Network Architecture and Mutational Sensitivity of the *C. elegans* Metabolome

Lindsay M. Johnson[1*], Luke M. Chandler[2*], Sarah K. Davies[3] and Charles F. Baer[1,2]

1 – Department of Biology, University of Florida, Gainesville, FL

2 – University of Florida Genetics Institute

3 - Faculty of Medicine, Department of Surgery & Cancer, Imperial College, London

* - these authors contributed equally

Email: lindsaymjohnson@ufl.edu; lukemchandler@ufl.edu; sarah.davies1@imperial.ed.uk

Correspondence to:

Charles F. Baer

Department of Biology

University of Florida

P. O. Box 118525

Gainesville, FL 32611-8525 USA

Email: cbaer@ufl.edu

Keywords: connectedness, metabolic network, mutation accumulation, mutational correlation, mutational variance

List of Figures and Tables

Figure 1. Metabolic Network

Table 1. List of definitions of network statistics

Table 2. Correlations between network statistics and mutational parameters

Supplementary Figure S1. Depiction of k-core

Supplementary Figure S2. Twelve distributions of r_M with six randomly chosen covariates.

Supplementary Figure S3. Distributions of random correlations between (a) r_M and shortest path length in the directed network, (b) $|r_M|$ and shortest path length, (c) r_M and shortest path length in the undirected network (i.e., shorter of the two path lengths between metabolites *i* and *j* in the directed network)

Supplementary Table S1. Network and mutational parameters of metabolites. (Word)

Supplementary Table S2. Table of discrepancies between MZ and YW methods (Word)

Supplementary Table S3. Mutational and environmental correlations (Excel)

Supplementary Table S4. Shortest network path lengths (Excel)

Supplementary Appendix A1. Metabolic network data (Excel)

Abstract

Background: A fundamental issue in evolutionary systems biology is understanding the relationship between the topological architecture of a biological network, such as a metabolic network, and the evolution of the network. The rate at which an element in a metabolic network accumulates genetic variation via new mutations depends on both the size of the mutational target it presents and its robustness to mutational perturbation. Quantifying the relationship between topological properties of network elements and the mutability of those elements will facilitate understanding the variation in and evolution of networks at the level of populations and higher taxa.

Results: We report an investigation into the relationship between topological properties of 29 metabolites in the *C. elegans* metabolic network and the sensitivity of those metabolites to the cumulative effects of spontaneous mutation. We find a positive correlation between network connectedness of a metabolite, as quantified by its core number, and sensitivity to mutation, as quantified by the mutational heritability. We further find a small but significant negative correlation between the shortest path length between a pair of metabolites and the mutational correlation between those metabolites.

Conclusions: The positive association between the connectedness of a metabolite and its mutational heritability is consistent with well-connected metabolites presenting a larger mutational target than sparsely-connected ones, and is inconsistent with well-connectedness conferring mutational robustness, at least *in toto*. The weakness of the correlation between shortest path length and the mutational correlation between pairs of metabolites suggests that network locality is an important but not overwhelming factor governing mutational pleiotropy. These findings provide necessary background against which the effects of other evolutionary forces, most importantly natural selection, can be interpreted.

Introduction:

The set of chemical reactions that constitute organismal metabolism is often represented as a network of interacting components, in which individual metabolites are the nodes in the network and the chemical reactions of metabolism are the edges [1]. Representation of a complex biological process such as metabolism as a network is conceptually powerful because it offers a convenient and familiar way of visualizing the system, as well as a well-developed mathematical framework for analysis.

If the representation of a biological system as a network is to be useful as more than a metaphor, it must have predictive power [2]. Metabolic networks have been investigated in the context of evolution, toward a variety of ends. Many studies have compared empirical metabolic networks to various random networks, with the goal of inferring adaptive features of global network architecture (e.g., [1, 3-7]. Other studies have addressed the relationship between network-level properties of individual elements of the network (e.g., node degree, connectedness) and evolutionary properties such as rates of protein evolution [8, 9] and within-species polymorphism [10].

One fundamental evolutionary process that remains essentially unexplored in the context of metabolic networks is mutation. Importantly, mutation is intimately - but not inextricably - associated with natural selection. On average, mutations are deleterious, and deleterious mutations disrupt the proper function of an organism. A properly functioning organism must maintain metabolic homeostasis, so it stands to reason that deleterious mutations must perturb the metabolic network in some way.

To see how mutation and selection are intertwined, consider a hypothetical example. A generic property of empirical networks, including metabolic networks, is that they are scale-free [1], which results in a network architecture with a few highly-connected nodes (hubs) and many weakly-connected nodes. If hubs, or reactions connected to hubs ("hub phenotypes") are, on average, more important to the function of the organism than non-hub phenotypes, purifying

selection will be stronger on mutations that affect hubs, so all else equal, there will be less standing genetic variation for hub phenotypes than for non-hub phenotypes.

However, it also stands to reason that the more highly-connected a node is, the greater the number of potential ways to affect the node, and its associated reactions. Again, all else equal, the more genes that are associated with a network element, the larger the mutational target, and the greater the genetic variation associated with that element. Thus, it may well be that mutation produces more genetic variation for hub phenotypes than for non-hub phenotypes.

