

Transcription in bacterial chromosomes

Characteristic lengths in bacterial genomes

Ivan Junier



Team: **G**enomics and **E**volution of **M**icroorganisms (GEM)

Transcription in bacterial chromosomes

Characteristic lengths in bacterial genomes

- How to identify **genomic motifs** associated to transcription regulation (and why)?
- Does some **universality** exist in the organisation of bacterial genomes?
- To which extent motifs are **compatible to other structural properties**?

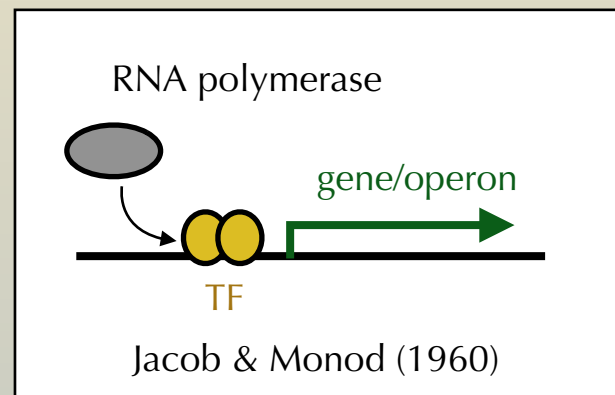


Olivier Rivoire
Collège de France, Paris

Regulation of transcriptional regulation in bacteria

Why bothering about genome organisation?

- How to identify **genomic motifs** associated to transcription regulation (and why)?



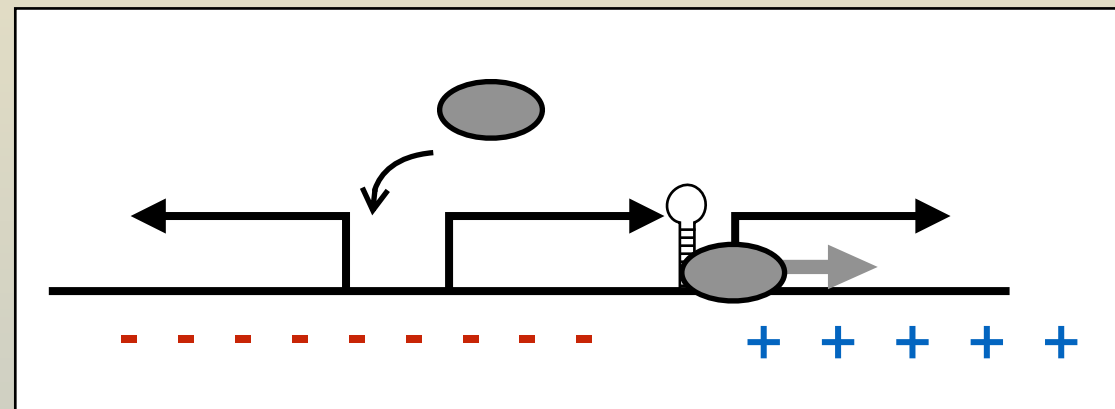
- *Escherichia coli*: < 40% genes regulated by TFs
 - 25% with strong evidence (database: regulonDB)
- **rapidly evolving** and **heterogeneous** regulatory factors
 - Browning & Busby, Nat. Rev. Micro., 2015

[...] **transcription regulation is a 'luxury'** for a bacterium, meaning that regulation **may not be essential for survival**.

By contrast, if a bacterium cannot 'handle' its DNA, it will not be viable, and handling here means **organising** and **compacting** it, as well as **replicating** and **expressing it in a reasonably reliable way**.

Visweswariah & Busby, Trends Microb., 2015

Expressing in a reliable way=
Basal coordination of transcription is due
to a mechanism of **facilitated co-transcription**



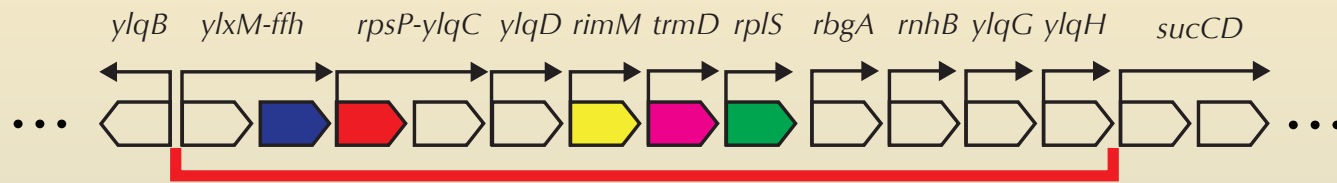
- How to identify **genomic motifs** associated to transcription regulation (and why)?

Junier & Rivoire, PLOS ONE, 2016

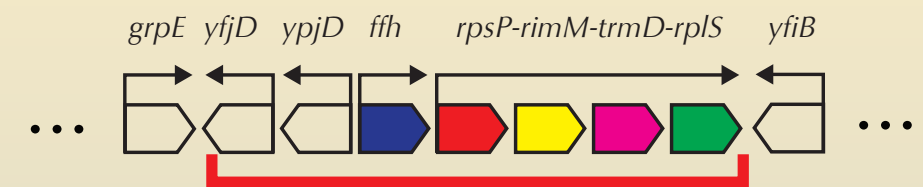
Junier, Unal, Yus, Lloréns-Rico, Serrano, Cell Systems, 2016

Synteny segments: array of genes with conserved proximity

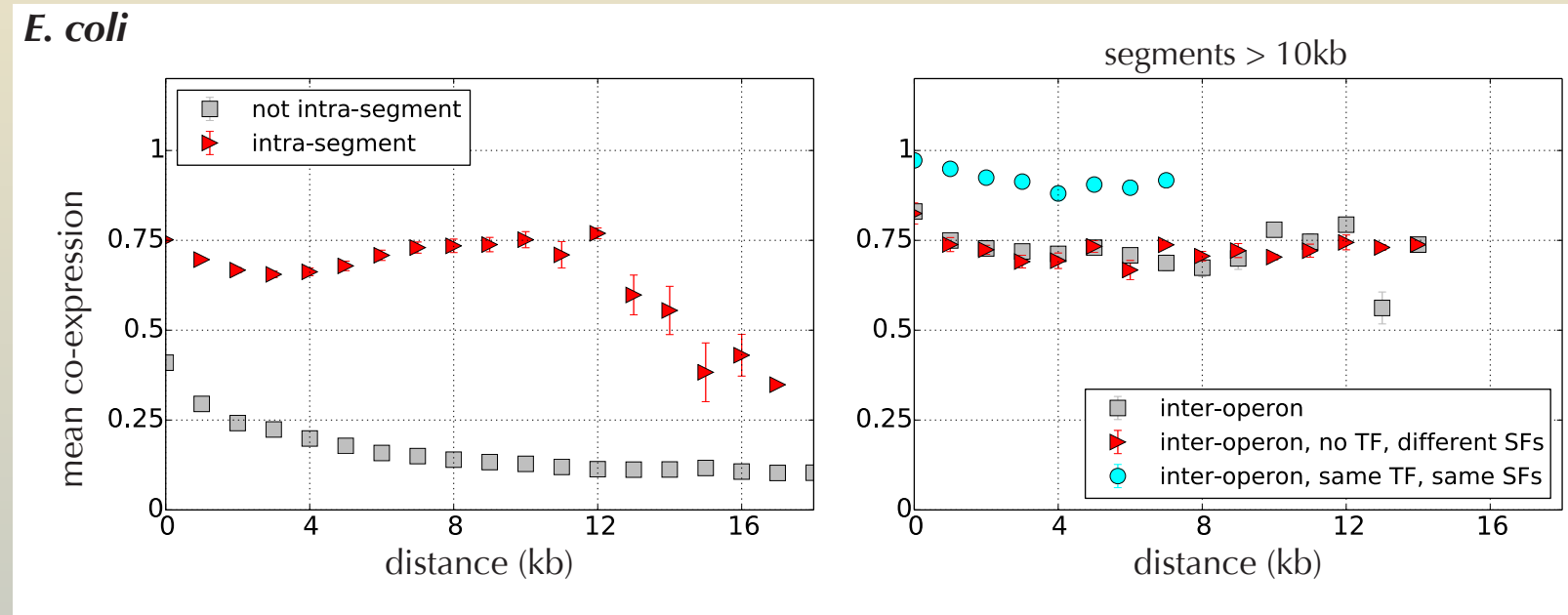
B. subtilis (661 segments)



