Molecular Microbiology (2003) **49**(4), 895–903 doi:10.1046/j.1365-2958.2003.03606.x

A strand-specific model for chromosome segregation in bacteria

Eduardo P. C. Rocha,[1,2] Joe Fralick,[3] Govindsamy Vediyappan,[3] Antoine Danchin[1,4] and Vic Norris[5]*

[1]*Unité Génétique des Génomes Bactériens, Institut Pasteur, 28 rue du Dr. Roux, 75724 Paris Cedex 15, France.*

[2]*Atelier de Bioinformatique, Université Pierre et Marie Curie, 12 Rue Cuvier, 75005 Paris, France.*

[3]*Health Science Center, Texas Tech University, Lubbock, TX, USA.*

[4]*HKU-Pasteur Research Centre, Dexter HC Man Building, 8, Sassoon Road, Pokfulam, Hong Kong.*

[5]*Laboratoire des Processus Intégratifs Cellulaires, UMR CNRS 6037, Faculté des Sciences and Techniques, Université de Rouen, 76821, Mont-Saint-Aignan, France.*

Summary

Chromosome separation and segregation must be executed within a bacterial cell in which the membrane and cytoplasm are highly structured. Here, we develop a *strand-specific* model based on each of the future daughter chromosomes being associated with a different set of structures or hyperstructures in an asymmetric cell. The essence of the segregation mechanism is that the genes on the same strand in the parental cell that are expressed together in a hyperstructure continue to be expressed together and segregate together in the daughter cell. The model therefore requires an asymmetric distribution of *classes* of genes and of binding sites and other structures on the strands of the parental chromosome. We show that the model is consistent with the asymmetric distribution of highly expressed genes and of stress response genes in *Escherichia coli* and *Bacillus subtilis*. The model offers a framework for interpreting data from genomics.

Introduction

Chromosome partitioning in bacteria comprises both the separation and segregation of chromosomes and is suspected to involve several mechanisms. The coupled transcription, translation and insertion – *transertion* – of nascent proteins into and through membrane is implicated in the separation of chromosomes which may depend on the existence of large transertion structures that couple genes to the membrane and that cannot diffuse past one another (Woldringh, 2002). It also depends on the folding of DNA and the involvement of numerous DNA-binding proteins such as FIS, HU, IciA, IHF, Dps and H-NS (Azam *et al*., 2000). The ongoing segregation of newly replicated origins of replication as they exit from a single replication factory (Dingman, 1974; Lemon and Grossman, 1998; Guzman *et al*., 2002) may be achieved by RNA polymerases transcribing and displacing the chromosomes in different directions (Dworkin and Losick, 2002). Segregation may also involve an overlay of more or less dedicated proteins, candidates for which include SMC (structural maintenance of the chromosome) proteins and related proteins such as MukB (Ohsumi *et al*., 2001) and FtsK (Margolin, 2000; Perals *et al*., 2001), as well as DNA gyrase (Duguet, 1997) and the sporulation-specific RacA protein (Ben-Yehuda *et al*., 2003). Here, we present a unifying model for chromosome partitioning that is compatible with these mechanisms and that relies on three postulates: first, a bacterium has a high degree of structural differentiation; second, this differentiation extends to the future daughter cells; and third, this differentiation results in part from the distribution of sequences on the chromosome.

Bacteria are asymmetric and it is therefore interesting to explore the significance of chromosome separation and segregation in an asymmetric context. This asymmetry exists at several levels and depends in part on the distribution of extended structures containing many macromolecules of different sorts (including genes, mRNA, proteins, lipids and ions) that are responsible for performing a particular function and that we have termed *hyperstructures* (Norris *et al*., 1999). Examples in *Escherichia coli* include the cluster of chemotaxis proteins at one pole (Maddock, 1993); a DNA compaction hyperstructure that may be related to the foci formed by the MukB protein (Ohsumi *et al*., 2001); a nucleolus-like hyperstructure for ribosome synthesis (Woldringh and Nanninga, 1985; Lewis *et al*., 2000); cell cycle hyperstructures responsible for the replication of the chromosome (Guzman *et al*., 2002; Norris *et al*., 2002) or for the sequestration of newly replicated origins of replication by the foci-forming protein SeqA (Onogi *et al*., 1999; Norris *et al*., 2000).

Structural and compositional asymmetry is still more

Accepted 1 May, 2003. *For correspondence. E-mail vjn@univ-rouen.fr; Tel. (+33) 235 14 69 08; Fax (+33) 235 14 70 20.