Clearly, understanding the evolution of biological networks requires an assessment of the effects of mutation independent from natural selection. Neither mutation nor selection can ever be turned off completely, but mutations with selective effects s less than approximately the reciprocal of the genetic effective population size N_e will be essentially invisible to natural selection [11]. A mutation accumulation (MA) experiment is an experiment designed to minimize the efficacy of selection by minimizing the genetic effective population size, N_e, thereby allowing all but the most strongly deleterious mutations to evolve as if they are invisible to selection [12].

Here we report results from a long-term MA experiment in the nematode *Caenorhabditis elegans*, in which replicate populations (MA lines) derived from a genetically homogeneous common ancestor (G0) were allowed to evolve under minimally effective selection ($N_e \approx 1$) for approximately 250 generations. We previously reported estimates from these MA lines of two key quantitative genetic parameters by which the cumulative effects of mutation can be quantified: the per-generation change in the trait mean (the mutational bias, ΔM) and the per-generation increase in genetic variation (the mutational variance, V_M) for the standing pools of 29 metabolites [13]; Supplementary Table S1. In this report, we interpret those results, and new estimates of mutational correlations (r_M), in the context of the topology of the *C. elegans* metabolic network.

Results and Discussion

Representation of the Metabolic Network – The metabolic network of *C. elegans* was estimated in two ways: *(i)* by the method of Ma and Zeng [14; MZ],and *(ii)* by the method of Yilmaz and Walhout [15; YW]. Details of the network construction are given in section I of the Methods; data are presented in Supplementary Appendix A1. In most cases, MZ and YW give identical results with respect to network topology; in the few cases in which there is a discrepancy (~1%; Supplementary Table S2), we use the MZ network. The resulting network is a directed graph including 646 metabolites, with 1203 reactions connecting nearly all metabolites (Figure 1).

Properties of networks can be quantified in many ways. The motivation for this study is exploratory data analysis, but we did not begin as agnostics – we suspected that highly-connected metabolites might be more susceptible to the cumulative effects of mutation than less connected metabolites, because they would seem to present a larger mutational target. On the other hand, it is also possible that well-connected metabolites are more robust to mutational perturbation than sparsely-connected metabolites. An analogous scenario is Kacser and Burns' [16] theory of dominance, in which metabolic pathways with many enzymatic steps are more robust to mutational perturbation than pathways with fewer steps. The relationship between network connectedness and robustness will depend on the functional (mathematical) relationship between the probability of perturbing a connection (i.e., the mutational target) and the functional (biochemical) consequences of the perturbation (e.g., a decrease in metabolite flux). Ultimately, the relationship between network architecture and mutational sensitivity is an empirical question.

We did not have a strong prior hypothesis about which specific network parameters would prove most informative in terms of a relationship with ΔM and/or V_M. Therefore, we assessed the relationship between mutational properties and several measures of network connectedness: betweenness, closeness, and degree centrality, in- and out-degree, and core number (depicted in Supplementary Figure S1). These parameters are highly correlated.

Definitions of the parameters are given in Table 1; correlations between the parameters are included in Table 2. For each of the six parameters, we calculated Spearman's correlation ρ between mutational statistics and the network parameter associated with the metabolite. The strict experiment-wide 5% significance level for these correlations is approximately P<0.002 (α=0.05/[6 network parameters x 4 mutational parameters]).

Mutational Parameters – Details of the MA experiment are reported in Baer et al. [17] and outlined in section II of the Methods. The experimental protocol by which metabolite concentrations were measured is reported in [13] and outlined in section III of the Methods; data are archived in Dryad at http://dx.doi.org/10.5061/dryad.2dn09/1. For each of the 29 metabolites, the cumulative effects of mutation are summarized by the mutational bias (ΔM), and the mutational variance (V_M). For a trait z, $\Delta M_z=\mu_G\alpha_z$, where μ_G is the genomic mutation rate and α_z is the average effect of a mutation on the trait; $V_M=\mu_G\alpha_z^2$ [18, p. 329]. Details of the estimation of mutational parameters are given in section IV of the Methods.

Comparisons of variation among traits or groups require that the variance be measured on a common scale. V_M is commonly scaled either relative to the trait mean, in which case V_M is the squared coefficient of variation and is often designated I_M, or relative to the residual variance, V_E; V_M/V_E is the mutational heritability, h_M^2. I_M and h_M^2 have different statistical properties and evolutionary interpretations [19], so we report both. I_M and I_E are standardized relative to the mean of the MA lines.

The relationship between network connectedness and sensitivity to mutation - Our results suggest that more highly connected metabolites are more susceptible to mutation (Table 2), although the statistical justification for that conclusion is marginal. Core number, defined as the largest value of a k-core that contains the node of interest, (a k-core is the is the largest subgraph that contains nodes of degree k), is positively correlated with h_M^2 (ρ=0.48, P<0.008) and less strongly positively correlated with I_M (ρ=0.30, P>0.11). The 29 metabolites in our data set have core number of either one or two (the maximum core number of any metabolite in the

network is two). Mean h_M^2 of metabolites of core number = 2 is approximately 2.5X greater than that of metabolites of core number = 1 (0.002 vs. 0.0008). To put that result in context, the average h_M^2 for a wide variety of traits in a wide variety of organisms is on the order of 0.001 [19].

We think it is not a coincidence that the signal of mutational vulnerability emerges most strongly for core number, in contrast to the other measures of network connectedness, and for h_M^2 rather than I_M as a measure of mutational variation. Core number is a categorical variable, whereas the other measures of network connectedness are continuous variables. In terms of the effect on power, quantifying connectedness in terms of core number is analogous to categorizing a set of size measurements into "small" and "large": power is increased, at the cost of losing the ability to discriminate between more subtle differences.