E. coli (740 segments)

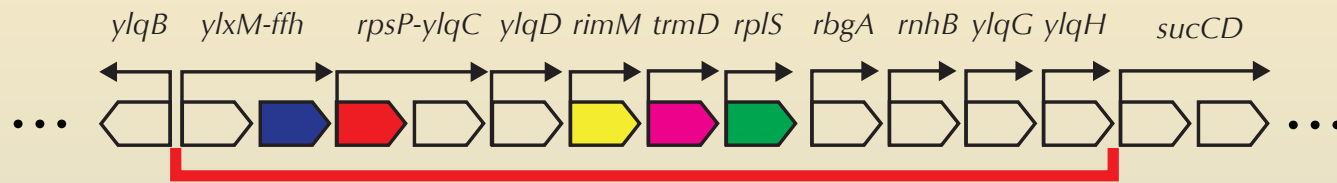


Expression data: M^{3D}
 Microarrays in > 400 conditions
 Faith et al., NAR, 2008

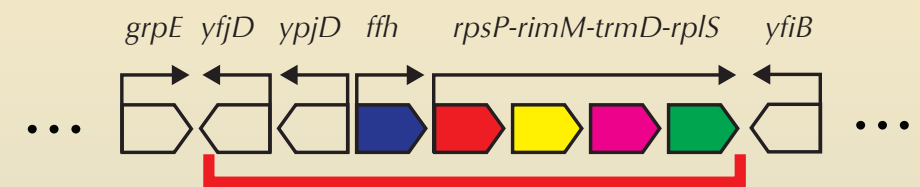


Synteny segments: array of genes with conserved proximity

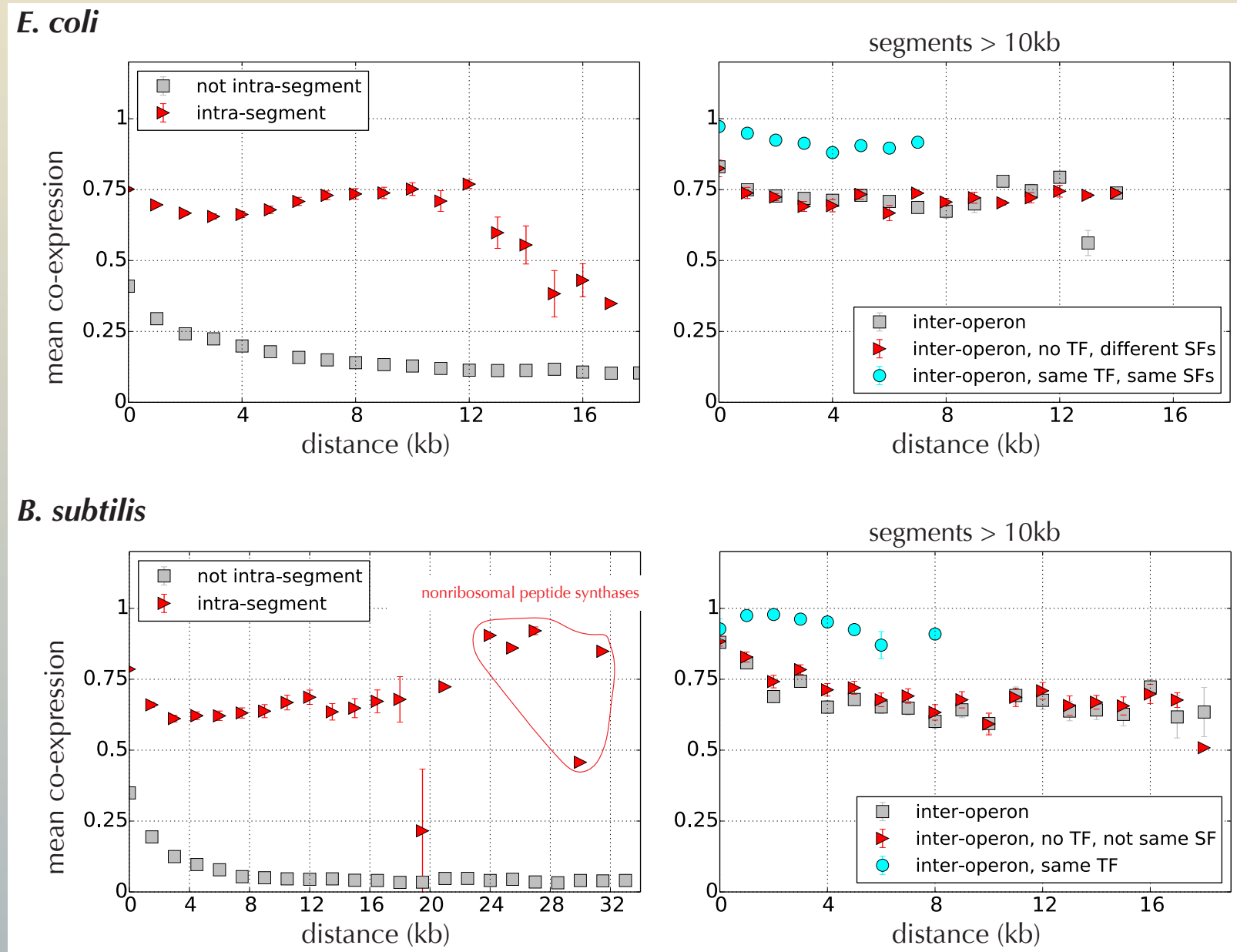
B. subtilis (661 segments)



E. coli (740 segments)



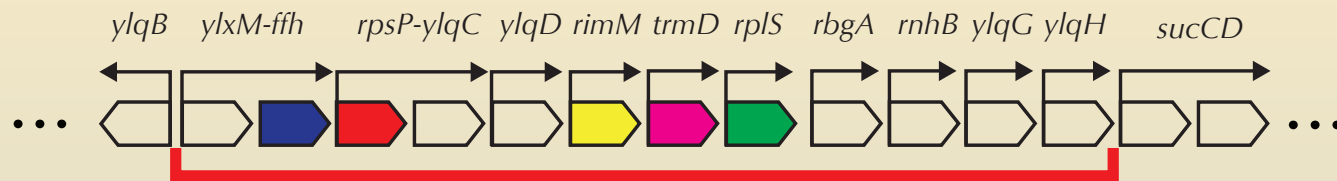
Expression data: M^{3D}
Microarrays in > 400 conditions
Faith et al., NAR, 2008



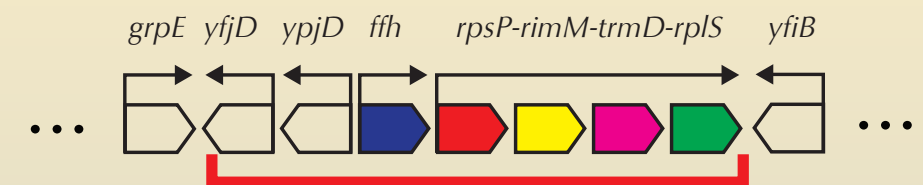
Expression data: BaSysBio
Microarrays in > 400 conditions
Nicolas et al., Science, 2012

