molecular systems biology

Metabolic modeling predicts endosymbiont genome reduction on a temporal scale

Keren Yizhak, Tamir Tuller, Balazs, Eytan Ruppin

Corresponding author: Eytan Ruppin, Tel Aviv University

Review timeline:

Submission date:	25 September 2010
Editorial Decision:	17 November 2010
Revision received:	30 December 2010
Acceptance letter:	08 February 2011
Additional correspondence:	15 February 2011
Accepted:	09 February 2011

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 17 November 2010

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the three referees whom we asked to evaluate your manuscript. As you will see from the reports below, the referees find the topic of your study of potential interest. However, they raise substantial concerns on your work, which, I am afraid to say, preclude its publication in its present form.

Ultimately, the reviewers were not yet entirely convinced that this work provided a significant advance in our understanding of the evolutionary forces that determine gene loss in endosymbionts. This issue seems to hinge heavily on concerns regarding the physiological relevance of key modeling assumptions:

1. The reviewers indicated that minimal media is unlikely to be physiologically realistic, and seemed to feel this cast doubt on the conclusions derived from the minimal media based analyses.

2. They also indicated that gene loss in endosymbionts is known to have not been gradual, an issue that the reviewers related to changes in the environmental conditions (i.e. media parameters) experienced by the organisms during different evolutionary periods.

Overall, addressing these concerns may require substantial additional analyses that look more closely at gene loss patterns under physiologically motivated conditions.

If you feel you can satisfactorily deal with these points and those listed by the referees, you may wish to submit a revised version of your manuscript. Please attach a covering letter giving details of the way in which you have handled each of the points raised by the referees. A revised manuscript will be once again subject to review and you probably understand that we can give you no guarantee

at this stage that the eventual outcome will be favorable.

Yours sincerely,

Section Editor
Molecular Systems Biology

REFEREE REPORTS

Reviewer #1 (Remarks to the Author):

This manuscript examines an interesting question: the extent to which the structural constraints inherent in a functioning metabolic network influence the course of evolution of the network. The results are of some interest in providing evidence that this may be the case, at least to some degree. Nevertheless:

1. The title of the paper appears to make a stronger claim than is actually demonstrated, i.e. that modelling the metabolic network can '... predict ... genome reduction on a temporal scale'. It seems that the temporal resolution offered by the phylogenetic reconstruction is not very high. Furthermore, in the abstract the claim is weakened to accounting for 33% of variability in time to loss.

2. This is not the only paper to consider the network constraints on loss of metabolic genes over evolutionary time scales. Though the context and methodology is different (genome reduction after whole genome duplications), it has been considered in the case of Paramecium by Gout, Duret and Kahn, and this is worth a mention.

3. The paper is rather hard to read. This is not entirely the authors' fault. The policy of Nature journals of placing all methodology at the end may be suitable for experimental papers using standard approaches, but is ill-suited for theoretical papers where the results make limited sense without knowing what has been calculated and how. In this case, some of the Supplementary Information is also necessary to understand what has been done and how the modelling methods behave. The net outcome is that the results section is very dense. A way round this within the journal rules would be to present the validation of the methodology as a result and move the corresponding Supplementary Figures into the main text, perhaps at the expense of Fig. 4, which is not very exciting and could go into Supplementary Information.

4. As a minor point, several of the references have duplicated entries.

Reviewer #2 (Remarks to the Author):

Title: Metabolic modeling predicts endosymbiont genome reduction on a temporal scale
Authors: Yizhak et al.

