History of Topology

Semester I, 2009-10

Graham Ellis
NUI Galway, Ireland

History of Topology Outline

- Lecture 1:
  What is topology?
  Euler circuits in graphs
- Lecture 2:
  Topological invariants:
  Euler-Poincaré characteristic
- Lecture 3:
  One recent application of topology in biology

Some Elementary Biology

- Charles Darwin & others aware of inheritance. Believed in *blended inheritance*: offspring are an average of their parents.

Some Elementary Biology

- Charles Darwin & others aware of inheritance. Believed in *blended inheritance*: offspring are an average of their parents.
- Gregor Mendel studied inheritance in pea plants (1860s). Noticed that biological variations only inherited as discrete traits.
  e.g. offspring of a blue eyed mother and brown eyed father has 3/4 chance of having brown eyes.

Some Elementary Biology

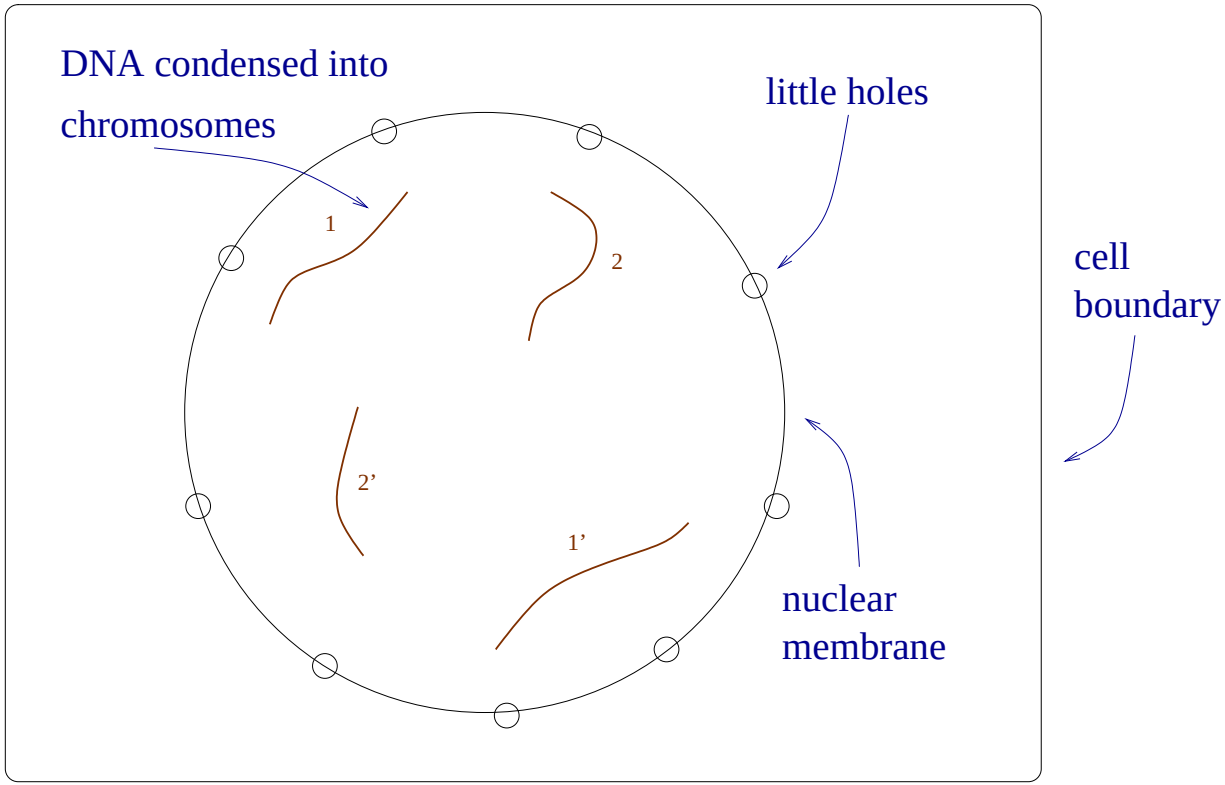
- Charles Darwin & others aware of inheritance. Believed in *blended inheritance*: offspring are an average of their parents.
- Gregor Mendel studied inheritance in pea plants (1860s). Noticed that biological variations only inherited as discrete traits.
  e.g. offspring of a blue eyed mother and brown eyed father has 3/4 chance of having brown eyes.
- In 1889 Hugo de Vries gave Mendel's inheritance mechanism the name *gene*. Had little understanding of the physical nature of a gene.

