Chromosomal arrangement of synthetic lethal gene pairs: repulsion or attraction?

Sayed-Rzgar Hosseini [1, *], Bishoy Wadie[1], Evangelia Petsalaki[1]

1. EMBL-EBI, Wellcome Genome Campus, CB10 1SD, Hinxton, Cambridgeshire, UK

*Author for Correspondence: Sayed-Rzgar Hosseini EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridgeshire, UK, (+44) 01223 494 468, rzgar@ebi.ac.uk

Abstract

Synthetic lethal interactions are of paramount importance both in biology and in medicine, and hence increasing efforts have been devoted to their systematic identification. A recent computational study found that synthetic lethal genes tend to be further away in chromosomes than random (i.e. repulsion), which was shown to provide bacterial genomes with greater robustness to large-scale DNA deletions. To evaluate this observation using published experimental data and to test its generalizability in eukaryotic genomes, we leveraged the wealth of experimentally determined synthetic lethal genetic interactions curated in BioGRID (Biological General Repository for Interaction Datasets) and CGIdb (Cancer Genetic Interaction db) databases. In support of the previous study, we observed repulsion of synthetic lethal gene pairs, manifested as striking depletion of closely located synthetic lethal gene pairs (i.e. with genomic distance < 25 KB) on *E. coli K12* chromosome. However, our observations indicate that this feature is specific to bacterial genomes and is not generalizable to eukaryotic ones. Moreover, in *S. cerevisiae* and the human genome we observed an opposite trend that is the genomic proximity of synthetic lethal gene pairs (i.e. attraction). Therefore, we conclude that repulsion of synthetic lethal gene pairs does exist in *E. coli* and likely also in other bacterial genomes. In contrast, we raise the possibility of genomic attraction of synthetic lethal genes in eukaryotes. Moreover, we showed that the observed attraction in eukaryotic genomes is not simply a consequence of gene duplication events, and so characterization of its evolutionary origin calls for further research.

Keywords: Synthetic lethality, Genome organization, Deletional robustness, Gene duplication, Prokaryotes, Eukaryotes

Significance statement: Unraveling the organizing principles underlying gene arrangements on prokaryotic and eukaryotic genomes is one of the most fundamental questions of research in evolutionary biology. One understudied aspect of this organization is the relative chromosomal arrangement of synthetic lethal gene pairs. In this study, by analyzing a wealth of recently available genetic interaction data on several model organisms and cancer cell lines, we provide evidence that synthetic lethal gene pairs tend to be repulsed in bacterial genomes, while they tend to be attracted to each other on eukaryotic chromosomes. These intriguing observations can open the door towards detailed understanding of the evolutionary forces driving gene arrangements both in prokaryotic and eukaryotic genomes.

Introduction

Synthetic lethal interactions occur when simultaneous perturbations of two genes cause cellular or organismal death, whilst perturbation of individual genes alone does not affect viability (Nijman 2011). Their profound impact on cellular viability and organismal survival, raises the possibility that various aspects of synthetic lethal gene interactions might have been optimized under strong selective pressure in the course of evolution (Dixon et al. 2008). Previously, some features of synthetic lethal gene pairs such as their co-functionality (Srivas et al. 2016), subcellular colocalization (Wong et al. 2004) and conserved patterns of protein-protein interactions (Benstead-Hume et al. 2019) have been investigated. However, chromosomal arrangement of synthetic lethal gene pairs remains poorly studied, in spite of the well-known functional relevance of the arrangement of genes on chromosomes both in prokaryotes (Adams et al. 1992) and in eukaryotes (Trowsdale 2002).

In a previous study (Hosseini & Wagner 2018), the so-called "repulsion of synthetic lethal gene pairs" was observed in bacterial genomes. This implies that synthetic lethal gene pairs are rarely located in close proximity but rather tend to be positioned far from each other on bacterial chromosomes. However, this observation was limited to metabolic genes and synthetic lethal interactions were predicted solely using a computational method called Flux Balance Analysis (FBA) (Orth et al. 2010), without direct experimental evidence. Recent technological advancements have enabled large-scale genetic screens in *E.coli* (Typas et al. 2008; Butland et al. 2008), yeast (Costanzo et al. 2010; Frost et al. 2012; Costanzo et al. 2016) and human cells (Wong et al. 2016; Du et al. 2017; Shen et al. 2017; Horlbeck et al. 2018) resulting in thousands of experimentally validated synthetic lethal interactions that can be used for checking the validity of this computational analysis.

The previous study (Hosseini & Wagner 2018) provided evidence that the repulsion of synthetic lethal genes might be an evolutionary adaptation to pervasive gene deletions (Mira et al. 2001; Kunin & Ouzounis 2003; Nilsson et al. 2005; Lee et al. 2012; Koskiniemi et al. 2012; Albalat & Cañestro 2016) that exert significant selective pressure on bacterial genomes (Sung et al. 2016). If a pair of synthetic lethal genes is closely located on a chromosome (Figure 1a), they are likely to be hit simultaneously by a single deletional event and thus cause lethality. In contrast, if they are located far from each other (i.e. repulsed, Figure 1b), only one of the genes is likely to be hit during a deletional event, which consequently cannot cause lethality on its own, and so the cell remains robust to such deletional events.