This manuscript is aimed at describing the metabolic dynamics during the process of genome shrinkage in Buchnera aphidicola, the primary symbiotic bacterium of aphids. These dynamics are important as they can shed light on the parameters governing the way and extent of reduced genomes. Authors correlate the dynamics of gene-by-gene loss and the timing of these losses with the process of genome shrinkage and compare their approximation with phylogenetic ones. The main conclusions of the study are: i) metabolic in silico predictions of gene loss coincide with the ones identified through phylogenetic inferences; ii) These predictions are better than the ones estimated from comparative genomics studies; iii) despite the stochastic process of gene loss, metabolism seems to account for 33% of gene loss dynamism.
The study is sound and authors perform an exhaustive and rigorous analysis of gene loss dynamics

introducing metabolic constraints. Although I find many of the manuscripts published by some of the authors of this study to be highly insightful, I struggle to see how this study involves a significant advance. In other words, the study is correct but I cannot see in what way it unearths unexpected or meaningful patterns.
1) To start with, I cannot find the need for comparing phylogenetic inferences with metabolic constraints in the inference of the timing and dynamics of gene loss-both types of approximations are completely different and expected to be biased. Why should phylogenetic approaches be a good control to compare the significance of metabolism in inferring true events?
2) I would have found very useful simulating a hypothetical genome shrinkage process in which reactions are predefined and seeing how the metabolic constraints can predict the order in which genes have been lost
3) Simulations are performed in minimum media!! Buchnera, and very likely its ancestor, was contained in a very rich medium. How would that affect the dynamics of gene loss? I would simply expect that genes that are essential--for example, with K=1, or 2 or with a great number of PPIs--be lost more readily in rich media. Maybe I missed the point but authors should elaborate on this.
4) Was the difference in the correlation values between random media and real media significant?
5) It is known that in Buchnera aphidicola we can distinguish two main events and dynamics of gene loss, each of which presents a particular rate. For example, if we take the split between Buchnera Acyrthosiphon pisum and Schizaphis graminum we can identify two main periods of gene loss: one predating the split--that is mainly characterized by the rapid loss of genes-and another postdating the split, which was followed by a tremendous genome stasis. Could the authors not introduce this into their models to improve predictions?
6) As a minor comment, I have to say that the section "Flux coupling analysis" is rather tedious to follow. Please rewrite that section as to make it easy to understand. In fact I understood what that section was about by looking at the results.

In sum, I find the manuscript well executed and everything in it is sound, however authors have not convinced me about the importance or novelty of the results. I'm not saying the manuscript should be rejected, however, authors would need to put much more effort on making the manuscript sweeter for the readership.

Reviewer #3 (Remarks to the Author):

In this manuscript Yizhak et al. deals with the important issue of examining the role played by metabolism in the genome reduction process of Buchnera aphidicola, primary endosymbiont of aphids. The authors apply the strategy of measuring the correlation coefficient between the number of inferred genes deleted in the phylogenetic history of four Buchnera clones and the average number deleted ones that still allows a 'viable' metabolism based on a huge amount of simulations (starting with minimal media, but also considering and exploring other media and biomass functions, including random ones). The major conclusion is a robust result where, no matter the many evolutionary factors involved in the genome reduction of endosymbiotic bacteria, metabolism necessarily determines one third of the losses occurred.

Major comments
1. As also carried out by Pal et al. (2006) the in silico experiments performed by the authors starts with the E. coli genome and a minimal media. To reinforce the correlations observed between in-silico experiments and gene deletions detected during phylogenetic history of Buchnera (Fig. 1a) the authors also carried out four additional analyses: i) random media and the E. coli biomass function (Fig. 1b), ii) random media and random biomass function (Fig. 1c), iii) an intriguing simulated annealing (SA), 'aimed at searching for the environment/biomass function that maximized our target correlation between in silico and reconstructed loss times' (Supp. Fig. 3), and iv) the biological media used by Thomas et al. (2009). The major conclusion of the authors is that the correlations obtained in points i-iv are not better, in general, than those reported in Fig. 1a. The control strategy followed by the authors lacks of logic or, in better positive terms, is not clearly explained. The SA appears to be a better first approach to deal with the issue of looking for the best correlation and, probably, and additional simulation that has not been explored by the authors. Why the authors assume that a minimal media is the better starting point to perform simulation? The process of

symbiotic integration involves an important environmental function that changes completely in terms of media composition through evolution. Depending on the stage of integration, media requirements for the symbiont are different (and biomass functions will be, correspondingly, different too) and, in broad terms, the media will change from minimal to complex ones. One way to overcome this situation may be to use a media with increasing number of essential products, something that may mimics better the symbiotic integration into the eukaryotic host. Alternatively, the authors should explore not only minimal media but also reach media, as Thomas et al (2009). In pag. 7, lines 10-12, authors state that 'Interestingly, the biomass functions found in solutions obtaining high correlation values all share essential amino acids....'. Are these correlations higher than the ones reported in Fig. 1a? The major criticism to this work is that a better knowledge of the media allowing to survive to symbionts at different stages of symbiotic integrations, and properly introduced in the in-silico study, would allow an increase in the correlations obtained.