Synteny segments: fundamental units of facilitated co-transcription

B. subtilis (661 segments)

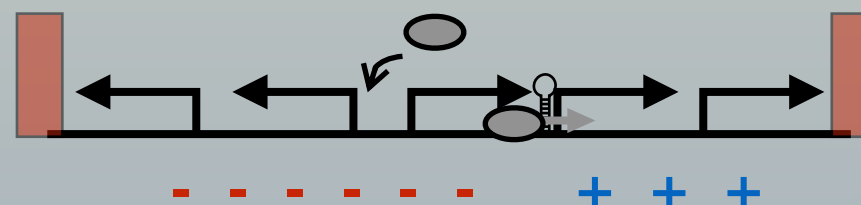


E. coli (740 segments)



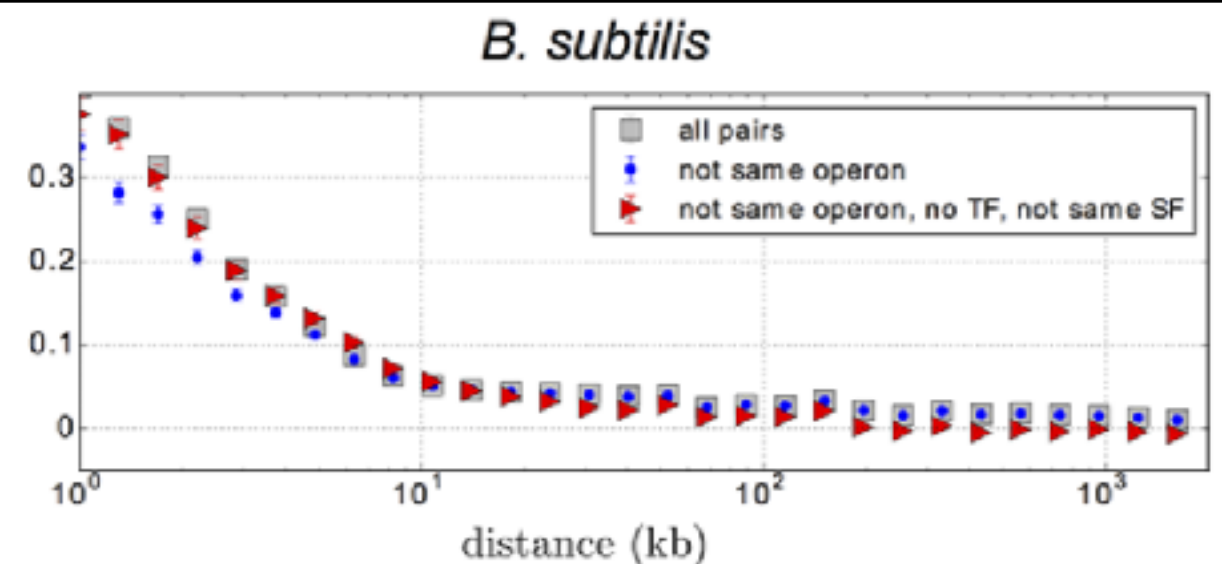
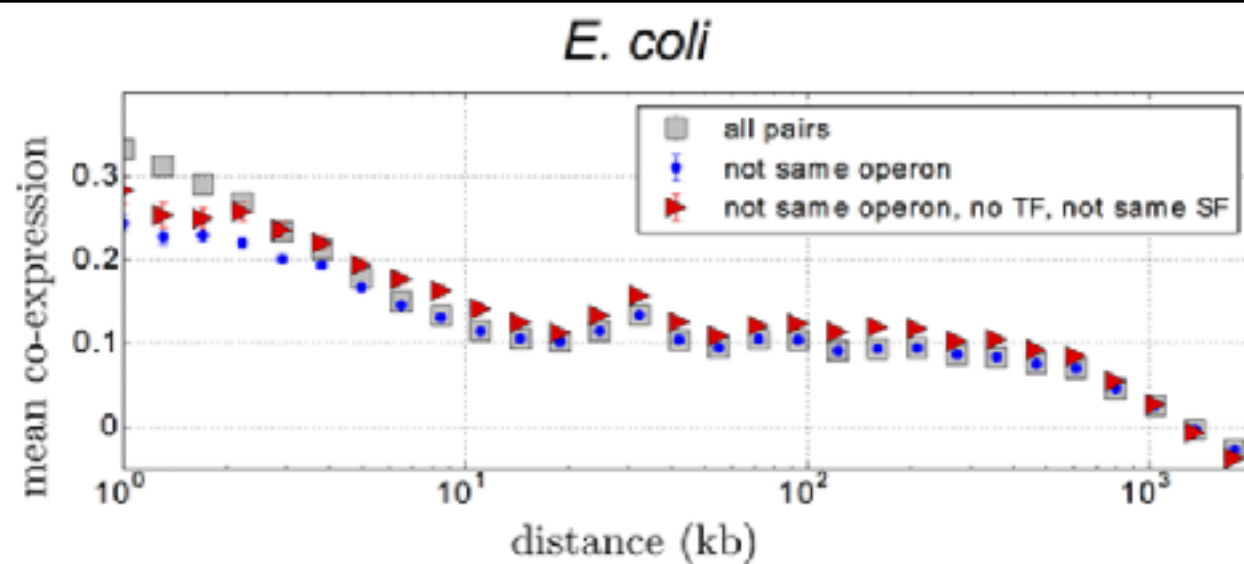
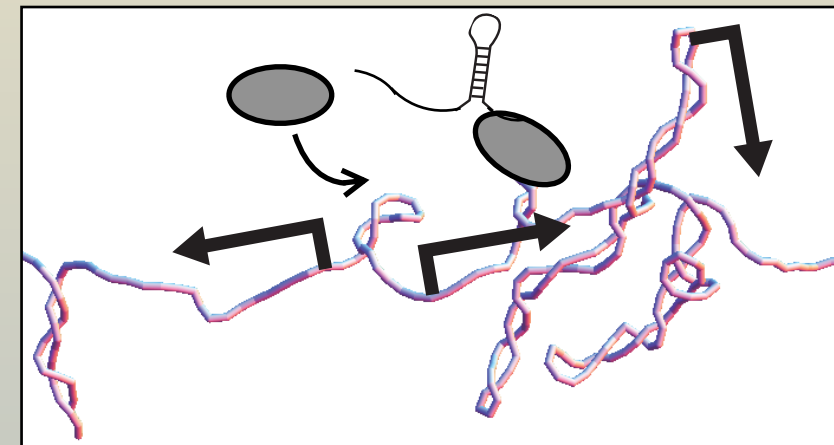
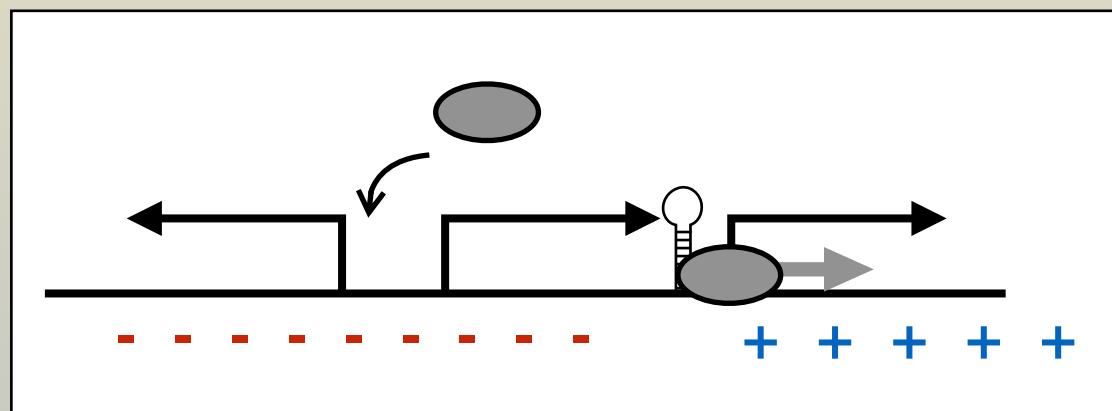
- specific operon organisation
- transcriptional read-through
- parsimonious regulation by TFs
- delimited by NAP DNA binding

3-operons	<i>E. coli</i> (45 segments)
	39% ($z=2.4$) 24%
	50% ($z=3.6$) 25%
	8% ($z=-2.5$) 27%
	3% ($z=-3.1$) 24%