The breadth of evidence on the role of large DNA deletions in the evolution of eukaryotic genomes is not comparable to that of prokaryotes. Therefore, we hypothesized that repulsion of synthetic lethal genes would not be observed on eukaryotic chromosomes. Furthermore, eukaryotic genomes are strongly influenced by gene duplication events (Friedman & Hughes 2001; Qian & Zhang 2014), which generate closely located paralog gene pairs accounting potentially for a significant proportion of the synthetic lethal gene interactions (De Kegel & Ryan 2019). This raises the possibility of observing an opposite trend, which is the enrichment of closely located synthetic lethal gene pairs (i.e. attraction), in eukaryotic chromosomes.

Therefore, in this study, we tested these hypotheses by using the synthetic lethal interactions curated in BioGRID (Oughtred et al. 2018) and CGIdb (Han et al. 2019) databases.

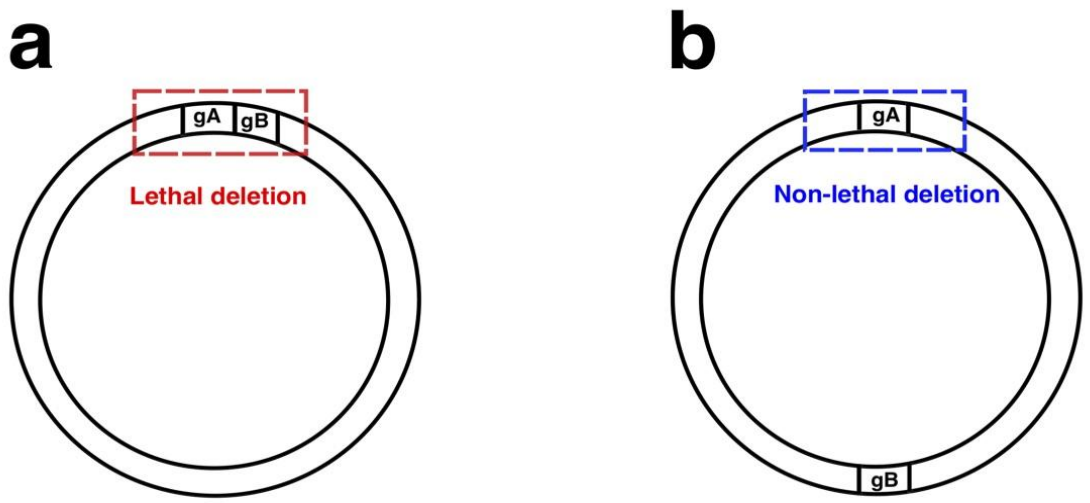

Fig. 1. Repulsion of synthetic lethal gene pairs and deletional robustness. The circles represent bacterial chromosomes in which a given pair of synthetic lethal genes are in attraction (panel **a**) and repulsion (panel **b**). Deletion of a DNA segment of a given size displayed as red dotted rectangle in panel **a**, which hits both genes (gA and gB) causes the cell to die (lethal deletion), while deletion of a DNA segment of the same size displayed as blue dotted rectangle in panel **b**, which hits only one of the synthetic lethal gene pair (gA) cannot cause lethality (Non-lethal gene deletion).

Methods

Synthetic lethal gene interactions were obtained from the BioGRID database (Biological General Repository for Interaction Datasets; release 3.5.184) (Oughtred et al. 2018). We focused on model organisms having more than a thousand synthetic lethal interactions including *S. cerevisiae* (17,046 gene pairs), *S. pombe* (1,570 gene pairs), and *H. sapiens* (2,036 gene pairs). Considering the fact that human chromosome size is on average more than a hundred times larger than that of yeast, we did not find the sample size of human synthetic lethal genes sufficient for our analysis, so we complemented it with the synthetic lethal interactions (58,268 gene pairs) identified in various cancer cell lines from CGIdb (Cancer Genetic Interaction db (Han et al. 2019)), resulting in a collection of 59,873 unique synthetic lethal gene pairs in human cells. There are no reported synthetic lethal interactions for prokaryotes, so instead we used negative genetic interactions in *E-coli K-12 substr. W3110* (89,465 gene pairs) in the BioGRID database, especially those (25,812 gene pairs) with high negative genetic interaction scores (S scores < -20) as a proxy for the synthetic lethal interactions. S scores are generated from Synthetic Genetic Arrays (SGA) as described in (Butland et al. 2008).

The chromosomal location and genomic coordinates for each gene were retrieved using rentrez R package v1.2.2 (Winter 2017) based on the entrez_id provided in the BioGRID and CGIdb datasets. The whole genome annotation for each organism was obtained using the Bioconductor R packages, AnnotationDBI based on the corresponding TxDB and org.db for each organism (Gentleman et al. 2004; Pages et al. 2018).

To generate the null distribution for a given species, we enumerated all of its possible gene pairs and calculated the genomic distance among the gene pairs located on the same chromosome. To calculate the expected number of synthetic lethal gene pairs located on the same chromosome, under the null hypothesis, we multiplied the fraction of gene pairs on the same chromosome in the null distribution by the number of synthetic lethal gene pairs in the given species. Similarly, to calculate the expected number of synthetic lethal gene pairs located in a genomic distance of shorter than a given threshold (e.g. 25 KB), under the null hypothesis, we multiplied the fraction of gene pairs with genomic distance shorter than the given threshold among those on the same chromosome, by the number of synthetic lethal gene pairs located on the same chromosome. Finally, to calculate the statistical significance of the difference between the expected and the observed values, we used Chi-squared test.

