Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Investigating the genomic basis of metabolic robustness through *in silico* flux analysis

Marcin Imielinski, Niels Klitgord, Calin Belta

***Abstract*— We employ a novel implementation of flux balance analysis to investigate the role of genome structure in the maintenance of metabolic robustness. We propose the hypothesis that the genomic organization of a bacterium buffers its metabolome against random gene deletion. To test this hypothesis, we use a novel implementation of producibility analysis to determine the metabolomic impact of gene deletions in the *E. coli* iJR904 genome-scale metabolic model. From these results, we determine *metabolomic fragility*, which we compute as the average number of metabolites knocked out across all gene deletions of a given size in a given nutrient media. We apply this analysis for three deletion window sizes (4000, 8000, 16000bp) across the length of the *E. coli* genome. We compare these results to those obtained from several null distributions of permuted genomes to assess the impact of *E. coli* genome organization on its metabolic robustness. Our results strongly suggest that the arrangement of genes on the *E. coli* genome buffers metabolite producibility against random gene deletion. Our results have interesting implications for the understanding of metabolic network evolution. Future work includes examining our hypothesis for a wider range of deletion sizes and nutrient environments and extending our results to the metabolic networks of other species.**

I. INTRODUCTION

The metabolic network is the biochemical machinery with which a cell transforms a limited set of nutrients in its environment into the multitude of molecules required for growth and survival. It consists of hundred to thousands of small molecule species intricately linked by an even larger set of biochemical reactions. The expansive and highly connected nature of this important cellular system greatly limits the degree of insight that may be gained from the isolated study of a single component or module. The first step towards systems-level understanding of metabolism is the construction of a model that captures what is known regarding an organism's small molecule biochemistry and its underlying genetics. The advent of sequencing technology combined with general improvements in the organization of biological information [7], [10] has allowed the building of such *genome-scale metabolic models* for numerous microbial organisms, including *E. coli*, *S. cereviseae*, *H. pylori*, and *S. aureus* [13], [14], [4], [5], [11], [12], [8].

Genome-scale metabolic modeling enables the *in silico* study of the relationships of biological components and systems-level functions. It also allows for the examination of global features of biological systems that may not be evident through the study of isolated genes or pathways. One such systems-level feature is that of *robustness*, which represents biological systems ability to function in a wide range of environments and in the context of component failure. One particular important aspect of metabolic network robustness is its ability to buffer essential functions of the organism against random gene deletion.

Flux balance analysis provides a powerful tool to examine metabolic network robustness at the genome-scale [10]. A variant of flux balance analysis, called *producibility* analysis, employs linear programming to identify the metabolite knockouts that are predicted to result from a gene knockout, given the genome-scale model and a nutrient media [6]. This set of metabolite knockouts resulting from a gene deletion provides a global measure of that gene deletion's effect on network function, which we term as the *metabolomic impact*.

Producibility analysis in *E. coli* shows its biosynthetic function to be highly robust to single gene deletion in rich media. Alternatively stated, most single-gene deletions in this strain and nutrient media have no metabolomic impact [6]. This robustness is thought to arise at three levels: gene, protein, and pathway. Robustness at the gene level is attributed to gene duplication. Robustness at the protein level results from multiple enzymes performing identical functions. Pathway-based robustness occurs when multiple pathways in the metabolic network achieve the same objective.

In this study, we propose a new layer of mechanisms underlying *E. coli* metabolic robustness at the genome-scale. Namely, we postulate that the *position* of genes in the genome has evolved to buffer the organism against random deletions. To test this hypothesis, we apply a novel and efficient implementation of producibility analysis to evaluate the biosynthetic robustness of the *E. coli* metabolic network to random genomic deletion. By comparing these results to those obtained from "permuted genomes", we demonstrate that the position of genes in *E. coli* significantly protects metabolites against gene deletion. This result has interesting implications for the understanding of metabolic network evolution.

