Res. Microbiol. 150 (1999) 725–733
© 1999 Éditions scientifiques et médicales Elsevier SAS. All rights reserved

Functional and evolutionary roles of long repeats in prokaryotes

Eduardo P.C. Rocha[a,b,*], Antoine Danchin[b], Alain Viari[a]

[a]*Atelier de bioInformatique, université Paris VI, 12 Rue Cuvier, 75005 Paris, France*
[b]*Unité de regulation de l'expression génétique, Institut Pasteur, 75724 Paris cedex 15, France*

Abstract — Most recently published complete bacterial genomes have revealed unexpectedly high numbers of long strict repeats. In this article we discuss the various functional and evolutionary roles of these repeats, focusing in particular on their role in terms of genome stability, gene transfer, and antigenic variation. © 1999 Éditions scientifiques et médicales Elsevier SAS

gene transfer / natural transformation / antigenic variation

1. Introduction

Due to their compact genomes, prokaryotes have been thought to lack long repeats. From here to conclude that any redundant sequence would be counterselected was a too easily warranted conclusion. Even if bacteria strive to attain minimal functional genomes [1], the action of transposable elements alone continuously introduces repeats in the sequence. Moreover, innovation brought about by repeats (be it only in terms of spacing or timing) may confer significant selective advantages. As we shall discuss at length in this article, long repeats are representative of important evolutionary mechanisms that allow bacteria to adapt faster to environmental change. These mechanisms involve increased ability to transfer genes or perform antigenic variation and are widespread in the bacterial lineage.

In recent years, various types of repeated DNA were discovered in many prokaryotes, either included in genes, in intergenic sequences, or in transposable elements. Repeats in genes usually involve duplicated or functionally related genes, such as the rRNA operons (reviewed in this issue by Hill), or protein domains [2]. Interspersed repetitive sequences such as bacterial interspersed mosaic elements (BIMEs) or intergenic repeat units (IRUs) [3, 4] are a common feature in genomes of enterobacteria and presumably reflect regulatory or structural requirements of the bacterial chromosome, although no clear-cut function has yet been ascribed to them [5]. DNA elements such as insertion sequences (ISs) and transposons are major evolutionary actors in the genome since they mediate genome rearrangements, plasmid integration, and gene transfer [6, 7].

In this review, we shall focus our attention upon long strict repeats in bacterial genomes. In recent years, the analysis of completely sequenced genomes revealed that long repeats are ubiquitous in archaea and eubacteria [8-12]. Although many of these repeats still have unknown roles, our knowledge of their functions has increased significantly with the recent genomic exploratory analysis. The definition of what is a 'long' repeat arises from statistical methods enabling assignment of the probability of the occurrence of a repeat of a certain length given simple stochastic models [13, 14]. This statistically significant length depends upon the genome size and its nucleotide composition. It typically ranges from 22 to 26 nucleotides for bacterial genomes [14]. Since repeats of this length or longer are very unlikely to exist by

* Correspondence and reprints
Tel.: +33 (0) 44 27 65 136; fax: +33 (0) 44 27 63 12;
erocha@alsi.jussieu.fr

chance alone (P value <1 ‰), they are most certainly meaningful in terms of the biology of the organism.

In the first part of this article we review current knowledge of the number and spatial distribution of repeats in completely sequenced genomes. Then we focus on the use of exact repeats by pathogenic organisms in order to evade the immune system of the host. In the third part we describe recent work on the very peculiar distribution of repeats in *B. subtilis*, and our proposition of a model of genetic transfer in this organism. We conclude with a general discussion on the lessons that the study of repeats can provide in terms of genome stability and bacteria adaptation strategies.

2. Distribution of long strict repeats in bacterial genomes

Using a statistic of extremes developed by Karlin and Ost [13] and an efficient algorithm to search for repeats [15, 16], we have proceeded to an extensive analysis of repeats in eight prokaryotic genomes [14]. This was done after excluding simple repeated motifs and rRNA operons (often present in multiple copies). Contrary to expectation, all analysed genomes possess a large number of repeats, from a minimum of 139 in *Mycoplasma genitalium* to a maximum of 552 in *Mycoplasma pneumoniae*. However, when the different sizes of the genomes were taken into account, we observed that the largest genomes (nonpathogenic) had the smallest density of repeats (40 repeats/Mb for *Bacillus subtilis* and 86 repeats/Mb for *Escherichia coli*) and the smallest (pathogenic) genomes had the highest density (676 repeats/Mb for *M. pneumoniae*, and 240 repeats/Mb for *M. genitalium*). This is surprising in the light of the minimal genome hypothesis for the *Mycoplasma*, but becomes quite understandable from the knowledge of their antigenic variation strategies (see below). The average size of the repeats goes from 52 bp in *Methanococcus jannaschii* up to 100 in *Helicobacter pylori*, though the longest repeats are typically larger than 400 bp, with a maximum of 1890 bp in *H. pylori* and 1856 bp in *Methanobacterium thermoautotrophicum*.

