

Comparison of segments of DNA

In the last few years a lot of progress has been made in the determination of the human DNA; think of the Human Genome project that tries to determine the function of different segments of chromosomes of the human DNA.

Suppose there are two segments of genes, and they have a segment of certain length in common. Then, do these segments of genes have the same origin? Segments of DNA are said to be related if these segments come from the same person or from two family members.

A segment of DNA consists of amino acids, and they can have four values (Adenine, Thymine, Guanine, Cytosine). A segment of DNA consists of a long sequence of these 4 letters. For more information, see the website <http://www.dnaftb.org/dnaftb/1/concept/index.html>

A method to compare segments of DNA is to look at the longest segment in common. If this segment is long, then one expects that the two segments of DNA are related. On the other hand, if the two segments of DNA are sufficiently long, but unrelated, it is also likely that there are segments in common (due to randomness). The question is: when is the longest segment in common too long?

Suppose we have two segments of DNA, both of the same length and this length is 10^3 . If we compare them, we find that the longest segment in common has length 8. Is this too long, or could we also expect this for two unrelated segments? What can we say if we also allow that the two segments may be shifted (relative to each other). If then the longest segment in common has length 10, is that coincidence or not?

In reality segments of DNA are of course much longer, say of length 10^6 . What is now a reasonable length for the largest segment in common (in case of shifting or not)?