II. METHODS

A. Genome scale metabolic models

Notation For $n, i \in \mathbb{N}$, we use I_n to denote the $n \times n$ identity matrix, and $e^{n,i} \in \mathbb{R}^n$ to denote the i-th element of the Euclidean basis in $\mathbb{R}^n$. Given $m, n \in \mathbb{N}$, we use the notation $M = \{1, \ldots, m\}$ and $N = \{1, \ldots, n\}$. For a set C, we use $|C|$ to denote its cardinality. If $A \in \mathbb{R}^{m \times n}$ and

M. Imielinski is with the University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA imielins@mail.med.upenn.edu
N. Klitgord and C. Belta are with the Boston University Graduate Program in Bioinformatics, Boston, MA 02115, USA {niels,cbelta}@bu.edu

978-1-4244-3124-3/08/$25.00 ©2008 IEEE

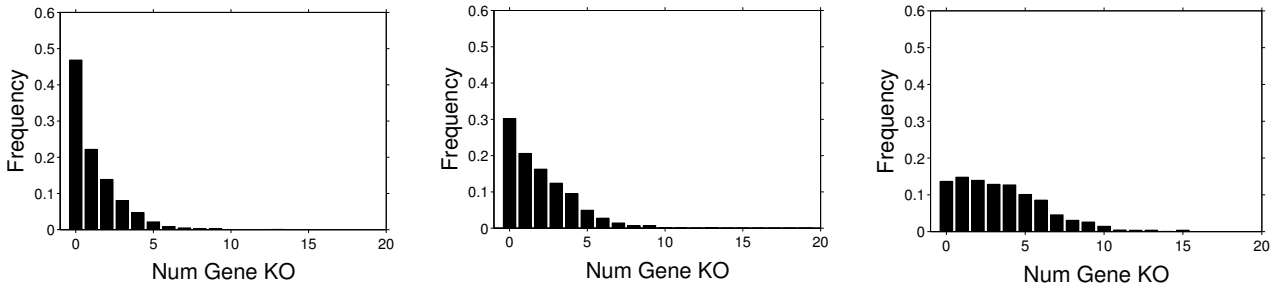

Fig. 1. Distributions of metabolic genes knocked out applying the three deletion window sizes in the original E. coli genome. (a) 4000 bp deletion size (b) 8000 bp deletion size (c) 16000 bp deletion size.

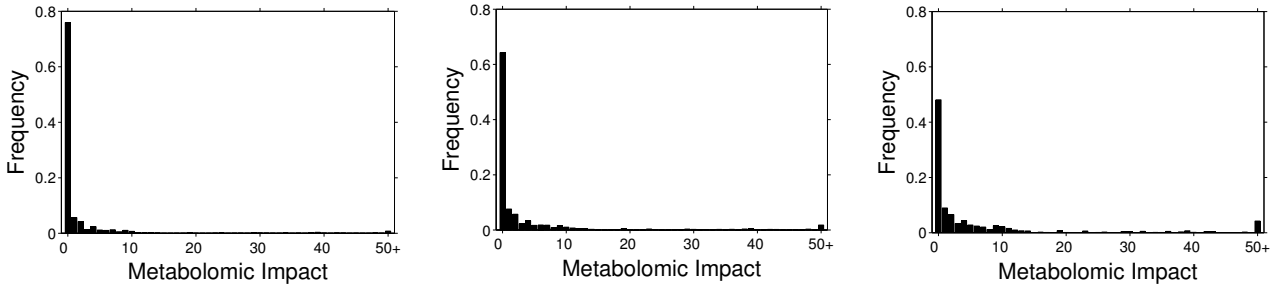

Fig. 2. Metabolomic impact histograms for the three deletion window sizes in the original E. coli genome. (a) 4000 bp deletion size (b) 8000 bp deletion size (c) 16000 bp deletion size.

$U \subseteq M$, then A_U denotes the submatrix of A containing the rows with indices in the set U. Therefore, if $x \in \mathbb{R}^n$, $i \in N$, and $U \subset N$, then x_i and $x_U \in \mathbb{R}^{|U|}$ denote its ith component and the vector formed by taking components with indices in set U, respectively. The inequality $x \geq 0$ is interpreted componentwise, *i.e.,* $x_i \geq 0$, $i = 1, \ldots, n$, while the inequality $x > 0$ is interpreted as $x \geq 0,\ x \neq 0$.