We also observed that the spatial distribution of the two copies of the repeats is very different between the genomes. In fact, in *E. coli* and *M. jannaschii,* the two copies of a repeat lie at a nearly random distance from each other. In contrast, in the two nonspecific competent organisms *B. subtilis* and *M. thermoautotrophicum*, more than half of the copies lie at less than 50 kb from each other [14]. Moreover, less than 5% of the repeats have occurrences spaced by more than 10% of the chromosome length, indicating a very strong counter selection of distant occurrences (see section 5). Although *E. coli* possesses a large number of ISs and repetitive ssRNA, this was found not to be the cause for the different bias [14].

3. Adaptation of pathogens

Bacteria present very diverse evolutionary strategies linked to pathogenesis, such as: pathogenicity islands in *Salmonella* and *E. coli* [17], and *Vibrio cholerae* [18], gene uptake systems in *V. cholerae* [19], tandem repeats for phase variation in *Haemophilus influenzae* [20], tandem repeats linked with contingency loci in *H. pylori* [10, 21], and long repeats for antigenic variation in *Mycoplasma* [9, 22], *B. burgdorferi* [23], and *M. tuberculosis* [12].

Long DNA repeats play a very important role in the strategies of antigenic variation, and represent an important fraction of some small genomes such as the *Mycoplasma*. In *M. genitalium*, the smallest known genome, repeats of the three-gene operon that encodes one of the major surface proteins, the adhesin MgPa, represent more than 4% of the genome [8, 24]. The percentage of identity between these repeats ranges from 78 to 90% [8], which is close enough for homologous recombination between the most similar regions, but far enough for complete gene conversion. The genome of *M. pneumoniae* is also composed of repetitive DNA elements that constitute up to 8% of the genome. These repeats code for polypeptides that have high similarity to the adhesins of the tip

structure that adheres to the host cell [9, 25]. In particular, a large number of open reading frames (ORFs) are similar to those of the P1 operon, including the P1 and ORF6 proteins that are essential for adherence to the host cell. In spite of the existence of multiple, but not exactly similar, copies of this operon, preliminary experiments have shown that none of these ORFs is expressed under standard laboratory conditions, though presumably they exchange genetic material by partial gene conversion [9]. Therefore both *Mycoplasma* use similar strategies for antigenic variation, though *M. pneumoniae* seems to possess a larger group of repeats [14, 26]. In both cases these repeats share sufficient homology to allow recombination, and therefore new solutions of antigenic presentation can be found when the immune environment changes. Furthermore, this homology is never very extensive, so that gene conversion cannot make all genes identical, which would place the bacteria at an evolutionary dead end, since no new solutions could be found by simple homologous recombination. *M. pneumoniae* possesses nine groups of repeated elements in its genome. They lie on cooriented genes [9], and share long strict repeats with all other groups, therefore producing a maximal set of recombination possibilities [14]. Finally, there are almost no anti-oriented repeats, and therefore the recombination between the repeated regions probably leads to frequent deletions of genetic material. This may help to explain the reduced size of *Mycoplasma* genomes.

Figure 1 presents an update of our previous results. We hypothesised that strategies linked to pathogenesis should involve large number of repeats in pathogenic bacteria [14]. However, this does not seem to be a general strategy, since *Chlamydia trachomatis* (the agent of several human infections) and *Rickettsia prowazekii* (the agent of epidemic typhus) present very few repeats. Both organisms are intracellular obligatory parasites [27, 28]. *C. trachomatis* has a cryptic plasmid, which possesses 22 tandem repeats and may therefore be used for the regulation of pathogenicity [29]. Interestingly, the closely related *C. pneumoniae* presents a much larger number of repeats. *R. prowazekii* is the sole bacterial genome sequenced to date to possess a large percentage of noncoding regions (24%), which is apparently a consequence of the genome reduction taking place in this organism [28]. It is not clear at the moment whether these facts are related to a different evolutionary strategy related to pathogenesis.

The case of *Borrelia burgdorferi* (the agent of Lyme disease) is, in this respect, particularly interesting. Though the genome analysis has revealed very few repeats in the chromosome, the plethora of linear and circular plasmids that add up to more than 600 kb of genetic material contain a very large amount of repeated sequences [23]. These plasmids possess lower densities of genes (70%) [23] of which many code for surface proteins, therefore paving the way for a global strategy of antigenic variation *(figure 1)*. The accumulation of recombinant genetic material in plasmids has the important advantage of removing most of the genetic rearrangements from the chromosome, thereby stabilising it, which may be particularly important for this linear chromosome. In fact the genome of *B. burgdorferi* is significantly polarised (65% of the genes are in the leading strand), and genes in different replicating strands exhibit a remarkably contrasted codon usage, indicating very little rearrangement of the chromosome [30, 31].