- In 1910 Thomas Hunt Morgan showed that genes resided in *chromosomes* in cells.

- In 1910 Thomas Hunt Morgan showed that genes resided in *chromosomes* in cells.
- In 1944 it was shown that Deoxyribonucleic acid, or DNA, contains the genes of an organism. DNA is a long polymer made from repeating units called *nucleotides*.

- In 1910 Thomas Hunt Morgan showed that genes resided in *chromosomes* in cells.
- In 1944 it was shown that Deoxyribonucleic acid, or DNA, contains the genes of an organism. DNA is a long polymer made from repeating units called *nucleotides*.
- In 1953 Crick and Watson demonstrated the molecular structure of DNA.

A typical (animal) cell

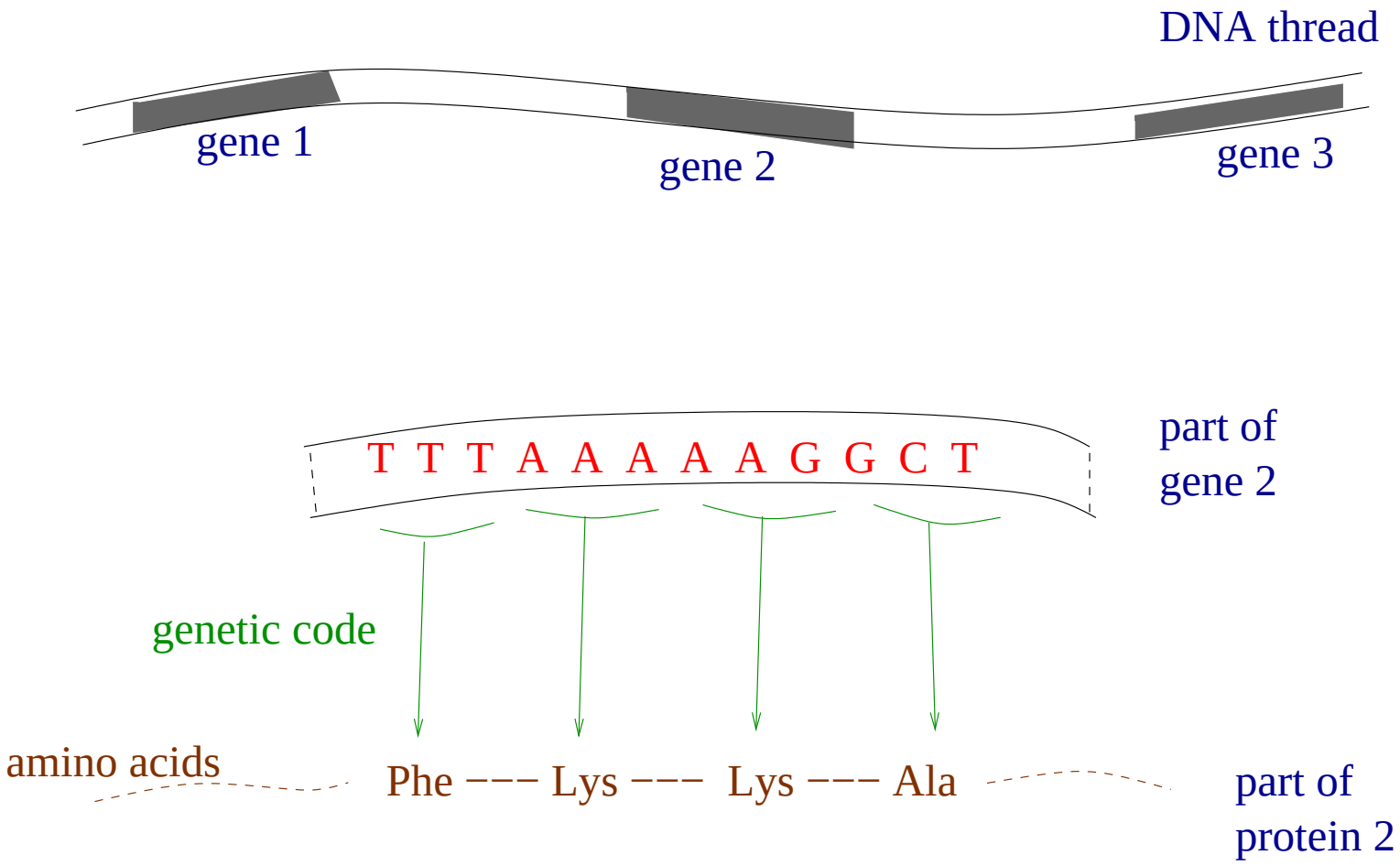


Chromosomes contain a mixture of protein and DNA. They often come in pairs.

Humans cells contain 46 chromosomes in 22 homologous pairs plus the non-homologous X and Y chromosomes that determine sex.

What is DNA?

DNA is a string of four different nucleotides: **A**denine, **G**uanine, **C**ytosine and **T**hymine.



What is DNA?

We can view DNA as a very, very long string or word in the four-letter alphabet A, T, G and C.

What is DNA?

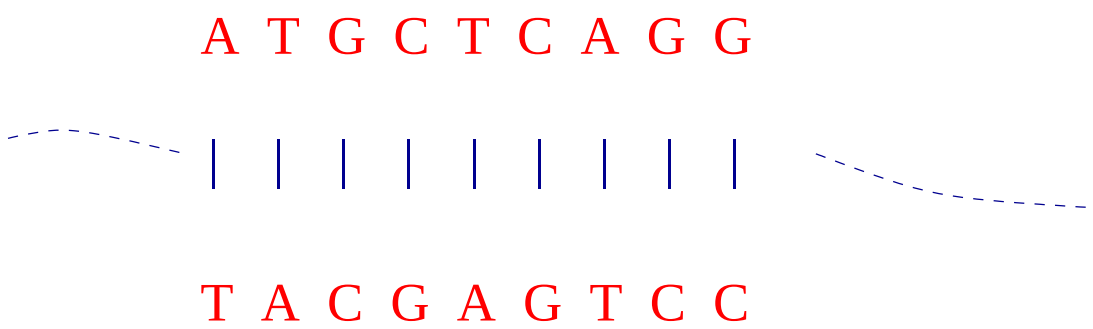
We can view DNA as a very, very long string or word in the four-letter alphabet A, T, G and C.

DNA is usually double stranded with one strand being the *Crick-Watson complement* of the other. (T pairs with A, C with G.)

What is DNA?

We can view DNA as a very, very long string or word in the four-letter alphabet A, T, G and C.

DNA is usually double stranded with one strand being the *Crick-Watson complement* of the other. (T pairs with A, C with G.)