We represent a mass-balanced metabolic network of n chemical reactions involving m metabolites in a stoichiometry matrix $S \in \mathbb{R}^{m \times n}$. We assume that matrix S also incorporates stoichiometric information about all exchange reactions (uptake and secretion) and about the maintenance and growth reactions. Each entry S_{ij} specifies the stoichiometric coefficient for metabolite i in reaction j, which is negative for substrates and positive for products. We represent the flux distribution through the reactions of the network by $v \in \mathbb{R}^n$, where a component v_j corresponds to the flux of reaction complex passing through reaction j. We assume that all reactions are irreversible, *i.e.,* $v \geq 0$. Note that this assumption is not restrictive since, at the price of (significantly) increasing the number of reactions, each reversible reaction can be replaced by two irreversible ones. The concentrations of species in the system at time t are denoted by $x(t) \in \mathbb{R}^m_+$. Under these assumptions, the rate of change in time of species concentrations is given by:

$$\dot{x} = Sv,\ v \geq 0. \tag{1}$$

Metabolic reactions occur at a fast rate with respect to cell regulatory and environmental changes. When modelling at the slower time scale it is reasonable to apply the *quasi-steady state assumption*, under which we have:

$$Sv = 0,\ v \geq 0. \tag{2}$$

B. Producibility analysis

A metabolite $j \in M$ is called *producible* by the metabolic network (1) if the network can sustain its synthesis under the steady state and thermodynamic constraints in equation (2). To test producibility of metabolite j, we add a "fictitious" chemical reaction that uses metabolite j, and then test whether the network can produce strictly positive flux through this chemical reaction, while observing the steady state and thermodynamic constraints. Formally, this corresponds to an augmentation of the stoichiometry matrix S with $e^{m,j}$ on the right, and the flux vector v with $w \in \mathbb{R}$, $w \geq 0$ at the bottom. In other words, w corresponds to the flux through the additional chemical reaction. To characterize the production capabilities of a metabolic network globally, we can add one test chemical reaction for each metabolite, and correspondingly augment the stoichiometry matrix S with I_m on the right and v with a $w \in \mathbb{R}^n$, $w \geq 0$ at the bottom. We will denote the augmented matrix by $\bar{S} \in \mathbb{R}^{m \times (n+m)}$ and the augmented vector by $\bar{v} \in \mathbb{R}^{n+m}$.

In this framework, producibility of metabolite $j \in M$ can be decided by solving the following linear program (LP):

$$\max_{\bar{v} \in \mathbb{R}^{n+m}} \bar{v}_{j+n} \ \ s.t\ \ \bar{S}\bar{v} = 0,\ \bar{v} \geq 0,\ \bar{v}_{j+n} \leq \alpha, \tag{3}$$

where $\alpha > 0$ is an arbitrary constant that keeps the problem

bounded. If the optimal value of (3) is zero, then metabolite j is not producible. Otherwise, metabolite j is producible*. To test the producibility of a metabolite under metabolic gene knockouts, one can simply add the equality constraints $\bar{v}_J = 0$ to (3), where J is the set of reactions knocked out by the gene knockouts.

In order to asses the robustness of metabolism, we are interested in studying the producibility of each metabolite in a metabolic network under various metabolic gene knockouts. With the framework we presented so far, this would imply solving a number of LP equal to the number of gene perturbations multiplied the number of metabolites. Since the size of each LP is in the order of thousands, and the number of gene perturbations is in the order of millions, such an approach would be infeasible computationally. In this paper, we propose the reduce the amount of computation by using both problem (3) and its dual, whose solutions are available simultaneously if an interior-point-type optimization method (such as the one implemented in SeDuMi [1]) is used. In the following we show that, in most of the cases, it is enough to solve one LP to decide the producibility of all metabolites.