The raw mutational variance, V_M, appears in the numerator of both h_M^2 and I_M; the difference lies in the denominator, which is the residual variance V_E for h_M^2 and the square of the trait mean for I_M. For some replicates of some metabolites, estimated metabolite concentrations were atypically low and near zero; I_M is more sensitive to low outliers than is h_M^2. However, the correlation between I_M and the trait mean is small (r = -0.11) and not significantly different from zero. Alternatively, it is possible that V_M does not vary consistently with metabolite connectedness, but that metabolites with low connectedness (core number = 1) are more susceptible to random microenvironmental variation ("noise") than are metabolites with high connectedness (core number = 2), in which case V_E would be greater for lowly-connected metabolites and h_M^2 would be lower. Unfortunately, the variance is correlated with the trait mean, so the least biased way to address that question is by comparing the residual squared coefficients of variation, I_E. There is no hint of correlation between core number and I_E (ρ=0.025, P>0.89; Table 2), and I_E is uncorrelated with the trait mean (r = -0.12, P>0.54), so the

association between h_M^2 and core number cannot obviously be attributed to differential sensitivity to random noise.

<u>The relationship between mutational correlation (r_M) and shortest path length</u> - The cumulative effects of mutations on a pair of traits *i* and *j* may covary for two, nonexclusive reasons [20]. More interestingly, individual mutations may have consistently pleiotropic effects, such that mutations that affect trait *i* also affect trait *j* in a consistent way. Less interestingly, but unavoidably, individual MA lines will have accumulated different numbers of mutations, and if mutations have consistently directional effects, as would be expected for traits correlated with fitness, lines with more mutations will have more extreme trait values than lines with fewer mutations, even in the absence of consistent pleiotropy. Estes et al. [20] simulated the sampling process in *C. elegans* MA lines with mutational properties derived from empirical estimates from a variety of traits and concluded that sampling is not likely to lead to large absolute mutational correlations in the absence of consistent pleiotropy ($|r_M| \leq 0.25$).

Ideally, we would like to estimate the full mutational (co)variance matrix, ***M,*** from the joint estimate of the among-line (co)variance matrix. However, with 25 traits, there are (25x26)/2 = 325 covariances, and with only 43 MA lines, there is insufficient information to jointly estimate the restricted maximum likelihood of the full ***M*** matrix. To proceed, we calculated mutational correlations from pairwise REML estimates of the among-line (co)variances, i.e., $r_M = \frac{COV_L(X,Y)}{\sqrt{VAR_L(X)VAR_L(Y)}}$ [21, 22]. Pairwise estimates of r_M are shown in Supplementary Table S3. To assess the extent to which the pairwise correlations are sensitive to the underlying covariance structure, we devised a heuristic bootstrap analysis. For a random subset of 12 of the 300 pairs of traits, we randomly sampled six of the remaining 23 traits without replacement and estimated r_M between the two focal traits from the joint REML among-line (co)variance matrix. For each of the 12 pairs of focal traits, we repeated the analysis 100 times.

There is a technical caveat to the preceding bootstrap analysis. Resampling statistics presume that the variables are exchangeable [23], which metabolites are not. For that reason, we do not present confidence intervals on the resampled correlations, only the distributions. However, we believe that the analysis provides a meaningful heuristic by which the sensitivity of the pairwise correlations to the underlying covariance structure can be assessed.

Distributions of resampled correlations are shown in Supplementary Figure S2. In every case the point estimate of r_M falls on the mode of the distribution of resampled correlations, and in 11 of the 12 cases, the median of the resampled distribution is very close to the point estimate of r_M. However, in six of the 12 cases, some fraction of the resampled distribution falls outside two standard errors of the point estimate. The most important point that the resampling analysis reveals is this: given that 29 metabolites encompass only a small fraction of the total metabolome of *C. elegans* (<5%), even had we been able to estimate the joint likelihood of the full 29x30/2 ***M***-matrix, the true covariance relationships among those 29 metabolites could conceivably be quite different from those estimated from the data.

Correlations are properties of pairs of variables, so we expect *a priori* that network parameters that apply to pairs of elements are more likely to contain information about the mutational correlation between a pair of metabolites than will the pairwise average of a parameter that applies to individual elements of a network. The shortest path length is the simplest network property that describes the relationship between two nodes, although since the metabolic network is directed, the shortest path from element i to element j is not necessarily the same as the shortest path from j to i. For each pair of metabolites i and j, we calculated the shortest path length from i to j and from j to i, without repeated walks (Supplementary Table S4). We then calculated Spearman's correlation ρ between the mutational correlation r_M and the shortest path length.

Statistical assessment of the correlation between mutational correlations (r_M) and shortest path length presents a problem of nonindependence, for two reasons. First, all

correlations including the same variable are non-independent. Second, even though the mutational correlation between metabolites *i* and *j* is the same as the correlation between *j* and *i*, the shortest path lengths need not be the same, and moreover, the path from *i* to *j* may exist whereas the path from *j* to *i* may not. To account for non-independence of the data, we devised a parametric bootstrap procedure; details are given in section V of the Methods. Three metabolites (L-tryptophan, L-lysine, and Pantothenate) lie outside of the great strong component of the network [24] and are omitted from the analysis.