evident during cell division in differentiating bacteria such as *Bacillus subtilis* and *Caulobacter crescentus* where the results are a spore plus mother cell and a swarmer plus stalk cell respectively. In single cells of *Deinococcus radiodurans* there are chromosomes with different structures appropriate to different environmental conditions (Levin-Zaidman *et al*., 2003). In fact, such generation of phenotypically different individuals may be a widespread strategy in bacterial populations, even in those not considered to undergo significant differentiation (Tolker-Nielsen *et al*., 1998; Elowitz *et al*., 2002). One of the ways it can arise is via positive feedback circuits. It has been argued that positive feedback results in one daughter chromosome tending to have one pattern of gene expression and the other daughter chromosome having a different pattern (Norris and Madsen, 1995). The idea is that the intracellular environment is a competitive place where access to RNA polymerases, ribosomes and to the phospholipid membrane is limited (see for example Shepherd *et al*., 2001). In this environment, cooperation and the *status quo* effect (or 'first come, first served' effect) tend to result in one copy of a gene being expressed and the other copy not being expressed. Cooperation between genes may take the form of (i) sharing a binding site for a protein, such that sites and proteins are co-localised via the phenomenon of local concentration (Müller-Hill, 1998) or (ii) encoding RNA or proteins with specific affinities for other copy cellular constituents (including other macromolecules or lipids or ions) such that co-localisation of these constituents is mutually reinforcing. The *status quo* effect means that when the expression of a group of genes begins in one chromosome the chance of the identical genes on the other chromosome being expressed is decreased (Norris and Madsen, 1995). In other words, genes that are already expressed stand a greater chance of continuing to be expressed than their silent rivals.

Codon preferences and other analyses of the bacterial chromosome suggest that the orientation and distribution of genes on the strands are also factors in the creation of intracellular heterogeneity (Danchin *et al*., 1999; Rocha *et al*., 1999). To put it differently, the strand asymmetry in the distribution of genes and other sequences helps determine the location of hyperstructures within the cell. Here, we propose a model for partitioning based on both asymmetry in the distribution of sequences in the chromosome and asymmetry in the distribution of hyperstructures within the cell (we stress that these hyperstructures generally have primary functions that are unrelated to partitioning). The model provides a unifying framework for interpreting the plethora of results from genomics.

A strand-specific model for chromosome partitioning

In our model, before chromosome replication begins, there are four classes of genes each physically associated with a particular hyperstructure or set of hyperstructures (Fig. 1A). These four classes differ in terms of either replichore or chirochore. Different chirochores are defined by the asymmetrical nature of DNA replication. This asymmetry leads to the continuous replication of a strand (the leading strand) and the semicontinuous replication of the other (the lagging strand) by means of the Okasaki fragments. Because replication in bacteria is bi-directional from the origin to the terminus of replication, the chromosome is also divided into two replichores. By convention we shall call them the left replichore and the right replichore. Therefore genes can be classed into either lead (D) or lag (G), depending on whether they are on the leading or lagging strand, and into either left (L) or right (R), depending on their positions in the genome with respect to the origin of replication. LG, 'left' of the origin of replication and RD, 'right' of the origin, are both encoded by the same strand and are transcribed in the 'clockwise' direction (this strand is coloured blue in Figs 1–4); LD and RG, to the left and right of the origin, respectively, are transcribed in the 'anticlockwise' direction (this strand is

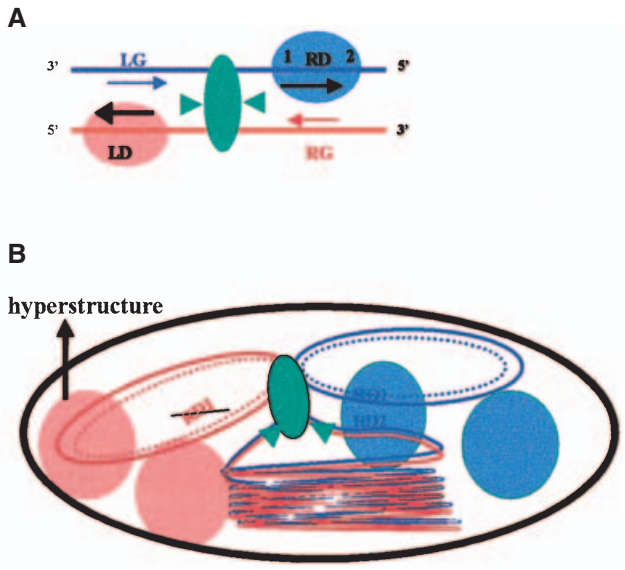

Fig. 1. Chromosome segregation driven by DNA replication, transcription and translation results in each of the original blue and red (or Watson and Crick) strands going into a different daughter cell. The translucid, blue or red filled ovals are hyperstructures.
A. Distribution of the four classes of genes R and L, right and left of origin, and D, leaDing strand, and G, laGging strand. The single giant replication factory (filled green oval; green arrowheads indicate movement of DNA through the factory) is about to replicate sets of genes that are transcribed in the directions indicated by arrows.
B. Distribution of hyperstructures and daughter chromosomes. Genes on the same strand influence one another's expression. Shortly after replication, RD1 genes are expressed from the original blue strand in the blue daughter chromosome while RD2 genes continue to be expressed from the original strand (dotted lines are newly replicated strands). RD1 genes are not expressed from the newly replicated red strand in the red daughter chromosome.

