

Article

# Testing the Domino Theory of Gene Loss in *Buchnera aphidicola*: The Relevance of Epistatic Interactions

David J. Martínez-Cano <sup>1</sup>, Gil Bor <sup>2</sup>, Andrés Moya <sup>3,4</sup>  and Luis Delaye <sup>1,\*</sup> 

<sup>1</sup> Departamento de Ingeniería Genética, CINVESTAV Irapuato, Km. 9.6 Libramiento Norte Carretera Irapuato-León, 36821 Irapuato, Guanajuato, Mexico; boost07@yahoo.com

<sup>2</sup> CIMAT, A.P. 402, Guanajuato 36000, Gto., Mexico; gil@ciamat.mx

<sup>3</sup> Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana (FISABIO)-Salud Pública, Avenida de Catalunya 21, 46020 València, Spain; andres.moya@uv.es

<sup>4</sup> Institute for Integrative Systems Biology, Universitat de València, Calle Catedrático José Beltrán 2, 46980 Paterna, València, Spain

\* Correspondence: luis.delaye@cinvestav.mx; Tel.: +52-462-623-9669

Received: 6 May 2018; Accepted: 25 May 2018; Published: 29 May 2018



**Abstract:** The domino theory of gene loss states that when some particular gene loses its function and cripples a cellular function, selection will relax in all functionally related genes, which may allow for the non-functionalization and loss of these genes. Here we study the role of epistasis in determining the pattern of gene losses in a set of genes participating in cell envelope biogenesis in the endosymbiotic bacteria *Buchnera aphidicola*. We provide statistical evidence indicating pairs of genes in *B. aphidicola* showing correlated gene loss tend to have orthologs in *Escherichia coli* known to have alleviating epistasis. In contrast, pairs of genes in *B. aphidicola* not showing correlated gene loss tend to have orthologs in *E. coli* known to have aggravating epistasis. These results suggest that during the process of genome reduction in *B. aphidicola* by gene loss, positive or alleviating epistasis facilitates correlated gene losses while negative or aggravating epistasis impairs correlated gene losses. We interpret this as evidence that the reduced proteome of *B. aphidicola* contains less pathway redundancy and more compensatory interactions, mimicking the situation of *E. coli* when grown under environmental constraints.

**Keywords:** endosymbiotic bacteria; genome reduction; gene interactions; correlated evolution

## 1. Introduction

Mutualistic endosymbiotic bacteria from insects show the most reduced genomes of all cells [1]. The genomes of these bacteria are the outcome of an evolutionary process driven by the symbiotic interaction with its host [2]. One of the best studied cases of genome reduction among these prokaryotes is that of *Buchnera aphidicola*, the endosymbiotic bacteria from aphids [3]. The basis of this mutualistic interaction is nutritional. *B. aphidicola* provides essential amino acids lacking in the diet of its host [4–9].

Genome reduction in these bacteria is a complex process that is only partially understood. The sequence of events leading to small genomes has been described by sequencing the genomes of bacteria in different stages of genome reduction [10]. However, the role played by evolutionary forces like selection and genetic drift during genome reduction have been more difficult to untangle. It is generally accepted that these genomes accumulate slightly deleterious mutations through Muller's ratchet process [11–13]. However, there is clear evidence that selection also plays a role in the evolution of these cells [14–17].

One hypothesis proposed to explain how these genomes shrink is the *domino theory* of gene loss [18]. The domino theory is based on the premise that the products of genes interact and that this interaction affects the fitness of the organism. Accordingly, the domino theory states that when some particular gene loses its function and severely cripples or causes the total loss of a cellular function, selection will relax in all functionally related genes, which may trigger the non-functionalization and loss of these genes.

There is evidence showing that genes whose products interact, tend to be lost together in symbiotic bacteria. In *Paulinella chromatophora*, genes coding for the ABC phosphate transporter *pstSACB* provide an example of correlated loss [19]. The ABC phosphate transporter is coded by four different genes: *pstS*, *pstC*, *pstA*, and *pstB*. The *pstS* gene codes for a phosphate binding periplasmic protein; *pstB* is an ATP binding protein located on the cytoplasmic side of the membrane; while *pstC* and *pstA* are two cytoplasmic transmembrane permeases [20]. The chromatophore of *P. chromatophora* M0880/a contains all four genes, with *pstC*, *pstA*, and *pstB* located in one locus and *pstS* found in a very different location in the genome. However, in the chromatophore of another strain (*P. chromatophora* FK01), precisely these four genes were jointly lost. Another example of coordinated gene loss is provided by *Mycobacterium leprae*. For instance, it was suggested that the loss of genes coding for the two-component system involved in hypoxia response (DevR and DevS) triggered a domino effect of gene loss in this bacterium [21]. Lastly, in *B. aphidicola* BCt it was found that the gene *pgi* coding for glucose-6-phosphate isomerase has been lost. As a consequence, this endosymbiont can only use fructose-6-phosphate for glycolysis. Associated to this loss, *B. aphidicola* BCt lacks *pgl* and *zwf* genes of the oxidative pentose pathway but retains the genes necessary to obtain ribose-5-phosphate from fructose-6-phosphate [22].

Gene loss has been carefully studied in a set of five different *B. aphidicola* genomes [22]. Thanks to extreme synteny (gene order has been conserved for at least 50 million years) and almost total lack of horizontal gene transfer, it is relatively easy to identify orthologs and reconstruct the history of gene losses among the *Buchnera* [6,23]. In accordance with the domino theory, it was found that more recent gene losses tend to occur in pathways already affected by previous gene losses [22].

An intriguing question is the relationship between the domino theory and epistasis. One definition of epistasis (*functional epistasis*) addresses “the molecular interactions that proteins have to one another” [24]. This definition of epistasis is expressed in the same terms as the domino theory of gene loss. Accordingly, epistasis can be measured in terms of the observed versus the expected fitness effects of mutations where these mutations are independent, for instance in two different loci. The expected outcome is calculated by using a multiplicative model whereby  $\omega_i$  and  $\omega_j$  are the fitness effects of deleting loci *i* and *j* independently; and  $\omega_{ij}$  is the fitness effect of deleting both loci at the same time. Then, epistasis is defined as:  $\epsilon_{ij} = \omega_{ij} - \omega_i \omega_j$ . The sign of  $\epsilon_{ij}$  determines if there is alleviating (positive) or aggravating (negative) epistasis. Positive epistasis is also known as alleviating gene interaction, while negative epistasis is known as aggravating gene interaction.

Babu et al. (2011) [25] studied epistatic interactions among genes involved in cell envelope biogenesis pathways in *Escherichia coli*. They studied the fitness effects of all possible digenic mutant combinations of 821 genes. These genes were divided into 128 essential biosynthetic genes (represented by partial loss of gene function alleles) and 683 non-essential protein coding genes plus 10 small non-coding regulatory RNAs. They tested these interactions for growing *E. coli* in rich and minimal medium. They were able to describe the gene interaction network and the modular architecture of the *E. coli* cell envelope biosynthetic machinery.

*B. aphidicola* is a close relative of *E. coli*. In fact, most genes in *Buchnera* have their counterparts in *E. coli* [26]. This offers an opportunity to study the domino theory of gene loss in *Buchnera* in relation to the epistasis found in *E. coli*. Of course, under the assumption that epistatic interactions in *E. coli* are conserved, at least approximately, in *Buchnera*. In this work, we test the role played by epistatic interactions during genome reduction by gene loss among the *Buchnera*.