Question

Given a long strand of physical DNA, how can a biologist determine the sequence of letters

$$\ldots ACCGATTGCAGTAGC \ldots$$

representing the DNA?

Answer

In 1994 a California based biotechnology company, Affymetrix, built the first commercial DNA array for tackling the problem.

Answer

In 1994 a California based biotechnology company, Affymetrix, built the first commercial DNA array for tackling the problem.

Such an array contains all possible single stranded DNA strings of length L (for example L=10). The 4^L strings are called probes.

Answer

In 1994 a California based biotechnology company, Affymetrix, built the first commercial DNA array for tackling the problem.

Such an array contains all possible single stranded DNA strings of length L (for example L=10). The 4^L strings are called probes.

Given a long single stranded target thread, the target will bind (or hybridize) to precisely those probes which are the Crick-Watson complement of fragments of the target thread.

Hybridization

For example, the probe

$$ACCGTGGA$$

will hybridize with the target

$$CCCTGGCACCTA$$

Hybridization

For example, the probe

$$ACCGTGGA$$

will hybridize with the target

$$CCCTGGCACCTA$$

since the probe's Crick-Watson complement is

$$TGGCACCT.$$

The upshot

Using arrays of probes of a given length L, biologists can determine all the letter sequences of length L contained in a target DNA thread.

A mathematical question

How can we determine a DNA letter sequence from a knowledge of all L-letter fragments of the sequence?