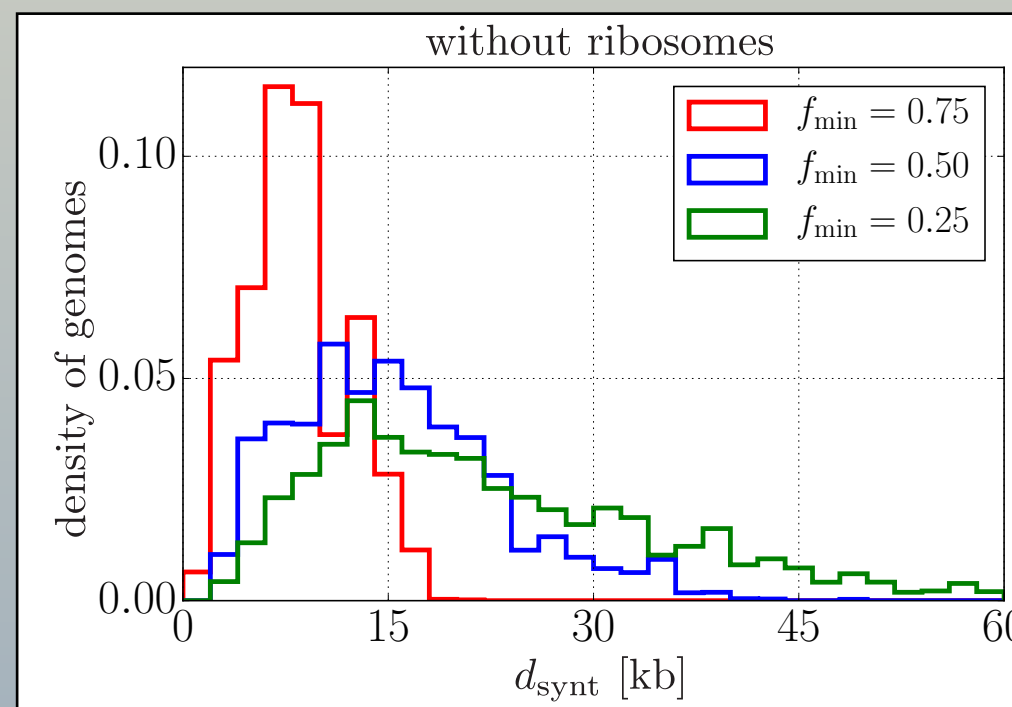
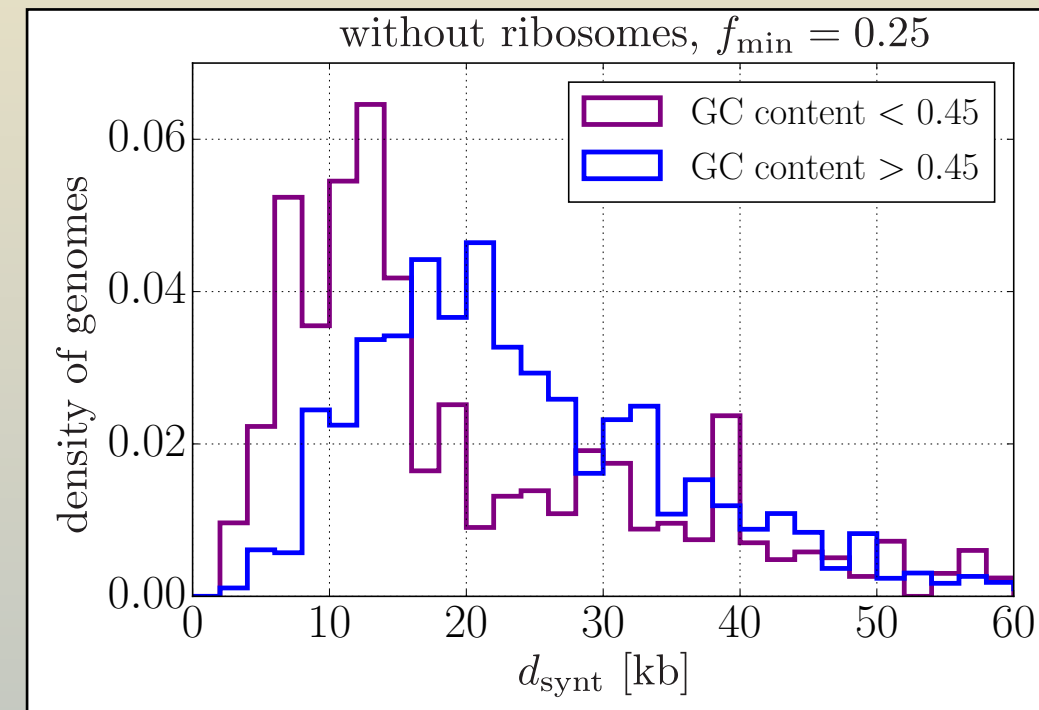
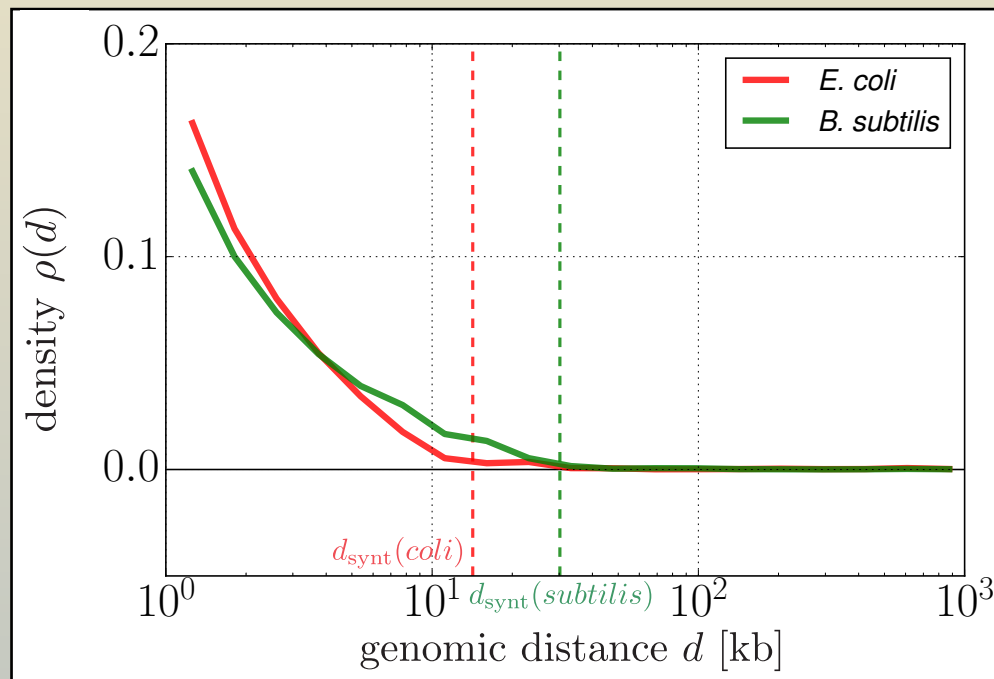
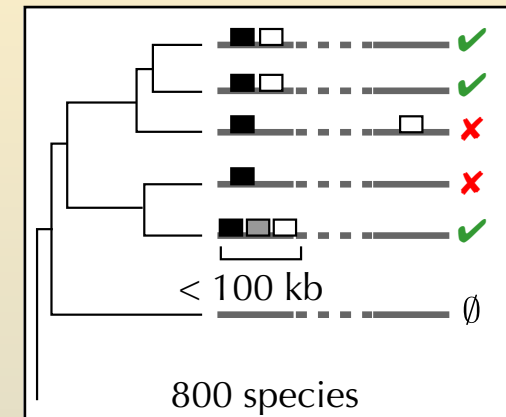
- How to identify **genomic motifs** associated to transcription regulation (and why)?
- Does some **universality** exist in the organisation of bacterial genomes?

Yes. But what about associated **characteristic lengths**?



Characteristic lengths associated to synteny properties

- 800 different bacterial species (NCBI+STRING database)
- given a genome:
what is the **fraction of pair of genes** remaining **below 100 kb**
in a certain **fraction (f_{\min}) of other genomes**



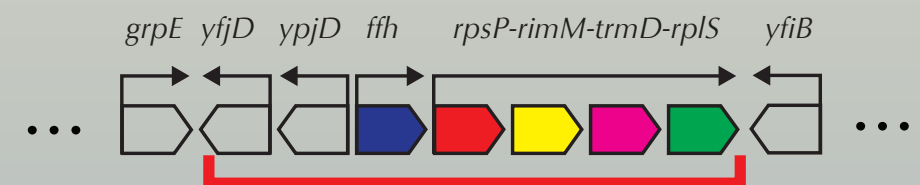
10-20 kb characteristic length

- How to identify **genomic motifs** associated to transcription regulation (and why)?
- Does some **universality** exist in the organisation of bacterial genomes?
- To which extent motifs are **compatible to other structural properties**?

B. subtilis (661 segments)



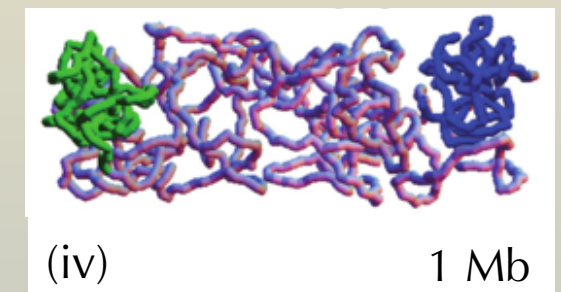
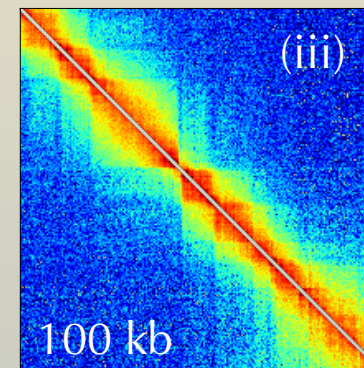
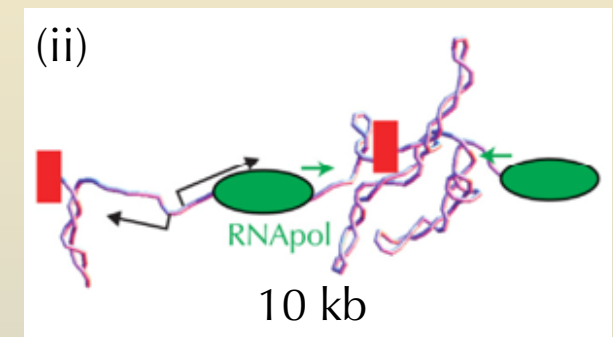
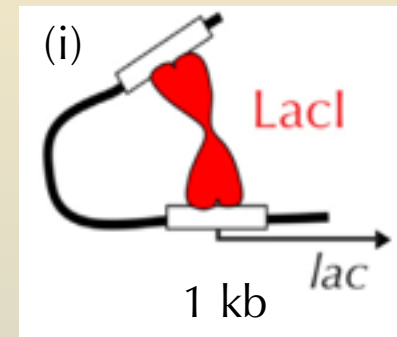
E. coli (740 segments)