4. Gene acquisition in *B. subtilis*

The *B. subtilis* chromosome is the genome exhibiting the smallest density of repeats *(figure 1)*. One third of these are related to prophage SPβ, 8% to intergenic elements in rRNA operons, and less than 10% to other elements such as putative ssRNA and rho-independent terminators. The remaining 50% constitutes a homogeneous set comprising repeats whose copies are very close together, at an average distance of 10 kb [14]. Though most of the occurrences of these repeats are located inside genes, they do not seem to be related to simple mechanisms of gene duplication. Moreover the analysis of spatial distribution of paralogues in *B. subtilis* re-

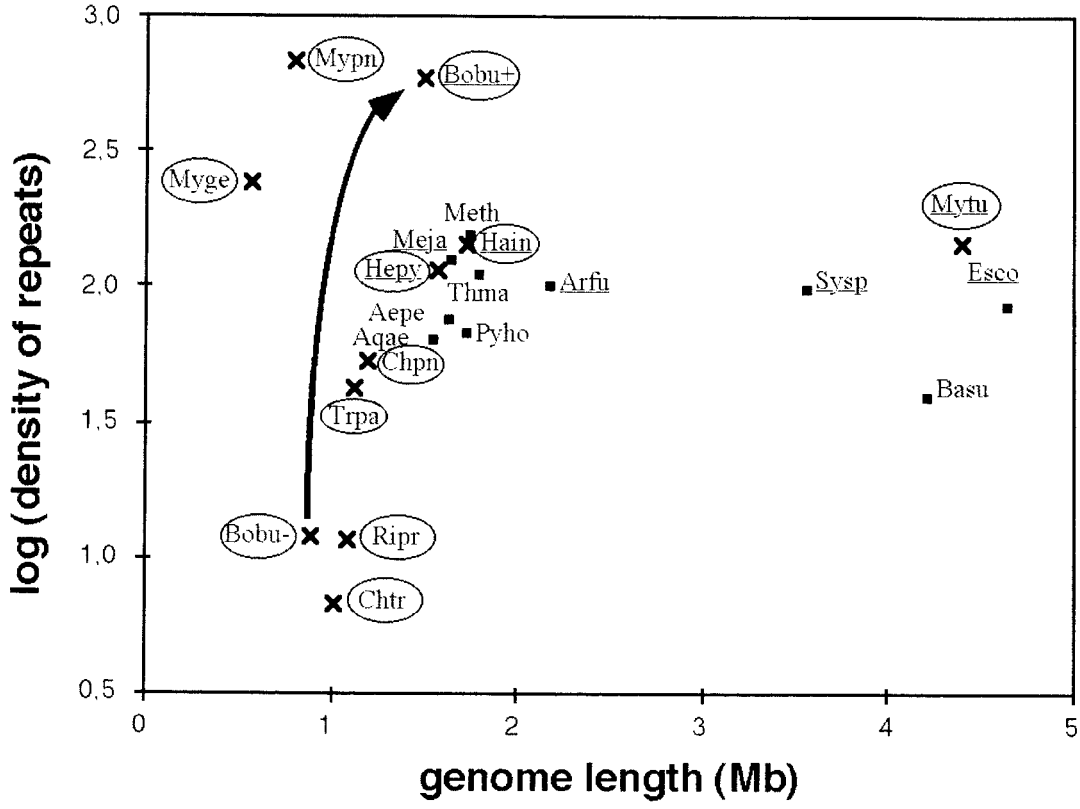

Figure 1. Relationship between the density of repeats in the genome and the genome length. The smallest genomes tend to have the largest and smallest number of repeats. Names underlined indicate genomes with ISs and names in encircled indicate pathogenic organisms (both tend to have large densities of repeats). The arrow represents the increase in the density of repeats of the genome of *B. burgdorferi* when its plasmids are included in the analysis. Abbreviations: *A. aeolicus*(aqae), *A. fulgidus* (arfu), *A. pernix* (aepe), *B. subtilis* (basu), *B. burgdorferi* (bobu), *C. pneumoniae* (chpn), *C. trachomatis* (chtr), *E. coli* (esco), *H. influenzae* (hain), *H. pylori* (hepy), *M. jannaschii* (meja), *M. thermoautotrophicum* (meth), *M. genitalium* (myge), *M. pneumoniae* (mypn), *M. tuberculosis* (mytu), *P. horikoshii* (pyho), *R. prowazekii* (ripr), *Synechocystis* sp (sysp), *T. pallidum* (trpa).

vealed no tendency towards spatial proximity of these genes (unpublished results). This random distribution is consistent with previous findings that *M. genitalium* and *H. influenzae* have randomly distributed paralogues [32].

B. subtilis is capable of gaining access to new genetic information by becoming competent, and its nonclonality [33] indicates that it is frequently subjected to interspecific gene exchange and recombination. However, it does not possesses ISs, nor any kind of transposons, the transformation of monomeric plasmids without chromosomal inserts occurs at very low rates, and conjugation is virtually undetectable [34]. Since most known gene transfer mechanisms involve the action of ISs (transfer from bacteriophages also frequently involves ISs or involve processes reminiscent to IS integration), we still do not know how *B. subtilis* proceeds in order to incorporate nonhomologous genetic information into its chromosome.

