Genomic landscapes

Jean R. Lobry

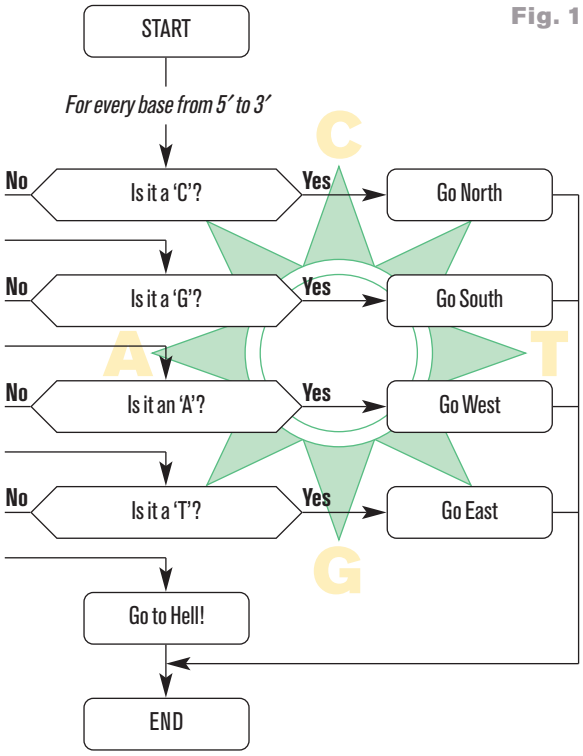


Fig. 1.

DNA walks

The recent availability of complete microbial genome sequences to the scientific community has opened the door to undoubtedly useful new bioinformatic approaches such as playing DNA music or drawing DNA walks. Since I hate noise, I will focus on DNA walks. To obtain a DNA walk all you need is a DNA sequence, a sheet of paper and a pencil to draw with using the instructions given in the chart on the left (Fig. 1).

Read the DNA sequence and walk into the plane according to the four directions defined by the four different bases. For a complete genome, this repetitive task could be extremely tedious but fortunately a computer is not easily bored. Now look at the output obtained with the 3,573,470 bases of the complete genome of *Synechocystis* PCC 6803, a cyanobacterium (Fig. 2).

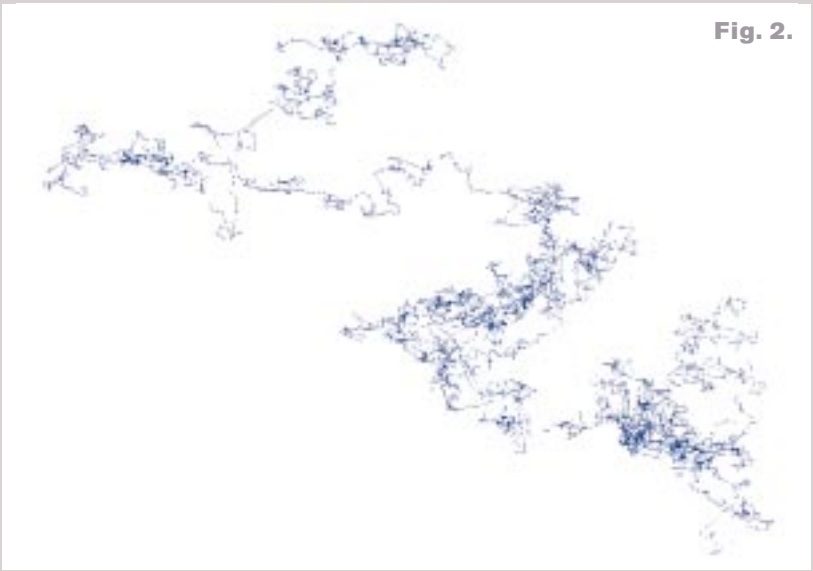


Fig. 2.

Here we have a simple graphical display corresponding to the complete genome sequence of this bacterium. What does it suggest to you? Yes, I know, a 3-year-old child could do that (see Fig. 3!!).