10-20 kb characteristic length

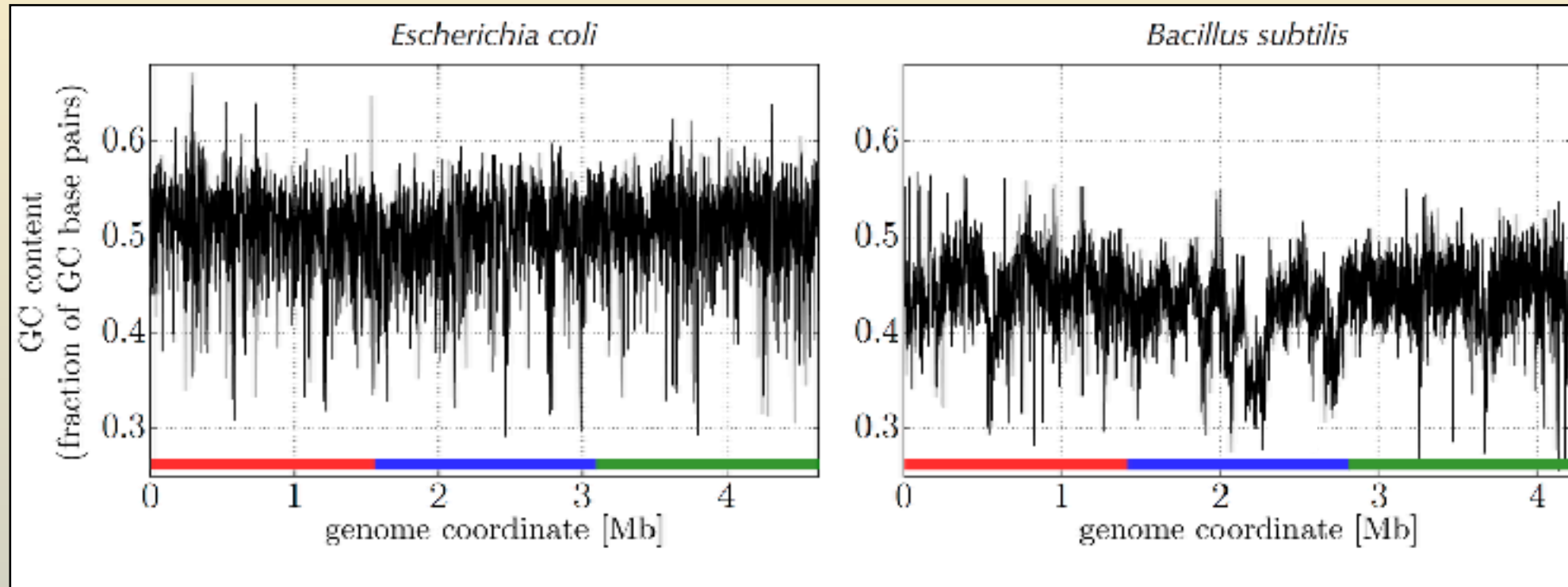
Known length scales associated with genome/chromosome structuring in bacteria

- (i) 1 kb scale: DNA loops
- (ii) 10 kb scale: supercoiling domains
- (iii) 100 kb scale: chromosome interaction domains
- (iv) 1000 kb (1 Mb) scale: macrodomains

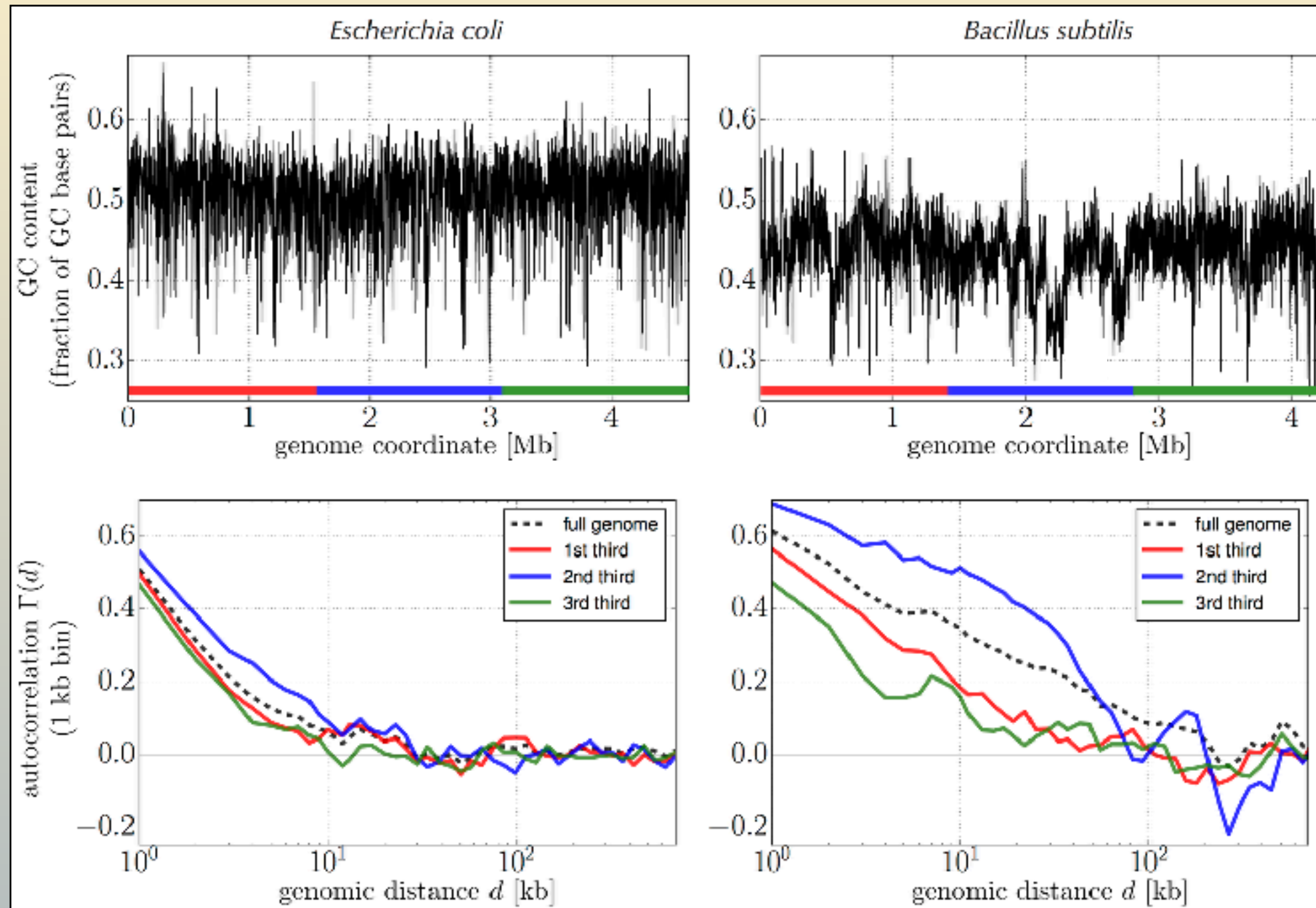


Question: is there an **associated sequence signature** (...GATTACA...)