## 2. Materials and Methods

**Genomic sequences.** Complete genomes from all *Buchnera aphidicola* and one *Escherichia coli* strains were retrieved from the NCBI database (<ftp://www.ncbi.nlm.nih.gov/genomes/refseq/>): *E. coli* str. K-12 substr. MG1655 (chromosome: NC\_000913.3); *B. aphidicola* str. F009 (chr: CP002703.1, plasmid pLeu: CP002704.1); *B. aphidicola* str. G002 (chr: CP002701.1, plasmid pLeu: CP002702.1); *B. aphidicola* str. W106 (chr: CP002699.1, plasmid pLeu: CP002700.1); *B. aphidicola* str. USDA (chr: CP002697.1, plasmid pLeu: CP002698.1); *B. aphidicola* str. Ak (chr: CP002645.1, plasmid pLeu: CP002646.1, plasmid pTrp: CP002647.1); *B. aphidicola* str. JF99 (chr: CP002302.1); *B. aphidicola* str. TLW03 (chr: CP002301.1); *B. aphidicola* str. 5A (chr: CP001161.1); *B. aphidicola* str. Tuc7 (chr: CP001158.1); *B. aphidicola* str. APS (chr: BA000003.2, plasmid pLeu: AP001071.1, plasmid pTrp: AP001070.1); *B. aphidicola* str. LSR1 (chr: NZ\_ACFK01000001.1); *B. aphidicola* str. JF98 (chr: CP002303.1); *B. aphidicola* str. LL01 (chr: CP002300.1); *B. aphidicola* str. Sg (chr: AE013218.1); *B. aphidicola* str. Ua (chr: CP002648.1, plasmid pLeu: CP002649.1, plasmid pTrp: CP002650.1); *B. aphidicola* str. Bp (chr: AE016826.1, plasmid pBBp1: AF492591.1); *B. aphidicola* str. BCc (chr: CP000263.1, plasmid pLeu: AY438025.1); *B. aphidicola* str. Ct (chr: CP001817.1).

**Ortholog identification.** Gene families in the 19 genomes were initially identified by using the GET\_HOMOLOGUES.pl software [27]. Options used were: -M (orthoMCL algorithm), -t 0 (identify all protein families) -C 75 (minimum percent coverage in BLAST alignments) and -E 1e-05 (maximum E-value). This allowed us to identify those *E. coli* proteins having orthologs in *Buchnera* among the 811 protein coding genes studied by [25]. To double check identification of gene families, we performed BLAST searches of the 811 *E. coli* protein coding sequences against the families identified by GET\_HOMOLOGUES.pl. We carefully inspected the results to check whether each one of the *E. coli* proteins with homologs in *Buchnera* identified by GET\_HOMOLOGUES.pl is indeed the best BLAST hit. By this, we were able to identify 102 protein coding genes with clear orthologs in *Buchnera*. Of these 102 *E. coli* proteins, 51 were single copy orthologs and 51 had one-to-many homology relationships. In all cases, the orthology relationship identified by GET\_HOMOLOGUES.pl corresponded to the best BLAST hit. In addition, there were four in-paralogs (genes undergoing duplications after speciation, in this case from the lineage leading to *Buchnera*) in *E. coli*: *phoE*, *ompF*, *ompC* and *ompM*. These four genes duplicated after *E. coli* diverged from *Buchnera* (phylogenetic analysis not shown), therefore, we considered only *phoE* for the analysis of correlated gene losses and epistasis (see below).

To inspect the conservation of *Buchnera* protein families we compared their domain composition to that of *E. coli* homologs. Domain composition of protein families was inferred by HMMER3 [28] and Pfam database [29]. Visualization of domain structure of protein families in a phylogenetic tree was done with the ETE Toolkit [30]. The protein domain composition of each one of the protein families is shown in Supplementary Material 1.

**Phylogenetic reconstruction.** To reconstruct the phylogenetic tree of *Buchnera*, we identified all protein families sharing protein domain composition and conserved synteny. For this, we used GET\_HOMOLOGUES.pl with parameters: -D - M and the option -s in the script compare\_clusters.pl (the script is part of the GET\_HOMOLOGUES distribution). By this, we identified 54 protein families. Next, we aligned each one of the families with MUSCLE [31] and retrotranscribed in silico the aligned protein sequences to the corresponding coding genes with PAL2NAL [32]. Phylogenetic reconstruction was done with MrBayes3.2 [33]. Parameters used were: GTR + GI model of evolution (lset nst = 6 rates = invgamma), 1e + 06 generations sampling every 500 (mcmc ngen = 1,000,000 samplefreq = 500 printfreq = 500 diagnfreq = 5000) and discarded 25% of samples from the cold chain (relburnin = yes burninfrac = 0.25). The analysis converged (average standard deviation < 0.01). To summarize the tree, we discarded 25% of the samples: sumt relburnin = yes burninfrac = 0.25. Eleven nodes had a posterior probability of 1.0 and four nodes had a posterior probability > 0.9. The phylogenetic tree with posterior probabilities is shown in Supplementary Material 1.

**Analysis of correlated gene loss.** To identify pair of genes that were lost in a correlated fashion along *Buchnera* evolution, we used the algorithm *Discrete* from the software BayesTraitsV2 [34,35];

<http://www.evolution.rdg.ac.uk/BayesTraitsV2.html>). For that purpose, we used the phylogenetic tree of *Buchnera* and *E. coli* reconstructed with MrBayes3.2 and a matrix of the presence/absence of genes per gene family per genome. This matrix was reconstructed from the results of GET\_HOMOLOGUES.pl and consisted of 85 gene families (rows) per the 19 genomes (columns). Briefly, “Discrete” algorithm tests for correlated character evolution (codified by two binary traits) along a phylogenetic tree. It does so by comparing two models. In the first model, characters are assumed to evolve independently. In the second model, it is assumed that the characters evolve in a correlated fashion. In our case, the characters were presence (1) or absence (0) of a given gene in a gene family. We compared whether the pattern of gene losses in a given gene family indicates correlated evolution with the pattern of gene loss in another family. The models are compared by a likelihood ratio test (LRT) which is asymptotically distributed as a  $\chi^2$  distribution with the number of degrees of freedom equal to the difference of the number of parameters between the two models ( $df = 8 - 4$ ). Since we incurred in multiple testing (we tested for correlated evolution between 85 gene families) we applied false discovery rate (FDR) to control for false positives. In the first case, we applied an FDR of 10% to the list of *p*-values, and in the second, case we used an FDR of 45%. In our analysis, an FDR of 45% correspond approximately to *p*-values < 0.05.

*Epistasis analysis.* In order to study the relationship between epistasis and correlated gene loss, we compared our results of correlated gene pairs to those reported for *E. coli* by [25]. We assume that *E. coli* represents an “ancestral state” for functional gene interactions. This is based on the fact that the gene content of *Buchnera* is almost a subset of *E. coli* the gene content [26]. Specifically, empirical gene interactions data of *E. coli* was obtained from the supplementary Table S3 of [25]. We used the same criteria to define positive and negative interactions as [25]. Positive and negative interactions are defined according to an E-score threshold. All empirical interactions were catalogued as positive epistasis if the E-score was above 2 and as negative epistasis if the E-score was below  $-2$ . Scores between  $-2$  and 2 were considered as neutral. In supplementary Table S3 from [25], there are 235,031 interactions in rich medium and 129,313 interactions in minimal medium. Due to the experimental procedure, the interactions in minimal medium are a subset of the interactions in rich medium. We compared our results of correlated gene loss analysis to all 129,313 interactions from minimal medium and to the corresponding 129,313 interactions from rich medium.