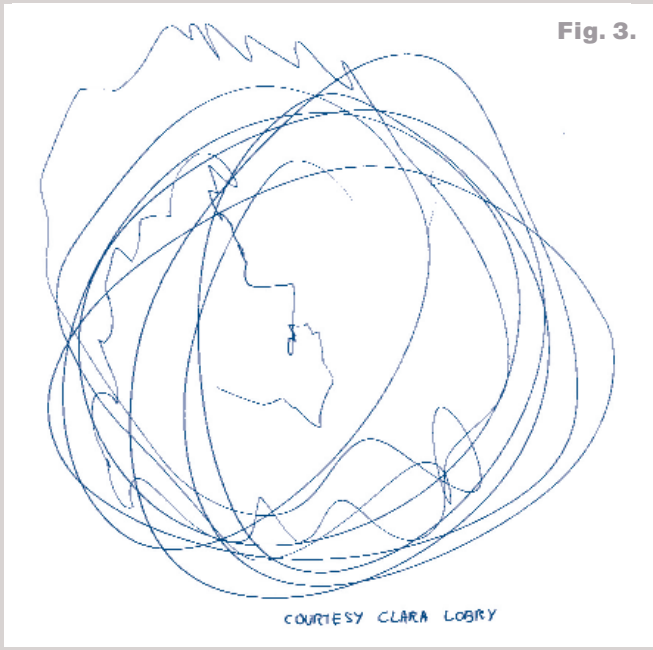


Fig. 3.

Hum. I agree that for *Synechocystis* it does look like nothing. But now look at *Borrelia burgdorferi*, a spirochaete (Fig. 4). The genome of this species is so incredible that the first time I analysed it I wondered if there were bugs in my program. Since then, the genome of *B. burgdorferi* has been analysed independently by other people and I am sure that this picture is not an artefact.

Fig. 4.

The pattern is much more regular than the previous random coils. Now if you plot, say, the '*y*' co-ordinate of the walk, corresponding to the 'C-G' axis, against the map position on the chromosome you obtain the graph shown in Fig. 5.

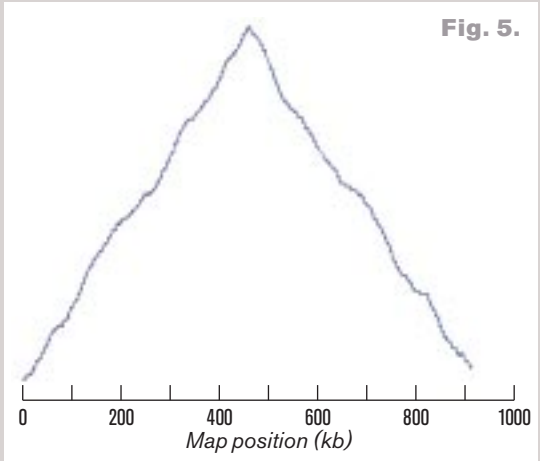


Fig. 5.

Now do you get it? This is a mountain – a *genome landscape*! The title of the article becomes comprehensible. But what is the meaning of this mountain? Simply that there are systematic statistical biases for the base composition along the chromosome. Travelling from 5′ to 3′ you start to climb the mountain, meaning that on average there is an excess of C over G in the sequence. Then when you start to go down the mountain it means exactly the opposite, i.e. there is an excess of G over C in the sequence. As you can see, the mountain slopes are approximately the same on both sides, meaning that these biases are symmetrical: their intensity is the same, the sign is just switched at the top of the mountain.

Climbing mountains

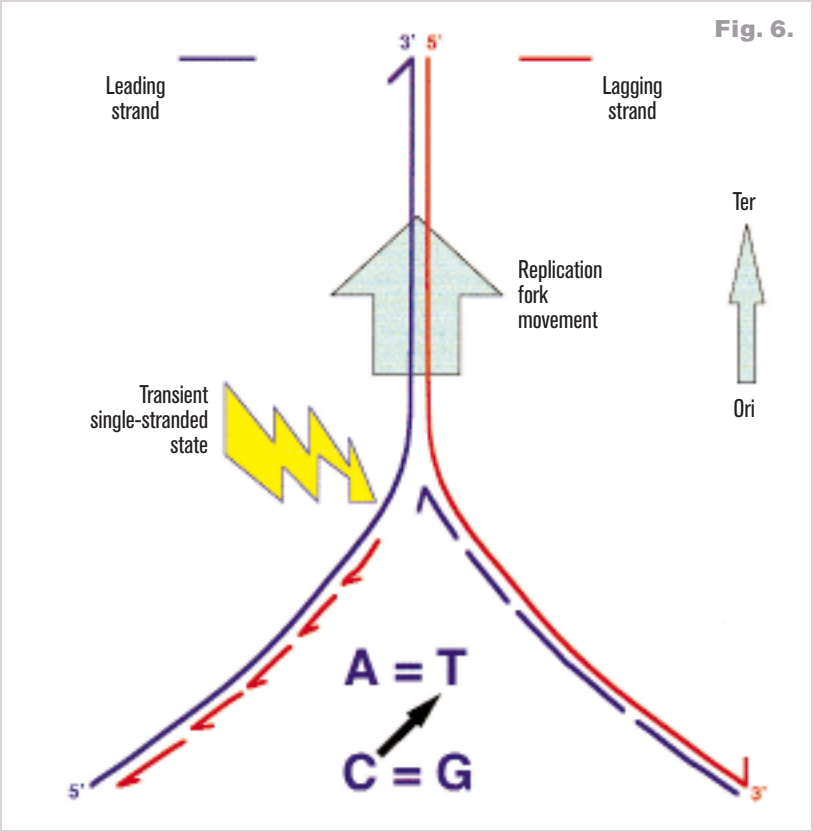
Do you know what is at the top of this mountain? It is the origin of replication of the *B. burgdorferi* chromosome that, incidentally, has just been experimentally mapped. As a consequence, when you are climbing the mountain you are reading the lagging strand for replication and when you are going down the mountain you are reading the leading strand for replication. What is usually found, but not always if you remember *Synechocystis*, is that the leading strand is enriched in keto (G or T) bases and that the lagging strand is enriched in amino (A or C) bases. Interestingly, these biases have been reported for both eubacteria and archaea, mitochondria, chloroplasts, viruses and plasmids, but not for eukaryotes up to now. In *B. burgdorferi* the biases are so important that they affect the amino acid content of proteins and you can guess if a protein was encoded on the leading or the lagging strand solely from its amino acid content.