Characteristic lengths associated to GC content (or how to circumvent the absence of translational invariance)

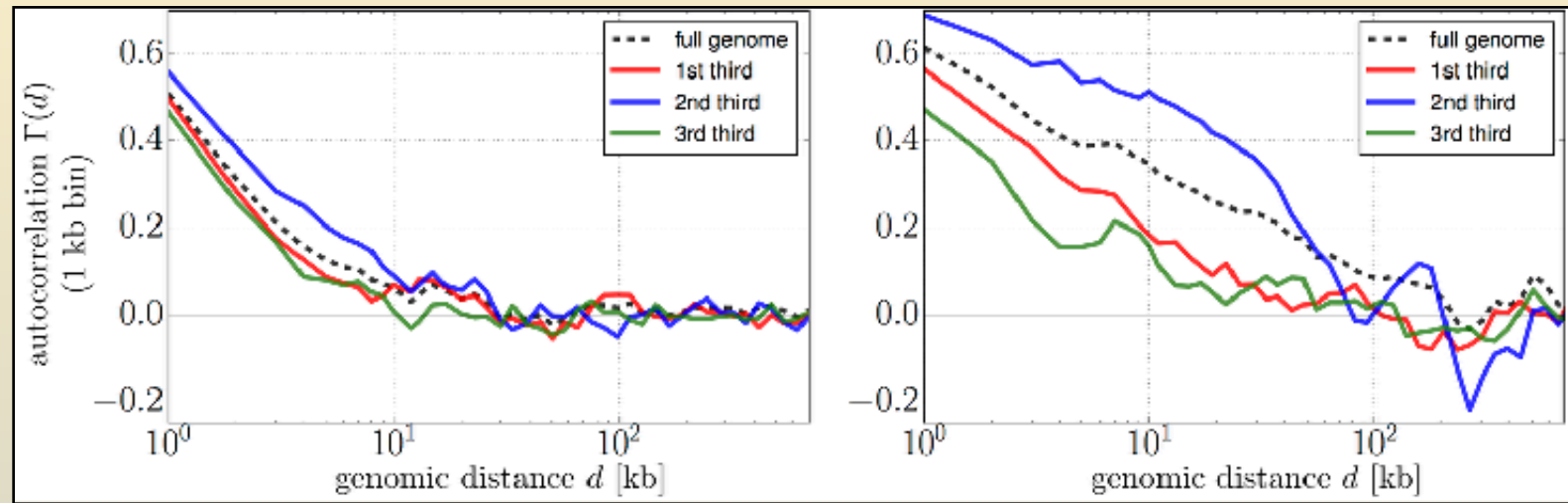


Characteristic lengths associated to GC content (or how to circumvent the absence of translational invariance)

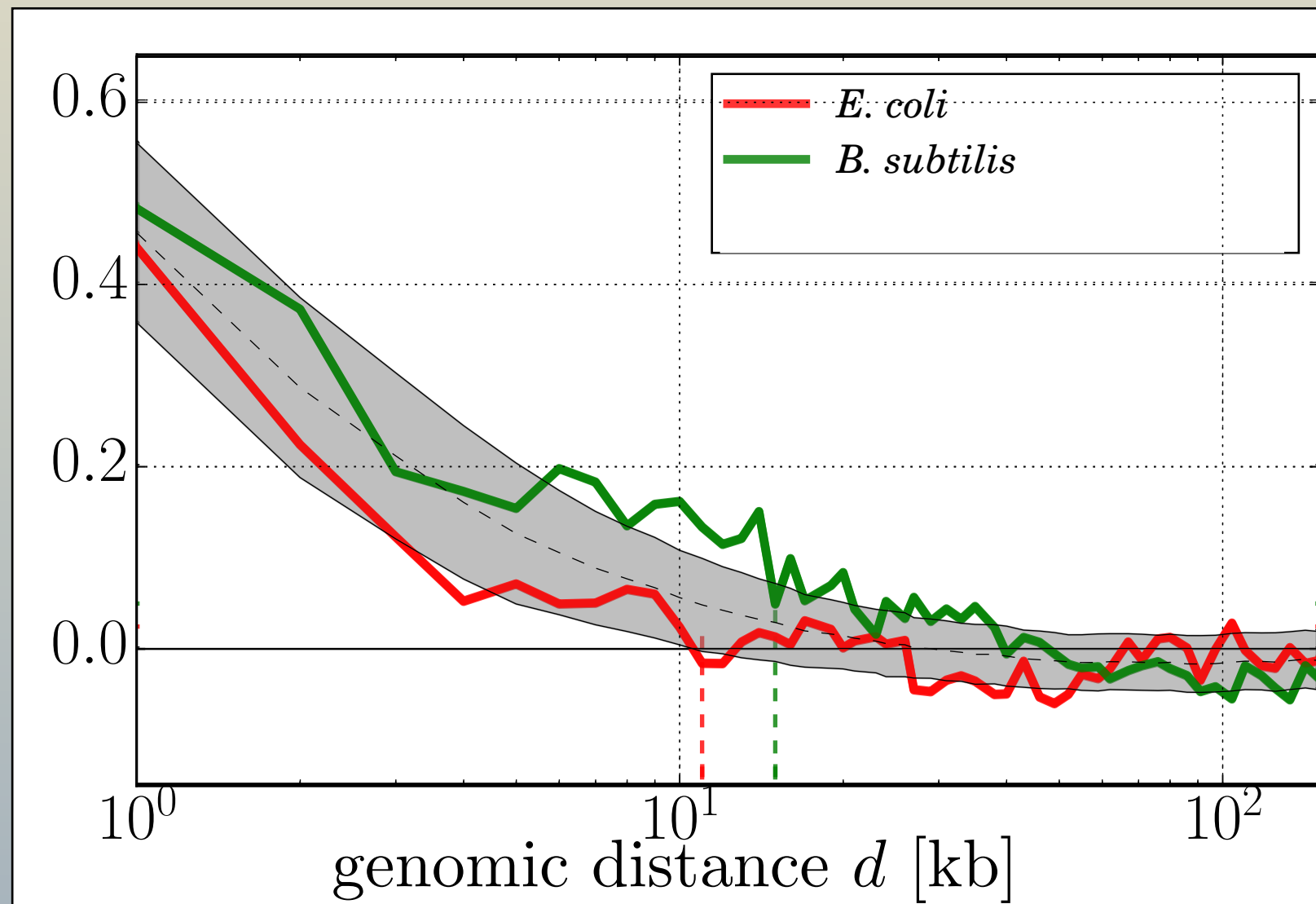


$$\Gamma(d) = \frac{1}{N\sigma_x^2} \sum_{i=0}^{N-1} (x[i] - m_x)(x[i + d/\ell] - m_x),$$

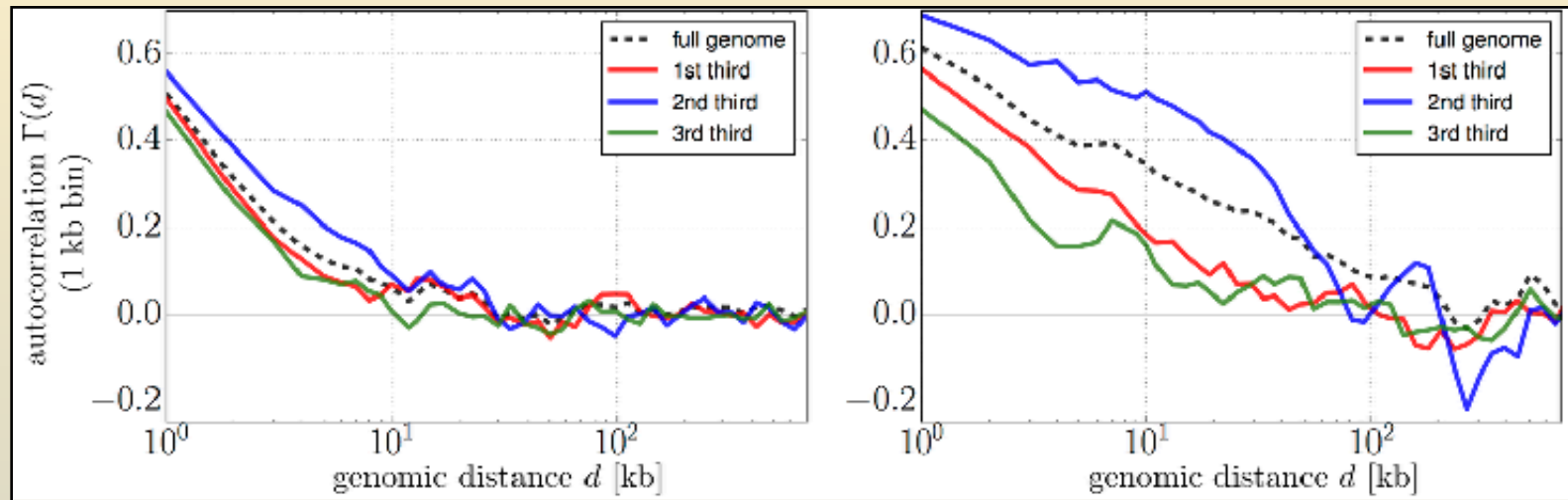
Workaround: use the median instead of the mean...