Paralogous genes were obtained from PANTHER classification system (Protein ANalysis THrough Evolutionary Relationships; release 15.0) (Mi et al. 2019). UniprotKB identifiers from PANTHER dataset were mapped to entrez_id using "Retrieve/ID mapping" web-based tool from Uniprot (Consortium 2019), except for *S. pombe,* where Pombase ID were used directly from the PANTHER platform.

All the statistical analyses and data visualization were done using R software version 3.6.1.

Results

Evidence for the "repulsion" of synthetic lethal gene pairs on the chromosome of *E. coli K-12*

First, we aimed to check the validity of the repulsion of synthetic lethal genes in prokaryotes as proposed in (Hosseini & Wagner 2018). To do so, we checked the distribution of genomic distances among the synthetic lethal gene pairs in *E-coli K-12* (25,812 pairs with S-score < -20; see methods). We observed that the overall distribution of the synthetic lethal gene pairs is almost the same as the null distribution (i.e. distribution of genomic distances among all possible gene pairs in *E-coli K-12;* Figure 2a). However, when we zoom in on the short-ranged distances (lower than 100 KB), where the so-called repulsion can potentially lie, we found a significant depletion of synthetic lethal gene pairs in very short genomic distances (<25 KB) (Figure 2b). Only 9 synthetic lethal gene pairs had a genomic distance < 25 KB, compared to 292.85 gene pairs expected by chance (Figure 2c). Therefore, in line with the previous study, we observed a significant repulsion of synthetic lethal genes in genomic distances of up to 25 KB in *E-coli K-12* ($\chi^2 = 275.12$ and p-value $< 10^{-61}$). Interestingly, when we repeated this analysis using all negative genetic interactions, including even the mild ones (89,465 gene pairs with S score < 0), we can still observe a significant repulsion of such gene pairs, albeit in a shorter range (up to 10 KB; Figure 2d-f).