2.Considering the amount of empirical work done in the field of insect endosymbiosis during the last ten years, the introduction does not reflect at all the state of the art. Of particular interest is to gain some knowledge of the media requirements (and, concomitantly the associated biomass function) of free living bacteria, ancestors fo the endosymtions, secondary endosymbionts and primary endosymbionts with medium at extreme reduced genome. There are few reviews published by Baumann, Moran et al. and Moya et al. that should be read it a properly introduced. In addition, authors use four Buchnera genomes and, consequently, the original papers where those genomes where reported should be cited.

3. Introduction. Second paragraph. The authors have used a protocol 'for simulating the gradual loss, during evolution, of metabolic enzymes'. However, it is well known that the process of genome reduction is not gradual. As also occurs in Pal et al. (2006), the authors should tell us something on the final outcome of the metabolism not matter if genes are lost one by one or following a different distribution (ie, probably big losses at the beginning of endosymbiotic integration and gene by gene in later stages of genome reduction).

4. Results. First paragraph, pag. 4. The authors declare that 'genes are still actually lost in a consistent and coordinated fashion, reflecting the role of necessity'. This is an interesting observation that deserves further development, probably in the discussion section, absolutely very short considering the nature of the issue raised in this manuscript. What is the relationship between 'necessity' and 'selection'? What type of selection: purifying or positive? The issue very pertinent to disentangle if natural selection is playing or not a major role, through metabolism, in the genomic reduction process. In the discussion section, first paragraph, pag. 7, the authors state that 'while remaining variation is likely to be determined by other factors'. What other factors? 'History' and 'chance' are very vague references.

Minor comments

5. Introduction. First paragraph. The authors state that 'It is believed that Buchnera has evolved in an accelerated rate...'. There is ample evidence on that.

6. Introduction. Second paragraph. Wigglesworthia is introduced with no reference at all to what type of symbiont it is.

7. Results. First paragraph. Could you please say a little more on the 'minimal conditions' used in the E. coli model based on Buchnera empirical results?

1st Revision - authors' response

Please find enclosed a revised version of our manuscript, now entitled "Metabolic modeling of endosymbiont genome reduction on a temporal scaleî.

We were pleased to see that the reviewers had found interest in our manuscript and would like to

thank them for their comments, which have helped us in further improving it. Following their and your suggestions, the analysis and validation were significantly extended and various parts were rewritten. Addressing the main comments highlighted in your cover letter, we have: (1) Extended our results and tested them in an array of other possible growth media, found both by mining the literature and by searching the media space ñ reassuringly, the results found in these media, which are now reported in the main text, reinforce our previous findings. (2) We now simulated an additional two-stage gene loss process, starting from a phase of block deletions and continuing with individual gene deletions, whose results are additionally reported and discussed in the paper. The specific comments of the reviewers are addressed below, as follows.

Reviewer #1:

1. "The title of the paper appears to make a stronger claim than is actually demonstrated, i.e. that modelling the metabolic network can '... predict ... genome reduction on a temporal scale'. It seems that the temporal resolution offered by the phylogenetic reconstruction is not very high. Furthermore, in the abstract the claim is weakened to accounting for 33% of variability in time to loss."

We thank the reviewer for his comment. Accordingly, we have changed the title of the manuscript, as above.

2. "This is not the only paper to consider the network constraints on loss of metabolic genes over evolutionary time scales. Though the context and methodology is different (genome reduction after whole genome duplications), it has been considered in the case of Paramecium by Gout, Duret and Kahn, and this is worth a mention."

Thanks for drawing our attention to this paper. We have now added a pertaining reference in the Introduction (page 3, paragraph 2).