Based on the analysis of repeats we have proposed a mechanism that: i) does not require ISs; ii) makes use of the nonspecific competent character of *B. subtilis;* and iii) explains the presence of the closely spaced repeats in the chromosome [14]. We suggest that repeats are remnants of recent horizontal transfer events into competent cells via a Campbell-like integration process *(figure 2)*. In *B. subtilis*, compe-

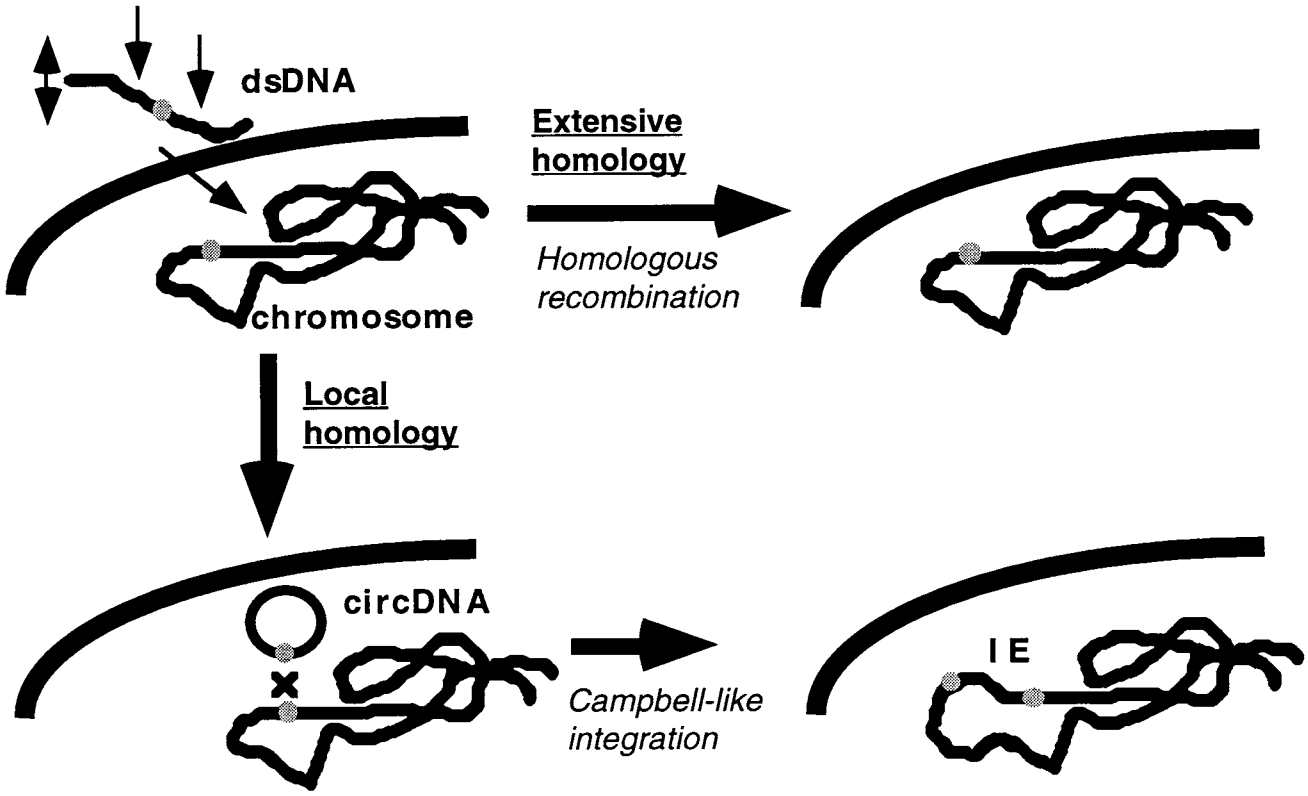

Figure 2. Proposed evolutionary mechanisms for horizontal transfer in *B. subtilis*. The competent cell uptakes DNA attached to the membrane, after being made single stranded and restricted in fragments of an average size of 10 kb. If the DNA does not share similarity to any DNA on the genome it is simply degraded. If the DNA shares extensive similarity to the chromosome it can proceed to integrate through homologous recombination, partly replacing the previous sequence. If the DNA shares a small region of homology it can circularise and integrate in a Campbell-like way. This produces the integration of the sequence in the chromosome, with the addition of the information flanked by the two occurrences of the repeat.

tence is reached in the stationary phase when the cell risks death by starvation [35]. In this situation a strategy of integration of foreign DNA may be advantageous by allowing the acquisition of new functions, such as antibiotic production or detoxification. DNA enters *B. subtilis* cells single-stranded, after a nonspecific interaction with the membrane, where it is cut into fragments [35]. The size of these fragments has been investigated by physical measurements [36] and by electron microscopy [37] giving averages of 8.5 kb and 11 kb, respectively. The average size of the 16 proposed inserted elements (IE) in the genome is 10.6 kb, therefore in agreement with the hypothesis of horizontal transfer into a competent bacteria. Heterologous single-stranded DNA entering the cell must circularise in order to integrate through a Campbell-like mechanism. This has been extensively studied for *B. subtilis* integrative plasmids, since these elements are also made single-stranded and cut into pieces before entering the cell. Though the integration of monomeric plasmids is very difficult in *B. subtilis*, the existence of a region of strong similarity between the donor 'plasmid' molecule and the chromosome allows facilitated plasmid transformation [38]. DNA synthesis and ligation convert the linear single-stranded DNA into a circular molecule, from which the second strand is synthesised. This process requires RecA [39], which raises its concentration 14-fold upon induction of compe-

tence in *B. subtilis* [40]. Once circular, the foreign DNA element is indistinguishable from any nonreplicative plasmid such as the ones used for cloning genes into *B. subtilis* [35, 41]. The length of the strict similarity necessary for the action of RecA recombination is optimal for values larger than 70 bp [42], but can be as small as 24 bp in *B. subtilis* [43]. Therefore, the length of repeats flanking all IEs is compatible with the model.