The gene interaction web was implemented with Cytoscape3.5.1 [36]. The synteny map was done with genoPlotR [37]. All statistical analyses were done in R (<https://www.r-project.org>).

### 3. Results

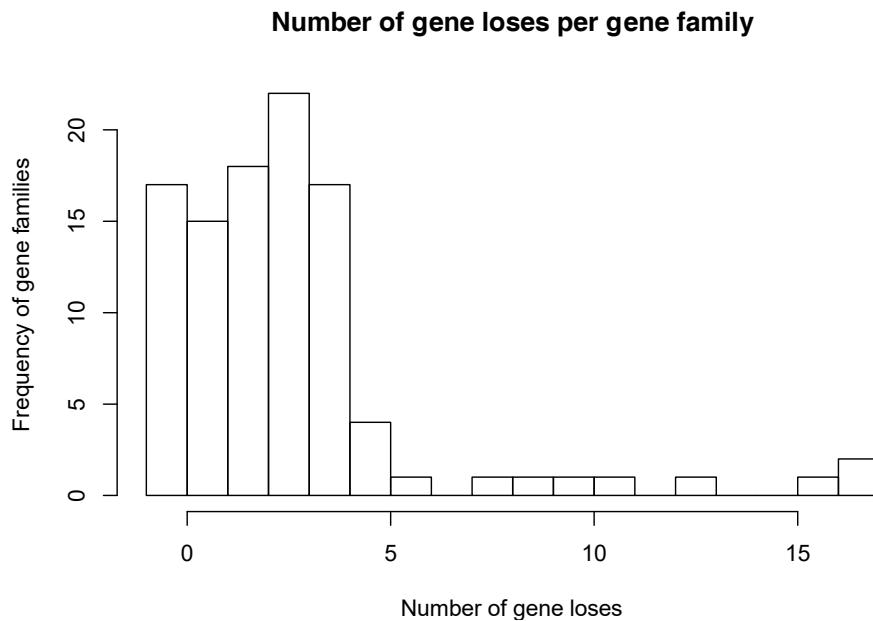
#### 3.1. Ortholog Identification between *Buchnera* and *E. coli*

Among the genes studied by [25], we identified 102 with clear orthologs in *Buchnera* (Table S1 and Supplementary Material 1). These genes participate in diverse pathways in *E. coli* (Figure 1). Of these 102 families, 17 are conserved between *E. coli* and all 18 *Buchnera* genomes. The remaining 85 families show varying degrees of conservation among *Buchnera* providing information for testing the role of epistasis in relation to the domino theory of gene loss.

General COG category	Cellular Compartment					Specific COG category
	Outer membrane	Periplasm	Inner membrane	Cytoplasm	Not specified	
Metabolism	0	0	3	1	0	Energy production and conversion (C)
	0	0	1	1	0	Amino acid transport and metabolism (E)
	0	0	0	2	0	Nucleotide metabolism and transport (F)
	0	0	1	5	0	Carbohydrate transport and metabolism (G)
	0	0	0	10	0	Coenzyme transport and metabolism (H)
	0	0	1	6	0	Lipid transport and metabolism (I)
	0	0	0	1	0	Lipid transport and metabolism & Secondary metabolites biosynthesis, transport, and catabolism (IQ)
	0	0	0	1	0	Lipid transport and metabolism & Secondary metabolites biosynthesis, transport, and catabolism & General function prediction only (IGR)
	0	0	0	1	0	Inorganic ion transport and metabolism (P)
	0	0	2	4	0	Cell cycle control, cell division, chromosome partitioning (D)
Cellular processes and signaling	5	1	6	11	2	Cell wall/membrane/envelope biogenesis (M)
	0	1	0	0	0	Cell motility & Post-translational modification, protein turnover, and chaperones (NO)
	0	0	1	0	0	Cell motility & Intracellular trafficking, secretion, and vesicular transport (NU)
	0	2	1	2	2	Post-translational modification, protein turnover, and chaperones (O)
	0	0	0	1	0	Signal transduction mechanisms (T)
	0	0	6	1	0	Intracellular trafficking, secretion, and vesicular transport (U)
	0	0	5	0	0	Defense mechanisms (V)
	0	0	0	1	0	Lipid transport and metabolism & Cell wall/membrane/envelope biogenesis (IM)
	0	1	0	1	0	Post-translational modification, protein turnover, and chaperones & Energy production and conversion (OC)
	0	0	0	3	1	Translation, ribosomal structure and biogenesis (J)
Metabolism & Cellular processes and signaling	0	1	0	0	0	
Information storage and processing	0	1	1	2	0	General function prediction only (R)
Poorly characterized	0	0	0	0	2	Function unknown (S)
Not in COG	0	0	2	0	0	Not in COG (-)

**Figure 1.** Functional classification of 102 genes from *E. coli* with orthologs in *Buchnera*. Each cell denotes the number of genes classified in each COG category. We also show the localization of the corresponding protein product in four cellular compartments. The classification of these genes follows that of Babu et al. (2011).

The numbers of gene losses vary between families and genomes. In some of these families, there is only a single gene loss, while in other families there are as many as 17 gene losses (Figure 2). Most of the families show 4 or fewer gene losses. In addition, gene losses among these 85 gene families occur unevenly among *Buchnera* genomes. As shown in Figure 3, *B. aphidicola* Cc and *B. aphidicola* Ct accumulate most gene losses. These are followed by *B. aphidicola* JF98 and *B. aphidicola* Bp.

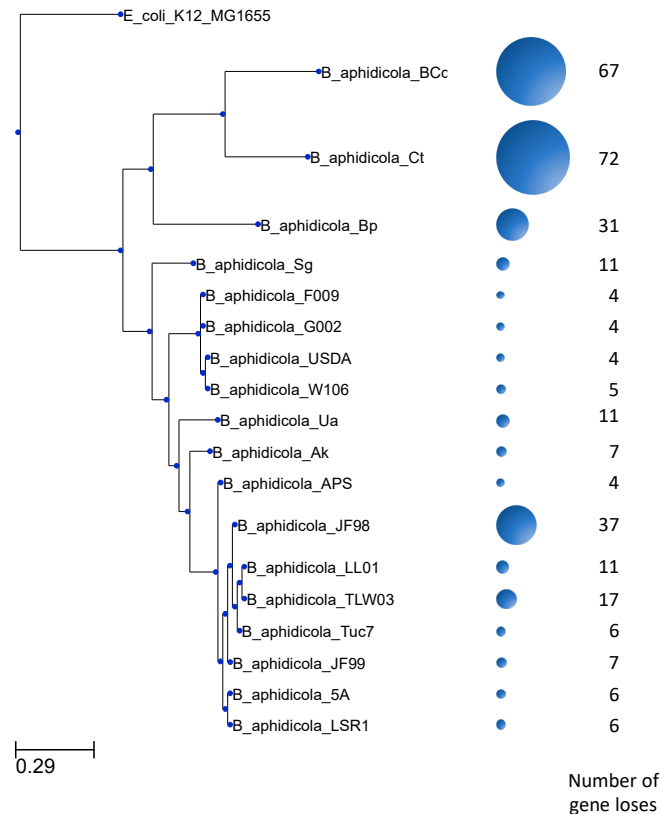


**Figure 2.** Distribution of number of gene losses per gene family. For each one of the 102 gene families studied here, we counted the number of gene losses among the 18 *Buchnera* genomes.