There is a weak, but significant, negative correlation between r_M and the shortest path length between the two metabolites (ρ = -0.128, two-tailed P<0.03; Supplementary Figure S2a), whereas $|r_M|$ is not significantly correlated with shortest path length (ρ = -0.0058, two-tailed P>0.45; Supplementary Figure S2b). The correlation between r_M and the shortest path in the undirected network is similar to the correlation between r_M and the shortest path in the directed network (ρ = -0.105, two-tailed P>0.10; Supplementary Figure S2c).

An intuitive possible cause of the weak negative association between shortest path length and mutational correlation would be if a mutation that perturbs a metabolic pathway toward the beginning of the pathway has effects that propagate downstream in the same pathway, but the effect of the perturbation attenuates. The attenuation could be due either to random noise or to the effects of other inputs into the pathway downstream from the perturbation (or both). The net effect would be a characteristic pathway length past which the mutational effects on two metabolites are uncorrelated, leading to an overall negative correlation between r_M and path length. The finding that the correlations between r_M and the shortest path length in the directed and undirected network are very similar reinforces that conclusion. The negative correlation between r_M and shortest path length is reminiscent of a finding from Arabidopsis, in which sets of metabolites significantly altered by single random gene knockouts are closer in the global metabolic network than expected by chance [25].

Conclusions and Future Directions

The proximate goal of this study was to find out if there are topological properties of the *C. elegans* metabolic network – connectedness, shortest path length, etc. – that are correlated with a set of statistical descriptions of the cumulative effects of spontaneous mutations (ΔM, V_M, r_M). Ultimately, we hope that a deeper understanding of those mathematical relationships will shed light on the mechanistic biology of the organism. Statistical fragility of the results notwithstanding (but not forgotten), we conclude:

(i) Network connectedness is associated with mutational sensitivity (V_M), not mutational robustness ($1/V_M$). The most plausible explanation is that metabolites that are more highly connected present a larger mutational target than do metabolites that are less connected. However, although $1/V_M$ is a meaningful measure of mutational robustness [26], it does not necessarily follow that highly-connected metabolites are therefore more robust to the effects *individual* mutations [27, 28].

(ii) Pleiotropic effects of mutations affecting the metabolome are predominantly local, as evidenced by the significant negative correlation between shortest path length between a pair of metabolites and the mutational correlation, r_M, between that pair of metabolites. That result is not surprising in hindsight, but the weakness of the correlation suggests that there are other important factors that underlie pleiotropy beyond network proximity.

To advance understanding of the mutability of the *C. elegans* metabolic network, three things are needed. First, it will be important to cover a larger fraction of the metabolic network. Untargeted mass spectrometry of cultures of *C. elegans* reveals many thousands of features (Art Edison, personal communication); 29 metabolites are only the tip of a large iceberg. For example, our intuition leads us to believe that the mutability of a metabolite will depend more on its in-degree (mathematically, the number of edges leading into a node in a directed graph; biochemically, the number of reactions in which the metabolite is a product) than its out-degree. The point-estimate of the correlation of h_M^2 with in-degree is twice that of the correlation of h_M^2 with out-degree (Table 2), although the difference is not statistically significant.

Second, to more precisely partition mutational (co)variance into within- and among-line components, more MA lines are needed. If our estimate of an average of 80 mutations per MA line is close to the mark, it means that the phenotypic (co)variance observed here is the result of about 3500 total mutations, distributed among 43 MA lines. In this case, the MA lines were a preexisting resource, and the sample size was predetermined. It is encouraging that we were able to detect significant mutational variance for 25/29 metabolites (Supplementary Table S1b), but only 14% (42/300) of pairwise mutational correlations are significantly different from zero at the experiment-wide 5% significance level, roughly corresponding to $|r_M|>0.5$ (Supplementary Table S3); 18 of the 42 significant mutational correlations are not significantly different from $|r_M| = 1$. It remains uncertain how sensitive estimates of mutational correlations are to the underlying covariance structure of the metabolome. It also remains to be seen if the mutability of specific features of metabolic networks are genotype or species-specific, and the extent to which mutability depends on environmental context.

Third, it will be important to quantify metabolites (static concentrations and fluxes) with more precision. The metabolite data analyzed in this study were collected from large cultures (n>10,000 individuals) of approximately age-synchronized worms, and were normalized relative to an external quantitation standard [13]. Ideally, one would like to characterize the metabolomes of single individuals, assayed at the identical stage of development. That is not yet practical with *C. elegans*, although it is possible to quantify hundreds of metabolites from a sample of 1000 individuals [29], and preliminary studies suggest it will soon be possible to reduce the number of individuals to 100 or even ten (M. Witting, personal communication). Minimizing the number of individuals in a sample is important for two reasons; (1) the smaller the sample, the easier it is to be certain the individuals are closely synchronized with respect to developmental stage, and (2) knowing the exact number of individuals in a sample makes normalization relative to an external standard more interpretable. Ideally, data would be

normalized relative to both an external standard and an internal standard (e.g., total protein; [21]).

This study provides an initial assessment of the relationship between mutation and metabolic network architecture. To begin to uncover the relationship between metabolic architecture and natural selection, the next step is to repeat these analyses with respect to the standing genetic variation (VG). There is some reason to think that highly-connected metabolites will be more evolutionarily constrained (i.e., under stronger purifying selection) than less connected metabolites [8], in which case the ratio of the mutational variance to the standing genetic variance (VM/VG) will increase with increasing connectedness.