The IE account for nearly 5% of the *B. subtilis* chromosome, and most likely reflect recent acquisitions of genetic material, because repeats are erased by random drift in the absence of strong functional constraints. Substantiating this interpretation, the codon usage bias of the genes in IE is very peculiar and more than 50% of them are classified as horizontally transferred in the FCA (factorial correspondence analysis) classification of the *B. subtilis* genes (they represent only 13% of the genes in the chromosome) [11, 44]. Nearly 60% of these genes have unknown function and are not similar to any gene in the databases, suggesting that they are not essential housekeeping genes. The functional classification of the remaining genes in IE indicates overrepresentation of some genes frequently transferred horizontally [45], such as genes involved in competence, antibiotic production, flagellins, ABC transporters, restriction modification, and repair [14].

Since *M. thermoautotrophicum* reveals a spatial distribution of occurrences of repeats similar to that of *B. subtilis*, while it is also nonspecifically competent and does not possess ISs, we think it is likely that this strategy for gene transfer may be widespread in the bacterial world.

5. Adaptability versus genome stability

Natural isolates of enterobacteria have large ranges of chromosome size distribution, and some isolates of *E. coli* differ by as much as 1 Mb of genetic material [46], which seems to be due to horizontal transfer and not to large-scale duplications. ISs may mediate horizontal transfer and gene duplication, but also gene loss by recombination between two cooriented copies. This probably explains why the amount of ISs in the genome of the different strains does not correlate with the chromosome length [46]. Nevertheless, genomes lacking ISs are almost certainly much more stable, since rearrangements, insertions, and deletions of information should be much less frequent. In short, ISs do not seem to contribute directly to the genome augmentation, but to genome size dynamics, thereby increasing intraspecies variability. The absence of ISs in *B. subtilis* is surprising, since a related organism such as *B. cereus* and related bacteria possess ISs [47]. It will be interesting to compare the genome organisation of the latter to investigate for processes allowing to select against IS insertion or spread.

Strand bias is known to be a general feature of microbial genomes, influencing the relative composition in nucleotides [48], codons [30, 31], and even amino acids [31] of each replicating strand. This has led to the development of several methods for the determination of origins of replication based on the contrast between the leading and lagging strands. If one analyses the maximal accuracy of the method using linear discriminant analysis [31], then one obtains values close to 1 for genomes with high strand bias and stable genome structure, and close to 0.5 for genomes which either do not possess such bias or shuffle genes frequently. *Figure 3* shows the relationship between the density of repeats in a genome and the accuracy of the discriminant analysis, suggesting that genomes with less repeats allow more efficient discrimination of the replicating strands. This is most certainly related to the high stability of genomes lacking repeats, since they should suffer much smaller levels of intramolecular recombination. In particular, chromosomes with few repeats such as *C. trachomatis* and the spirochaetes have very contrasted genes in the leading and lagging strand, whereas genomes with many repeats such as *M. jannaschii* and *M. pneumoniae* have genes with nearly similar composition in both strands. This tendency is partially independent on the level of ISs in the genomes, since the *Mycoplasma* and *M. thermoautotrophicum* lack such elements.