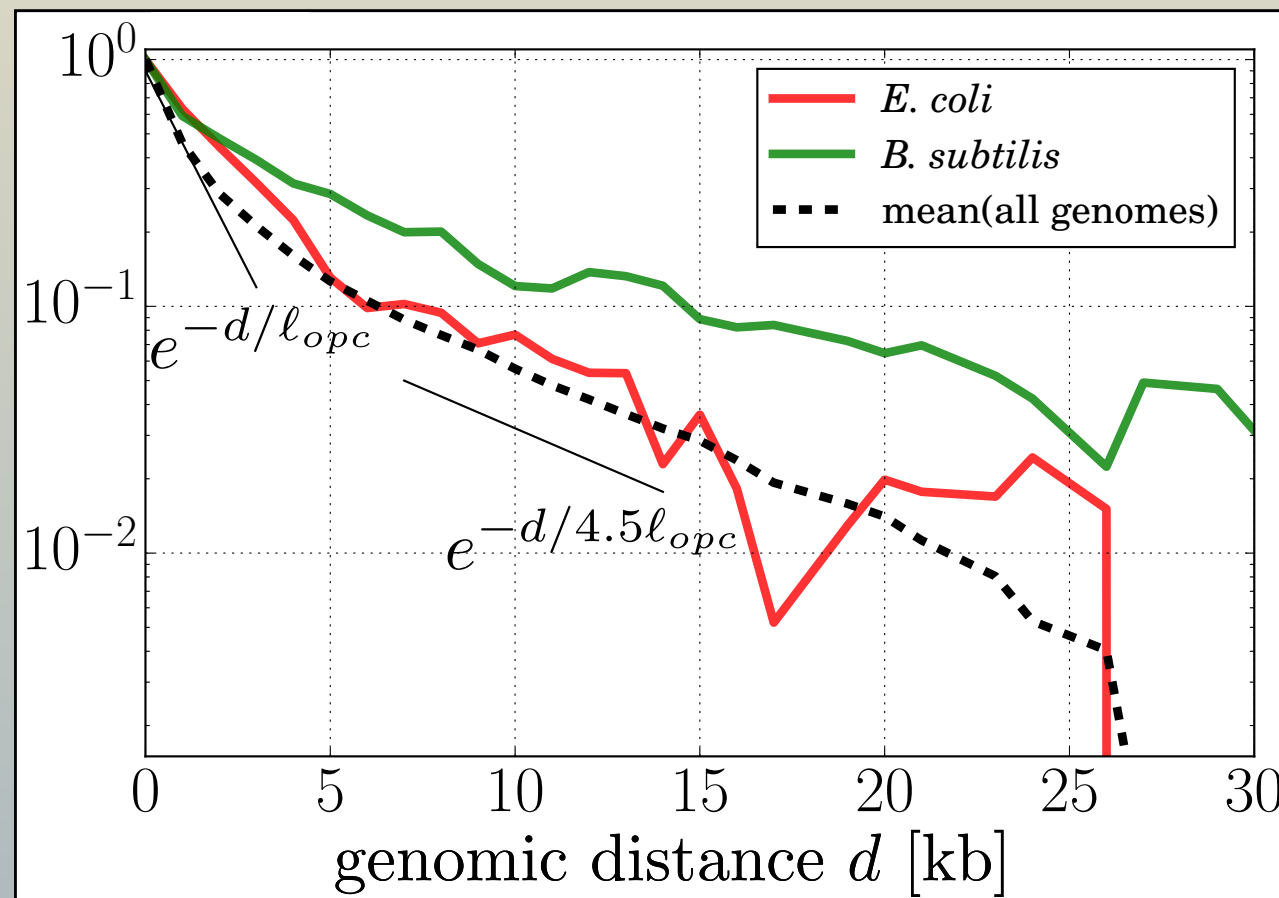
median autocorrelation $\Gamma_{\text{med}}(d) = \text{med}(\{\Gamma_{\alpha}(d)\}_{\alpha})$



Workaround: use the median instead of the mean...

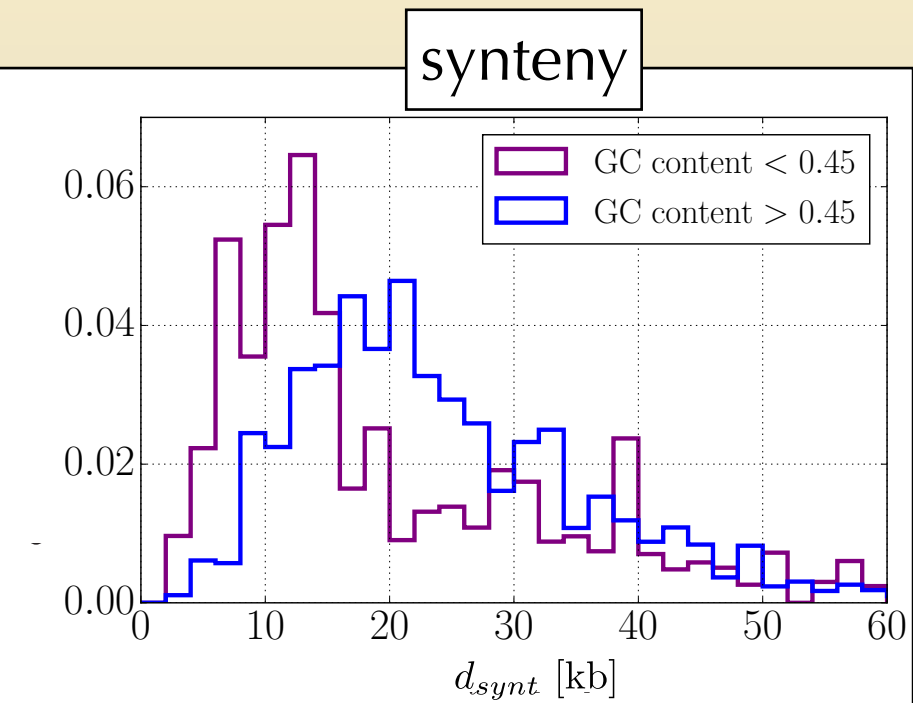
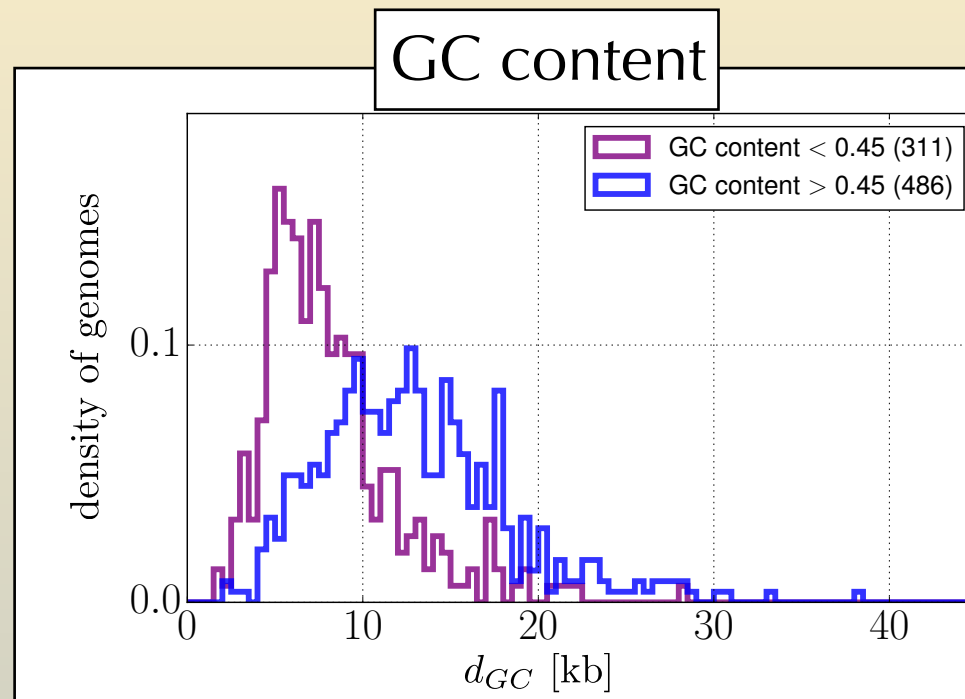
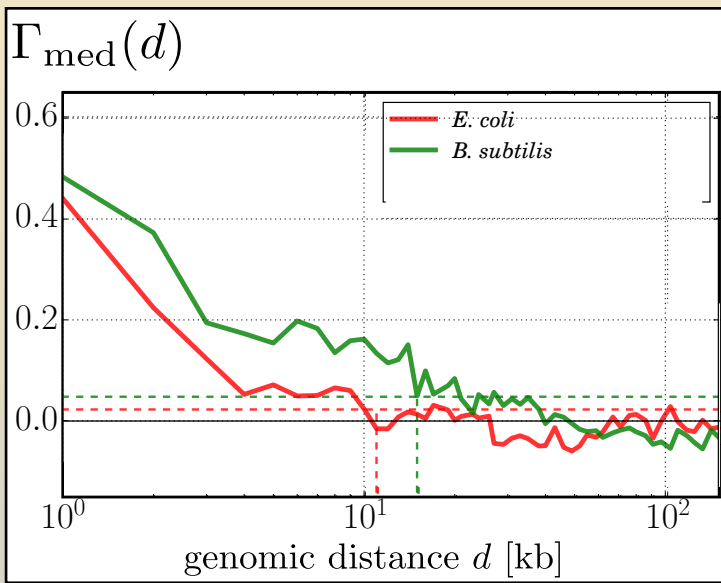


