AT content in Bacterial Genomes

Dave Ussery
Comparative Microbial Genomics Workshop
Pathumthani, Thailand
8 March, 2005
Outline

Local AT content & Promoters

Global AT content in chromosomes
LECTURE 9 a.m.
The DNA sequence contains information.

But what kind of information?
pVir Atlas

G Content

A Content

T Content

C Content

Annotations:

AT Skew

GC Skew

Percent AT

Resolution: 15

http://www.cbs.dtu.dk/
Center for Biological Sequence Analysis

BASE ATLAS
pVir genome Atlas

Intrinsic Curvature

Stacking Energy

Position Preference

Annotations:

CDS +
CDS -
rRNA
tRNA

Global Direct Repeats

Global Inverted Repeats

GC Skew

Percent AT

Resolution: 15

pVir 37,468 bp

http://www.cbs.dtu.dk/
Center for Biological Sequence Analysis

Resolution: 15
Camylobacter jejuni, plasmid pVir

AF226280  37,468 bp

A) Intrinsic Curvature

B) Stacking Energy

C) Position Preference

D) Annotations:

E) Global Direct Repeats

F) Global Inverted Repeats

G) GC Skew

H) Percent AT

Resolution: 2

Range: 31000 .. 36000

Center for Biological Sequence Analysis
http://www.cbs.dtu.dk/
AF226280_copolymer #7

G(x) (kcal/mole)

nucleotide position

From Craig Benham, U.C. San Diego, 6 Sept. 2003
Promoter cartoon

DNA curvature, flexibility important here

DNA melting important here

rigid

-35

-10

β / β'
Promoter structure

<table>
<thead>
<tr>
<th>Organism</th>
<th>Region</th>
<th>Curvature</th>
<th>Stacking energy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mycobacterium tuberculosis</td>
<td>CDS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34% AT</td>
<td>← →</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>→ ←</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M. thermoautotrophicum</td>
<td>CDS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50% AT</td>
<td>← →</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>→ ←</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Synechocystis sp.</td>
<td>CDS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>52% AT</td>
<td>← →</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>→ ←</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacillus subtilis</td>
<td>CDS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56% AT</td>
<td>← →</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>→ ←</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aquifex aeolicus</td>
<td>CDS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57% AT</td>
<td>← →</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>→ ←</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlamydia trachomatis</td>
<td>CDS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>59% AT</td>
<td>← →</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>→ ←</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helicobacter pylori</td>
<td>CDS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>61% AT</td>
<td>← →</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>→ ←</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Campylobacter jejuni</td>
<td>CDS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>69% AT</td>
<td>← →</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>→ ←</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rickettsia prowazekii</td>
<td>CDS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>71% AT</td>
<td>← →</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>→ ←</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AT content upstream and downstream of translation starts in 158 bacterial chromosomes

What is responsible for this strange pattern of AT content within the beginning of coding regions?
Codon Usage

Firmicutes  43 segments
Nucleotide Bias in triplet position
Firmicutes    43 segments

Average Base Frequency

1:A
3:U
1:C
3:G
1:G
3:C
1:G
3:A
2:G
2:C
2:U
1:U
0.39
0.34
0.30
0.26
0.21
0.17
0.13
Sequence logo for Shannon information of aligned codons for *Mycoplasma mycoides* strain SC
Get around in ENGLISH

Lesson Twenty Five
How to be Polite

1. Wrong

HELP!

2. Right

Excuse me, Sir. I'm terribly sorry to bother you, but I wonder if you would mind helping me a moment, as long as it's no trouble, of course.
Figure 2. Curvature profiles. The profiles have been smoothed using a running average with a window size of 250 kilo-base-pairs (kbp), and values have furthermore been normalized and are in units of standard deviations from genomic average. Upper panel, curvature profile of the entire E. coli chromosome. The location of the origin of replication (oriC) and the terminus region (with terminator sites TerA, TerB, TerC, TerD, TerE, TerF, and TerG) is indicated. Note the broad curvature peak in the terminus region, centered on TerC (which is believed to be the most frequently used terminator site). Lower panel, curvature profile of the entire B. subtilis chromosome. The origin of replication is at position zero kbp. The location of the terminus region (with terminator sites TerI-TerVII) is indicated.
Average AT content of different regions within 158 bacterial chromosomes
Summary

1. Upstream regions are more AT rich, curved, rigid, and easily melted.

2. On a chromosomal scale, replication terminus regions tend to be more AT rich, origins less AT rich.