Example of the question

Suppose that a DNA array analysis of a piece of single-standed target DNA, using probes of length $L = 3$, shows that the length 3 fragments of the target DNA are precisely:

$$F = \{ATG, TGG, TGC, GTG, GGC, GCA, GCG, CGT\}$$

Example of the question

Suppose that a DNA array analysis of a piece of single-standed target DNA, using probes of length $L = 3$, shows that the length 3 fragments of the target DNA are precisely:

$$F = \{ATG, TGG, TGC, GTG, GGC, GCA, GCG, CGT\}$$

What can we say about the target DNA?

Some topology

To answer this question let's build the graph whose vertices are the $4^2 = 16$ pairs $AA, AC, AG, AT, CA, \ldots GT, TT$.

Some topology

To answer this question let's build the graph whose vertices are the $4^2 = 16$ pairs $AA, AC, AG, AT, CA, \ldots GT, TT$.

We'll add an edge between vertex XY and vertex YZ if the triple XYZ is in the collection F of fragments.

Some topology

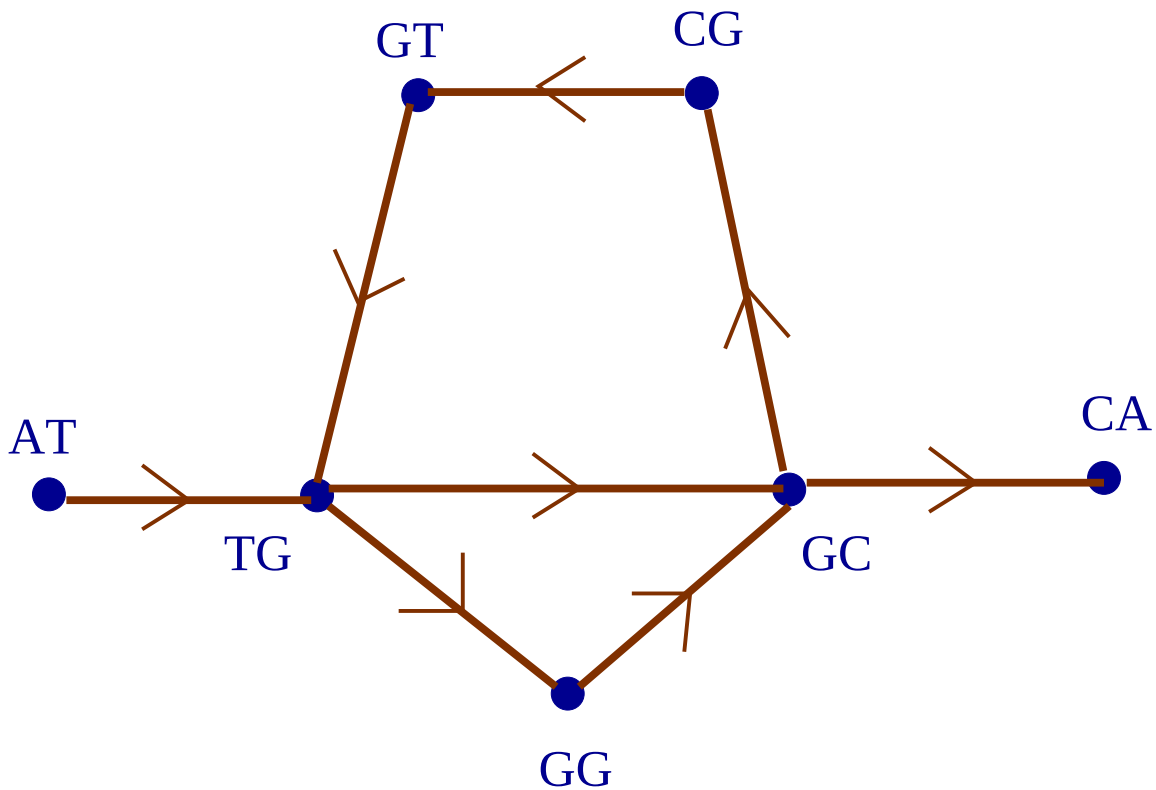
To answer this question let's build the graph whose vertices are the $4^2 = 16$ pairs $AA, AC, AG, AT, CA, \ldots GT, TT$.