3. "The paper is rather hard to read. This is not entirely the authors' fault. The policy of Nature journals of placing all methodology at the end may be suitable for experimental papers using standard approaches, but is ill-suited for theoretical papers where the results make limited sense without knowing what has been calculated and how. In this case, some of the Supplementary Information is also necessary to understand what has been done and how the modelling methods behave. The net outcome is that the results section is very dense. A way round this within the journal rules would be to present the validation of the methodology as a result and move the corresponding Supplementary Figures into the main text, perhaps at the expense of Fig. 4, which is not very exciting and could go into Supplementary Information."

Thanks. Following the reviewer's advice, we have now added a paragraph at the beginning of the Results section providing an overview of what was done from a methodological standpoint (page 4, paragraph 2). We additionally moved Supp. Figure 3 to the main paper (now named Figure 3) and moved Figure 4 to the Supp. Material.

4. "As a minor point, several of the references have duplicated entries."

Thanks and we apologize for this error, which has now been remedied.

Reviewer #2:

1. "To start with, I cannot find the need for comparing phylogenetic inferences with metabolic constraints in the inference of the timing and dynamics of gene loss-both types of approximations are completely different and expected to be biased. Why should phylogenetic approaches be a good control to compare the significance of metabolism in inferring true events? "

While phylogenetic reconstructions approaches are not flawless or free of any biases, they are still considered one of the mainstays of the modern study of evolution, to which tremendous efforts have been devoted in recent years in computational biology. These data-driven approaches aim to track down the gene loss order as a combined end result of the rich multitude of evolutionary forces taking part in determining gene loss. As we were primarily interested here in determining the relative contribution of metabolic constraints per se in relation to these more general selection forces, the comparison between the metabolic, model-driven estimation to this general data-driven one is informative and called for. Following the refereeís question this rationale is further clarified now in brief in the main text (page 5, paragraph 2).

2. *"I would have found very useful simulating a hypothetical genome shrinkage process in which reactions are predefined and seeing how the metabolic constraints can predict the order in which genes have been lost"*

This is indeed the case and the goal of our paper, where we examine to what extent metabolic constraints can predict the order of gene loss. The simulations have now been done in two test-scenarios by considering an evolutionary process of single-gene deletions and of block deletions. Reassuringly, both resulted with a significant correlation of 0.54 and 0.45, respectively (page 7, paragraph 1 and page 9, paragraph 2). Additionally, following the reviewerís comment we tested our metabolic simulation results against randomly generated orders of hypothetical gene loss and find that, indeed, the correlations obtained by the model predictions with these random loss orders are negative and hence significantly and markedly lower than those obtained with the phylogenetically reconstructed order (Supp. Table 1.e).

3. *"Simulations are performed in minimum media!! Buchnera, and very likely its ancestor, was contained in a very rich medium. How would that affect the dynamics of gene loss? I would simply expect that genes that are essential--for example, with K=1, or 2 or with a great number of PPIs--be lost more readily in rich media. Maybe I missed the point but authors should elaborate on this."*