First, whenever we find a solution $\bar{v}$ of (3), we search it for all components for which $\bar{v}_{n+i} > 0$. If such a component is found, then we label metabolite i as producible. Second, the dual of (3) is given by:

$$\min_{g \in \mathbb{R}^m, s \in \mathbb{R}} \alpha s \;\; s.t. \;\; \bar{S}^T g + s e^{n+m, j+n} \geq e^{n+m, j+n}, \; s \geq 0, \tag{4}$$

where s is a "slack" variable. It is easy to see that the primal problem (3) can only produce two optimal values of its objective: 0 and α. Therefore, the optimal value of the dual problem (4) can only be 0 or α, which implies that every dual solution $g \in \mathbb{R}^m$ satisfies $\bar{S}^T g \geq 0$. In our previous work [6], by employing a simple variation of the Farkas' Lemma, we showed that g satisfying $(\bar{S}^T g)_{n+i} > 0$, $\bar{S}^T g \geq 0$ certifies the infeasibility of the set $\{\bar{S}\bar{v} = 0, \bar{v} \geq 0, \bar{v}_{n+i} > 0\}$. Therefore, by a simple inspection of the solution g of the dual problem (4), we can classify all metabolites i for which $(\bar{S}^T g)_{n+i} > 0$ as "non-producible".

C. E. coli Genome-Scale Metabolic Model

In this study, we employ the *E. coli* iJR904 genome scale model [12], which has 761 metabolites involved in 931 reversible and irreversible chemical reactions. We model rich media by supplementing the network with 143 extracellular nutrient fluxes, 5 intracellular nutrient fluxes representing supply of cofactors and carrier proteins from outside of small molecular metabolism, and 761 species sink fluxes for each species in the system. This yields a stoichiometry matrix S (Equation 2) of dimension 761×2085.

Reactions in S represent the inflow, outflow, and interconversion of small-molecule chemical species in an *E. coli* cell grown in a rich nutrient media. Each species "sink" reaction represents its growth mediated dilution and macromolecular consumption. Of the 761 species, 49 correspond to "biomass components" that are considered to be essential substrates for survival and growth in the original Reed *et al* specification [12].

The *E. coli* genome-scale metabolic model contains a detailed mapping of gene, protein, and reaction (GPR) relations that facilitate the mapping of gene knockouts to reaction knockouts. Furthermore, each *E. coli* metabolic gene has a precise start and end site in the genome according to the *E. coli* genome build [2]. Finally, *E. coli* genes are organized into transcriptional units called operons that comprise multiple genes that are controlled by a single promoter.

D. Metabolomic impact and fragility

We define the *metabolomic impact* of a genomic deletion as the number of metabolites it renders non-producible. We compute metabolomic impact through a multi-step process: First, we determine the operons contained in the deletion region, given its base pair location and extent. Then for each operon, we determine the genes that lie downstream to the deletion and label these as knocked out. For that given gene deletion combination, we use GPR associations to calculate the set of reaction knockouts. Finally, given the set of reaction knockouts, we apply producibility analysis to determine that gene deletion's metabolomic impact.

We define the metabolomic fragility of a genome to a given size deletion as the mean metabolomic impact across all deletions of a given base pair size. We compute metabolomic fragility with a "sliding window" approach. Namely, we map all possible genomic deletion of a given size to the reactions that they are predicted to knockout. We then apply producibility analysis of each unique reaction knockout set to assess its metabolomic impact on the network. Taking the mean of metabolomic impact over all possible deletion windows yields a measure of "metabolomic fragility" of that genome to that given deletion size. In addition to the latter, we also compute "biomass fragility", which we derive as the mean number of *biomass* metabolites knocked out per deletion. In this study we apply such to assess metabolomic fragility to 4000, 8000, and 16000 bp genomic deletions.

E. Genomic permutations

To evaluate our main hypothesis, we compare the metabolomic fragility of the *E. coli* genome against that of randomized genomes obtained under a variety of permutation schemes. For each permutation schemes, we generate 1000 randomized genomes, from which we generate "null distributions" of metabolomic fragility. Each null distributions allows us to statistically infer the importance of a particular aspect of genomic organization for metabolic robustness.

In permutation scheme 1, we shuffle the assignment of metabolic genes and their positions, while maintaining the original position distribution of genes in the *E. coli* genome. This shuffling dissociates the grouping of genes into operons,

*Note that producibility characterizes the capacity of a metabolic network to produce a metabolite, and does not guarantee that the metabolite is actually produced by the network. However, if a metabolite in not supplied as a nutrient, and is not producible, then it is guaranteed to be absent from the cell.