Methods and Materials:

I. Metabolic Network:

The metabolic network of *C. elegans* was estimated in two ways: *(i)* by the static, purely graphical method of Ma and Zeng ([14]; updated at http://www.ibiodesign.net/kneva/; we refer to this method as MZ), and *(ii)* by the dynamical, flux-balance analysis (FBA) method of Yilmaz and Walhout ([15]; http://wormflux.umassmed.edu/; we refer to this method as YW). Subnetworks that do not contain at least one of the 29 metabolites were excluded from downstream analyses. The MZ method uses as its source data the KEGG LIGAND database (http://www.genome.jp/kegg/ligand.html), and includes several *ad hoc* criteria for retaining or omitting specific metabolites from the analysis (criteria are listed on p. 272 of [14]). In most cases, MZ and YW give identical results; in the few cases in which there is a discrepancy (Supplementary Table S2), we chose to use the MZ network because we used the MZ criteria for categorizing currency metabolites (defined below).

To begin, the 29 metabolites of interest were identified and used as starting sites for the network. Next, all forward and reverse reactions stemming from the 29 metabolites were incorporated into the subnetwork until all reactions either looped back to the starting point or

reached an endpoint. Currency metabolites were removed following the MZ criteria (a currency metabolite is roughly defined as a molecule such as water, ATP, NADH, etc., that appears in a large fraction of metabolic reactions but is not itself an intermediate in an enzymatic pathway).

A graphical representation of the network was constructed with the Pajek software package (http://mrvar.fdv.uni-lj.si/pajek/) and imported into the networkX Python package [30], which was used to generate network statistics. Proper importation from Pajek to networkX was verified by visual inspection.

II. Mutation Accumulation Lines

A full description of the construction and propagation of the mutation accumulation (MA) lines is given in [17]. Briefly, 100 replicate MA lines were initiated from a nearly-isogenic population of N2-strain *C. elegans* and propagated by single-hermaphrodite descent at four-day (one generation) intervals for approximately 250 generations. The long-term N_e of the MA lines is very close to one, which means that mutations with a selective effect less than about 25% are effectively neutral [31]. The common ancestor of the MA lines ("G0") was cryopreserved at the outset of the experiment; MA lines were cryopreserved upon completion of the MA phase of the experiment. Based on extensive whole-genome sequencing [32; A. Saxena and CFB, in prep], we estimate that each MA line carries about 80 mutant alleles in the homozygous state.

At the time the metabolomics experiments reported in [13] were initiated, approximately 70 of the 100 MA lines remained extant, of which 43 ultimately provided sufficient material for Gas Chromatography/Mass Spectrometry (GC-MS). Each MA line was initially replicated five-fold, although not all replicates provided data of sufficient quality to include in subsequent analyses; the mean number of replicates included per MA line is 3.9 (range = 2 to 5). The G0 ancestor was replicated nine times. However, the G0 ancestor was not subdivided into "pseudolines" [33], which means that inferences about mutational variances and covariances are necessarily predicated on the assumption that the among-line (co)variance of the ancestor

is zero. Each replicate consisted of age-synchronized young-adult stage worms taken from a single 10 cm agar plate.

III. Metabolomics:

Details of the extraction and quantification of metabolites are given in [13]. Briefly, samples were analyzed using an Agilent 5975c quadrupole mass spectrometer with a 7890 gas chromatograph. Metabolites were identified by comparison of GC-MS features to the Fiehn Library [34] using the AMDIS deconvolution software [35], followed by reintegration of peaks using the GAVIN Matlab script [36]. Metabolites were quantified and normalized relative to an external quantitation standard. 34 metabolites were identified, of which 29 were ultimately included in the analyses. Normalized metabolite data are archived in Dryad (http://dx.doi.org/10.5061/dryad.2dn09).

IV. Quantitative Genetic Analyses: There are three quantitative genetic parameters of interest: *(i)* the per-generation proportional change in the trait mean, referred to as the mutational bias, ΔM; *(ii)* the per-generation increase in the genetic variance, referred to as the mutational variance, V_M; and *(iii)* the genetic correlation between the cumulative effects of mutations affecting pairs of traits, the mutational correlation, r_M. Details of the calculations of ΔM and V_M are reported in [13]; we reprise the basic calculations here.

(i) Mutational bias (ΔM) – The mutational bias is the change in the trait mean due to the cumulative effects of all mutations accrued over one generation. $\Delta M_z = \mu_G \alpha_z$, where μ_G is the per-genome mutation rate and α_z is the average effect of a mutation on trait z, and is calculated as $\Delta M_z = \frac{\bar{z}_{MA} - \bar{z}_0}{t\bar{z}_0}$, where $\bar{z}_{MA}$ and $\bar{z}_0$represent the MA and ancestral (G0) trait means and t is the number of generations of MA.

(ii) Mutational variance (V_M) - The mutational variance is the increase in the genetic variance due to the cumulative effects of all mutations accrued over one generation. $V_M = \mu_G \alpha_z^2$ and is calculated as $V_M = \Delta V_L = \frac{V_{L,MA} - V_{L,G0}}{2t}$, where $V_{L,MA}$ is the variance among MA lines, $V_{L,G0}$ is the

among-line variance in the G0 ancestor, and t is the number of generations of MA [18, p. 330]. In this study, we must assume that $V_{L,G0} = 0$.

(iii) Mutational correlation, r_M – Pairwise mutational correlations were calculated from the among-line components of (co)variance, which were estimated by REML as implemented in the in the MIXED procedure of SAS v. 9.4, following Fry [37]. Statistical significance of individual correlations was assessed by Z-test, with a global 5% significance criterion of approximately $P<0.000167$.