### 3.2. Essential Genes in *E. coli* Tend to be Retained in the Genomes of *Buchnera*

As expected, we found that genes that are reported as essential in *E. coli* are less likely to belong to families showing gene losses among the *Buchnera*. The odds ratio of this association is 2.7 and the *p*-value is 0.06 (95% CI = 0.95 to ∞, one sided Fisher exact test). Although the probability of finding

this association at random is slightly larger than 0.05, this shows a trend that suggests that essential genes in *E. coli* tend to be essential also in *Buchnera*, despite the differences in biology and environment between the two species. It also suggests that gene function in *E. coli* is conserved among *Buchnera*, i.e., lessons learned in *E. coli* can be extrapolated to *Buchnera* with some degree of confidence.



**Figure 3.** Distribution of gene losses among *Buchnera* genomes. The number of gene losses in the 102 gene families studied here is unevenly distributed among *Buchnera* genomes. The size of the circles corresponds to the number of gene losses.

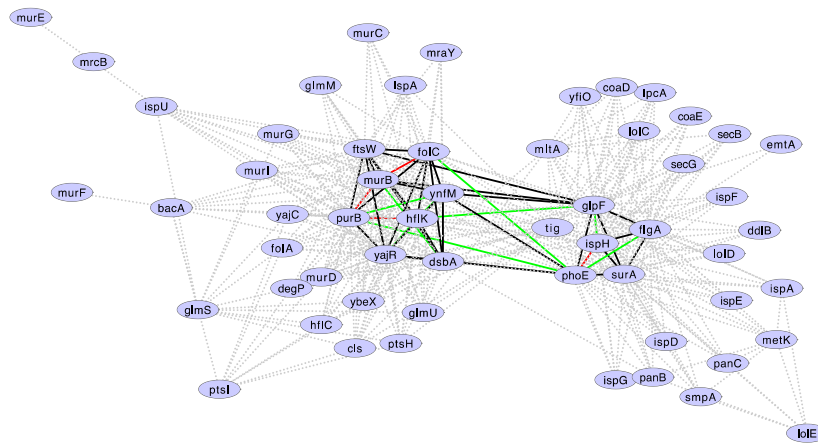
### 3.3. Correlated Gene Losses along *Buchnera* Phylogeny

To investigate the relationship between correlated gene loss and epistasis, we first studied if some of the 85 variable families show evidence of coevolution. Using the BayesTraits software, we estimated the probability that the pattern of gene conservation between two given families is the result of random gene losses. If this probability is smaller than 0.05, we assumed that the similarity in gene losses is unlikely to be due to chance alone. That is, genes were lost in a correlated fashion, presumably because their protein products interact [35].

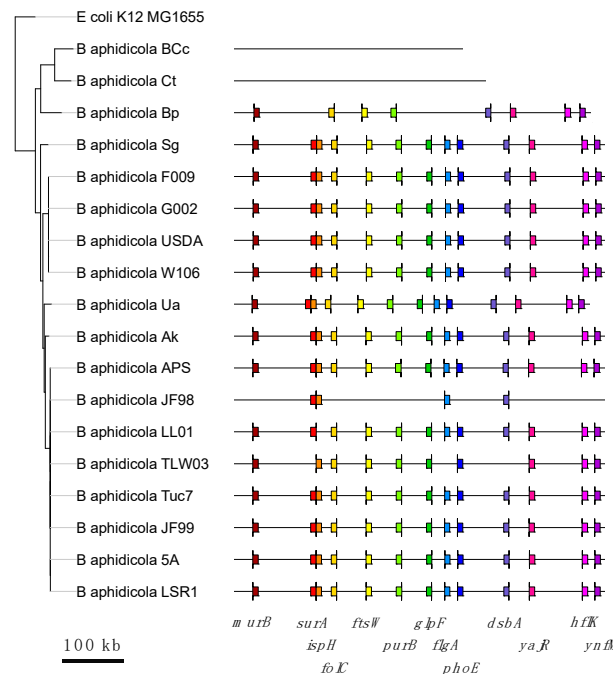
BayesTraits analysis identified 356 pairs of gene families whose similarity in their pattern of gene losses along the phylogeny of *Buchnera* would be found by chance less than 5% of the time ( $p$ -value < 0.05). Of course, this probability is under the null hypothesis of no correlated gene losses. These 356 pairs are conformed by 57 gene families.

Since we performed several statistical tests (i.e., we calculated the probability of correlated evolution, under the null hypothesis of no association, between all the pairs that can be formed with 85 gene families), we expect more false positives than those specified by a  $p$ -value of 0.05. Therefore, it is necessary to introduce a correction to the  $p$ -value. In this case, we used the false discovery rate (FDR or Benjamini-Hochberg procedure). When controlling for a 10% (FDR < 0.1) of false positives, the above numbers reduced our original results to 43 pairs of correlated gene losses conformed by 13 gene families. These gene families and their correlations are shown in Figure 4. It is important

to notice that none of these genes are found adjacent to each other in *Buchnera* genomes (Figure 5), indicating that correlated gene loss is not simply due to proximity within the genome. We suggest that most pairs of gene families showing correlated gene loss (as defined by a  $p$ -value < 0.05 and FDR < 0.1) code for proteins that functionally interact.



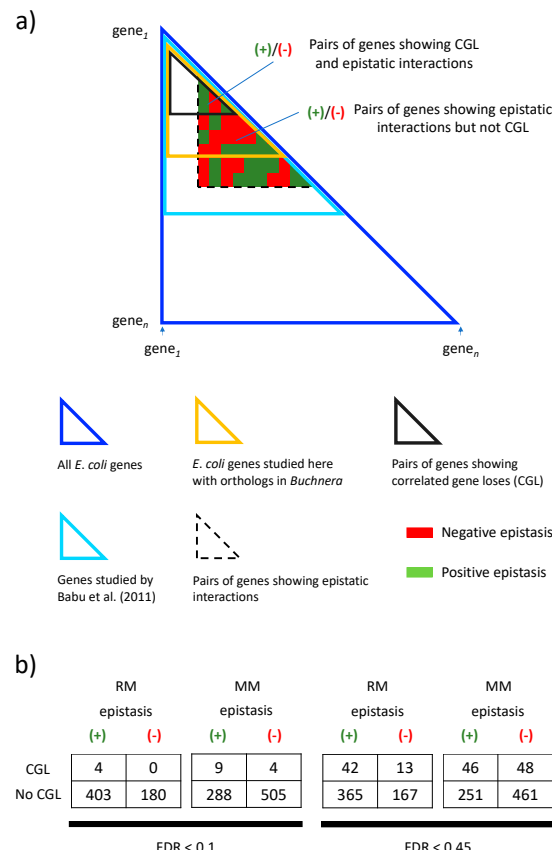
**Figure 4.** Gene families showing correlated gene losses along *Buchnera* phylogeny. Dotted lines: Correlated gene losses showing a  $p$ -value < 0.05. Solid lines: Correlated gene losses discovered by setting a FDR < 0.1. Green lines: Positive epistatic interactions in minimal medium. Red lines: Negative epistatic interactions in minimal medium. Epistatic interactions are according to Babu et al. (2011). For clarity, epistatic interactions are shown only for those correlated gene losses showing an FDR < 0.1. An FDR < 0.1 captures all interactions with a  $p$ -value < 0.0015.