Following the reviewer's comment we first repeated the process under an exhaustively rich medium (where each of the 299 exchange reactions in the ancestral E. coli model can be potentially active). The correlation obtained under this medium was fairly low (R=0.15, empirical P-value < 0.011), apparently testifying that the host media is not exhaustively rich. In the previous version of our paper we used a minimal medium following the procedure of (P·l et al, 2006), and we were reassured to see that it has given a strong signal. However, rethinking this following the refereeís comment, we agree that this important issue deserves deeper consideration and attention and we further performed two additional analyses, accordingly: (1) We first searched the literature for components known to exist in Buchnera's habitat. These components were then completed with the minimal number of metabolites that are essential for growth considering E. coli's biomass function, to form a literature based compilation that enables the growth of the Buchnera model (see Supp. Table 1.a, page 4, paragraph 2). (2) We next turned to improve our Simulated Annealing (SA) search by limiting the size of the media to several predefined magnitudes, as an unconstrained media space is simply too dauntingly large to enable an efficient search. Interestingly, we find that SA-identified media composed of 50 components achieve significantly higher correlation values compared to media of other magnitudes (hypergeometric P-value = 3.06e-6 and Methods, page 6-7, Supp. Tables 2.e and 2.f). Reassuringly, we have managed to improve the correlation to the phylogenetically reconstructed time loss values significantly in this test scenario to a mean Spearman correlation of 0.54 that is 63% of the maximal correlation with the end-point genes, and to 0.39 (44% of the maximal correlation) without the end-point genes (empirical P-value < 9.9e-4). This constitutes a significant improvement over the corresponding correlations found using the minimal medium that have been reported in the original version of the paper. We therefore modified the paper according to the refereeís suggestions such that: (a) we first establish the robustness of the evolutionary process and examine the corresponding correlations with the phylogenetic reconstruction based on the literature-based medium (page 4-6). (b) additional analyses throughout the paper (K-robustness, Flux Coupling Analysis, etc.) are now carried under the best media conditions found by the search process (pages 8-9), replacing all analyses performed using minimal medium that have been presented in the previous version.

4. "Was the difference in the correlation values between random media and real media significant?"

Thank you for this comment. Indeed, the difference in correlation between each of the random media examined and the real media (minimal or literature-based) is significant (Wilcoxon P-value < 1.7e-9). This result has now been added to the main paper (page 6, paragraph 2).

5. "It is known that in Buchnera aphidicola we can distinguish two main events and dynamics of gene loss, each of which presents a particular rate. For example, if we take the split between Buchnera Acyrthosiphon pisum and Schizaphis graminum we can identify two main periods of gene loss: one predating the split--that is mainly characterized by the rapid loss of genes-and another postdating the split, which was followed by a tremendous genome stasis. Could the authors not introduce this into their models to improve predictions?"

Thanks for this suggestion. Consequently, we have now simulated a process of block-deletions similar to that done in (P·l et al, 2006), where we remove a randomly chosen block of contiguous genes in the genome. When no further contiguous genes can be deleted, a single gene deletion process starts until no further genes can be deleted. The SA search was repeated anew under this test scenario as well, obtaining maximal correlations of 0.45 that is 53% of the maximal correlation with the end-point genes, and to 0.37 (43% of the maximal correlation) without the end-point genes (empirical P-value < 9.9e-4). These correlations are evidently smaller than those achieved via a single-phase individual gene loss process. This may be surprising at first sight, but in a second thought itís a fairly plausible outcome of the modelís restricted nature: In reality, the blocks lost in the Buchnera genome obviously contain many non-metabolic genes that play a central role in determining their evolutionary fate, while the effects of the latter remain completely out of the scope of a block loss simulation within the model. In contrast, the simulations of individual gene loss remain focused just on the metabolic genes, and hence better succeed in faithfully capturing their loss times. Therefore, we present the main analyses in the paper by considering the single-gene deletion scenario, but in response to the reviewersí comments now additionally present the double-phase simulations and discuss their limits in the main text (page 9-10 ñ and a similar approach was taken by (P·l et al, 2006).

6. "As a minor comment, I have to say that the section "Flux coupling analysis" is rather tedious to follow. Please rewrite that section as to make it easy to understand. In fact I understood what that section was about by looking at the results."

Thanks. We have rewritten this part in the Methods section such that the concept of directionally flux coupled reactions is better explained (page 14). Accordingly, we now explain that: "Flux-coupled genes were identified by applying the previously developed flux coupling algorithm (Burgard et al, 2004) on the E. coli metabolic reconstruction (Feist et al, 2007). Namely, we look for directionally coupled genes, where a non-zero flux for v1 implies a non-zero flux for v2 but not necessarily the reverse. In order to test whether genes encoding directionally coupled reactions have a tendency to be lost later, we applied a binomial test with p=0.5 testing for the significance of the observed deviation of loss times from the theoretically expected distributionî.