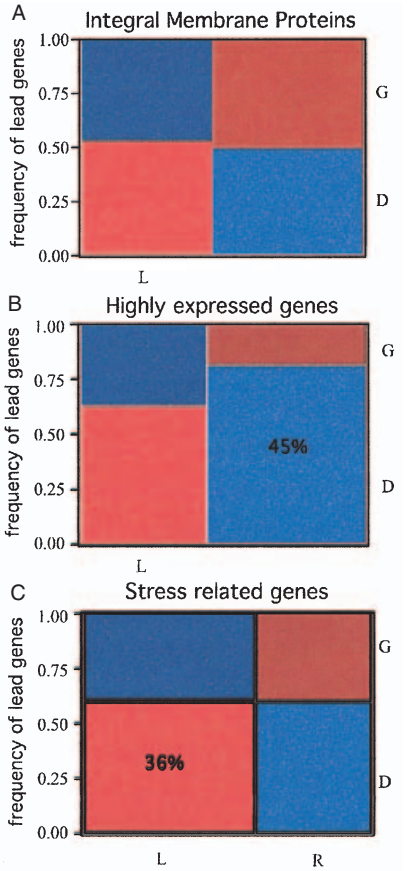

Fig. 2. Analysis of the distribution of genes in the four *Escherichia coli* classes. Classes are defined according to the side of the origin of replication (L or R) and the replication strand (D-leaDing, G-laGging). The colour (red/blue) of each square corresponds to the class. The lighter colours correspond to leading strand genes, the darker colours to lagging strand genes. The area of each square is proportional to the relative frequency of the genes in the class. When statistically significant, the most abundant class is labelled with the relative frequency. Three sets of genes are considered: (A) genes encoding integral membrane proteins, (B) highly expressed genes and (C) genes encoding stress-related proteins.

coloured red). Chromosome replication of, for example, the RD genes means that some of these genes on the as-yet-unreplicated parental strand – RD2 – are still being transcribed in a hyperstructure while neighbouring RD1 genes have already been replicated and are therefore present in two copies (Fig. 1B). In a system in which the expression of genes diverges spontaneously, the question is which of these two sets of newly replicated copies of RD1 genes will be expressed first? We propose that it is the copy on the original (conserved) strand that is expressed. There are several reasons for this:

(i) Before replication, there is only one copy of RD1 in the cell to be expressed and this is on the blue strand. If replication does not disrupt its transcription completely then this same copy of RD1 should continue to be expressed after replication. In other words, it benefits from a *status quo* effect.

(ii) Before replication, the single copies of RD1 and RD2 are adjacent on the blue strand and are both expressed. After replication of RD1 but before replication of RD2, the physical link between RD1 and RD2 (both are on the blue strand) means that the continued expression of the as-yet-to-be-replicated RD2 favours expression from the physically close RD1. Put differently, genes that are linked physically by being on the same strand have a greater chance of staying together and being expressed together in

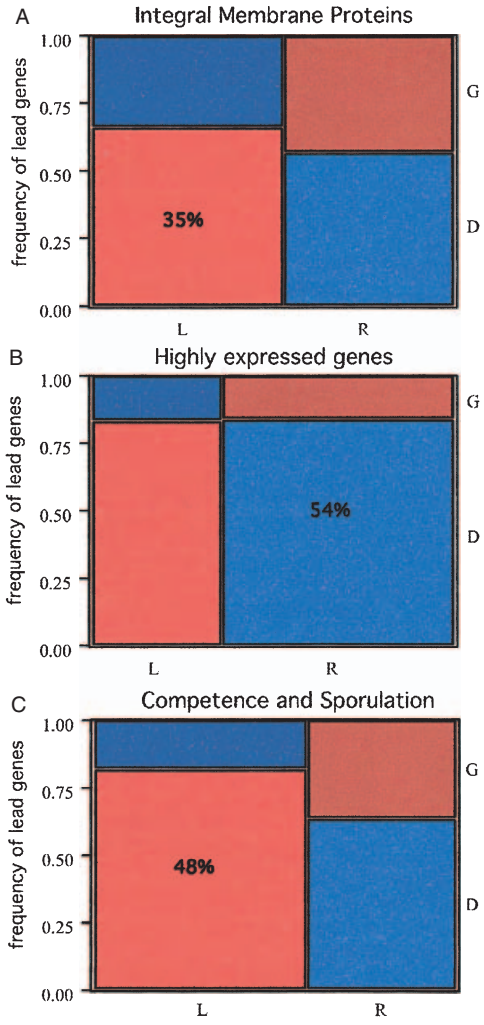

Fig. 3. Analysis of the distribution of genes in the four *Bacillus subtilis* classes. Classes were defined according to the side (L or R) and replicating strand (D-leaDing, G-laGging). The colour (red/blue) of each square corresponds to the class. The lighter colours correspond to leading strand genes, the darker colours to lagging strand genes. The area of each square is proportional to the relative frequency of the genes in the class. When statistically significant, the most abundant class is labelled with the relative frequency. Three sets of genes are considered: (A) genes encoding integral membrane proteins, (B) highly expressed genes and (C) sporulation and competence genes.