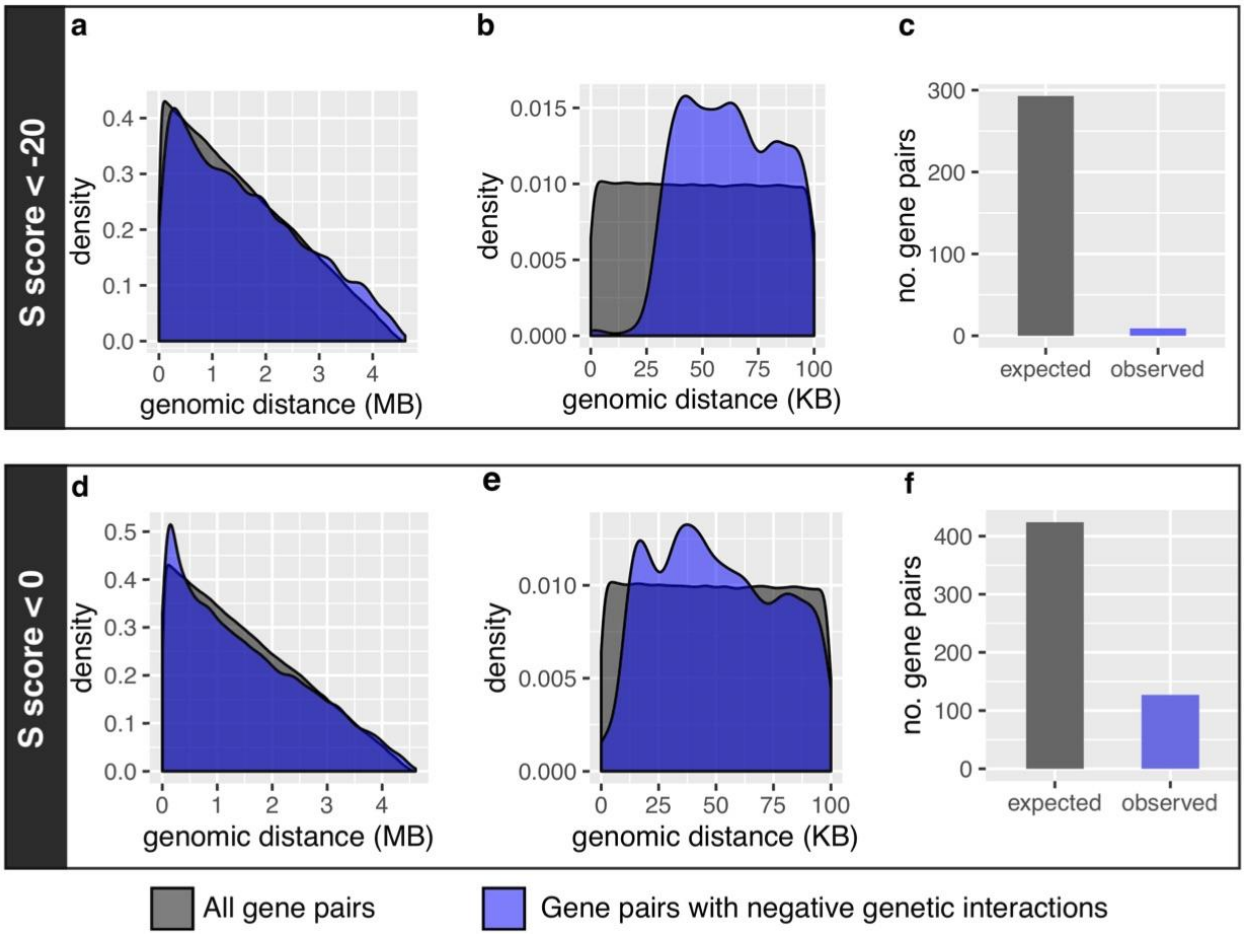

Fig. 2. Chromosomal arrangement of synthetic lethal gene pairs in prokaryotes (*E. coli K12*). Upper panels (a-c) correspond to synthetic lethal gene pairs (genetic interaction S-score < -20), while lower panels (d-f), correspond to all gene pairs with negative genetic interactions (S-score < 0). Distribution of the genomic distance among all potential pairs of genes (grey) versus gene pairs with negative genetic interaction (blue) in **a, d)** all genomic distances and **b, e)** short range genomic distances (up to 100 KB). It is important to notice the very low frequency of negatively interacting gene pairs within genomic distance of shorter than 25 KB and 10 KB respectively in panels b and e. Expected (grey) versus observed (blue) number of negatively interacting gene pairs **c)** within genomic distance of shorter than 25 KB (expected: 292.85, observed: 9, $\chi^2 = 275.12$ and p-value $< 10^{-61}$), and **f)** within genomic distance of shorter than 10 KB (expected: 423.9, observed: 127, $\chi^2 = 207.95$ and p-value $< 10^{-46}$).

No repulsion of synthetic lethal gene pairs was observed on *S. pombe* chromosomes

Next, we aimed to investigate the arrangement of synthetic lethal gene pairs in eukaryotic chromosomes. First, we examined the simplest one in our dataset, namely *S. pombe*, which has only three chromosomes. We initially aimed to check whether synthetic lethal gene pairs tend to be located on different chromosomes or on the same one (i.e. repulsion or attraction on the inter-chromosomal level). We did not find significant depletion or enrichment of synthetic lethal gene pairs located on the same chromosome. We observed that 590/1570 (37.58%) synthetic lethal gene pairs, compared to 580.16 (36.95%) expected by the null distribution (i.e. all possible gene pairs), were located on the same chromosome (Figure 3a), which indicates no significant repulsion (or attraction) on the inter-chromosomal level ($\chi^2 = 0.16$ and p-value=0.68). Furthermore, analysis of the synthetic lethal pairs on the same chromosome indicates that the observed distribution of synthetic lethal pairs is not significantly different from the null distribution, both overall (Figure 3b), and also in the short-range genomic distances (Figures 3c-d). Thus, synthetic lethal gene pairs are neither repulsed nor attracted in *S. pombe* chromosomes.

Evidence for the "attraction" of synthetic lethal gene pairs on *S. cerevisiae* chromosomes

Next, we examined *S. cerevisiae*, which has 16 distinct chromosomes, and initially checked whether synthetic lethal gene pairs tend to be located on the same chromosome or on different ones. We observed that 1418/17046 (8.32%) synthetic lethal gene pairs, compared to 1227.9 (7.2%) expected by the null distribution (i.e. all possible gene pairs), were located on the same chromosome (Figure 3e). Thus, synthetic lethal gene pairs are to some extent attracted to each other on the inter-chromosomal level in *S. cerevisiae* ($\chi^2 = 29.43$ and p-value $< 10^{-7}$). Next, we checked the distribution of genomic distance of synthetic lethal gene pairs located on the same chromosome. Although the overall distribution of the synthetic lethal gene pairs looks almost the same as the null distribution (Figure 3f), we observed significant enrichment of synthetic lethal interactions in shorter genomic distances (Figure 3g). Specifically, 186 gene pairs are located in close proximity (i.e. with genomic distance <25 KB), which is significantly higher than what is expected (81.95) from the null distribution ($\chi^2 = 132.11$and p-value $< 10^{-29}$; Figure 3h). Thus, our observations not only invalidated the repulsion of synthetic lethality in *S. cerevisiae,* but also revealed an opposite trend (attraction).

Strong evidence for the "attraction" of synthetic lethal gene pairs on *H. sapiens* chromosomes

Finally, we checked the distribution of synthetic lethal genes on human chromosomes. We identified 4297/59873 (7.17%) synthetic lethal gene pairs located on the same chromosome, which is significantly higher than what is expected (3091.1 (5.16%)) from the null distribution ($\chi^2 = 470.46$ and p-value $< 10^{-103}$. Figure 3i). Thus, synthetic lethal genes are attracted to each other on the inter-chromosomal level in human cells. We then checked the distribution of the genomic distance of synthetic lethal genes located on the same chromosome. We observed striking enrichment of closely located synthetic lethal gene pairs (Figures 3j and 3k). Specifically, we identified 143 synthetic lethal gene pairs in close proximity (i.e. with genomic distance <25 KB), which is substantially higher than expected (5.79) from the null distribution ($\chi^2 = 3249.07$ and p-value $< 10^{-300}$; Figure 3l). Thus, our results indicate that synthetic lethal gene pairs are attracted to each other on human chromosomes.