7. In sum, I find the manuscript well executed and everything in it is sound, however authors have not convinced me about the importance or novelty of the results

The main achievements of this paper are twofold: (a) we show for the first time that it is possible to obtain a fine-grained view of the time course of an organismís evolution towards symbiosis, and (b) obtain a quantitative estimate of the role of metabolic considerations in determining this process. Overall, these issues are quite important and basic ñ they shed an interesting light on the role of chance vs. necessity in evolution, and enabling us to obtain provide a bound on evolutionary necessity stemming from metabolic constraints in a scenario which is widely considered to be a paradigmatic case of symbiotic evolution.

8. authors would need to put much more effort on making the manuscript sweeter for the readership.

We thank the reviewer for his comment. Accordingly we have improved the manuscript by: (1) Extending the introduction with a paragraph describing the nutritional basis of the symbiosis. (2) We have added a paragraph at the beginning of the Results section that provides an overview of what was done from a methodological standpoint. (3) We additionally moved Supp. Figure 3 to the main paper (now named Figure 3), providing an overview of the Simulated Annealing search underlying much of the work presented in this paper. (4) We have extended the conclusions part of the paper with a more comprehensive discussion concerning the role of chance vs. necessity, and finally, we reorganized some parts of the paper to better fit the global changes requested by the referees (in particular, the emphasis on media search) and to allow for an overall better flow for the readers.

Reviewer #3:

1. "As also carried out by Pal et al. (2006) the in silico experiments performed by the authors starts with the E. coli genome and a minimal media. To reinforce the correlations observed between in-silico experiments and gene deletions detected during phylogenetic history of Buchnera (Fig. 1a) the authors also carried out four additional analyses: i) random media and the E. coli biomass function (Fig. 1b), ii) random media and random biomass function (Fig. 1c), iii) an intriguing simulated annealing (SA), 'aimed at searching for the environment/biomass function that maximized our target correlation between in silico and reconstructed loss times' (Supp. Fig. 3), and iv) the biological media used by Thomas et al. (2009). The major conclusion of the authors is that the correlations obtained in points i-iv are not better, in general, than those reported in Fig. 1a. The control strategy followed by the authors lacks of logic or, in better positive terms, is not clearly explained. The SA appears to be a better first approach to deal with the issue of looking for the best correlation and, probably, and additional simulation that has not been explored by the authors. Why the authors assume that a minimal media is the better starting point to perform simulation? The process of symbiotic integration involves an important environmental function that changes completely in terms of media composition through evolution. Depending on the stage of integration, media requirements for the symbiont are different (and biomass functions will be, correspondingly, different too) and, in broad terms, the media will change from minimal to complex ones. One way to overcome this situation may be to use a media with increasing number of essential products, something that may mimics better the symbiotic integration into the eukaryotic host. Alternatively, the authors should explore not only minimal media but also reach media, as Thomas et al (2009). In pag. 7, lines 10-12, authors state that 'Interestingly, the biomass functions found in solutions obtaining high correlation values all share essential amino acids....'. Are these correlations higher than the ones reported in Fig. 1a? The major criticism to this work is that a better knowledge of the media allowing to survive to symbionts at different stages of symbiotic integrations, and properly introduced in the in-silico study, would allow an increase in the correlations obtained."

Indeed, we definitely agree with the reviewer that the issue of growth media deserved careful attention. Following this and similar comments raised by referee No. 2, we have repeated the analysis in a medium that was constructed based on the information available in the literature and performed an extensive search in the space of potential media (and biomass functions) employing a size-constrained Simulated Annealing approach (please see pages 4-8 and our response to comment no. 3 of reviewer no.2). Indeed, the correlation achieved by this comprehensive SA search (R=0.54, empirical P-value < 9.9e-4) is higher than that achieved under the four media mentioned above by the reviewer. Reassuringly, as we now report in the main text, this medium is also enriched with media components derived from the literature (hyper geometric P-value = 0.03).

2. "Considering the amount of empirical work done in the field of insect endosymbiosis during the last ten years, the introduction does not reflect at all the state of the art. Of particular interest is to gain some knowledge of the media requirements (and, concomitantly the associated biomass function) of free living bacteria, ancestors fo the endosymtions, secondary endosymbionts and primary endosymbionts with medium at extreme reduced genome. There are few reviews

published by Baumann, Moran et al. and Moya et al. that should be read it a properly introduced. In addition, authors use four Buchnera genomes and, consequently, the original papers where those genomes where reported should be cited."