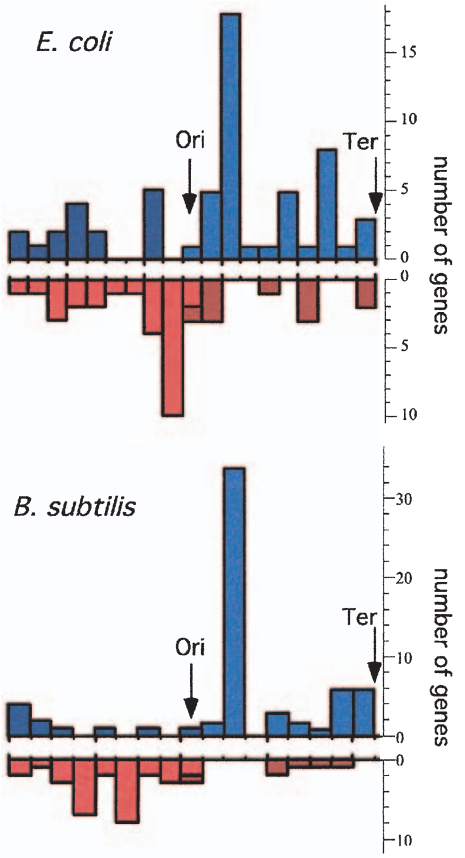

Fig. 4. Distribution of highly expressed genes along the genomes of *E. coli* (upper) and *B. subtilis* (bottom), relative to the four classes. Bars above the *X* axis correspond to one DNA strand, bars below the *X* axis to the other DNA strand. The position of the origin of replication, Ori, separates the replichores. The colour (red/blue) of each square corresponds to the class as in the other figures. The lighter colours correspond to leading strand genes, the darker colours to lagging strand genes. The genomes were re-centred so that the first position corresponds to the terminus of replication in both cases and the origin to approximately the middle of the graph.

the same hyperstructure where the appropriate transcription factors are concentrated. In other words, RD1 benefits from a next-door-neighbour effect. (This is obvious when the genes are in the same operon but we are proposing a more general idea.)

(iii) DNA modifications that distinguish between the newly replicated daughter chromosomes may favour continued expression from the original strand. For example, RD sequences on the blue strand that are important in transcription may be methylated before and after replication while the copy on the red strand has to be methylated after replication. This may create a window in which expression from the blue copy can continue while expression from the red copy cannot start.

Now consider a bacterium in which different hyperstructures occupy different parts of the cell. The RD genes (on the blue strand) might code for a set of hyperstructures that lies in one half of the bacterium and the LD or RG genes (on the red strand) code for another set that lies in the other half. The continued existence of these hyperstructures during chromosome replication means that they can steer the daughter chromosomes towards them via the pre-existing expression pattern of the parental strands. For example, after replication of RD1, the continued expression of RD2 in the hyperstructure in one half of the cell favours expression and localization of the blue strand copy of RD1 in the blue half of the cell (Fig. 1B). Reciprocally, after replication of LD1, the continued expression of LD2 in the red half of the cell favours expression and localization of the red strand copy of LD1 in the red half.

Predictions

The main prediction of the model is that there are clear differences between the distribution of genes in some of the four classes LG, RD, LD and RG. For example, there may be a difference between LG/RD and LD/RG on the basis of strands or a difference between RD and all the others such that the particular makeup of one class suffices to ensure partitioning. These differences may take the following molecular forms:

(i) Membrane proteins. Transertion structures are likely to be important both to achieve separation and to maintain it (Woldringh, 2002). One of the classes may contain more of genes encoding membrane proteins than the other classes. Alternatively, there may be differences in the phospholipid preferences of these proteins such that those encoded by RD genes have an affinity for lipids with one acyl chain length while those encoded by LD have a different affinity (for example Weiner *et al.*, 1984; Ksenzenko and Brusilow, 1993; Arechaga *et al.*, 2000 and for other references see Sueoka, 1998).

(ii) Abundant proteins. Genes that are highly expressed are more likely to be important in separation than those that are rarely expressed. A simple prediction is that the discrimination is made just on the basis of highly expressed genes such that these genes are located on either one strand (e.g. RD+LG) or indeed in just one class. Because the initial segregation of the origins of replication is likely to be important in steering the separation of the rest of the chromosomes, an asymmetric distribution of abundant genes in terms of classes should be accompanied by the location of these highly expressed genes near the origin.

(iii) Exponential versus stationary phase or sporulation. Consider that one of the halves of the cell is destined to become an exponential phase cell that can grow

rapidly and the other a stationary phase cell that can resist stresses. An interesting prediction is that the genes encoding the exponential phase proteins are in one class (RD) while those encoding the abundant stress proteins are in another (LD or RG).

(iv) Chromosome inversions. Inversions of the chromosome that exchange, for example, RD genes on the blue strand with RG genes on the red strand should perturb partitioning.

(v) Chromosome partitioning. Perturbing hyperstructure formation by inhibiting transcription and translation should lead to an abnormal partitioning of the nucleoids.

Analysis of gene distribution

Here, we focus on the distribution and orientation of genes with respect to the origin of replication in the leading Gram-negative and Gram-positive model organisms, *E. coli* and *B. subtilis*. The tables with these analyses can be found at http://wwwabi.snv.jussieu.fr/people/erocha/segregation/.