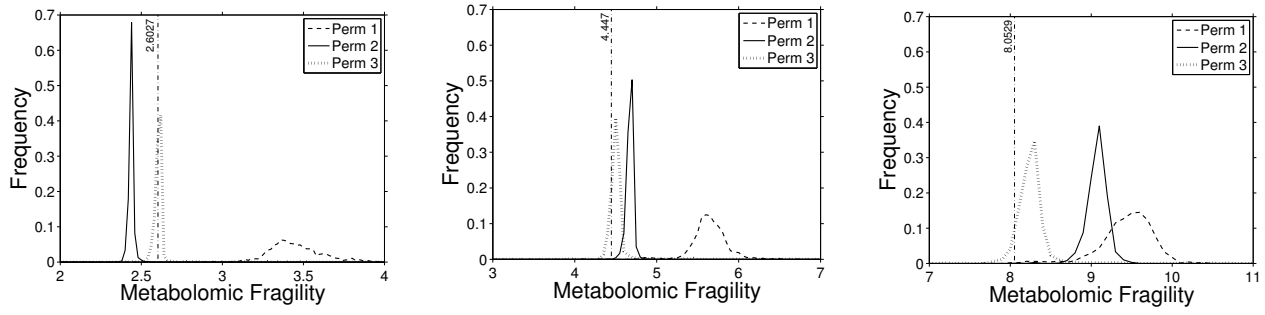

Fig. 3. Null distributions of metabolomic fragility for the three permutation schemes: gene-position shuffling (Perm 1), gene position randomization (Perm 2), and operon position randomization (Perm 3). Panels a-c show results for the following deletion sizes: (a) 4000 bp deletion size (b) 8000 bp deletion size (c) 16000 bp deletion size. The horizontal dotted line in each figure represents the metabolomic fragility of the original *E. coli* genome with respect to the given deletion size. Each distribution shown is generated from 1000 permuted genomes.

allowing us to evaluate the importance of operon composition for metabolomic fragility.

In permutation scheme 2, we assign metabolic genes random positions in the genome. This produces an approximately uniform distribution of metabolic genes across the genome, and disrupts gene-operon groupings. Computing metabolomic fragility across these permuted genomes allows us to infer the importance of positional gene clustering for metabolomic fragility.

In permutation scheme 3, we shuffle operon position while maintaining gene to operon mapping. This produces an approximately uniform distribution of operons across the genome, while allowing non-uniform positional gene clustering. In this permutation scheme, we determine the importance of operon clustering for metabolomic fragility.

III. Results and Discussion

A. E. coli metabolomic fragility in rich media

We computed the metabolomic fragility of the *E. coli* genome for 3 deletion window sizes (4000bp, 8000bp, 16000bp) in rich media. We chose these three sizes as they span the range of biologically realistic deletion events. As shown in Figure 1, these three genomic deletion sizes produce qualitatively different distributions of gene knockouts, allowing us to richly probe *E. coli* metabolic robustness.

Computation of metabolomic impact across all k bp windows in the genome yielded the frequency plot shown in (Figure 2). We refer to this plot as a "metabolomic impact histogram", which represents a probability distribution of metabolomic impact assuming that random gene-deletion events are uniformly distributed in the genome. As intuitively expected, *E. coli* metabolomic fragility increased with genomic deletion size. On average, 4000, 8000, and 16000 bp deletions knocked out 2.60, 4.45, and 8.05 of 539 total producible metabolites, respectively. Analysis of biomass fragility found 4000, 8000, and 16000 bp deletions to knock out 0.20, 0.27, and 0.61 of 49 total biomass metabolites, respectively. Though biomass metabolites did not appear to be more likely to be knocked out by 4000 bp deletions, they were relatively more fragile to larger (8000bp and 16000bp) deletions.

B. Gene-operon grouping promotes metabolomic robustness

We compared the metabolomic fragility of the *E. coli* genome to a permuted null distribution obtained by shuffling position-gene assignment. By maintaining the original genomic distribution of genes but effectively altering gene-operon groupings, we aimed to ascertain whether the distribution of genes into operons helps buffer the metabolome against gene deletion. As shown in Table I, the metabolomic fragility of every random genome obtained through position-gene assignment shuffling was greater than the metabolomic fragility of the original *E. coli* genome. These results were statistically significant ($p < 0.001$), allowing us to reject the null hypothesis that gene-operon organization does not influence metabolomic fragility. A trace of the null distribution obtained from gene-position shuffling is shown in Figure 3.