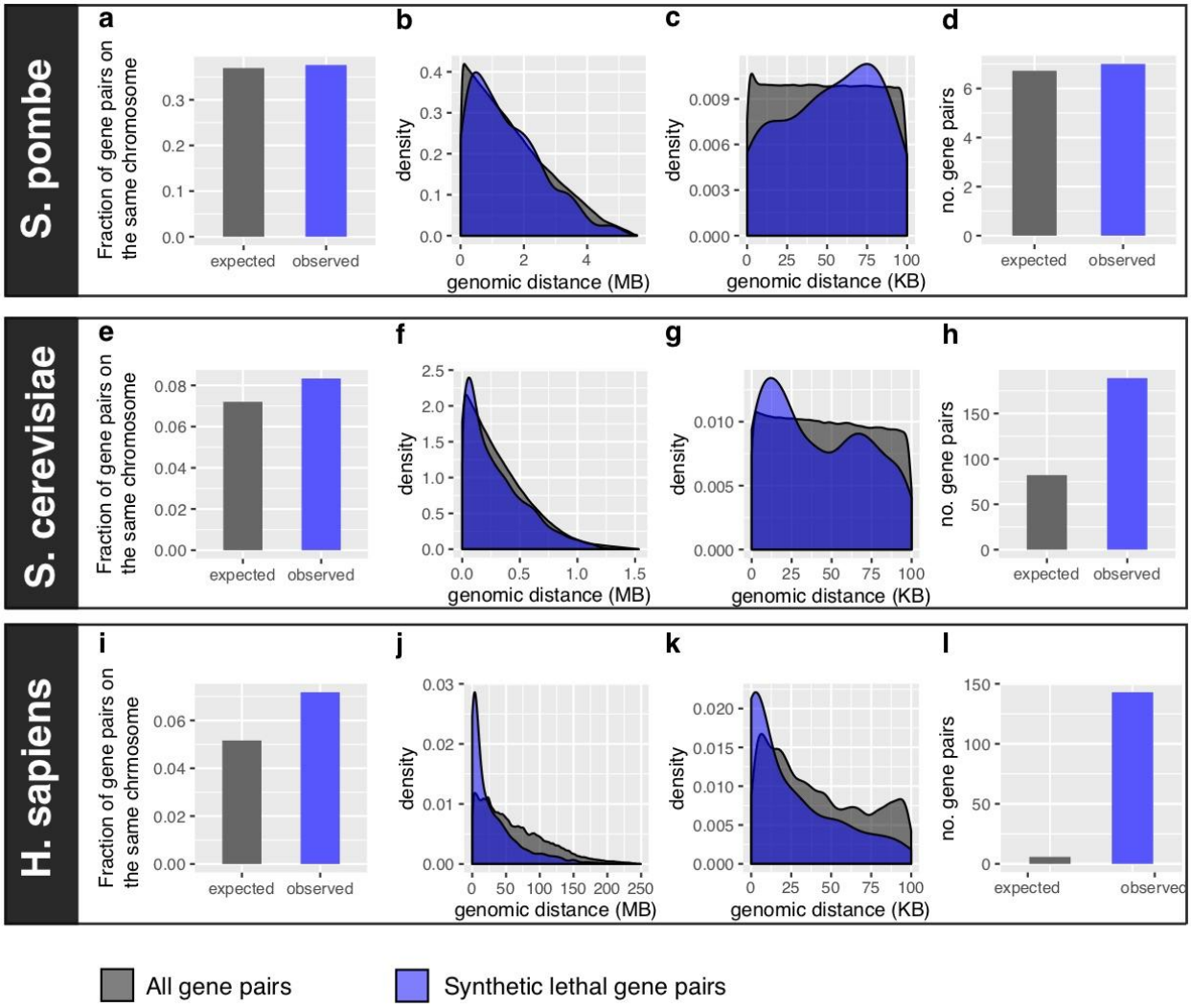

Fig. 3. Chromosomal arrangement of synthetic lethal gene pairs in eukaryotes (yeast and human). Upper panels **(a-d)**, middle panels **(e-h)** and lower panels **(i-k)** correspond to *S. pombe*, *S. cerevisiae*, and *H. sapiens* respectively. The leftmost panel in each row shows the expected (grey) versus the observed (blue) fraction of synthetic lethal gene pairs on the same chromosome: **a)** expected: 580.16 pairs, observed: 590 (among total 1,570 synthetic lethal pairs), $\chi^2 = 0.16$ and p-value $= 0.68$ **e)** expected: 1,227.9 pairs, observed: 1,418 (among total 17,046 synthetic lethal pairs), $\chi^2 = 29.43$ and p-value $< 10^{-7}$, and **i)** expected: 3091.1 pairs, observed: 4,297 (among total 59,873 synthetic lethal pairs), $\chi^2 = 470.46$ and p-value $< 10^{-103}$. The middle panels in each row illustrate the distribution of the genomic distance among all potential pairs of genes (grey) versus synthetic lethal gene pairs (blue) in all genomic distances **(b, f, j)** and short-range genomic distances (up to 100 KB) **(c, g, k)**. The rightmost panel in each row shows the expected (grey) versus the observed (blue) number of synthetic lethal gene pairs within genomic distance of shorter than 25 KB: **d)** expected: 6.72, observed: 7, and $\chi^2 = 0.01$ and p-value $= 0.91$, **h)** expected: 81.95, observed: 186, $\chi^2 = 132.11$ and p-value $< 10^{-29}$, and **l)** expected: 5.79, observed: 143, $\chi^2 = 3249.07$ and p-value $< 10^{-300}$.