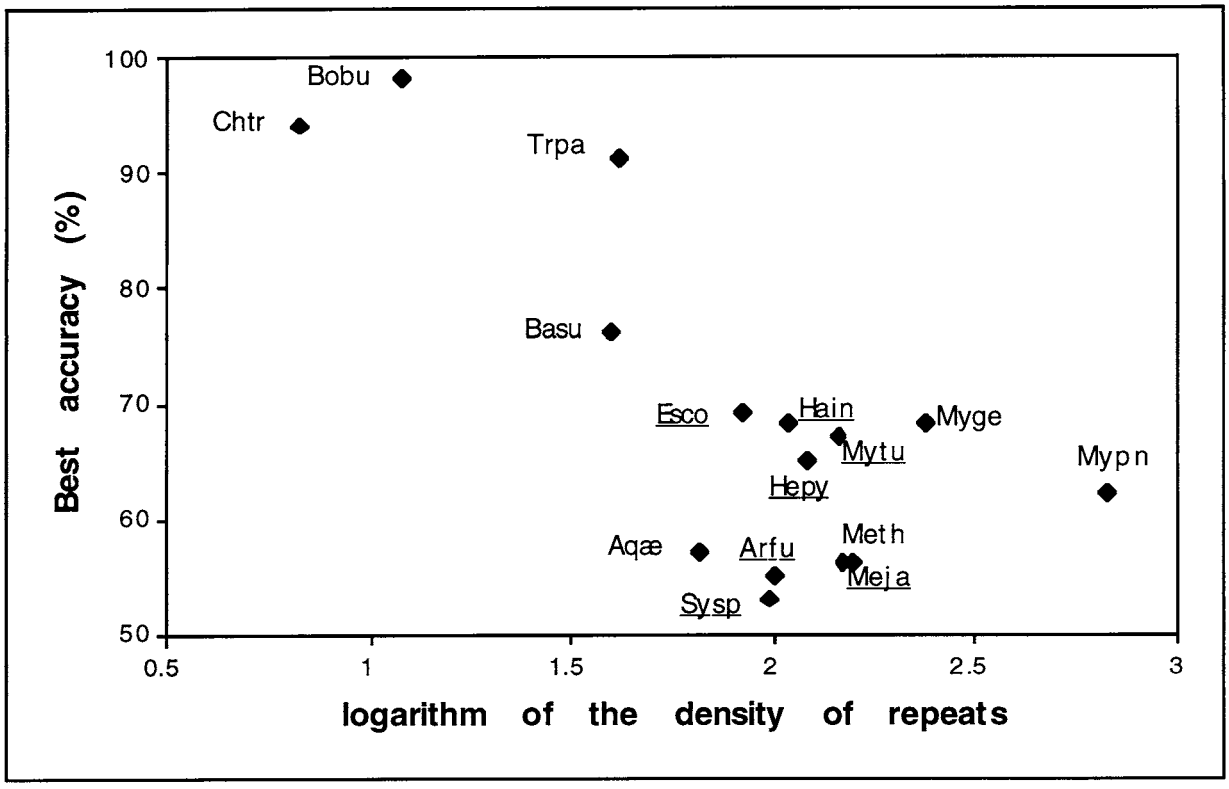

Figure 3. Relationship between the accuracy of gene strand discrimination and the density of repeats of each genome. Clearly, genomes with few repeats exhibit better accuracy, which is probably the result of genome stability. Underlined names indicate genomes with ISs, thereby indicating that this trend does not depend exclusively on its existence (e.g., the *Mycoplasma* do not seem to have ISs). See legend, *figure 1*, for abbreviations.

In contrast to the diversity of the chromosome sizes found in different *E. coli* isolates, many studies with other *B. subtilis* strains have revealed similar physical maps and chromosome lengths [49]. This agrees with the hypothesis that *B. subtilis* is more stable because it lacks long repeats. Itaya and coworkers have introduced two long repeats 300 kb apart in the *B. subtilis* genome, with a plasmid origin of replication between them. By recombination the genome was spliced into two subgenomes that were both maintained and required for survival [50]. Since the small subgenome is unstable, easily lost, and leads to very slow growth rates [50, 51], one may conclude that repeats lying far apart in the chromosome introduce significant instability in the genome. This may justify the absence of such repeats, and incidentally this indicates that experiments involving many laboratory constructions should always be considered with some caution.

It is somehow surprising that the two nonspecific competent bacteria sequenced so far do not have ISs. Because these organisms take up DNA of any organism, they should be 'contaminated' frequently by such selfish DNA, unless an efficient mechanism discriminates between self and foreign DNA. One may suppose that the instability produced in the chromosome by largely spaced repeats is partly responsible for their avoidance, since this would force an IS to remain single copy. However, many insertion sequences and transposons occur as single-copy [6], and therefore this explanation does not seem fully satisfactory. The analysis of the *B. subtilis* genome has revealed that putative trans-

posases in the genome are pseudogenes with many frameshifts [11]. All this leads us to wonder whether *B. subtilis* possesses a mechanism to eliminate ISs such as the ones known to exist in some eukaryotes [52]. The presence of ISs would confer some selective advantage if it allowed bacteria to adapt quickly by facilitating horizontal transfer or gene duplication. However, if *B. subtilis* has found an evolutionary mechanism that makes ISs obsolete, then the evolutionary advantages they confer also become obsolete. All these observation make determination of the genome sequence of a *B. cereus* isolate most interesting. Indeed, it has been shown not only that this organism harbours ISs, but also that the length of its genome varies widely [53]: comparison with the *B. subtilis* genome will allow us to have much deeper insights into the processes underlying IS propagation and stability in Gram-positive bacteria.

6. Conclusion

The availability of complete genomes is changing our perspective on bacterial genetics in many respects, and quantifying our ignorance about these genomes is not one of its minor contributions. The increasing types of repeats that have been found through genome programmes and other works reveal that repeated DNA is by no means junk DNA. In this article we have tried to discuss some of the latest findings on the subject of long nonrepetitive DNA repeats. Though still at the beginning, these studies have already paved the way for evolutionary studies regarding horizontal gene transfer, a major enemy in the fight against important human pathogens.

Acknowledgments

This work was partially funded by the EU grant Biotech BIO4-CT96-0655. E. R. acknowledges the support of PRAXIS XXI, through the grant BD/9394/96.