**Figure 5.** Localization of genes showing correlated gene loss (FDR < 0.1) along *Buchnera* genomes. The size of the genomes is shown with horizontal lines and the relative localization of the genes showing correlated gene loss with colored squares. The vertical lines in each one of the genes indicate the direction of transcription. For clarity, the genome of *E. coli* is not shown. None of the genes are coded contiguously. *surA* and *ispH* look contiguous because of the scale, but are separated by more than 7000 bp.

### 3.4. Is Correlated Gene Loss in *Buchnera* Explained by Epistasis Data in *E. coli*?

We then decided to investigate whether the correlations identified with BayesTraitsV2 in *Buchnera* are associated with known epistatic interactions among these genes in *E. coli* [25]. In particular, we would expect positive (or alleviating) epistasis to facilitate correlated gene losses (due to reduction in selective costs), while negative (or aggravating) epistasis would impair correlated gene losses (because of increased selective costs). If so, correlated gene losses in *Buchnera* should occur more often among gene families whose *E. coli* orthologs show positive epistasis and if correlated gene losses among *Buchnera* occurs less often among gene families whose *E. coli* orthologs show negative epistasis (Figure 6a).



**Figure 6.** Comparison of correlated gene losses (CGL) identified in *Buchnera* with epistatic interactions in *E. coli* detected by Babu et al. (2011). **(a)** Dark blue triangle: All *E. coli* gen pairs are accommodated in an  $n \times n$  matrix (only “half” the matrix is needed, below its main diagonal). Light blue triangle: Pairs of genes studied by Babu et al. (2011). Gold triangle: Pairs of *E. coli* genes with orthologs in *Buchnera* studied here. Gray triangle: Pairs of genes showing correlated gene lose in *Buchnera*. Dashed triangle: Pairs of genes showing epistatic interactions by Babu et al. (2011). Among the set of genes showing epistatic interactions in *E. coli*, some have orthologs in *Buchnera* showing CGL and some of them do not. **(b)** Frequency of gene families showing positive and negative epistatic interactions and showing (or not) CGL. The frequency is shown for an FDR < 0.45 and for a FDR < 0.1. The frequency is also shown for minimal and rich media. An FDR < 0.45 comprises all CGL showing a  $p$ -value < 0.05. The size of the triangles is not to scale.

We tested the above-mentioned hypotheses by using epistasis data from [25]. The data, was obtained from growing *E. coli* in minimal and rich medium. We found that for minimal medium, the probability of observing actual data (Figure 6b) is less than 5% ( $p$ -value  $\approx$  0.02) under the null hypothesis of no association between correlated gene lose and epistasis (Table 1). We find the same

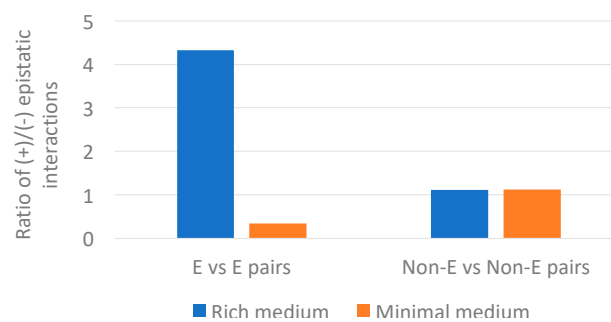


result if we relax our false discovery rate to  $FDR < 0.45$ , this to include all  $p$ -values from BayesTratisV2 analysis smaller than 0.05. That is, we found that the fraction of gene pairs showing positive epistasis *versus* negative epistasis is different in gene pairs showing correlated gene loss than in those not showing it. In other words, positive epistasis seems to favor correlated gene losses while negative epistasis seems to disfavor it (see below). However, it is important to notice that we did not find this association for epistasis data obtained from *E. coli* growing in rich medium.

**Table 1.** The proportion of gene pairs showing positive and negative epistasis is different among the set of gene pairs showing correlated gene loss and those not showing it. However, the probability of finding this association under the null hypothesis of no association is less than 0.05 only for epistasis data from growing *E. coli* in minimal medium. The probability was calculated with a two-sided Fisher exact test. We calculated the probability for two FDR cutoffs: an  $FDR < 0.1$  comprises all pairs of genes showing a  $p$ -value  $< 0.01$  under the null hypothesis of no correlated gene loss and an  $FDR < 0.45$  comprises all pairs of genes with a  $p$ -value  $< 0.05$  under the null hypothesis of no correlated gene loss. In all cases, the null hypothesis is True odds ratio is greater than 1. RM, rich medium; MM, minimal medium.

Correlated Losses Gene Set Defined by	Media	$p$ -Value	95% CI	Odds Ratio
FDR $< 0.1$	RM	0.31	0.29 to $\infty$	Inf.
( $p$ -value $< 0.01$ )	MM	0.02	1.09 to 17.7	3.93
FDR $< 0.45$	RM	0.28	0.75 to 3.08	1.48
( $p$ -value $< 0.05$ )	MM	0.01	1.11 to 2.78	1.76

It is intriguing that we only find the association between epistasis and gene loss for minimal and not for rich medium data. In *E. coli*, essential genes showed more positive epistatic interactions than non-essential genes in rich medium. While in minimal medium, essential genes showed more negative epistatic interactions than non-essential genes [25]. Our analysis mirrors the pattern (Figure 7 and Supplementary Material 2). Essential genes also showed more positive interactions in rich medium ( $148/34 \approx 4.4$ ) than non-essential genes ( $60/52 \approx 1.2$ ); and in minimal medium, essential genes showed more negative interactions ( $66/196 \approx 0.3$ ) than non-essential genes ( $69/58 \approx 1.2$ ). Interestingly, non-essential genes showed about the same proportion of positive versus negative epistasis in rich and minimal medium. This suggests that the  $p$ -value differences between rich and minimal medium, reported in Table 1, are driven mostly by changes in the proportion of positive versus negative epistatic interactions in essential genes. According to [25], the increase of negative epistatic interactions between essential genes in minimal medium suggest the emergence of compensatory relationships under environmental constraints. That is, under minimal medium there are less redundant pathways. This decrease of redundant pathways causes a concomitant increase of negative interactions.



**Figure 7.** Ratio of positive versus negative epistatic interactions for pairs of essential (E) genes and pairs of non-essential (Non-E) genes in rich and minimal medium. The epistasis data is from *E. coli* genes having clear orthologs in *Buchnera* genomes.