To examine the nature of the buffering effect of gene-operon organization in more detail, we computed a "mean metabolomic impact histogram" across all permuted genomes. These results showed random gene deletions in the original *E. coli* genome to be significantly more likely to have a metabolomic impact of 0, relative the average permuted genome. Alternatively stated, this implies that gene-operon shuffling brings together distantly separated and mutually buffering gene combinations into close proximity. Such "buffering" gene combinations can correspond to gene duplicates, isoenzymes, and parallel pathways.

C. Gene position distribution heterogenously impacts E. coli metabolome fragility

By shuffling gene-operon assignment, we maintain the overall positional distribution of genes across the genome. This gene-position distribution is highly non-uniform, resulting primarily from the clustering of genes into operons. We applied a permutation approach to examine how this parameter influences metabolomic fragility.

In this analysis, we constructed permuted genomes in which we assigned *E. coli* genes random positions. Comparing the metabolomic fragility of the original *E. coli* genome

Permutation Scheme 1			
Deletion size	Metabolic Fragility (Original Genome)	Metabolomic Fragility (Permutation Mean +/- SD)	Percentile
4000 bp	2.6027	3.466 (+/- 0.180)	0.0 %
8000 bp	4.45	5.674 (+/- 0.209)	0.0 %
16000 bp	8.05	9.490 (+/- 0.333)	0.0 %

TABLE I

PERMUTATION RESULTS FOR GENE-POSITION SHUFFLING FOR THE THREE DELETION SIZES TESTED. THIS PERMUTATION SCHEME EFFECTIVELY SHUFFLES GENE-OPERON ASSIGNMENT WHILE MAINTAINING GENE POSITION DISTRIBUTION

Permutation Scheme 2			
Deletion size	Metabolic Fragility (Original Genome)	Metabolomic Fragility (Permutation Mean +/- SD)	Percentile
4000 bp	2.6027	2.448 (+/- 0.015)	97.2 %
8000 bp	4.45	4.697 (+/- 0.045)	0.01 %
16000 bp	8.05	9.123 (+/- 0.126)	0.0 %

TABLE II

PERMUTATION RESULTS FOR GENE-POSITION RANDOMIZATION FOR THE THREE DELETION SIZES TESTED. THIS PERMUTATION SCHEME RESULTS IN ROUGHLY UNIFORM DISTRIBUTION OF GENE POSITIONS ACROSS THE GENOME.

to the null distribution computed from these genomes, we observed that genomic position was protective against longer deletions (8000bp, 16000bp) but not protective for shorter deletions (4000bp) . For 4000bp deletions, the metabolomic fragility of the original genome (2.60) was significantly above the 95% confidence interval of metabolomic fragility across permuted genomes (2.448 +/- .015). However for 8000bp and 16000bp deletions, we observed that gene position had a statistically significant ($p < .001$) *protective* effect on the metabolome (Table II). A trace of the null distribution obtained from gene position randomization is shown in Figure 3. Analysis of mean metabolomic impact histograms showed that 4000 bp deletions were significantly more likely to have 0 metabolomic impact in the *E. coli* genome with randomized gene positions. Meanwhile, for both 8000 and 16000 bp deletion sizes, we observed significant enrichment in the frequency of deletions with 0 metabolomic impact in the original *E. coli* genome.

These results show that uniform distribution of genes in the genome brings functionally buffering groups of genes into close proximity, however the size of the deletion has to be sufficient to span the mean length separating neighboring genes. The qualitative difference of effect seen between 4000 and 8000 bp gene likely arises due to the uncoupling of both gene position and gene-operon grouping in this permutation scheme. Since gene-operon relationships are abolished with this shuffling, a deletion event only affects the genes that lie in that region, without knocking out nearby downstream genes. For each deletion size, the average number of genes knocked out by a gene deletion in this scheme is thus smaller than for the original *E. coli* genome. At small deletion sizes, this greatly decreases the average metabolomic impact of a deletion event. However, given a sufficient deletion size, "quality" overcomes "quantity", resulting in complex gene knockouts that interact epistatically to disable multiple parallel pathways in the *E. coli* network.