We'll add an edge between vertex XY and vertex YZ if the triple XYZ is in the collection F of fragments.

Moreover, we'll place an arrow on the edge going from XY towards YZ.

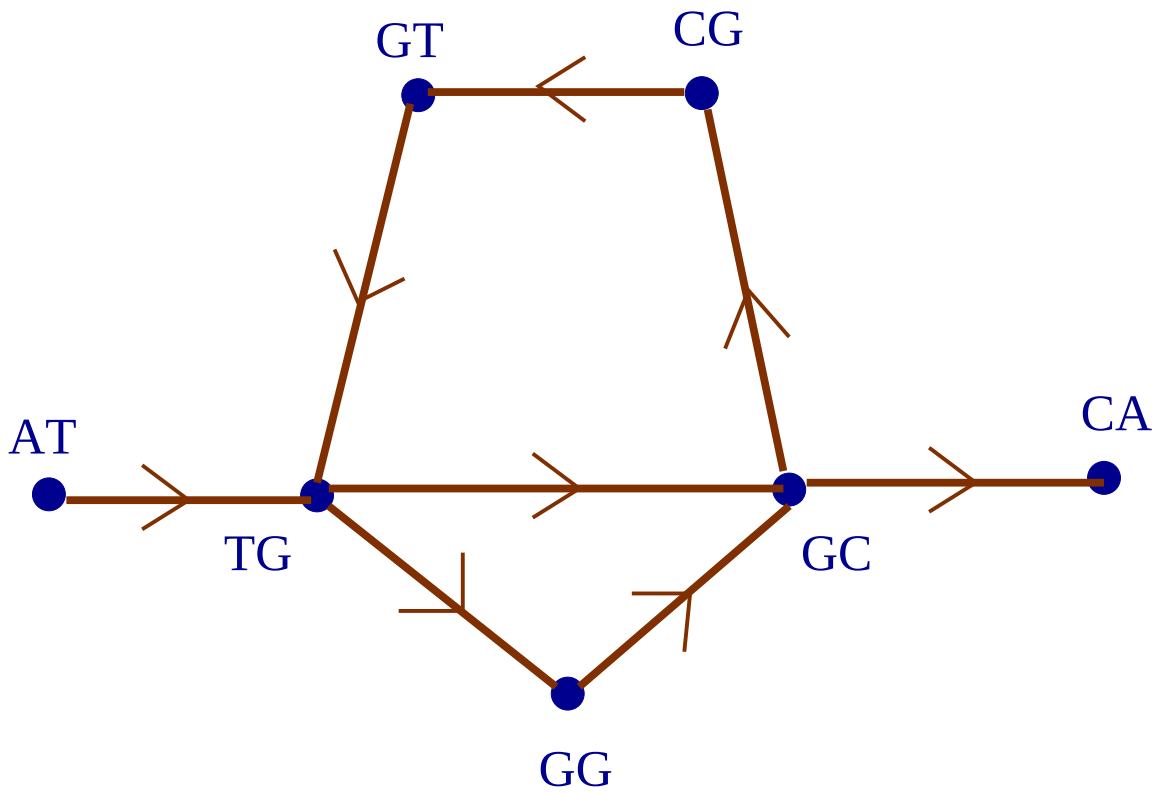
Some topology

For $F = \{ATG, TGG, TGC, GTG, GGC, GCA, GCG, CGT\}$ the graph's eight edges are as follows.