Gene duplication in yeast and humans is not the only source of the observed attraction of synthetic lethal gene pairs

Next we aimed to check whether the observed attraction of synthetic lethal gene pairs in eukaryotes (*S. cerevisiae* and human) can be explained by the prominence of gene duplication events in these organisms. We identified all paralog gene pairs for the three eukaryotic species in this study (1,045 in *S. pombe*, 1,595 in *S. cerevisiae*, and 33,971 in the human. Table 1). We confirmed that paralog pairs indeed tend to be located on the same chromosome, most notably in the human, where 30.1% of paralog pairs are located on the same chromosome, which is considerably higher than expected (5.16%) by chance (Table 1). Moreover, paralog pairs located on the same chromosome in general tend to be close to each other. Specifically, the fraction of paralog pairs located within genomic distance shorter than 25 KB is considerably higher (4.2% in *S. pombe*, 24.8% in *S. cerevisiae*, and 7.9% in human; Table 1) than expected by chance (1.1% in *S. pombe*, 5.8% in *S. cerevisiae*, and 0.1% in human).

Next, we tried to quantify the contribution of paralog genes on synthetic lethality and their observed attraction. We noticed that a considerable percentage of paralog gene pairs, most notably in *S. cerevisiae* (38.4%), are synthetic lethal in our dataset. However, the paralog genes account only for about 4% of the total synthetic lethal gene pairs in all three species (Table 1). More importantly, paralog gene pairs account for 3.6% and 12% of the synthetic lethal genes located on the same chromosome, respectively in *S. cerevisiae* and the human (Table 1). Thus, even if we simply remove paralog gene pairs from the pool of known synthetic lethal genes, the tendency of synthetic lethal gene pairs to be located on the same chromosome remains statistically significant (observed: 1370, expected: 1183.8, $\chi^2 = 28.35$ and p-value $< 10^{-6}$ in *S. cerevisiae*, and observed: 3781, expected: 2972.1, $\chi^2 = 220.2$ and p-value $< 10^{-49}$ in human). Similarly, we observed that only 7/186 (3.8%) and 23/143 (16.1%) of the closely located synthetic lethal gene pairs (with genomic distance < 25 KB) are paralogs, and hence if we remove them from the pool of known synthetic lethal genes in both species, the attraction of synthetic lethal gene pairs still remains statistically significant (observed: 79, expected: 179, and $\chi^2 = 126.58$ and p-value $< 10^{-28}$ in *S. cerevisiae*, and observed: 5.1, expected: 120, $\chi^2 = 2588.63$ and p-value $< 10^{-300}$ in human).

Thus, we conclude that although paralog gene pairs tend to be among synthetic lethal genes and tend to be located on the same chromosome and are more closely located than expected by chance, the observed attraction of synthetic lethal gene pairs on *S. cerevisiae* and human chromosomes cannot be explained solely by gene duplication.

Species	*S. pombe*	*S. cerevisiae*	*H. sapiens*
Synthetic lethal gene pairs			
Number of synthetic lethal gene pairs	1,570	17,046	59,873
Number of synthetic lethal gene pairs located on the same chromosome	590	1,418	4,297
Percentage of synthetic lethal gene pairs that are located on the same chromosome	37.5%	8.3%	7.2%
Number of synthetic lethal gene pairs located within genomic distance <25 KB	7	186	143
Percentage of synthetic lethal gene pairs located within genomic distance <25 KB (among synthetic lethal gene pairs located on the same chromosome)	1.2%	13.1%	1.7%
Paralog gene pairs			
Number of paralog pairs	1,045	1,595	33,971
Number of paralog pairs located on the same chromosome	400	153	10225
Percentage of paralog pairs that are located on the same chromosome	38.3%	9.6%	30.1%
Number of paralog pairs located within genomic distance <25 KB	17	38	813
Percentage of paralog pairs located within genomic distance <25 KB (among paralog pairs located on the same chromosome)	4.2%	24.8%	7.9%
Synthetic lethal gene pairs vs. paralogs			
Number of synthetic lethal paralog pairs	67	612	2,305
Percentage of paralog pairs that are synthetic lethal	6.4%	38.4%	6.8%
Percentage of synthetic lethal gene pairs that are paralogs	4.3%	3.6%	3.8%
Number of synthetic lethal paralog pairs located on the same chromosome	36	51	516
Percentage of synthetic lethal paralog pairs that are located on the same chromosome	53.7%	8.3%	22.4%
Percentage of synthetic lethal gene pairs located on the same chromosome that are paralog	6.1%	3.6%	12.0%
Number of synthetic lethal paralogs located within genomic distance <25 KB	1	7	23
Percentage of synthetic lethal gene pairs located within genomic distance <25 KB that are paralogs	14.3%	3.8%	16.1%

Table 1. Contribution of paralogs on chromosomal arrangements of synthetic lethal gene pairs in three eukaryotic species (*S. pombe*, *S. cerevisiae* and *H. sapiens*).