D. Operon position distribution protects the metabolome against large deletions

We compared *E. coli* metabolomic fragility against a distribution of genomes with random operon position. In this permutation scheme, we maintained gene-operon assignment and relative positions of genes within operons. For 4000bp and 8000bp deletions, the metabolomic fragility of the original genome was contained within the 95% confidence interval of the null distribution, suggesting no significant relationship between operon position and metabolomic fragility (Figures 2 and Figure 3). However for 16000 bp deletions, metabolomic fragility of the *E. coli* genome was less than 95% of permuted genomes (Table III).

These results suggest that operon position distribution in the *E. coli* genome separates mutually buffering gene combinations from each other. However, this effect manifests at larger deletion sizes than for gene-position shuffling. This size-dependence occurs most likely because operons are fewer in number and thus a larger distances separates neighboring operons after position randomization. Though this effect appears only at larger deletion sizes, our results clearly suggest the that operon position has functional systems-level consequences for metabolism.

E. Implications for metabolic network evolution

Our results strongly suggest an association between sequence organization of the *E. coli* genome and the robustness of its metabolism. This effect is specifically arises from the grouping of genes into operons and the positional distribution of operons and genes across the genome. The resulting genomic structure physically separates epistatically interacting gene combinations that contribute to parallel and robust biosynthetic functions. When these gene combinations are placed into physical proximity through gene shuffling in our simulation, these robust functions become targets for genomic deletion.

We interpret the observed robustness properties to suggest that genomic proximity in *E. coli* mirrors metabolic network

Permutation Scheme 3			
Deletion size	Metabolic Fragility (Original Genome)	Metabolomic Fragility (Permutation Mean +/- SD)	Percentile
4000 bp	2.6027	2.655 (+/- 0.3355)	21.7 %
8000 bp	4.45	4.573 (+/- 0.3758)	7.8 %
16000 bp	8.05	8.305 (+/- 0.1876)	3.7 %

TABLE III

PERMUTATION RESULTS FOR GENE-POSITION RANDOMIZATION FOR THE THREE DELETION SIZES TESTED. THIS PERMUTATION SCHEME RESULTS IN ROUGHLY UNIFORM DISTRIBUTION OF OPERON POSITIONS ACROSS THE GENOME.

proximity. Proximity in the metabolic network, in classical biochemical terms, implies nearness in the context of a multi-step linear or branched "pathway". In genome-scale metabolic network parlance, we would restate this claim to suggest that genes sharing elementary modes will lie close to each other on the genome. Though the latter is accepted as fact for operons, which very often consist of multiple genes encoding steps in a pathway (e.g. biotin synthesis), it has not been explored for larger scale genomic structures.

If such larger-scale relationships exist, they may be evolutionary consequences of bacterial horizontal gene transfer. In horizontal gene transfer, modules of functionally related genes were swapped between bacteria through transformation, conjugation, and bacteriophage-mediated transduction. This mechanism is used to explain positional clustering of functionally related genes into operons, but may also underly larger functional clustering in the genome. Interestingly, there has been evidence shown to suggest existence of gene context conservation between bacterial species on an "uber-operonic" scale [9], [3]. Functional correlation between genes in these larger conserved regions may provide support for the role of such large scale horizontal gene transfer in metabolic network evolution.

IV. CONCLUSION

We have employed a novel and efficient implementation of producibility analysis to investigate the robustness properties of the *E. coli* metabolic network to random gene deletion. Comparing the metabolomic fragility of the *E. coli* genome to several null distributions obtained from analysis of permuted genomes, we have found that genome-structure has a significant impact on metabolic network robustness. From these results, we conclude that epistatically interacting gene combinations in the *E. coli* network tend to be physically far apart in the genome. These results have interesting implications for metabolic network evolution. Future work includes probing a larger range of deletion sizes and extending our work to other prokaryotic and eukaryotic organisms.