E. coli

Genes encoding membrane proteins (Fig. 2A). Integral membrane proteins (IMP) were identified from the complete *E. coli* K-12 genome (Blattner, 1997) as those with a hydrophobicity score greater than 0.4 on the Kyte and Doolittle scale (Kyte and Doolittle, 1982; Boyd *et al.*, 1998). This corresponds to 613 proteins among the 4286 of the complete genome. We found no significant difference in the distribution of the genes coding for these proteins between the four classes ($P = 0.47$, Pearson's test). However, this class is very large and heterogeneous and this does not exclude the possibility that an important role is played by particular membrane complexes, such as the chemotaxis apparatus or ATP synthetase (which is encoded by genes in a single operon), or by those with pronounced lipid affinities (a characteristic that is largely unknown). Given the possible importance of the degree of expression, we analysed the 40 most highly expressed genes encoding IMP. More of these are located in the RD class (14 of 40) than the other classes but the number is too low for the result to be considered significant. Nevertheless, it may be worth noting that these genes include those encoding the subunits of ATP synthetase where 10 of its 11 genes are in RD. Peripheral membrane proteins and secreted proteins, which are not IMP, are not considered in this analysis. For example, the 40 genes annotated as involved in flagella biosynthesis are mostly clustered in two regions in the chromosome and as a result their distribution is very biased ($P < 0.025$), with 40% of genes in RG.

Highly expressed genes (Fig. 2B). We used the data of Wei *et al.* (2001) to examine the distribution of 97 of the most highly expressed genes. Even though the prediction is that one strand or one class will contain a disproportionate number of highly expressed genes it is nonetheless surprising that 45% of these lie in just one class, the RD (leading strand) class (expected 25%, $P < 0.05$) (Fig. 2C). Moreover, these highly expressed genes are asymmetrically located close to the origin of replication where they can play an early role in partitioning (Fig. 4). A similar distribution is found for rDNA genes since five of the seven rDNA operons are in class RD. Given that many of the previous genes encode proteins involved in exponential phase growth, one interpretation is that one of the daughter chromosomes is steering its corresponding future daughter cell towards exponential growth while the other chromosome is steering its future daughter towards another growth pattern (e.g. resistance to hazardous conditions).

Genes encoding responses to stresses (Fig. 2C). We wondered whether this other growth pattern might be resistance to stresses and we therefore examined the distribution of genes encoding stress proteins. We used a dataset of 296 genes that had previously been compiled on *E. coli* stress response genes (Rocha *et al.*, 2002a, b). Analysis of this set shows that the LD class contains 36% of these genes ($P < 0.001$, χ^2 test) and that these are therefore not on the strand that has the class in which the highly expressed genes are overrepresented (Fig. 2C).

B. subtilis

Genes encoding integral membrane proteins (Fig. 3A). These were identified as for *E. coli*. The analysis of the distribution of these genes shows some asymmetry in their distribution ($P < 0.001$). However, part of this asymmetry is caused by the much larger fraction of genes in the leading strand of *B. subtilis* (75%) than in *E. coli* (55%) (Rocha, 2002). Thus *B. subtilis*, like *E. coli*, does not seem to have a strong bias in the distribution of integral membrane proteins.

Highly expressed genes (Fig. 3B). These could not be defined as in *E. coli* because of the lack of equivalent published work on transcriptome data and were therefore identified using codon usage bias and the CAI index (Sharp and Li, 1987). We used the 100 genes with highest CAI values to build the set of highly expressed genes. Here the finding is similar to that for *E. coli* and one of the classes, RD, contains 54% of the highly expressed genes ($P < 0.001$). As in *E. coli*, highly expressed genes are close to the origin of replication and, again as in *E. coli*, in a highly asymmetric way (Fig. 4) (Rocha *et al.*, 2000).

A similar distribution is found for rDNA genes in *B. subtilis* since nine of the 10 rDNA operons are in class RD.

Genes encoding responses to stresses (Fig. 3C). *B. subtilis* has two important responses to stresses, namely sporulation and competence. We thus analysed the distribution of the genes of these two responses regarding the four positional classes. Using the Subtilist database (Moszer *et al.*, 1995), we identified 164 sporulation-related genes and 25 competence-related genes. The results for both categories are qualitatively similar so we pooled them. As in *E. coli*, these *B. subtilis* stress response genes show a distribution opposite to the ones of highly expressed genes, with 48% of genes in the LD strand ($P < 0.001$), thus consistent with the hypothesis of differential adaptability of the two daughter cells (Fig. 3C).

Discussion

It is clear that bacteria have evolved a number of sophisticated molecular mechanisms for partitioning their plasmids and chromosomes. It is also clear that bacteria possess a high degree of internal structure which must be affected by chromosome partitioning. The question we have addressed here is the reciprocal one of whether structures associated directly with the strands of the parental chromosome might drive partitioning. We have therefore developed a *strand-specific* model based on each of the strands of the chromosome being associated with a different set of structures or hyperstructures in an essentially asymmetric cell. The two parental strands continue to be attached to their respective hyperstructures during replication because neighbouring genes on the same strand are physically associated in the hyperstructure. During replication and partitioning, the pattern of hyperstructures associated with the parental strand therefore determines the pattern of hyperstructures associated with the daughter chromosome and the daughter cell. The model requires an asymmetric distribution of *classes* of genes and of binding sites and other structures on the strands of the parental chromosome and we have therefore looked at the distribution of classes of genes in two model organisms, *E. coli* and *B. subtilis*, to test the predictions.