References

[1] Maniloff J., The minimal cell genome: 'on being the right size', Proc. Natl. Acad. Sci. USA 93 (1996) 10004–10006.
[2] Ohno S., Epplen J.T., The primitive code and repeats of base oligomers as the primordial protein-encoding sequence, Proc. Natl. Acad. Sci. USA 80 (1983) 3391–3395.
[3] Versalovic J., Koeuth T., Lupski J.R., Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes, Nuceic Acids Res. 19 (1991) 6823–6831.
[4] Bachellier S., Clément J.M., Hofnung M., Gilson E., Bacterial interspersed mosaic elements (BIMEs) are a major source of sequence polymorphism in *E. coli* intergenic regions including specific associations with a new insertion sequence, Genetics 145 (1997) 551–562.
[5] Versalovic J., Lupski J.R., Bacterial Genomes, in: Bruijn F.J.D., Lupski J.R., Weinstock G.M. (Eds.), Chapman & Hall, 1998, pp. 38–48.
[6] Mahillon J., Chandler M., Insertion sequences, Microbiol. Mol. Biol. Rev. 62 (1998) 725–774.
[7] Syvanen M., in: Bruijn F.J.D., Lupski J.R., Weinstock G.M. (Eds.), Bacterial Genomes, Chapman & Hall, 1998, pp. 213–220.
[8] Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A., Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G. et al., The minimal gene complement of *Mycoplasma genitalium,* Science 270 (1995) 397–403.
[9] Himmelreich R., Hilbert H., Plagens H., Pirki E., Li B.C., Herrmann R. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*, Nucleic Acids Res. 24 (1996) 4420–4449.
[10] Tomb J.F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G., Fleischmann R.D.E.T Al The complete genome sequence of the gastric pathogen *Helicobtacter pylori,* Nature 388 (1997) 539–547.
[11] Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bertero M., Bessieres P., Bolotin A., Borchert S. et al., The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis,* Nature 390 (1997) 249–256.
[12] Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K. et al., Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, Nature 393 (1998) 537–544.
[13] Karlin S., Ost F.,in: Cam L.M.L., Olshen R.A. (Eds.), Proceedings of the Berkeley Conference in honor of Jerzy Neyman and Jack Kiefer, Vol. I, Wadsworth Inc., 1985, pp. 225–243.
[14] Rocha E., Danchin A., Viari A., Analysis of long repeats in bacterial genomes reveals alternative evolutionary mechanisms in *Bacillus subtilis* and other competent prokaryotes, Mol. Biol. Evol. 16 (1999) 1219–1230.
[15] Karp R.M., Miller R.E., Rosenberg A.L., Proceedings 4th Annual ACM Symposium Theory of computing, ACM 1972, pp. 125–136.
[16] Soldano H., Viari A., Champesme M., Searching for flexible repeated patterns using a non-transitive relation, Patt. Recogn. Lett. 16 (1995) 233–246.
[17] Groisman E.A., Ochman H., How Salmonella became a pathogen, Trends Microbiol. 5 (1997) 343–349.
[18] Lin W., Fullner K.J., Clayton R., Sexton J.A., Rogers M.B., Calia K.E., Calderwood S.B., Fraser C., Mekalanos J.J., Identification of a vibrio cholerae RTX toxin gene cluster that is tightly linked to the cholera toxin prophage, Proc. Natl. Acad. Sci. USA 96 (1999) 1071–1076.
[19] Mazel D., Dychinco B., Webb V.A., Davies J., A distinctive class of integron in the *Vibrio cholerae* genome, Science 280 (1998) 605–608.
[20] Hood D.W., Deadman M.E., Jennings M.P., Bisercic M., Fleischmann R.D., Venter J.C., Moxon R., DNA repeats identify novel virulence genes in *Haemophilus influenzae,* Proc. Natl. Acad. Sci. USA 93 (1996) 11121–11125.
[21] Saunders N.J., Peden J.F., Hood D.W., Moxon E.R., Simple sequence repeats in the *Helicobacter pylori* genome, Mol. Microbiol. 27 (1998) 1091–1098.