Thanks for drawing our attention to this issue. We extended the introduction with a paragraph describing the nutritional basis of the symbiosis, drawing upon and citing the pertaining review papers (page 3, paragraph 1). In addition, we searched the literature and reconstructed a literature-based media under which we perform further analysis (page 4-5). Moreover, the corresponding references for the four Buchnera strains were added to the main text (page 3, paragraph 2).

3. "Introduction. Second paragraph. The authors have used a protocol 'for simulating the gradual loss, during evolution, of metabolic enzymes'. However, it is well known that the process of genome reduction is not gradual. As also occurs in Pal et al. (2006), the authors should tell us something on the final outcome of the metabolism not matter if genes are lost one by one or following a different distribution (ie, probably big losses at the beginning of endosymbiotic integration and gene by gene in later stages of genome reduction)."

Thanks. Indeed, we added the information on the remaining metabolic capabilities to the results part and to the Material (page 6, first paragraph and Supp. Table 1.f). Accordingly, we now report that: "the model accurately predicts that the most preserved pathways are those involved in essential amino acid metabolism, and in central metabolism, including the Pentose Phosphate Pathway, Glycolysis etc., while genes associated with cell envelope synthesis, lipopolysaccharides synthesis and membrane lipid metabolism constitute a smaller fraction of the final networks (Supp. Table 1.f). It is reassuring to see that these predictions match the reports known from the literature, where it is known that Buchnera aphidicola lacks genes for the biosynthesis of cell-surface components, including lipopolysaccharides and phospholipids (Shigenobu et al, 2000). Furthermore, the extensive loss of transport capabilities and conservation of essential amino acids biosynthetic pathways are prime characteristics of the aphid symbiont (van Ham et al, 2003)." The non-gradual, bi-phasic nature of gene loss in Buchnera aphidicola, is now further studied and discussed in the revised version of the paper (and please see our detailed response to comment 5 of referee No 2.).

4. "Results. First paragraph, pag. 4. The authors declare that 'genes are still actually lost in a consistent and coordinated fashion, reflecting the role of necessity'. This is an interesting observation that deserves further development, probably in the discussion section, absolutely very short considering the nature of the issue raised in this manuscript. What is the relationship between 'necessity' and 'selection'? What type of selection: purifying or positive? The issue very pertinent to disentangle if natural selection is playing or not a major role, through metabolism, in the genomic reduction process. In the discussion section, first paragraph, pag. 7, the authors state that 'while remaining variation is likely to be determined by other factors'. What other factors? 'History' and 'chance' are very vague references."

Following the reviewer's comment, we've extended the discussion accordingly (page 10-11). It now reads: ì`Necessityí in our context refers to selection forces acting to conserve metabolic genes whose contribution to the symbiontís growth within the host is significant. However, the weight of 'necessity' estimated by our framework serves as a lower bound on its actual magnitude, since the true contributions of metabolic genes may go well beyond those captured in the model due to their contribution to other, non-metabolic cellular functions (as many genes may have diverse pleiotropic effects). 'Chanceí, albeit, reflects the true randomness in order of gene losses occurring in evolution that may be due to the workings of a variety of stochastic effects occurring in nature including, e.g., randomness in mutational processes driving gene loss and variations in the hostís environment etcî. In response to the referee's comment we now explicitly mention in the text that our simulations focus on the loss of genes reflecting the act of purifying selection (page 10).

5. "Introduction. First paragraph. The authors state that 'It is believed that Buchnera has evolved in an accelerated rate...'. There is ample evidence on that."

Thanks, we removed this sentence from the introduction.

6. "Introduction. Second paragraph. Wigglesworthia is introduced with no reference at all to what type of symbiont it is."

Thanks, we added a reference to that type of symbiont as well (page 3, paragraph 2).

7. "Results. First paragraph. Could you please say a little more on the 'minimal conditions' used in the E. coli model based on Buchnera empirical results?"