The relatively uniform distribution of genes encoding inner membrane proteins does not support the idea that transertion alone might be important in a strand-specific segregation mechanism although the seductive possibility remains of the importance of a few, specific, highly expressed genes near the origin encoding proteins with lipid preferences (such as the *atp* genes). The distribution of the highly expressed genes is, however, consistent with our hypothesis as most are grouped in a single class on one leading strand, the RD class. Moreover, most genes in the RD class are located near the origin where they may be expected to have the greatest effect on segregation (Dworkin, 2002). If the pattern of expression of one daughter chromosome confers a coherent phenotype appropriate for one sort of environment – for example, exponential growth – then the pattern of expression of the other daughter chromosome should also be coherent and appropriate for another environment; in our hypothesis, this should be reflected in the strand distribution of different sets of genes. The genes in the RD class do indeed encode products required for a coherent phenotype – the one required for exponential growth. We therefore searched for genes on the other strand that might confer a complementary phenotype, for example, one required for survival in adverse conditions. It may therefore be significant that genes encoding resistance to stresses are grouped on the other leading strand, the LD class.

One important prediction of our model is that mutants in which chromosome inversions transfer large numbers of highly expressed genes from the RD class near the origin to a class on the other strand will be defective in chromosome segregation. Inversions in *E. coli* that strongly change the symmetry of the chromosome or that shift genes between replicating strands have been found to be severely detrimental (Louarn *et al.*, 1985; Hill and Gray, 1988). This may be caused by different factors, such as alterations in gene dosage or increases in head-on collisions between DNA and RNA polymerases. Such inversions also have an important impact on segregation. After replication, the origins of replication in *E. coli* and *B. subtilis* are rapidly partitioned whereas the regions near the termini remain together at the cell centre (Niki *et al.*, 2000; Li *et al.*, 2002; Nanninga *et al.*, 2002). Inversions in *E. coli* that place the origin near the terminus or replication lead to segregation problems and production of anucleated cells (Niki and Yamaichi, 2000). Such problems seem to increase at faster growth rates. In sporulating *B. subtilis* cells, the genes to the left of the origin (i.e. LD) tend to be present in the pre-spore more often than the genes to the right (Wu and Errington, 2002). Moreover, the conservation of the chromosomal structure at this region is important for correct sporulation and rearrangements near the origin that change its distance from the LD genes lead to lower growth and sporulation rates. This is in agreement with our observation that sporulation genes are preferentially in the LD class. (Note that these genes are not within the class of highly expressed genes, so gene dosage does not explain these results.) Thus, although we have not extensively analysed rearrangements in genomes, the available published data is in good agreement with our model.

Another prediction of our model is that perturbing hyperstructure formation can perturb the partitioning of the

nucleoids. This is consistent with the contraction and fusion of already segregated nucleoids in cells or filaments treated with inhibitors of transcription or translation (van Helvoort *et al*., 1996; Dworkin and Losick, 2002; Woldringh, 2002) as these treatments should disrupt those hyperstructures that depend on transcription or translation for their existence. Targeted disruptions of individual hyperstructures may also perturb chromosome segregation, as argued for the SeqA hyperstructure (Norris *et al*., 2000), and such disruptions therefore afford a basis for future experiments.

There are of course other ways to explain the results presented here but these are still compatible with our hypothesis. Clustering of genes needed for exponential growth near the origin allows a gene dosage effect at high growth rates as the ratio of origin-to-terminus region increases. It has also been suggested that highly expressed genes are oriented to avoid collisions between replication and transcription complexes and the consequent production of hemicatenates (Brewer, 1988; Olavarrieta *et al*., 2002). However, the relevant variable in the first case is distance to the origin of replication and in the second it is the replicating strand. Neither of these alternatives is sufficient to explain the asymmetry between the gene composition of the two leading strands near the origin of replication. Bacterial genes tend to be grouped in operons for regulatory purposes. This may explain to a certain degree the clustering of some classes of genes that we observe. For example, genes encoding for ribosomal proteins (which are almost always highly expressed) are usually clustered in a small number of operons. This means that part of the effects we observe may not result initially from direct selective pressure for segregation. However, once hyperstructures are created they can be used for segregation.

In conclusion, the asymmetric distribution of genes reported here for *E. coli* and *B. subtilis* is compatible with a model in which the two different patterns of gene expression that characterise the strands of the parental chromosome create two different sets of hyperstructures that determine the segregation of chromosomes within an asymmetric cell. These two patterns of gene expression may reflect a compromise to generate daughter cells that can be both *efficient* so as to grow rapidly when times are good and *robust* so as to survive when times are hard (Norris and Madsen, 1995; Norris *et al*., 2002). It would not be surprising if the physically different states of DNA required for each chromosome to play its different role (Levin-Zaidman *et al*., 2003) were also to facilitate their separation (Bouligand and Norris, 2001).

