GOING DEEPER 7.1

DNA: The Basics

DNA, deoxyribonucleic acid, holds the genetic coding used for **protein** synthesis and its regulation, via a set of intermediate steps. It includes a linear array of **nucleotides**, which hold the code for the **amino acids** in proteins plus sites for regulation of genes for protein synthesis, initiation of reading of the code, and cessation of reading, among other functions. DNA is contained in structures known as **chromosomes**, which in animals and higher plants add up to the **genome**. The information imparted by the sequences used for amino acids and proteins is known as the **genetic code**. Transcription involves the reading of DNA and copying the code into the nucleic acid known as **RNA**, which is used for protein synthesis and other functions.

DNA occurs as an intertwined double helix of two DNA molecules. Each molecule contains a chain of **nucleotides**. The four nucleotides belong to two chemical classes:

Purines: adenine (A) or guanine (G)

Pyrimidines: cytosine (C) and thymine (T)

An important feature is that, within a DNA strand, a given nucleotide can be any of the four nucleotide types. But, at each location on the DNA, a given nucleotide in one DNA chain that is a purine will crosslink properly only to the other intertwined chain with a complementary pyrimidine, and vice versa. The stable crosslinks are AT and GC; this is known as **complementarity**, and the pairs are known as **base pairs**. The crosslinks are chemically hydrogen bonds, which allows the strands to be pulled apart fairly easily.

During protein synthesis, an initiation code on the DNA starts the attachment of complementary base pairs to one of the DNA strands with the aid of RNA polymerase, which binds "upstream" of the site where transcription is initiated, finally resulting in the construction of a **messenger RNA** molecule. This messenger RNA molecule is the template for protein synthesis. Each amino acid is coded by a specific sequence of three nucleotides (actually many amino acids have more than one code). The presence of a code of three nucleotides for an amino acid allows a ready **translation** of a DNA sequence into the amino acid sequence of a protein.

There is a lot more to know about the details of the process, but this introduction allows us to understand some of the technology used in this text. PCR, for example, involves the design of **primers** usually less than 50 nucleotides long. We need two primers, one near the beginning of the reading of the DNA that codes for a specific protein and one near the end. These primers are placed in a sample solution of messenger RNAs but will bind selectively only to those which have complementary nucleotide sequences to the primers. Then the reaction proceeds as mentioned in the text, and the primers combined with the thermal cycler technique result in the desired DNA sequence to be amplified. Once DNA is amplified, it then can be subjected to **DNA sequencing**. The DNA sequences may be slightly different, which reflects genetic variants within populations of a species, but they may be very different, which might reflect membership in different species. ■

to obtain sequences of the molecule 16S ribosomal RNA, which is ubiquitous in bacteria and higher organisms. One can use the same primers for all species because the primers correspond to parts of the gene that have not changed over the course of evolution. It is, therefore, very useful for understanding relationships by analyzing similarities of DNA sequences to other organisms. Using this approach, the investigators sampled waters in the Pacific of 100 m and 500 m depth. Previous studies at the surface had found more common bacteria, but the samples at depth produced DNA sequences of bacteria that could only be related to the Archaea, a very ancient group of bacteria that includes a number of species that live in extreme environments of high temperature (e.g., hot springs), acid environments such as the guts of cows, and petroleum.

In studies designed to learn about marine biodiversity, researchers often look for types of DNA sequence where the rate of evolution is high, so even closely related species will differ in sequence. Repetitive sequences known as **microsatellites** or sequences obtained by specialized DNA-cleaving enzymes, known as **amplified fragment length polymorphisms**, are often favored. But many other types of gene are also used in PCR. In **real-time PCR**, the *rate* of production of DNA product in the thermal cycler can be quantified. This allows more exact studies allowing identification of different DNA types or can even allow quantification of gene expression to study gene functioning by microbes performing a critical function such as nitrogen fixation.

Fluorescent Labeling

For rapid visualization, one technique involves the use of labeling probes with a fluorescent label and combining this with a natural sample. This allows direct cell counts in natural samples.[8]

[8] See DeLong and others, 1989, in Further Reading, Molecular Methods and Microorganismal Diversity.