V. Analysis of the correlation between mutational correlation (r_M) and shortest path length -

Each off-diagonal element of the 24x24 mutational correlation matrix ($r_{ij}=r_{ji}$) was associated with a random shortest path length sampled with probability equal to its frequency in the empirical distribution of shortest path lengths between all metabolites included in the analysis. Next, we calculated the Spearman's correlation ρ between r_M and the shortest path length. The procedure was repeated 10,000 times to generate an empirical distribution of ρ, to which the observed ρ can be compared. This comparison was done for the raw mutational correlation, r_M, the absolute value, $|r_M|$, and between r_M and the shortest path length in the undirected network (i.e., the shorter of the two paths between metabolites i and j).

Acknowledgments –

This work was initially conceived by Armand Leroi and Jake Bundy. We thank Art Edison, Dan Hahn, Tom Hladish and especially Hongwu Ma for their generosity and very helpful advice. Support was provided by NIH grant R01GM107227 to CFB and E. C. Andersen.

References Cited

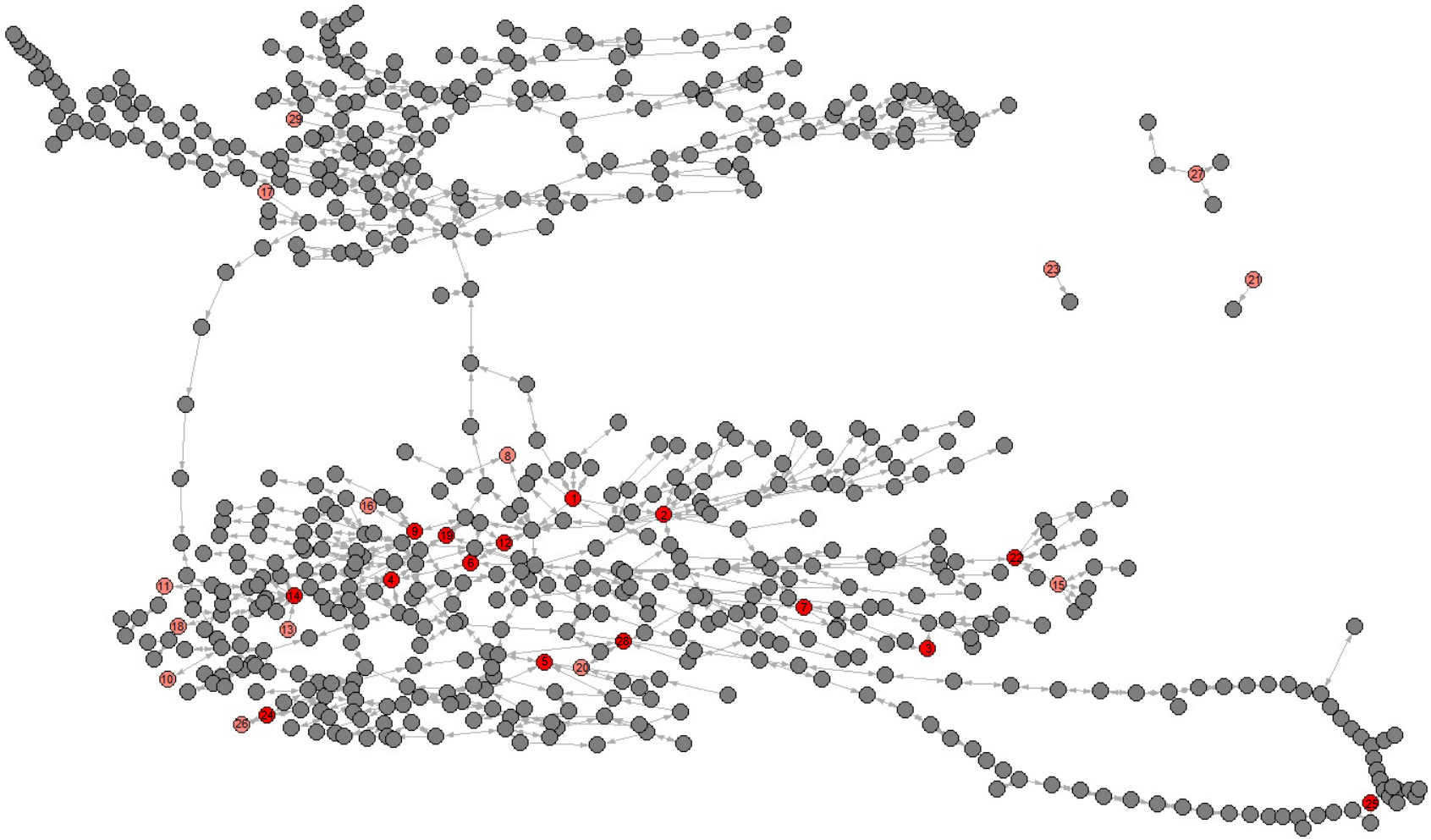
1. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL. The large-scale organization of metabolic networks. Nature. 2000;407(6804):651-4.
2. Winterbach W, Van Mieghem P, Reinders M, Wang HJ, de Ridder D. Topology of molecular interaction networks. Bmc Systems Biology. 2013;7. doi:10.1186/1752-0509-7-90.
3. Minnhagen P, Bernhardsson S. The Blind Watchmaker Network: Scale-Freeness and Evolution. PLoS One. 2008;3(2). doi:10.1371/journal.pone.0001690.
4. Wagner A, Fell DA. The small world inside large metabolic networks. Proceedings of the Royal Society B-Biological Sciences. 2001;268(1478):1803-10.
5. Fell DA, Wagner A. The small world of metabolism. Nature Biotechnology. 2000;18(11):1121-2. doi:10.1038/81025.
6. Bernhardsson S, Minnhagen P. Selective pressure on metabolic network structures as measured from the random blind-watchmaker network. New Journal of Physics. 2010;12. doi:10.1088/1367-2630/12/10/103047.
7. Papp B, Teusink B, Notebaart RA. A critical view of metabolic network adaptations. Hfsp Journal. 2009;3(1):24-35. doi:10.2976/1.3020599.
8. Vitkup D, Kharchenko P, Wagner A. Influence of metabolic network structure and function on enzyme evolution. Genome Biology. 2006;7(5). doi:10.1186/gb-2006-7-5-r39.
9. Greenberg AJ, Stockwell SR, Clark AG. Evolutionary Constraint and Adaptation in the Metabolic Network of Drosophila. Molecular Biology and Evolution. 2008;25(12):2537-46. doi:10.1093/molbev/msn205.
10. Hudson CM, Conant GC. Expression level, cellular compartment and metabolic network position all influence the average selective constraint on mammalian enzymes. Bmc Evolutionary Biology. 2011;11. doi:10.1186/1471-2148-11-89.
11. Kimura M. Evolutionary rate at the molecular level. Nature. 1968;217(5129):624-6.
12. Halligan DL, Keightley PD. Spontaneous mutation accumulation studies in evolutionary genetics. Annual Review of Ecology Evolution and Systematics. 2009;40:151-72. doi:10.1146/annurev.ecolsys.39.110707.173437.
13. Davies SK, Leroi AM, Burt A, Bundy J, Baer CF. The mutational structure of metabolism in *Caenorhabditis elegans*. Evolution. 2016. doi:10.1111/evo.13020.
14. Ma HW, Zeng AP. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics. 2003;19(2):270-7. doi:10.1093/bioinformatics/19.2.270.
15. Yilmaz LS, Walhout Albertha JM. A *Caenorhabditis elegans* genome-scale metabolic network model. Cell Systems. 2016;2(5):297-311. doi:10.1016/j.cels.2016.04.012.
16. Kacser H, Burns JA. The molecular basis of dominance. Genetics. 1981;97(3-4):639-66.
17. Baer CF, Shaw F, Steding C, Baumgartner M, Hawkins A, Houppert A et al. Comparative evolutionary genetics of spontaneous mutations affecting fitness in rhabditid nematodes. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(16):5785-90.
18. Lynch M, Walsh B. Genetics and Analysis of Quantitative Traits. Sunderland, MA.: Sinauer; 1998.
19. Houle D, Morikawa B, Lynch M. Comparing mutational variabilities. Genetics. 1996;143(3):1467-83.
20. Estes S, Ajie BC, Lynch M, Phillips PC. Spontaneous mutational correlations for life-history, morphological and behavioral characters in Caenorhabditis elegans. Genetics. 2005;170(2):645-53. doi:10.1534/genetics.104.040022.
21. Clark AG, Wang L, Hulleberg T. Spontaneous mutation rate of modifiers of metabolism in Drosophila. Genetics. 1995;139(2):767-79.