Some topology

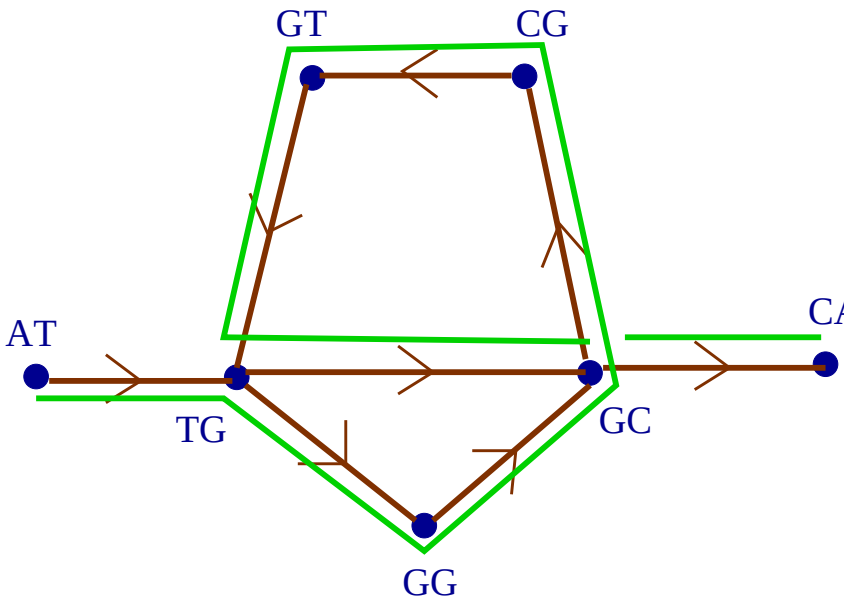
For $F = \{ATG, TGG, TGC, GTG, GGC, GCA, GCG, CGT\}$ the graph's eight edges are as follows.



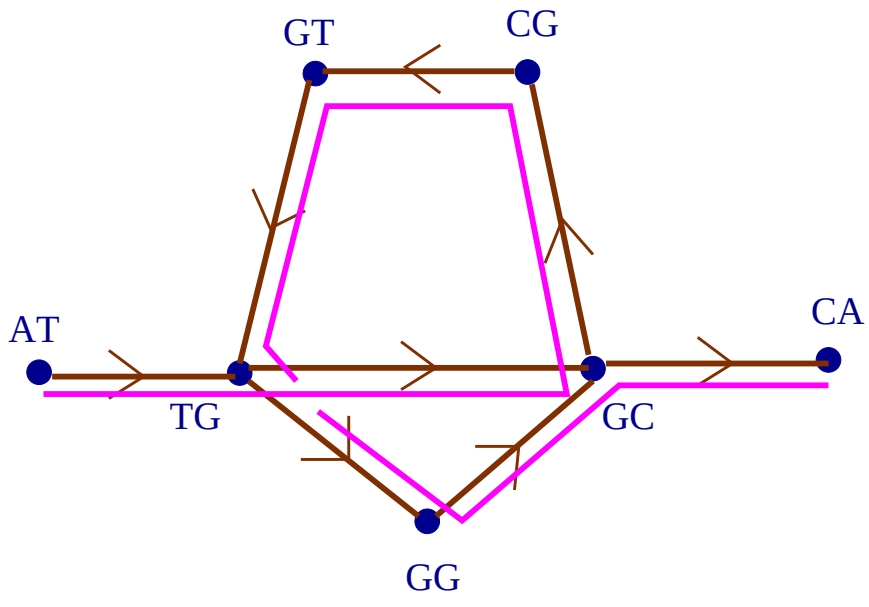
The target DNA corresponds to an Euler path in this graph (where the path direction is given by the arrows in the graph).

Some topology

There are just two possible Euler paths.