Acknowledgements

We thank Conrad Woldringh, Jean-Michel Louarn and an anonymous referee for helpful comments. This work was supported in part by the BIOSUPPORT grant from the Innovation and Technology Commission of the government of Hong Kong. J. F. wishes to thank the NIH for grant A148696.

Supplementary material

The following material is available from
http://www.blackwellpublishing.com/products/journals/suppmat/mmi/mmi3606/mmi3606sm.htm and
http://wwwabi.snv.jussieu.fr/people/erocha/segregation/

Table S1. Genes of *B. subitlis,* their position and their classification regarding expression levels.

Table S2. Genes of *E. coli*, their position and their classification regarding expression levels.

References

Arechaga, I., Miroux, B., Karrasch, S., Huijbegts, R., de Kruijff, B., Runswick, M.J., and Walker, J.E. (2000) Characterisation of new intracellular membranes in *Escherichia coli* accompanying large scale overproduction of the b subunit of F_1F_0 ATP synthase. *FEBS Microbiol Lett* **482:** 215–219.

Azam, T.A., Hiraga, S., and Ishihama, A (2000). Two types of localization of the DNA-binding proteins within the *Escherichia coli* nucleoid. *Genes Cells* **5:** 613–626.

Ben-Yehuda, S., Rudner, D.Z., and Losick, R. (2003) RacA., a bacterial protein that anchors chromosomes to the cell poles. *Science* **299:** 523–526.

Blattner, F.R., Plunkett, G., III, Bloch, C.A., Perna, N.T. Burland, V. Riley, M., *et al.* (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277:** 1453–1462.

Bouligand, Y., and Norris, V. (2001). Chromosome separation and segregation in dinoflagellates and bacteria may depend on liquid crystalline states. *Biochimie* **83:** 187–192.

Boyd, D., Schierle C., and Beckwith, J. (1998). How many membrane proteins are there? *Protein Sci* **7:** 201–205.

Brewer, B.J. (1988) When polymerases collide: replication and the transcriptional organization of the *E. coli* chromosome. *Cell* **53:** 679–686.

Danchin, A., Guerdoux-Jamet, P., Moszer I., and Nitschké, P. (1999) Mapping the bacterial cell architecture in the chromosome. *Phil Trans R Soc Lond B* **354:** 1–12.

Dingman, C.W. (1974) Bidirectional chromosome replication: some topological considerations. *J Theoret Biol* **43:** 187–195.

Duguet, M. (1997) When helicase and topoisomerase meet! *J Cell Sci* **110:** 1345–1350.

Dworkin, J., and Losick, R. (2002) Does RNA polymerase help drive chromosome segregation in bacteria? *Proc Natl Acad Sci USA* **99:** 14089–14094.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. *Science* **297:** 1183–1186.

Guzman, E.C.J., Caballero, L., and Jimenez-Sanchez, A. (2002) Ribonucleoside diphosphate reductase is a component of the replication hyperstructure in *Escherichia coli. Mol Microbiol* **43:** 487–495.

van Helvoort, J.M.L.M., Kool, J., and Woldringh, C.L. (1996) Chloramphenicol causes fusion of separated nucleoids in *Escherichia coli* K-12 cells and filaments. *J Bacteriol* **178:** 4289–4293.

Hill, C.W., and Gray, J.A. (1988). Effects of chromosomal inversion on cell fitness in *Escherichia coli* K-12. *Genetics* **119:** 771–778.

Ksenzenko, S.M., and Brusilow, W.S.A. (1993) Protein–lipid interactions of the proteolipid c subunit of the *Escherichia coli* proton-translocating adenosinetriphosphatase. *Arch Biochem Biophys* **305:** 78–83.

Kyte, J., and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157:** 105–132.

Lemon, K.P., and Grossman, A.D. (1998) Localization of bacterial DNA polymerase: evidence for a factory model of replication. *Science* **282:** 1516–1519.

Levin-Zaidman, S., Englander, J., Shimoni, E., Sharma, A.K., Minton, K.W., and Minsky, A. (2003) Ringlike structure of the *Deinococcus radiodurans* genome: a key to radioresistance? *Science* **299:** 254–256.

Lewis, P.J., Thaker, S.D., and Errington, J. (2000) Compartmentalization of transcription and translation in *Bacillus subtilis*. *EMBO J* **19:** 710–718.

Li, Y., Sergueev, K., and Austin, S. (2002) The segregation of the *Escherichia coli* origin and terminus of replication. *Mol Microbiol* **46:** 985–996.

Louarn, J.M., Bouche, J.P., Legendre, F., Louarn J., and Patte, J. (1985) Characterization and properties of very large inversions of the *E. coli* chromosome along the origin-to-terminus axis. *Mol Gen Genet* **201:** 467–476.

Maddock, J.R., and Shapiro, L. (1993) Polar location of the chemoreceptor complex in the *Escherichia coli* cell. *Science* **259:** 1717–1723.

Margolin, W. (2000) Themes and variations in prokaryotic cell division. *FEMS Microbiol Rev* **24:** 531–548.

Moszer, I., Glaser, P., and Danchin, A. (1995) SubtiList: a relational database for the *Bacillus subtilis* genome. *Microbiol* **141:** 261–268.

Müller-Hill, B. (1998) The function of auxiliary operators. *Mol Microbiol* **29:** 13–18.