22. Mezey JG, Houle D. The dimensionality of genetic variation for wing shape in *Drosophila melanogaster*. Evolution. 2005;59(5):1027-38.
23. Shaw RG. Comparison of quantitative genetic parameters - reply. Evolution. 1992;46(6):1967-9. doi:10.2307/2410047.
24. Ma HW, Zeng AP. The connectivity structure, giant strong component and centrality of metabolic networks. Bioinformatics. 2003;19(11):1423-30. doi:10.1093/bioinformatics/btg177.
25. Kim T, Dreher K, Nilo-Poyanco R, Lee I, Fiehn O, Lange BM et al. Patterns of Metabolite Changes Identified from Large-Scale Gene Perturbations in Arabidopsis Using a Genome-Scale Metabolic Network. Plant Physiology. 2015;167(4):1685-U890. doi:10.1104/pp.114.252361.
26. Stearns SC, Kawecki TJ. Fitness sensitivity and the canalization of life-history traits. Evolution. 1994;48(5):1438-50.
27. Houle D. How should we explain variation in the genetic variance of traits? Genetica. 1998;103:241-53.
28. Ho WC, Zhang JZ. Adaptive Genetic Robustness of Escherichia coli Metabolic Fluxes. Molecular Biology and Evolution. 2016;33(5):1164-76. doi:10.1093/molbev/msw002.
29. Witting M, Lucio M, Tziotis D, Wagele B, Suhre K, Voulhoux R et al. DI-ICR-FT-MS-based high-throughput deep metabotyping: a case study of the Caenorhabditis elegans-Pseudomonas aeruginosa infection model. Analytical and Bioanalytical Chemistry. 2015;407(4):1059-73. doi:10.1007/s00216-014-8331-5.
30. Hagberg AA, Schult DA, Swart PJ, editors. Exploring network structure, dynamics, and function using NetworkX. 7th Python in Science Conference (SciPy2008); 2008; Pasadena, CA USA.
31. Keightley PD, Caballero A. Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans*. Proceedings of the National Academy of Sciences of the United States of America. 1997;94(8):3823-7.
32. Denver DR, Wilhelm LJ, Howe DK, Gafner K, Dolan PC, Baer CF. Variation in base-substitution mutation in experimental and natural lineages of Caenorhabditis nematodes. Genome Biology and Evolution. 2012;4(4):513-22. doi:10.1093/gbe/evs028.
33. Teotonio H, Estes S, Phillips PC, Baer CF. Experimental Evolution with Caenorhabditis Nematodes. Genetics. 2017;206(2):691-716. doi:10.1534/genetics.115.186288.
34. Kind T, Wohlgemuth G, Lee do Y, Lu Y, Palazoglu M, S S et al. Mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. Analytical Chemistry. 2009 81:10038-48.
35. Halket JM, Przyborowska A, Stein SE, Mallard WG, Down S, Chalmers RA. Deconvolution gas chromatography/mass spectrometry of urinary organic acids -- potential for pattern recognition and automated identification of metabolic disorders. Rapid Communications in Mass Spectrometry. 1999;13:279-84.
36. Behrends V, Tredwell GD, Bundy JG. Gavin, a add-on to AMDIS, new GUI-driven version. Analytical Biochemistry. 2011;415 206–8.
37. Fry JD. Estimation of genetic variances and covariances by restricted maximum likelihood using PROC MIXED. In: Saxton AM, editor. Genetic Analysis of Complex Traits Using SAS. Cary, NC: SAS Institute, Inc.; 2004. p. 11-34.