The term 'chirochore' was coined to describe fragments of the genome corresponding to a mountainside, that is a DNA fragment more or less homogeneous for the base composition biases. This is a purely descriptive term without reference to any mechanism, reminiscent of 'isochore' for the description of DNA fragments with a homogeneous G+C content in some vertebrate chromosomes. On the other hand, the term 'replichore' was introduced to designate the two oppositely replicated halves of the chromosome between the origin and the terminus in bacteria. The good thing is that chirochore and replichore boundaries are the same in bacteria: the origins of replication are found at the top of the mountains while the termini are found at the bottom of the lowlands. This strongly suggests that these kinds of genome landscapes have something to do with replication.

Genomic tectonics

A simple model to explain the universality of the phenomenon is based on the spontaneous deamination of cytosines that induce C→T mutations. The rate of this deamination is highly increased in single-stranded DNA, probably because of greater accessibility to the solvent in this state than in the double-stranded state. During replication the lagging strand is continuously protected by the newly synthesized leading strand, but the leading strand has to maintain a transient single-stranded state while waiting for the next Okazaki fragment to be long enough to restore the double-stranded state (see Fig. 6).

Fig. 6.



This fundamental asymmetry of replication may explain the universality of the observed systematic biases in base composition; these are at least compatible with the hypothesis. The protection against cytosine deamination may differ between species and explain the variability in intensity of biases.

The cytosine deamination theory is nice but it is just a theory. The fundamental limit of bioinformatics is that without experimental data you can discuss *in silico* results endlessly and fruitlessly. I would be very curious to know the minimum inhibitory concentration (MIC) of chemicals such as bisulphite, which catalyses the deamination of cytosine, for bacteria in which compositional biases are important (for instance *B. burgdorferi*, *Treponema pallidum*, *Neisseria meningitidis*). These compounds may not only inhibit replication but also transcription because during transcription a transient single-stranded DNA state is necessary too. It would also be interesting to determine toxicological information for eukaryotic cells.

Dr Jean R. Lobry is Maître de Conférences Universitaire (Associate Professor), Département Biométrie et Biologie Evolutive CNRS UMR 5558, Université Claude Bernard de Lyon, 43 Boulevard du 11 Novembre 1918, F-69622 Villeurbanne Cédex, France
Tel. +33 472 43 12 87; Fax +33 478 89 27 19
e-mail lobry@biomserv.univ-lyon1.fr

Further reading

A good introduction to DNA music and courses on bioinformatics and DNA correlations, among other subjects, is Wentian Li's home page at **http://linkage.rockefeller.edu/wli/**

More about DNA walks, their transformations and analysis, is available on the smORFland home page at **http://smorfland.microb.uni.wroc.pl/**

The story of the *Borrelia burgdorferi* replication origin can be found in:
Picardeau, M., Lobry, J.R. & Hinnebusch, B.J. (1999). Physical mapping of an origin of bidirectional replication at the centre of the *Borrelia burgdorferi* linear chromosome. *Mol Microbiol* 32, 437–445.

The underlying mechanisms that may explain genomic landscapes were recently reviewed in:
Frank, A. C. & Lobry, J. R. (1999). Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms. *Gene* 238, 65–77.