Nanninga, N., Roos, M., and Woldringh, C.L. (2002) Models on stickiness of replicated *Escherichia coli Oric*. *Microbiol* **148:** 3327–3328.

Niki, H., Yamaichi, Y., and Hiraga, S. (2000) Dynamic organization of chromosomal DNA in *Escherichia coli*. *Genes Dev* **14:** 212–223.

Norris, V., and Madsen, M.S. (1995) Autocatalytic gene expression occurs *via* transertion and membrane domain formation and underlies differentiation in bacteria: a model. *J Mol Biol* **253:** 739–748.

Norris, V., Alexandre, S., Bouligand, Y., Cellier, D., Demarty, M., Grehan, G., *et al.* (1999) Hypothesis: hyperstructures regulate bacterial structure and the cell cycle. *Biochimie* **81:** 915–920.

Norris, V., Fralick, J., and Danchin, A. (2000) A SeqA. hyperstructure and its interactions direct the replication and sequestration of DNA. *Mol Microbiol* **37:** 696–702.

Norris, V., Demarty, M., Raine, D., Cabin-Flaman, A., and Le Sceller, L. (2002) Hypothesis: hyperstructures regulate initiation in *Escherichia coli* and other bacteria. *Biochimie* **84:** 341–347.

Ohsumi, K., Yamazoe, M., and Hiraga, S. (2001) Different localization of SeqA-bound nascent DNA clusters and MukF-MukE-MukB complex in *Escherichia coli* cells. *Mol Microbiol* **40:** 835–845.

Olavarrieta, L., Martinez-Robles, M.L., Hernandez, P., Krimer, D.B., and Schvartzman, J.B. (2002) Knotting dynamics during DNA replication. *Mol Microbiol* **46:** 699–707.

Onogi, T., Niki, H., Yamazoe, M., and Hiraga, S. (1999) The assembly and migration of SeqA-Gfp fusion in living cells of *Escherichia coli*. *Mol Microbiol* **31:** 1775–1782.

Perals, K., Capiaux, H., Vincourt, J.B., Louarn, J.M., Sherratt, D.J., and Cornet, F. (2001) Interplay between recombination, cell division and chromosome structure during chromosome dimer resolution in *Escherichia coli*. *Mol Microbiol* **39:** 904–913.

Rocha, E.P.C. (2002) Is there a role for replication fork asymmetry in the distribution of genes in bacterial genomes? *Trends Microbiol* **10:** 393–396.

Rocha, E.P., Danchin, A., and Viari, A. (1999) Universal replication biases in bacteria. *Mol Microbiol* **32:** 11–16.

Rocha, E.P., Moszer, C.I., Guerdoux-Jamet, P., Viari, A., and Danchin, A. (2000) Implication of gene distribution in the bacterial chromosome for the bacterial cell factory. *J Biotech* **78:** 209–219.

Rocha, E., Matic. I., and Taddei, V. (2002a) Over-representation of repeats in stress response genes: a strategy to increase versatility under stressful conditions? (Supplementary information). [WWW document]. URL http://wwwabi.snv.jussieu.fr/people/erocha/stress/

Rocha, E.P., Matic, C.I., and Taddei, F. (2002b) Over-representation of close repeats in stress response genes: a strategy to increase versatility under stressful conditions? *Nucleic Acids Res* **30:** 1886–1894.

Sharp, P.M., and Li, W.H. (1987) The codon Adaptation Index – a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **15:** 1281–1295.

Shepherd, N., Dennis, P., and Bremer, H. (2001) Cytoplasmic RNA polymerase in *Escherichia coli*. *J Bacteriol* **183:** 2527–2534.

Sueoka, N. (1998) Cell membrane and chromosome replication in *Bacillus subtilis*. *Prog Nucleic Acid Res Mol Biol* **59:** 35–53.

Tolker-Nielsen, T., Holmstrom, K., Boe, L., and Molin, S. (1998) Non-genetic population heterogeneity studied by *in situ* polymerase chain reaction. *Mol Microbiol* **27:** 1099–1105.

Wei, Y., Lee, J.-M., Richmond, C., Blattner, F.R., Rafalski, J.A., and La Rossal, R.A.L. (2001) High-density microarray-mediated gene expression profiling of *Escherichia coli*. *J Bacteriol* **183:** 545–556.

Weiner, J.H., Lemire, B.D., Elmes, M.L., Bradley, R.D., and Scraba, D.G. (1984) Overproduction of fumarate reductase in *Escherichia coli* induces a novel intracellular lipid-protein organelle. *J Bacteriol* **158:** 590–596.

Woldringh, C.L. (2002) The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation. *Mol Microbiol* **45:** 17–29.

Woldringh, C., and Nanninga, L. (1985) Structure of the nucleoid and cytoplasm in the intact cell. In *Molecular Cytology of Escherichia Coli.* Nanninga, N. (ed.). London: Academic Press, pp. 161–197.

Wu, L.J., and Errington, J. (2002) A large dispersed chromosomal region required for chromosome segregation in sporulating cells of *Bacillus subtilis*. *EMBO J* **21:** 4001–4011.