## 4. Discussion

### 4.1. Conservation of Epistatic Relationships between *E. coli* and *Buchnera*

The extrapolation of the experimental results found by [25] to *B. aphidicola* depends on several assumptions. First, orthologs have to be accurately identified, and the functions of these genes in *B. aphidicola* have to be the same as in *E. coli*. As mentioned before, ortholog identification between *E. coli* and *B. aphidicola* is relatively straightforward because of the evolutionary distance between these species and the high level of synteny among *Buchnera*. However, it is not always easy to evaluate if the ortholog gene in *B. aphidicola* is performing the same function as in *E. coli* or if it is functional at all. Often, the ortholog in *B. aphidicola* is lacking some domains present in the *E. coli* version of the protein. For instance, of all the 85 families studied here, only in 70 of them the orthologs in *B. aphidicola* have exactly the same domain composition in *Buchnera* as the *E. coli* proteins (Supplementary Material 1). Despite this lack of conservation of protein domains, essential genes in *E. coli* tend to be more conserved in *B. aphidicola* than non-essential genes ( $p$ -value = 0.06). This suggests that protein function tend to be conserved. The  $p$ -value of this association is slightly larger than 0.05. However, there is ample evidence showing that essential genes tend to be conserved in evolution more than non-essential genes [38–43]). Therefore, we consider that a  $p$ -value of 0.06 is in agreement with previous results.

This leads us to the second point, whether epistatic interactions in *E. coli* can be extrapolated to *B. aphidicola*. This is more difficult to evaluate, but existent data suggest that epistatic interactions are conserved to some degree between species sharing similar biology. For instance, *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* share between ~18% to 29% of their negative (synthetic lethal) genetic interactions [44,45]. This percentage increases above 50% for positive interactions when the involved genes code for physically associated proteins [44,46]). Remarkably, these interactions are shared despite several million years of divergence. The time of divergence between *S. pombe* and *S. cerevisiae* has been estimated to be from ~400 to ~1000 million years [47,48]. In comparison, the divergence time between *B. aphidicola* and *E. coli* is estimate to be between ~160 to 280 million years [49,50]). Therefore, it is likely that similar percentages of shared epistatic interactions are shown between the bacteria analyzed here.

Given all the difficulties discussed above associated to extrapolating epistasis data from *E. coli* to *B. aphidicola*, we consider that there is statistical evidence (Table 1) indicating that correlated gene loss occurs more often among gene pairs showing positive epistasis than among those showing negative epistasis. This indicates that the domino effect applies mostly to genes participating in the same pathway or protein complex [51]. Genes in the same pathway or protein complex tend to show positive epistasis.

It is important to notice that the domino theory, to better work as a hypothesis, requires a refinement. Ultimately, the selective advantage conferred by a given gene is dependent on the environment in which the bacteria live. Therefore, a gene loss mutation cannot by itself cause the relaxation in selective pressure of functionally related genes. To reach fixation, a domino-effect-gene-loss requires population bottlenecks to get rid of bacteria lacking the mutation by chance. Otherwise, bacteria that have retained the gene might outcompete the mutant strain. Alternatively, the domino-effect-gene-loss has to provide some selective advantage. In intracellular mutualistic bacteria, gene-loss mutations favoring the fitness of the holobiont will be favored by selection.

### 4.2. The Role of Epistasis in Evolution of Genome Reduction

“One of the more interesting long term questions in evolutionary biology is whether or not epistasis determines the path of evolutionary change” [24]. Here, we provide statistical support for the role played by epistasis in correlated gene loss in *B. aphidicola*. But, why is this interaction only is found only for epistasis data obtained from minimal and not from rich medium? If we consider that: (a) the relationship between epistasis and correlated gene loss is unlikely due to chance ( $p$ -value = 0.02, FDR = 10%) and (b) accept that negative epistasis impairs and positive epistasis facilitates gene loss.

This lead us to conclude (c) that somehow the medium on which *B. aphidicola* thrives causes a pattern of epistatic relationships similar to those of *E. coli* growing on minimal medium.

This result may seem counterintuitive. One would expect that intracellular conditions would be more similar to rich medium. However, *B. aphidicola* does not live in direct contact with the cytoplasm of the host. On the contrary, *B. aphidicola* grows inside symbiotic vesicles within bacteriocytes whose membranes are of host origin [3,52]. Therefore, the host seems to have direct control of what molecules are provided to *B. aphidicola* for nutrition. In fact, recent studies show that *B. aphidicola* population is carefully controlled along aphid life cycle [53] and that *Buchnera* may grow under nitrogen limiting conditions [9].

A more appealing explanation of this pattern is that in the reduced proteome of *B. aphidicola* there are few, if any, redundant pathways. Therefore, aggravating interactions will be more common than alleviating ones and, as mentioned above, this cellular condition is similar to that of *E. coli* growing on minimal medium, where essential genes showed more negative epistatic interactions than non-essential genes [25].

How widespread are the effects detected here on bacterial genome evolution? In a recent study [54] calculated the frequency of pairs of ortholog genes that co-occur across ~600 species of bacterial genomes. They also identified those pairs of orthologs that tend not to occur in the same genome. They called the former “correlogs” and the later “anti-correlogs”. They found that correlogs in these genomes organize in 483 subgroups. These subgroups associate to each other by “anti-correlogy”. Because genes within subgroups tend to participate in the same functional category it is likely that epistasis among them will tend to be positive, and that they will be retained by natural selection or, according to the results shown here, lost together once their function becomes dispensable.

## 5. Conclusions

*B. aphidicola* allow us to study the relationship between epistasis and the evolutionary fate of genes—under the assumption that epistatic interactions in *E. coli* are maintained in *B. aphidicola*. We provided statistical evidence indicating that, along genome reduction in *B. aphidicola*, positive or alleviating epistasis facilitates correlated gene loss, while negative or aggravating epistasis impairs correlated gene loss. Our analysis paves the way to more extensive analysis of the role of epistasis in genome evolution in *B. aphidicola* by including more genes [55,56], other factors like protein-protein interactions [57], metabolic networks [58], or by studying this phenomenon in other model organisms like yeast.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2075-1729/8/2/17/s1>, Supplementary Material 1. This file contains the Bayesian phylogenetic tree of the *Buchnera* used for BayesTraits. This file also contains a representation of each one of the gene families studied here on a phylogenetic context. Gene families also show coded protein domains according to Pfam database. Supplementary Material 2. This file contains Figure 6 in more detail, indicating the frequency of essential and non-essential genes. Table S1 contains the families of *Buchnera* and *E. coli* genes studied here. Tables S2 and S3 contain the pairs of genes showing epistasis and correlated gene losses for minimal and rich medium.

**Author Contributions:** Conceptualization, D.J.M.-C., A.M., G.B., and L.D.; Methodology, G.B., D.J.M.-C., and L.D.; Software, D.J.M.-C., and L.D.; Validation, D.J.M.-C. and L.D.; Formal Analysis, G.B., D.J.M.-C., and L.D.; Investigation, D.J.M.-C., and L.D.; Resources, D.J.M.-C. and L.D.; Data Curation, D.J.M.-C. and L.D.; Writing-Original Draft Preparation, D.J.M.-C. and L.D.; Writing-Review & Editing, D.J.M.-C., A.M., G.B., and L.D.; Funding Acquisition, A.M. and L.D.

**Funding:** This work was supported by grants to AM from the Spanish Ministry of Economy and Competitiveness (SAF2015-65878-R) and Generalitat Valenciana (project PrometeoII/2014/065) and co-financed by the European Regional Development Fund (ERDF). G.B. wishes to thank CONACYT, grant number 222870. D.J.M.-C would like to thank CONACYT for providing Scholarship. L.D. wishes to thank the Universitat de València for support during sabbatical leave.