The minimal conditions used in the paper are the ones reported in the paper of (Feist et al, 2007) to allow growth under E. coli's biomass function and appear in Supp. Table 2.a. More specifically, it includes Calcium, Chloride, Co2+, Fe2+, D-Glucose, Glutamate, Magnesium, Molybdate-MoO4, Ammonium, Phosphate, Tungstate, Cob(I)alamin, CO2, Cu2+, Fe3+. H20, Potassium, Mn2+, Sodium, O2, Sulfate and zinc. The pertaining details are provided in Supp. Table 2.a and refereed to from the main text (page 5, paragraph 3).

References:

Burgard AP, Nikolaev EV, Schilling CH, Maranas CD (2004) Flux Coupling Analysis of Genome-Scale Metabolic Network Reconstructions. Genome Research 14: 301-312.

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO (2007) A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 3.

P·l C, Papp B, Lercher MJ, Csermely P, Oliver SG, Hurst LD (2006) Chance and necessity in the evolution of minimal metabolic networks. Nature 440: 667-670.

Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H (2000) Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. Nature 407: 81-86.

van Ham RCHJ, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U, Fern·ndez JM, JimÈnez L, Postigo M, Silva FJ, Tamames J, Viguera E, Latorre A, Valencia A, Mor·n F, Moya A (2003) Reductive genome evolution in Buchnera aphidicola. Proceedings of the National Academy of Sciences of the United States of America 100: 581-586.

Acceptance letter 08 February 2011

Thank you again for sending us your revised manuscript. The reviewers are now satisfied with the modifications made and I am pleased to inform you that your paper has been accepted, in principle, for publication.

Before we can send this work to production, there are a number of remaining format and content issues that we would like to ask you address:

1. Reviewer #1 had one remaining minor point that may require some additional clarification (see below).

2. We would like to ask you to provide, if possible, the estimated loss times for each gene, from the in silico and phylogenetic analyses, as supplementary material, to aid researcher who may want to reuse this data. Ideally, this data should be provided as an additional supplementary table (in text or excel formats), and should be referenced in the main manuscript and listed in the Table of Contents at the beginning of the Supplementary Materials file.

3. Please specify in the legend of Fig. 4 what the error bars represent (i.e. standard deviation, standard error, etc.).

Thank you very much for submitting your work to Molecular Systems Biology.

Sincerely,

Section Editor
Molecular Systems Biology

REFEREE REPORTS

Reviewer #1 (Remarks to the Author):

The changes made by the author have addressed my points, and in my view have also addressed the major points of the other reviewers. The clarity of the manuscript has been increased and the findings have been more carefully delineated and better justified.

A minor remaining point is that on p. 9, para 2, line 12, there is a reference 'nodes 1 to 3 in Figure 1a, Supplementary Material' that I found unclear. Is this intended to mean 'nodes 1 to 3 in Figure 1a and Supplementary Material 3'?

Reviewer #2 (Remarks to the Author):

Authors have addressed satisfactorily my concerns and I have not other problems with the manuscript. I recommend accepting it in the present form.

Reviewer #3 (Remarks to the Author):

This reviewer has read carefully the new manuscript, the answers to my comments as well as the answers given to the other reviewers. Based on that I consider that the new version is a nice piece of science that deserves publication in this journal.

Additional correspondence 15 February 2011

Please find attached the following things:

1. Reviewer #1 had one remaining minor point that may require some additional clarification (see below).
Following this comment, we removed the reference to the Supp. Material at that place (page 9, second paragraph) as we indeed already refer to it at the end of that paragraph.

2. We would like to ask you to provide, if possible, the estimated loss times for each gene, from the in silico and phylogenetic analyses, as supplementary material.
We attach an Excel sheet with the required data. We refer to it as Supp. Table 2j and also from the main text (page 7, first paragraph).
Please let me know if you think it should be named otherwise.

3. Please specify in the legend of Fig. 4 what the error bars represent (i.e. standard deviation, standard error, etc.).
We added the required information to the figure's legend.

Please let me know if there is anything else I can do to improve the manuscripts.

