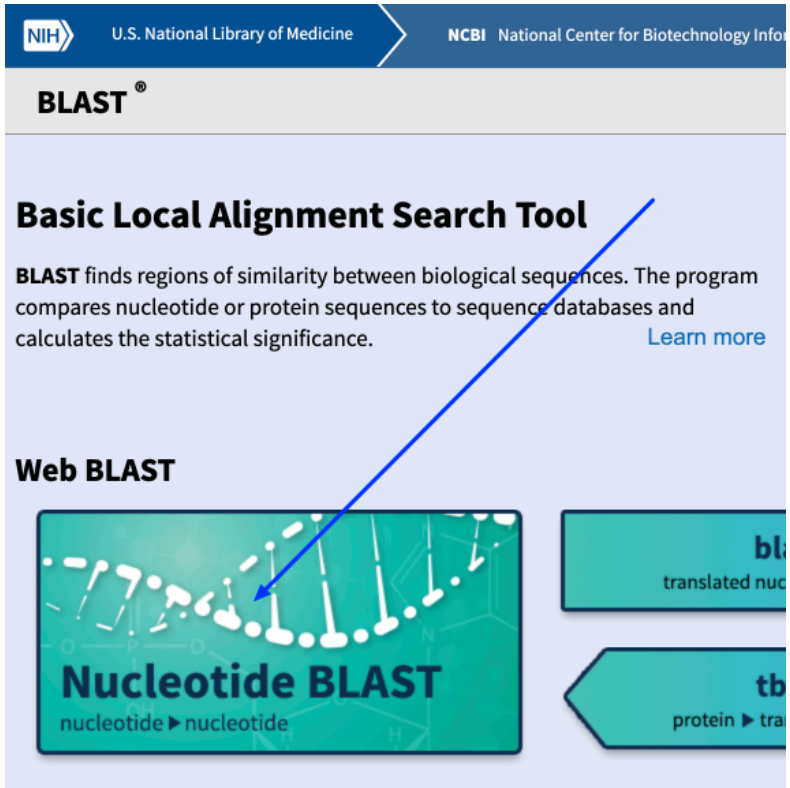
Start watching from 4:00 of 5:26 of the video. Harvard geneticist George Church shows Stephen Colbert a piece of paper with dot of synthetic DNA storing 20 million copies of Church's new book Regenesis. Church talks about converting the book's images and text to zeros and ones and then into the ATGC code of DNA.

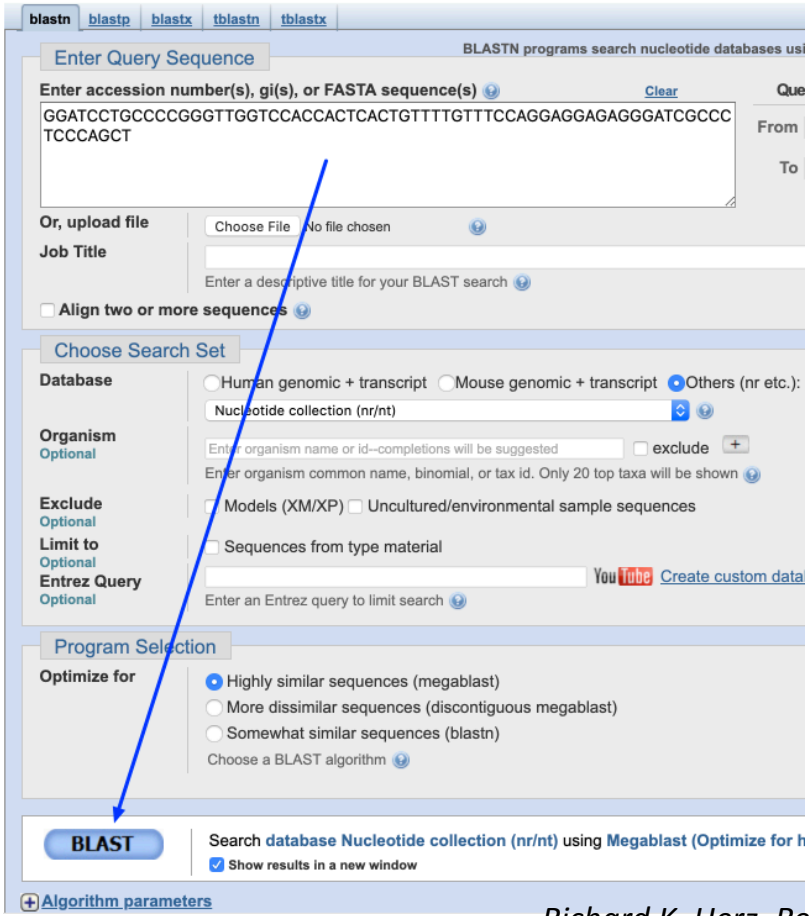
To search a DNA sequence for a match in organism genes, you can use the BLAST tool at the National Institutes of Health

https://blast.ncbi.nlm.nih.gov/Blast.cgi

For example, click the Nucleotide BLAST link on that page. then enter this sequence from a class assignment at the New York Medical College in the search field on the next page, then click the BLAST button and see what the search returns

GGATCCTGCCCCGGGTTGGTCCACCACTCACTGTTTTGTTTCCAGGAGGAGAGGGATCGCCCTCCCAGCT





Richard K. Herz, ReactorLab.net