**Acknowledgments:** We would like to thank Gabriel Moreno-Hagelsieb for bioinformatic support and useful comments to the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Lo, W.-S.; Huang, Y.-Y.; Kuo, C.H. Winding paths to simplicity: Genome evolution in facultative insect symbionts. *FEMS Microbiol. Rev.* **2016**, *40*, 855–874. [[CrossRef](#)] [[PubMed](#)]
2. Martínez-Cano, D.J.; Reyes-Prieto, M.; Martínez-Romero, E.; Partida-Martínez, L.P.; Latorre, A.; Moya, A.; Delaye, L. Evolution of small prokaryotic genomes. *Front. Microbiol.* **2015**, *5*, 742. [[CrossRef](#)] [[PubMed](#)]
3. Shigenobu, S.; Wilson, A.C.C. Genomic revelations of a mutualism: The pea aphid and its obligate bacterial symbiont. *Cell. Mol. Life Sci.* **2011**, *68*, 1297–1309. [[CrossRef](#)] [[PubMed](#)]
4. Sapp, J. Paul Buchner (1886–1978) and hereditary symbiosis in insects. *Int. Microbiol.* **2002**, *5*, 145–150. [[CrossRef](#)] [[PubMed](#)]
5. Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **2000**, *407*, 81–86. [[PubMed](#)]
6. Tamas, I.; Klasson, L.; Canbäck, B.; Näslund, A.K.; Eriksson, A.S.; Wernegreen, J.J.; Sandström, J.P.; Moran, N.A.; Andersson, S.G. 50 million years of genomic stasis in endosymbiotic bacteria. *Science* **2002**, *296*, 2376–2379. [[CrossRef](#)] [[PubMed](#)]
7. Van Ham, R.C.; Kamerbeek, J.; Palacios, C.; Rausell, C.; Abascal, F.; Bastolla, U.; Fernández, J.M.; Jiménez, L.; Postigo, M.; Silva, F.J.; et al. Reductive genome evolution in *Buchnera aphidicola*. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 581–586. [[CrossRef](#)] [[PubMed](#)]
8. International Aphid Genomics Consortium. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol.* **2010**, *8*, e1000313. [[CrossRef](#)]
9. Hansen, A.K.; Moran, N.A. Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 2849–2854. [[CrossRef](#)] [[PubMed](#)]
10. McCutcheon, J.P.; Moran, N.A. Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **2012**, *10*, 13–26. [[CrossRef](#)] [[PubMed](#)]
11. Wernegreen, J.J. Endosymbiont evolution: Predictions from theory and surprises from genomes. *Ann. N. Y. Acad. Sci.* **2015**, *1360*, 16–35. [[CrossRef](#)] [[PubMed](#)]
12. Wernegreen, J.J. Ancient bacterial endosymbionts of insects: Genomes as sources of insight and springboards for inquiry. *Exp. Cell. Res.* **2017**, *358*, 427–432. [[CrossRef](#)] [[PubMed](#)]
13. Wernegreen, J.J. In it for the long haul: evolutionary consequences of persistent endosymbiosis. *Curr. Opin. Genet. Dev.* **2017**, *47*, 83–90. [[CrossRef](#)] [[PubMed](#)]
14. Fares, M.A.; Moya, A.; Barrio, E. Adaptive evolution in GroEL from distantly related endosymbiotic bacteria of insects. *J. Evol. Biol.* **2005**, *18*, 651–660. [[CrossRef](#)] [[PubMed](#)]
15. Delmotte, F.; Rispé, C.; Schaber, J.; Silva, F.J.; Moya, A. Tempo and mode of early gene loss in endosymbiotic bacteria from insects. *BMC Evol. Biol.* **2006**, *6*, 56. [[CrossRef](#)] [[PubMed](#)]
16. Toft, C.; Fares, M.A. Selection for translational robustness in *Buchnera aphidicola*, endosymbiotic bacteria of aphids. *Mol. Biol. Evol.* **2009**, *26*, 743–751. [[CrossRef](#)] [[PubMed](#)]
17. Sabater-Muñoz, B.; Toft, C.; Alvarez-Ponce, D.; Fares, M.A. Chance and necessity in the genome evolution of endosymbiotic bacteria of insects. *ISME J.* **2017**, *11*, 1291–1304. [[CrossRef](#)] [[PubMed](#)]
18. Dagan, T.; Blekhman, R.; Graur, D. The “domino theory” of gene death: Gradual and mass gene extinction events in three lineages of obligate symbiotic bacterial pathogens. *Mol. Biol. Evol.* **2006**, *23*, 310–316. [[CrossRef](#)] [[PubMed](#)]
19. Reyes-Prieto, A.; Yoon, H.S.; Moustafa, A.; Yang, E.C.; Andersen, R.A.; Boo, S.M.; Nakayama, T.; Ishida, K.; Bhattacharya, D. Differential gene retention in plastids of common recent origin. *Mol. Biol. Evol.* **2010**, *27*, 1530–1537. [[CrossRef](#)] [[PubMed](#)]
20. McGrath, J.W.; Chin, J.P.; Quinn, J.P. Organophosphonates revealed: new insights into the microbial metabolism of ancient molecules. *Nat. Rev. Microbiol.* **2013**, *11*, 412–419. [[CrossRef](#)] [[PubMed](#)]
21. Tyagi, J.S.; Saini, D.K. Did the loss of two-component systems initiate pseudogene accumulation in *Mycobacterium leprae*? *Microbiology* **2004**, *150*, 4–7. [[CrossRef](#)] [[PubMed](#)]
22. Lamelas, A.; Gosalbes, M.J.; Moya, A.; Latorre, A. New clues about the evolutionary history of metabolic losses in bacterial endosymbionts, provided by the genome of *Buchnera aphidicola* from the aphid *Cinara tujaefilina*. *Appl. Environ. Microbiol.* **2011**, *77*, 4446–4454. [[CrossRef](#)] [[PubMed](#)]
23. Gómez-Valero, L.; Latorre, A.; Silva, F.J. The evolutionary fate of nonfunctional DNA in the bacterial endosymbiont *Buchnera aphidicola*. *Mol. Biol. Evol.* **2004**, *21*, 2172–2181. [[CrossRef](#)] [[PubMed](#)]