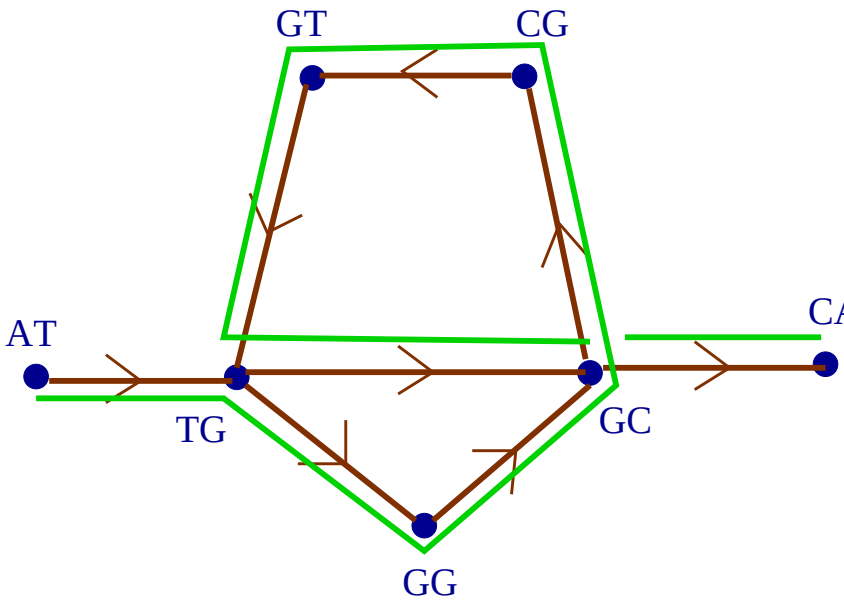
ATGGCGTGCA



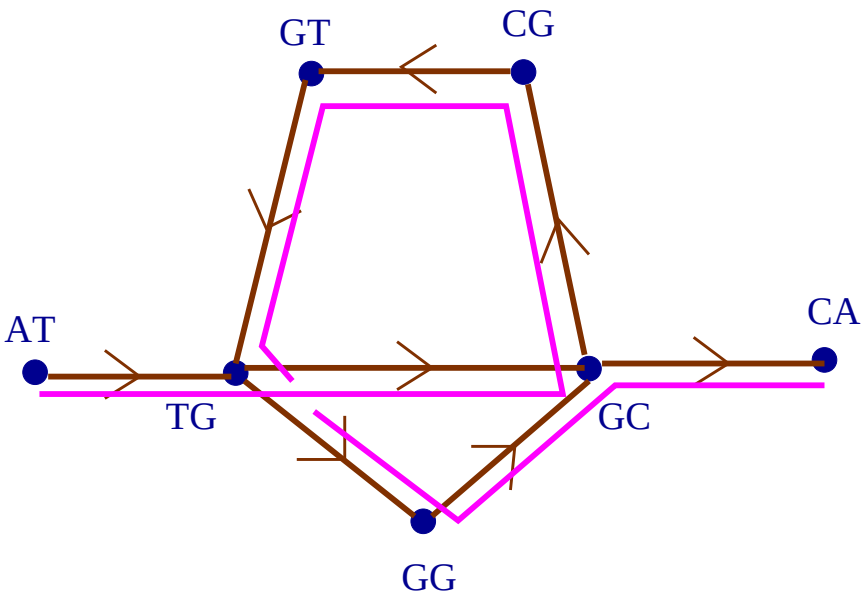
ATGCGTGGCA

Some topology

There are just two possible Euler paths.



ATGGCGTGCA



ATGCGTGGCA

So the possible sequences for the target DNA are: ATGGCGTGCA and ATGCGTGGCA.

Some topology

For sequencing DNA using arrays of DNA probes, biologists are faced with the following.

Question: Given a "fragment graph", does it contain any Euler path? (If not then maybe try a different length of probe.)

Some topology

For sequencing DNA using arrays of DNA probes, biologists are faced with the following.

Question: Given a "fragment graph", does it contain any Euler path? (If not then maybe try a different length of probe.)

Answer: Euler's theorem says that a graph has an Euler path if, and only if, exactly zero or exactly two vertices have odd degree.

Some topology

For sequencing DNA using arrays of DNA probes, biologists are faced with the following.

Question: Given a "fragment graph", does it contain any Euler path? (If not then maybe try a different length of probe.)

Answer: Euler's theorem says that a graph has an Euler path if, and only if, exactly zero or exactly two vertices have odd degree.

Question: Given a "fragment graph" in which just zero or two vertices have odd degree, how does one produce an Euler path?

Some topology

For sequencing DNA using arrays of DNA probes, biologists are faced with the following.

Question: Given a "fragment graph", does it contain any Euler path? (If not then maybe try a different length of probe.)

Answer: Euler's theorem says that a graph has an Euler path if, and only if, exactly zero or exactly two vertices have odd degree.

Question: Given a "fragment graph" in which just zero or two vertices have odd degree, how does one produce an Euler path?

Answer: Euler's proof of his theorem explains how to produce an Euler path.

Reference

P.A. Prevzner, *L*-tuple DNA sequencing: computer analysis, *Journal of Biomolecular Structure and Dynamics* 7 (1989), 65-73.