[22] Peterson S.N., Bailey C.C., Jensen J.S., Borre M.B., King E.S., Bott K.F., Hutchisson I.I.I., Ca., Characterisation of repetitive DNA in the *Mycoplasma genitalium* genome: possible role in the generation of antigenic variation, Proc. Natl. Acad. Sci. USA 92 (1995) 11829–11833.
[23] Fraser C.M., Casjens S., Huang W.M., Sutton G.S., Clayton R., Lathigra R., White O. et al., Genomic seqience of a Lyme disease spirochaete, *Borrelia burgdorferi* Nature 390 (1997) 580–586.
[24] Peterson S.N., Hu P.C., Bott K.F., Hutchisson C.A., A survey of the Mycoplasma genitalium genome by using random sequencing, J. Bacteriol. 175 (1993) 7918–7930.
[25] Razin S., Yogev D., Naot Y., Molecular biology and pathogenicity of Mycoplasmas, Microbiol. Mol. Biol. Rev. 62 (1998) 1094–1165.
[26] Himmelreich R., Plagens H., Hilbert H., Reiner B., Herrmann R., Comparative analysis of the genomes of the bacteria *Mycoplasma pneumoniae* and *Mycoplasma genitalium,* Nucleic Acids Res. 25 (1997) 701–712.
[27] Stephens R.S., Kalman S., Lammel C., Fan J., Marathe R., Aravind L., Mitchell W., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., Davis R.W., Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis* Science 282 (1998) 754–759.
[28] Andersson S.G.E., Zomorodipour A., Andersson J.O., Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Näslund A.K., Eriksson A.S., Winkler H.H., Kurland C.G., The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria Nature 396 (1998) 133–143.
[29] Thomas N.S., Lusher M., Storey C.C., Clacke I.N., Plasmid diversity in Clamydia Microbiology 143 (1997) 1847–1854.
[30] McInerney J.O., Replicational and transcriptional selection on codon usage in *Borrelia burgdorferi,* Proc. Natl. Acad. Sci. USA 95 (1998) 10698–10703.
[31] Rocha E.P.C., Danchin A., Viari A., Universal replication bias in bacteria, Mol. Microbiol. 32 (1999) 11–16.
[32] Coissac E., Maillier E., Netter P., A comparative study of duplications in bacteria and eukaryotes: the importance of telomeres, Mol. Biol. Evol. 14 (1997) 1062–1074.
[33] Graham J.B., Istock C.A., Genetic exchange in *Bacillus subtilis* in soil, Mol. Gen. Genet. 166 (1978) 287–290.
[34] Lorenz M.G., Wackernagel W., Bacterial gene transfer by natural genetic transformation in the environment, Microbiol. Rev. 58 (1994) 563–602.
[35] Dubnau D.I.N Sonenshein A.L., Hoch J.A., Losick R., (Eds.), *Bacillus subtilis* and other Gram-positive bacteria, American Society for Microbiology, Washington DC, 1993, pp. 555–584.
[36] Dubnau D., Cirigliano C., Fate of transforming deoxyrribonucleic acid after uptake by competent *Bacillus subtilis*: size and distribution of the integrated donor sequences, J. Bacteriol. 111 (1972) 488–494.
[37] Fornilli S.L., Fox M.S., Electron microscope visualisation of the products of *Bacillus subtilis* transformation, J. Mol. Biol. 113 (1977) 181–191.
[38] Canosi U., Iglesias A., Trautner T.A., Plasmid transformation in *Bacillus subtilis*: DNA in plasmid pC194, Mol. Gen. Genet. 181 (1981) 434–440.
[39] Christie P.J., Korman R.Z., Zahler S.A., Adsit J.C., Dunny G.M., Two conjugation systems associated with *Streptococcus faecalis* plasmid pCF10: identification of a conjugative transposon that transfers between *S. faecalis* and *Bacillus subtilis,* J. Bacteriol. 169 (1987) 2529–2536.
[40] Lovett C.M., Love P.E., Yasbin R.E., Competence-specific induction of the *B. subtilis* RecA protein analog: evidence for dual regulation of a recombination protein, J. Bacteriol. 171 (1989) 2318–2322.
[41] Mazza G., Galizzi A., Revised genetics of DNA metabolism in *Bacillus subtilis,* Microbiologica 12 (1989) 157–179.
[42] Watt V.M., Ingles C.J., Urdea M.S., Rutter W.J., Homology requirements for recombination in *E. coli,* Proc. Natl. Acad. Sci. USA 82 (1985) 4768–4772.
[43] Roberts M.S., Cohan F.M., The effect of DNA sequence divergence on sexual isolation in *Bacillus*, Genetics 134 (1993) 401–408.
[44] Moszer I., The complete sequence of *Bacillus subtilis*: from sequence annotation to data management and analysis, FEBS Lett. 430 (1998) 28–36.
[45] Syvanen M., Horizontal gene transfer: evidence and possible consequences, Annu. Rev. Genet. 28 (1994) 237–261.
[46] Bergthorsson U., Ochman H., Distribution of chromosome length variation in natural isolates of *Escherichia coli*, Mol. Biol. Evol. 15 (1998) 6–16.
[47] Leonard C., Chen Y., Mahillon J., Diversity and differential distribution of IS231, IS232 and IS240 among *Bacillus cereus*, *Bacillus thuringiensis* and *Bacillus mycoides*, Microbiology 143 (1997) 2537–2547.
[48] Lobry J.R., Asymetric substitution patterns in the two DNA strands of bacteria, Mol. Biol. Evol. 13 (1996) 660–665.
[49] Itaya M., Physical map of the *Bacillus subtilis* 166 genome, Microbiology 143 (1997) 3723–3732.
[50] Itaya M., Tanaka T., Experimental surgery to create subgenomes of *Bacillus subtilis* 168, Proc. Natl. Acad. Sci. USA 94 (1997) 5378–5382.
[51] Itaya M., Tanaka T., Fate of unstable *Bacillus subtilis* subgenome: re-integration and amplification in the main genome, FEBS Lett. 448 (1999) 235–238.
[52] Colot V., Rossignol J.L., Eukaryotic DNA methylation as an evolutionary device, Bioessays 21 (1999) 402–411.
[53] Carlson C.R., Kolsto A.B., A small *Bacillus cereus* chromosome corresponds to one conserved region of a larger *Bacillus cereus* chromosome, Mol. Microbiol. 13 (1994) 161–169.