median autocorrelation $\Gamma_{\text{med}}(d) = \text{med}(\{\Gamma_{\alpha}(d)\}_{\alpha})$

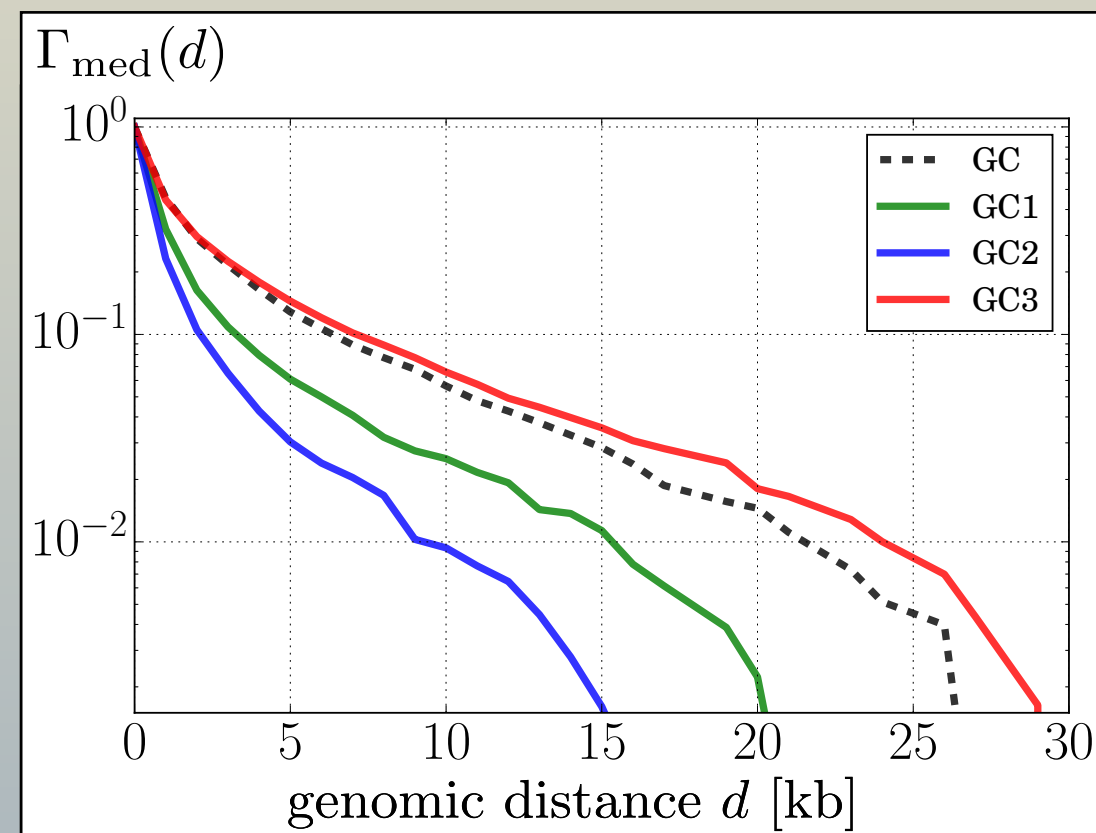
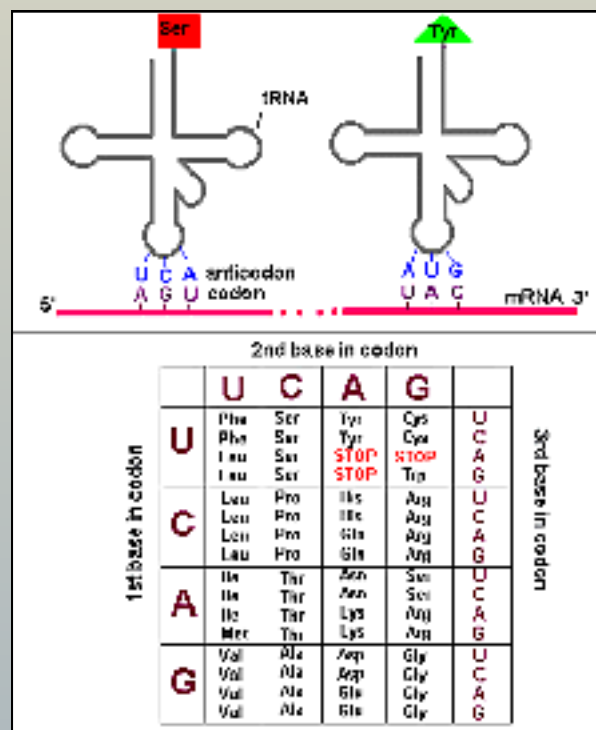
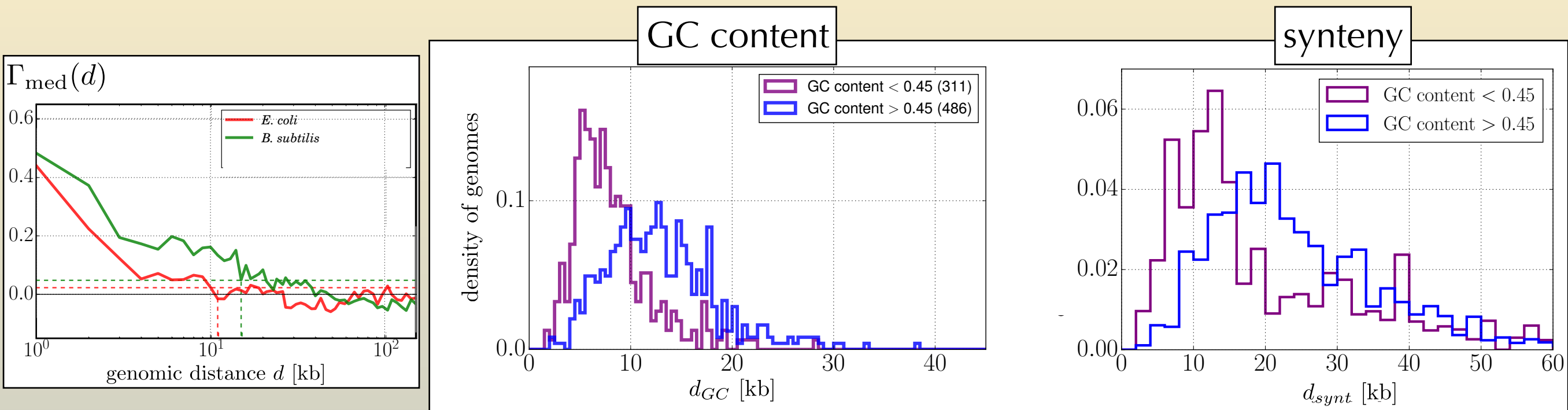


2 correlation lengths with the shortest one = operon distribution
(not a power law)

synteny vs GC content: similar characteristic lengths (similar segments)



synteny vs GC content: similar characteristic lengths (similar segments) driven by the correlation of the third codon base



synteny = coupling basal transcription AND translational coordination

Take-home

- **genomic motifs:**
 - synteny segments = basal coordination of transcription and translation
- **universality:**
 - gene/operon organisation
 - 10-20 kb characteristic lengths
- **relation with other structural properties:**
 - supercoiling-domains