REFERENCES

[1] "Sedumi." [Online]. Available: http://sedumi.mcmaster.ca/

[2] F. R. Blattner, G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao, "The complete genome sequence of escherichia coli k-12." *Science*, vol. 277, no. 5331, pp. 1453–1474, Sep 1997.

[3] D. Che, G. Li, F. Mao, H. Wu, and Y. Xu, "Detecting uber-operons in prokaryotic genomes." *Nucleic Acids Res*, vol. 34, no. 8, pp. 2418–2427, 2006. [Online]. Available: http://dx.doi.org/10.1093/nar/gkl294

[4] I. Famili, J. Forster, J. Nielsen, and B. O. Palsson, "Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network." *Proc Natl Acad Sci U S A*, vol. 100, no. 23, pp. 13 134–9, Nov 2003. [Online]. Available: http://dx.doi.org/10.1073/pnas.2235812100

[5] J. Forster, I. Famili, P. Fu, B. O. Palsson, and J. Nielsen, "Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network." *Genome Res*, vol. 13, no. 2, pp. 244–53, Feb 2003. [Online]. Available: http://dx.doi.org/10.1101/gr.234503

[6] M. Imielinski, C. Belta, A. Halász, and H. Rubin, "Investigating metabolite essentiality through genome-scale analysis of Escherichia coli production capabilities." *Bioinformatics*, vol. 21, no. 9, pp. 2008–16, May 2005. [Online]. Available: http://grasp.upenn.edu/ imielins/papers/bioinformatics.pdf

[7] P. D. Karp, C. A. Ouzounis, C. Moore-Kochlacs, L. Goldovsky, P. Kaipa, D. Ahrn, S. Tsoka, N. Darzentas, V. Kunin, and N. Lpez-Bigas, "Expansion of the BioCyc collection of pathway/genome databases to 160 genomes." *Nucleic Acids Res*, vol. 33, no. 19, pp. 6083–6089, 2005. [Online]. Available: http://dx.doi.org/10.1093/nar/gki892

[8] K. J. Kauffman, J. D. Pajerowski, N. Jamshidi, B. O. Palsson, and J. S. Edwards, "Description and analysis of metabolic connectivity and dynamics in the human red blood cell." *Biophys J*, vol. 83, no. 2, pp. 646–62, Aug 2002.

[9] W. C. Lathe, B. Snel, and P. Bork, "Gene context conservation of a higher order than operons." *Trends Biochem Sci*, vol. 25, no. 10, pp. 474–479, Oct 2000.

[10] N. D. Price, J. L. Reed, and B. O. Palsson, "Genome-scale models of microbial cells: evaluating the consequences of constraints." *Nat Rev Microbiol*, vol. 2, no. 11, pp. 886–97, Nov 2004. [Online]. Available: http://dx.doi.org/10.1038/nrmicro1023

[11] A. Raghunathan, N. D. Price, M. Y. Galperin, K. S. Makarova, S. Purvine, A. F. Picone, T. Cherny, T. Xie, T. J. Reilly, R. Munson, R. E. Tyler, B. J. Akerley, A. L. Smith, B. O. Palsson, and E. Kolker, "In Silico Metabolic Model and Protein Expression of Haemophilus influenzae Strain Rd KW20 in Rich Medium." *OMICS*, vol. 8, no. 1, pp. 25–41, 2004. [Online]. Available: http://dx.doi.org/10.1089/153623104773547471

[12] J. L. Reed, T. D. Vo, C. H. Schilling, and B. O. Palsson, "An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR)." *Genome Biol*, vol. 4, no. 9, p. R54, 2003. [Online]. Available: http://dx.doi.org/10.1186/gb-2003-4-9-r54

[13] C. H. Schilling and B. O. Palsson, "The underlying pathway structure of biochemical reaction networks." *Proc Natl Acad Sci U S A*, vol. 95, no. 8, pp. 4193–8, Apr 1998.

[14] C. H. Schilling, M. W. Covert, I. Famili, G. M. Church, J. S. Edwards, and B. O. Palsson, "Genome-scale metabolic model of Helicobacter pylori 26695." *J Bacteriol*, vol. 184, no. 16, pp. 4582–93, Aug 2002.