
The Genome of Theobroma Cacao

Genetic selection is a new tool for developing varieties of cacao with improved yield, quality and disease resistance.

Siela Maximova and Mark Guiltinan

The Pennsylvania State University

A genome is a blueprint containing the biological information for building, operating, maintaining and reproducing a living organism. Hence, determining and understanding whole genome sequences will ultimately lead to a better understanding of how organisms function.

The genome, consisting of long strings of chemicals called DNA sequence, includes all the genes of a given organism, but also DNA that is not part of a gene, or noncoding DNA sequence. Each gene contains instructions for assembly of proteins, which consist of strands of amino acids that fold into an infinite array of shapes and sizes, each carrying out specific functions in every living cell. Nearly every cell in a living organism contains a complete copy of its genome and since each organism consists of trillions of individual cells each organism contains trillions of copies of its genome.

The DNA in the cells is packaged in chromosomes. Most living cells contain two complete sets of paired (homologous) chromosomes or two sets of all genes; each one of these sets is contributed from one of the two parents. Many organisms, especially plants,

have more than two paired sets of chromosomes and may contain three (watermelon), four (cotton) or even eight sets (sugarcane). Genome sequencing is the process of determining the sequence of consecutive DNA “letters” spanning all of the chromosomes of a cell from start to finish (the four chemical “letters” of DNA are abbreviated A, C, T and G). A computer representation of the genome sequence of cacao appears as 10 long strings of these four letters (one string for each chromosome) for a total of 430 million letters. The recent development of new technologies has made DNA sequencing dramatically easier and cheaper, and the number of complete genome sequences is growing rapidly. The study of the global properties of genomes of organisms is usually referred to as *genomics*, which is different from genetics. Genetics generally studies the properties of single genes or groups of genes.

A single genome sequence does not capture the genetic diversity of a species. For example, a complete genome sequence of an individual in principle could be determined from just half of the DNA information from one cell, the set of genes from one parent. ➤



Siela Maximova, PhD, is a senior research associate and associate professor at The Pennsylvania State University. She is a member of the Cocoa Genome Sequencing Consortium which recently published the complete sequence of the cacao genome.