24. Phillips, P.C. Epistasis—The essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* **2008**, *9*, 855–867. [[CrossRef](#)] [[PubMed](#)]
25. Babu, M.; Díaz-Mejía, J.J.; Vlasblom, J.; Gagarinova, A.; Phanse, S.; Graham, C.; Yousif, F.; Ding, H.; Xiong, X.; Nazarians-Armavil, A.; et al. Genetic interaction maps in *Escherichia coli* reveal functional crosstalk among cell envelope biogenesis pathways. *PLoS Genet.* **2011**, *7*, e1002377. [[CrossRef](#)] [[PubMed](#)]
26. Tamas, I.; Klasson, L.M.; Sandström, J.P.; Andersson, S.G. Mutualists and parasites: how to paint yourself into a (metabolic) corner. *FEBS Lett.* **2001**, *498*, 135–139. [[CrossRef](#)]
27. Contreras-Moreira, B.; Vinuesa, P. GET\_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* **2013**, *79*, 7696–7701. [[CrossRef](#)] [[PubMed](#)]
28. Mistry, J.; Finn, R.D.; Eddy, S.R.; Bateman, A.; Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **2013**, *41*, e121. [[CrossRef](#)] [[PubMed](#)]
29. Finn, R.D.; Coghill, P.; Eberhardt, R.Y.; Eddy, S.R.; Mistry, J.; Mitchell, A.L.; Potter, S.C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; et al. The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **2016**, *44*, D279–D285. [[CrossRef](#)] [[PubMed](#)]
30. Huerta-Cepas, J.; Serra, F.; Bork, P. ETE 3: Reconstruction, analysis and visualization of phylogenomic data. *Mol. Biol. Evol.* **2016**, *33*, 1635–1638. [[CrossRef](#)] [[PubMed](#)]
31. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)] [[PubMed](#)]
32. Suyama, M.; Torrents, D.; Bork, P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **2006**, *34*, W609–W612. [[CrossRef](#)] [[PubMed](#)]
33. Ronquist, F.; Teslenko, M.; van der Mark, P.; Ayres, D.L.; Darling, A.; Höhna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst. Biol.* **2012**, *61*, 539–542. [[CrossRef](#)] [[PubMed](#)]
34. Pagel, M. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc. R. Soc. Lond. B.* **1994**, *255*, 37–45. [[CrossRef](#)]
35. Barker, D.; Pagel, M. Predicting Functional Gene Links from Phylogenetic-Statistical Analyses of Whole Genomes. *PLoS Comput. Biol.* **2005**, *1*, 24–31. [[CrossRef](#)] [[PubMed](#)]
36. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)] [[PubMed](#)]
37. Guy, L.; Kultima, J.R.; Andersson, S.G.E. genoPlotR: Comparative gene and genome visualization in R. *Bioinformatics* **2010**, *26*, 2334–2335. [[CrossRef](#)] [[PubMed](#)]
38. Decottignies, A.; Sanchez-Perez, I.; Nurse, P. *Schizosaccharomyces pombe* essential genes: A pilot study. *Genome Res.* **2003**, *13*, 399–406. [[CrossRef](#)] [[PubMed](#)]
39. Fang, G.; Rocha, E.; Danchin, A. How essential are nonessential genes? *Mol. Biol. Evol.* **2005**, *22*, 2147–2156. [[CrossRef](#)] [[PubMed](#)]
40. Hwang, Y.C.; Lin, C.C.; Chang, J.Y.; Mori, H.; Juan, H.F.; Huang, H.C. Predicting essential genes based on network and sequence analysis. *Mol. Biosyst.* **2009**, *5*, 1672–1678. [[CrossRef](#)] [[PubMed](#)]
41. Jordan, I.K.; Rogozin, I.B.; Wolf, Y.I.; Koonin, E.V. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Res.* **2002**, *12*, 962–968. [[CrossRef](#)] [[PubMed](#)]
42. Luo, H.; Gao, F.; Lin, Y. Evolutionary conservation analysis between the essential and nonessential genes in bacterial genomes. *Sci. Rep.* **2015**, *5*, 13210. [[CrossRef](#)] [[PubMed](#)]
43. Rancati, G.; Moffat, J.; Typas, A.; Pavelka, N. Emerging and evolving concepts in gene essentiality. *Nat. Rev. Genet.* **2017**, *19*, 34–49. [[CrossRef](#)] [[PubMed](#)]
44. Roguev, A.; Bandyopadhyay, S.; Zofall, M.; Zhang, K.; Fischer, T.; Collins, S.R.; Qu, H.; Shales, M.; Park, H.O.; Hayles, J.; et al. Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science* **2008**, *322*, 405–410. [[CrossRef](#)] [[PubMed](#)]
45. Dixon, S.J.; Fedyshyn, Y.; Koh, J.L.; Prasad, T.S.; Chahwan, C.; Chua, G.; Toufighi, K.; Baryshnikova, A.; Hayles, J.; Hoe, K.L.; et al. Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 16653–16658. [[CrossRef](#)] [[PubMed](#)]
46. Beltrao, P.; Ryan, C.; Krogan, N.J. Comparative interaction networks: Bridging genotype to phenotype. *Adv. Exp. Med. Biol.* **2012**, *751*, 139–156. [[PubMed](#)]

47. Sipiczki, M. Where does fission yeast sit on the tree of life? *Genome Biol.* **2000**, *1*, REVIEWS1011. [[CrossRef](#)] [[PubMed](#)]
48. Hedges, S.B. The origin and evolution of model organisms. *Nat. Rev. Genet.* **2002**, *3*, 838–849. [[CrossRef](#)] [[PubMed](#)]
49. Moran, N.A.; Munson, M.A.; Baumann, P.; Ishikawa, H. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc. R. Soc. B* **1993**, *253*, 167–171. [[CrossRef](#)]
50. Kuo, C.-H.; Ochman, H. Inferring clocks when lacking rocks: The variable rates of molecular evolution in bacteria. *Biol. Direct.* **2009**, *4*, 35. [[CrossRef](#)] [[PubMed](#)]
51. Beyer, A.; Bandyopadhyay, S.; Ideker, T. Integrating physical and genetic maps: from genomes to interaction networks. *Nat. Rev. Genet.* **2007**, *8*, 699–710. [[CrossRef](#)] [[PubMed](#)]
52. Houk, E.J.; Griffiths, G.W.; Hadjokas, N.E.; Beck, S.D. Peptidoglycan in the cell wall of the primary intracellular symbiote of the pea aphid. *Science* **1977**, *198*, 401–403. [[CrossRef](#)] [[PubMed](#)]
53. Simonet, P.; Duport, G.; Gaget, K.; Weiss-Gayet, M.; Colella, S.; Febvay, G.; Charles, H.; Viñuelas, J.; Heddi, A.; Calevro, F. Direct flow cytometry measurements reveal a fine-tuning of symbiotic cell dynamics according to the host developmental needs in aphid symbiosis. *Sci. Rep.* **2016**, *6*, 19967. [[CrossRef](#)] [[PubMed](#)]
54. Kim, P.J.; Price, N.D. Genetic co-occurrence network across sequenced microbes. *PLoS Comput. Biol.* **2011**, *7*, e1002340. [[CrossRef](#)] [[PubMed](#)]
55. Babu, M.; Arnold, R.; Bundalovic-Torma, C.; Gagarinova, A.; Wong, K.S.; Kumar, A.; Stewart, G.; Samanfar, B.; Aoki, H.; Wagih, O.; et al. Quantitative genome-wide genetic interaction screens reveal global epistatic relationships of protein complexes in *Escherichia coli*. *PLoS Genet.* **2014**, *10*, e1004120. [[CrossRef](#)] [[PubMed](#)]
56. Kumar, A.; Beloglazova, N.; Bundalovic-Torma, C.; Phanse, S.; Deineko, V.; Gagarinova, A.; Musso, G.; Vlasblom, J.; Lemak, S.; Hooshyar, M.; et al. Conditional Epistatic Interaction Maps Reveal Global Functional Rewiring of Genome Integrity Pathways in *Escherichia coli*. *Cell. Rep.* **2016**, *14*, 648–661. [[PubMed](#)]
57. Rajagopala, S.V.; Sikorski, P.; Kumar, A.; Mosca, R.; Vlasblom, J.; Arnold, R.; Franca-Koh, J.; Pakala, S.B.; Phanse, S.; Ceol, A.; et al. The binary protein-protein interaction landscape of *Escherichia coli*. *Nat. Biotechnol.* **2014**, *32*, 285–290. [[CrossRef](#)] [[PubMed](#)]
58. Yizhak, K.; Tuller, T.; Papp, B.; Ruppín, E. Metabolic modeling of endosymbiont genome reduction on a temporal scale. *Mol. Syst. Biol.* **2011**, *7*, 479. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).