Discussion

It is well known that the arrangement of genes on chromosomes both in prokaryotes and in eukaryotes is not random, but it is rather like a piece of art portrayed over the course of evolution (Adams et al. 1992; Trowsdale 2002). Synthetic lethal interactions have direct impact on cellular survival and so it is not unexpected that the relative arrangement of synthetic lethal gene pairs on chromosomes reveals special evolutionary patterns. In a previous study (Hosseini & Wagner 2018), using a computational approach, it was observed that in bacterial genomes, the synthetic lethal metabolic gene pairs are repulsed. In the present study, we provide experimentally validated evidence confirming that the repulsion of synthetic lethal gene pairs is indeed a property of *E. coli K12* genome. Therefore, the observed repulsion is not just a peculiarity of metabolic genes that the previous work was focused on or an artefact of the Flux Balance Analysis computational method (Orth et al. 2010) used in the previous study. Instead, our present analysis of *E. coli K12* genome, suggests the generalizability of this interesting observation, regardless of the gene function (metabolic or non-metabolic) and irrespective of the particular method used for synthetic lethality determination (experimental or computational).

Moreover, it is tempting to generalize this observation beyond the prokaryotic species, and so to check this we leveraged the wealth of synthetic lethality information accumulated in eukaryotic species, particularly in yeast and the human. Jacques Monod assertion on biological unity, "*anything found to be true of E. coli must also be true of Elephants*" (Friedmann 2004), seems encouraging for this generalization purpose, but it needs to be taken with a grain of salt, at least because exceptions pervade

biological rules in general. As a prominent example, the operon organization, an *E. coli* based discovery that earned Jacques Monod the Nobel prize, has rarely been observed in eukaryotes.

More importantly, the evolutionary forces operating on prokaryotic and eukaryotic genomes in some respects are fundamentally different. For example, whereas there is ample evidence attesting to the importance of large-scale gene deletions on the evolution of prokaryotic genomes (Mira et al. 2001; Kunin & Ouzounis 2003; Nilsson et al. 2005; Lee et al. 2012; Koskiniemi et al. 2012; Albalat & Cañestro 2016; Sung et al. 2016), such evidence for eukaryotic ones is rare. Therefore, if the repulsion of synthetic lethal gene pairs is a signature of adaptation to pervasive deletional events as the previous study claimed (Hosseini & Wagner 2018), it must remain as a particular characteristic of the prokaryotic genomes. Observing this phenomenon in eukaryotes would thus be an evidence against the previous "deletional robustness centered" claim (Hosseini & Wagner 2018) and would raise doubts that the observed repulsion could be a signature of an unknown selective force that must be common between prokaryotes and eukaryotes.

Our results indeed revealed evidence against the repulsion of synthetic lethal gene pairs in eukaryotes. Interestingly, we observed even an opposite trend in *S. cerevisiae* and the human, namely the enrichment of closely located synthetic lethal gene pairs (i.e. attraction). We hypothesize that gene duplication events, which result in closely located paralog gene pairs, might explain the attraction of synthetic lethal gene pairs in eukaryotic chromosomes. However, we showed that although paralog gene pairs tend to be synthetic lethal more often and are more closely located than expected by chance, gene duplication on its own cannot explain the attraction of synthetic lethal gene pairs.

Hence, the evolutionary origin of the observed "attraction" remains an open question. Specifically, our analysis is agnostic on whether the attraction of synthetic lethal gene pairs on its own confers selective advantage in eukaryotes or it is simply a by-product of other evolutionary processes, driven for example by the need for colocalization of functionally related genes or co-regulation of gene pairs that later became synthetic lethal. This is a question that is beyond the scope of our present analysis, and so is left for future studies to be systematically addressed.

In summary, our observations provide evidence in support of the repulsion of synthetic lethal gene pairs in prokaryotes (*E. coli*) most probably as an adaptation to ongoing gene deletions. In contrast, our findings raise the possibility of the attraction of synthetic lethal gene pairs in eukaryotes (at least in *S. cerevisiae* and human cells), which is not simply explained by the frequent gene duplication events in eukaryotes and thus calls for further research.

Acknowledgements

This work was supported by Open Targets [grant number OT2062 to S-R.H]; and EMBL-EBI [to E.P and B.W].

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

References

Adams RLP et al. 1992. The arrangement of genes. In: The Biochemistry of the Nucleic Acids. Springer Netherlands pp. 317–338. doi: 10.1007/978-94-011-2290-0_8.

Albalat R, Cañestro C. 2016. Evolution by gene loss. Nat. Rev. Genet. 17:379–391. doi: 10.1038/nrg.2016.39.

Benstead-Hume G et al. 2019. Predicting synthetic lethal interactions using conserved patterns in protein interaction networks. PLoS Comput. Biol. 15. doi: 10.1371/journal.pcbi.1006888.

Butland G et al. 2008. eSGA: E. coli synthetic genetic array analysis. Nat. Methods. 5:789–795. doi: 10.1038/nmeth.1239.

Consortium TU. 2019. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45:D158–D169.

Costanzo M et al. 2016. A global genetic interaction network maps a wiring diagram of cellular function. Science (80-.). 353:42–65. doi: 10.1126/science.aaf1420.

Costanzo M et al. 2010. The genetic landscape of a cell. Science (80-.). 327:425–431. doi: 10.1126/science.1180823.

Dixon SJ et al. 2008. Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes. Proc. Natl. Acad. Sci. U. S. A. 105:16653–16658. doi: 10.1073/pnas.0806261105.

Du D et al. 2017. Genetic interaction mapping in mammalian cells using CRISPR interference. Nat. Methods. 14:577–580. doi: 10.1038/nmeth.4286.

