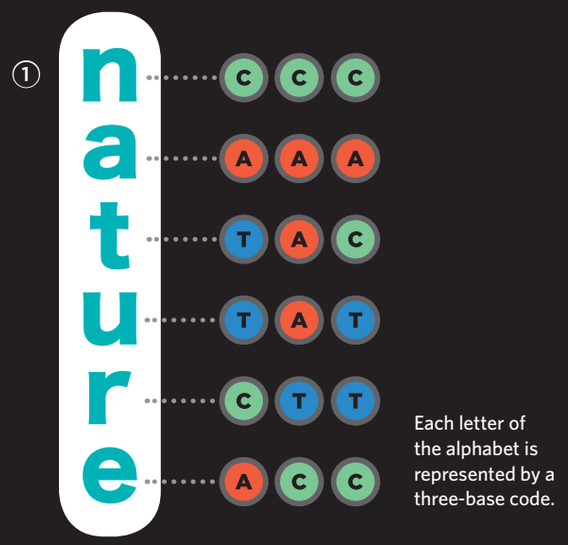
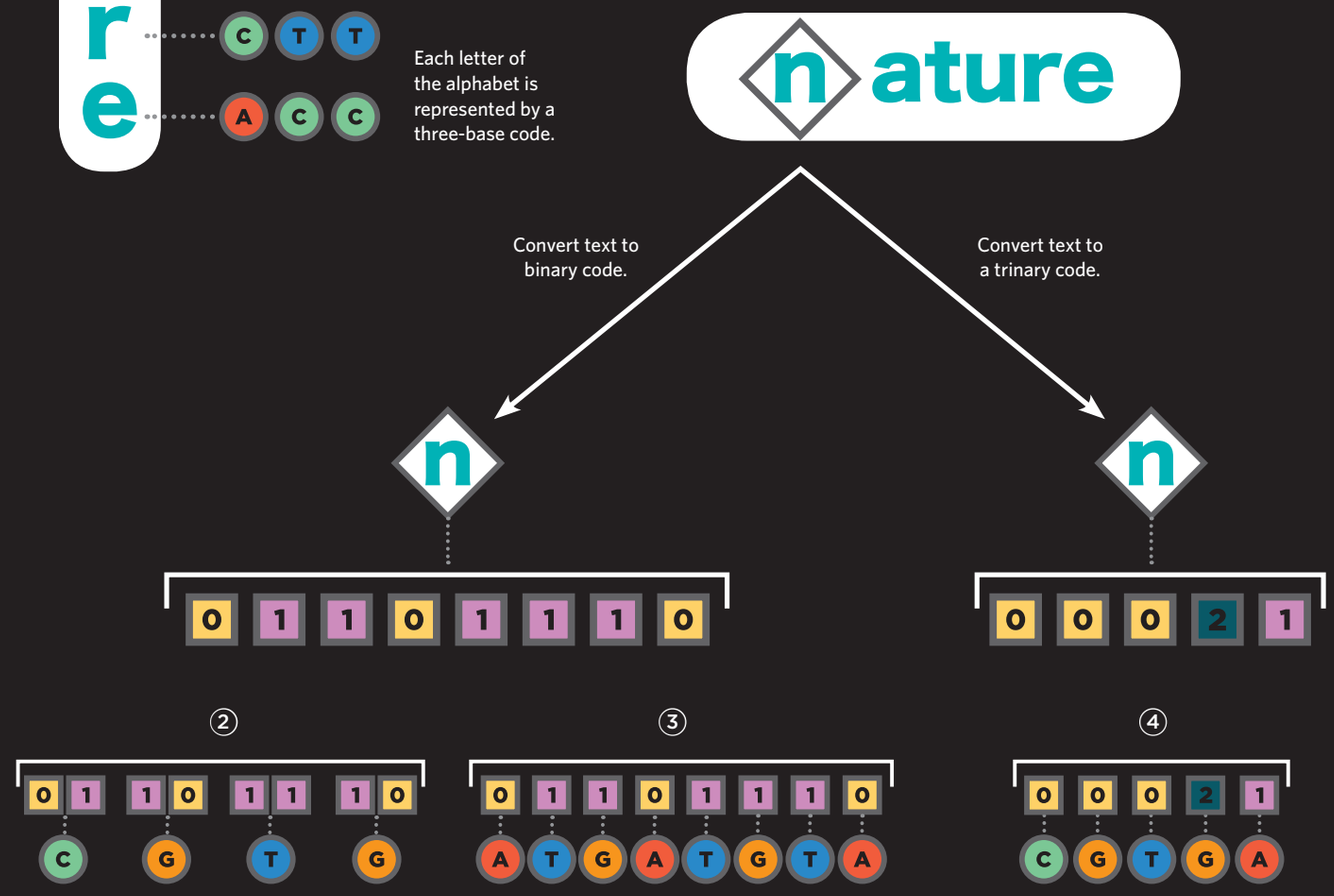
"THE GREATEST SINGLE ACHIEVEMENT OF nature TO DATE WAS SURELY THE INVENTION OF THE MOLECULE DNA."

—LEWIS THOMAS



Each letter of the alphabet is represented by a three-base code.



Each base encodes two binary digits.

Each binary digit can be encoded by one of two bases.

Each digit is encoded by a single base; which one depends on the base that immediately preceded it.

WRITING WITH DNA

There are many possible ways to encode information into DNA. If just encoding text, one way is to convert each letter of the alphabet into a three-letter code. Using three bases, such as A, C, and T, gives 27 combinations—enough for the English alphabet plus a space—with a code such as AAA = A, AAC = B, and so on (1). However, researchers often want to encode more than just text, so most current methods instead first translate data into binary code—the language of 1s and 0s used in electronic media. Using binary, the four bases of DNA could theoretically store up to two bits of information per nucleotide, with a code such as A = 00, C = 01, and so on (2).

In reality, though, biochemical features of nucleic acids make some combinations of bases more desirable than others. Particularly problematic are homopolymers—long strands of the same nucleotide—which are difficult to write and read using current methods. One way to avoid homopolymers is by allocating two bases to each binary digit; long runs of the same digit can then be encoded by alternating base pairs (3). A more efficient method is to convert text or other data into a code that employs three digits rather than two, and then write bases so that no base is used twice in a row—for example by encoding 0, 1, and 2 as C, G, and T after an A, but as G, T, and A after a C (4). Newer methods include more complex codes, as well as error-correcting techniques, to pack as much information as possible into DNA while maximizing the accuracy of information retrieval.

Sources for methods depicted: 1. Bancroft et al., 2001; 3. Church et al., 2012; 4. Goldman et al., 2013.

THE SCIENTIST STAFF