Friedman R, Hughes AL. 2001. Gene duplication and the structure of eukaryotic genomes [1]. Genome Res. 11:373–381. doi: 10.1101/gr.155801.

Friedmann HC. 2004. From 'Butyribacterium' to 'E. coli': An essay on unity in biochemistry. Perspect. Biol. Med. 47:47–66. doi: 10.1353/pbm.2004.0007.

Frost A et al. 2012. Functional repurposing revealed by comparing S. pombe and S. cerevisiae genetic interactions. Cell. 149:1339–1352. doi: 10.1016/j.cell.2012.04.028.

Gentleman RC et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5:R80. doi: 10.1186/gb-2004-5-10-r80.

Han Y et al. 2019. Genetic Interaction-Based Biomarkers Identification for Drug Resistance and Sensitivity in Cancer Cells. Mol. Ther. - Nucleic Acids. 17:688–700. doi: 10.1016/j.omtn.2019.07.003.

Horlbeck MA et al. 2018. Mapping the Genetic Landscape of Human Cells. Cell. 174:953-967.e22. doi: 10.1016/j.cell.2018.06.010.

Hosseini S-R, Wagner A. 2018. Genomic organization underlying deletional robustness in bacterial metabolic systems. Proc. Natl. Acad. Sci. U. S. A. 115:7075–7080. doi: 10.1073/pnas.1717243115.

De Kegel B, Ryan CJ. 2019. Paralog buffering contributes to the variable essentiality of genes in cancer cell lines Dudley, AM, editor. PLOS Genet. 15:e1008466. doi: 10.1371/journal.pgen.1008466.

Koskiniemi S, Sun S, Berg OG, Andersson DI, Boxer D. 2012. Selection-Driven Gene Loss in Bacteria Casadesús, J, editor. PLoS Genet. 8:e1002787. doi: 10.1371/journal.pgen.1002787.

Kunin V, Ouzounis CA. 2003. The Balance of Driving Forces During Genome Evolution in Prokaryotes. Genome Res. 13:1589–1594. doi: 10.1101/gr.1092603.

Lee M-C, Marx CJ, Lenski R, Sivam D, Lidstrom M. 2012. Repeated, Selection-Driven Genome Reduction of Accessory Genes in Experimental Populations Moran, NA, editor. PLoS Genet. 8:e1002651. doi: 10.1371/journal.pgen.1002651.

Mi H et al. 2019. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). Nat. Protoc. 14:703–721. doi: 10.1038/s41596-019-0128-8.

Mira A, Ochman H, Moran NA. 2001. Deletional bias and the evolution of bacterial genomes. Trends Genet. 17:589–96. http://www.ncbi.nlm.nih.gov/pubmed/11585665 (Accessed February 4, 2014).

Nijman SMB. 2011. Synthetic lethality: General principles, utility and detection using genetic screens in human cells. FEBS Lett. 585:1–6. doi: 10.1016/j.febslet.2010.11.024.

Nilsson AI et al. 2005. Bacterial genome size reduction by experimental evolution. Proc. Natl. Acad. Sci. U. S. A. 102:12112–6. doi: 10.1073/pnas.0503654102.

Orth JD, Thiele I, Palsson BØ. 2010. What is flux balance analysis? Nat. Biotechnol. 28:245–8. doi: 10.1038/nbt.1614.

Oughtred R et al. 2018. The BioGRID interaction database: 2019 update. Nucleic Acids Res. 47:529–541. doi: 10.1093/nar/gky1079.

Pages H, Carlson M, Falcon S, Li N. 2018. AnnotationDbi: annotation database interface. R Packag. version 1.42.1. 1471–2164. doi: 10.18129/B9.bioc.AnnotationDbi.

Qian W, Zhang J. 2014. Genomic evidence for adaptation by gene duplication. Genome Res. 24:1356–1362. doi: 10.1101/gr.172098.114.

Shen JP et al. 2017. Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions. Nat. Methods. 14:573–576. doi: 10.1038/nmeth.4225.

Srivas R et al. 2016. A Network of Conserved Synthetic Lethal Interactions for Exploration of Precision Cancer Therapy. Mol. Cell. 63:514–525. doi: 10.1016/j.molcel.2016.06.022.

Sung W et al. 2016. Evolution of the Insertion-Deletion Mutation Rate Across the Tree of Life. G3: Genes|Genomes|Genetics. 6:2583–2591. doi: 10.1534/g3.116.030890.

Trowsdale J. 2002. The gentle art of gene arrangement: The meaning of gene clusters. Genome Biol. 3:comment2002.1. doi: 10.1186/gb-2002-3-3-comment2002.

Typas A et al. 2008. High-throughput, quantitative analyses of genetic interactions in E. coli. Nat. Methods. 5:781–787. doi: 10.1038/nmeth.1240.

Winter DJ. 2017. rentrez: An R package for the NCBI eUtils API. R J. 9:520–526. doi: 10.32614/rj-2017-058.

Wong ASL et al. 2016. Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. Proc. Natl. Acad. Sci. U. S. A. 113:2544–2549. doi: 10.1073/pnas.1517883113.

Wong SL et al. 2004. Combining biological networks to predict genetic interactions. Proc. Natl. Acad. Sci. U. S. A. 101:15682–15687. doi: 10.1073/pnas.0406614101.