Figure Legends

Figure 1. Graphical depiction of the metabolic network including all 29 metabolites. Pink nodes represent included metabolites with core number = 1, red nodes represent included metabolites with core number = 2. Gray nodes represent metabolites with which the included 29 metabolites directly interact. Metabolite identification numbers are listed in Supplementary Table S1.

Figure 1



Network Parameter	Heuristic Definition	Formal Definition
In Degree (**IN°**), $deg^{+}(v)$	The number of incoming edges to node v in a directed graph.	self-explanatory
Out Degree (**OUT°**), $deg^{-}(v)$	The number of outgoing edges from node v in a directed graph.	self-explanatory
Shortest Path Length, $d(v,u)$	Shortest distance from node v to another node u with no repeated walks	self-explanatory
Betweenness Centrality (**BET**), $c_B(v)$	Betweenness centrality of node v is the sum of the fraction of all-pairs shortest paths that pass through v. The greater $c_B(v)$, the greater the fraction of shortest paths that pass through node v.	$\frac{c_B(v)}{(n-1)(n-2)}$, where $c_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t\|v)}{\sigma(s,t)}$, *V* is the set of nodes, $\sigma(s,t)$ is the number of shortest paths from node *s* to node *t*, $\sigma(s,t\|v)$ is the number of paths from *s* to *t* that pass through node v, and *n* is the number of nodes in the graph. The denominator (*n*-1)(*n*-2) is the normalization factor for a directed graph that scales $c_B(v)$ between 0 and 1.

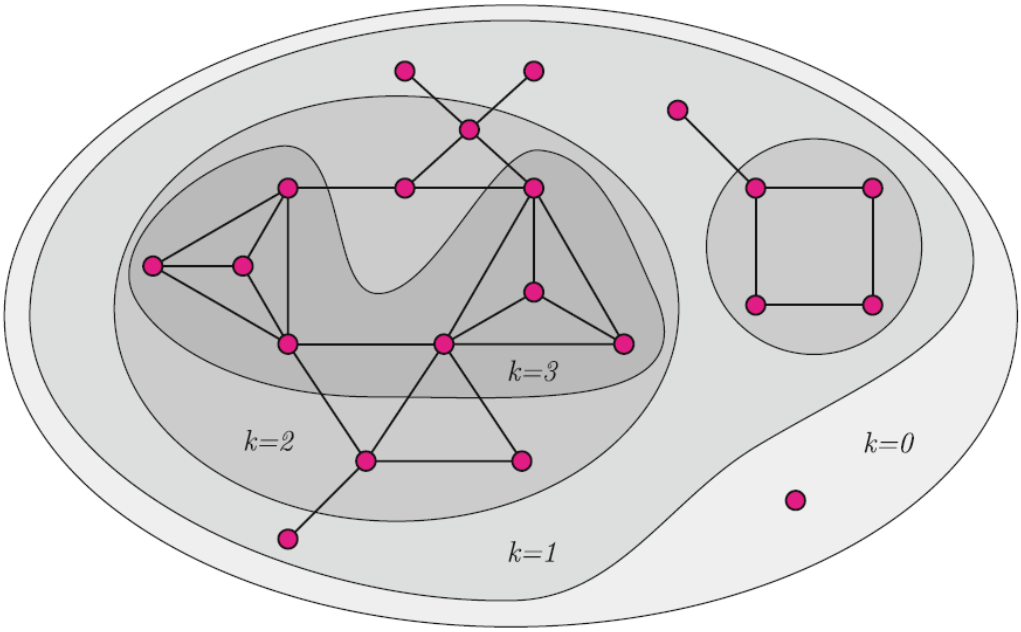
Network Parameter	Heuristic Definition	Formal Definition
Closeness Centrality (**CLO**), *C(*v*)*	Closeness centrality of node v is the reciprocal of the sum of the shortest path lengths to all *n*-1 other nodes, normalized by the sum of minimum possible distances *n-1*. The greater C(v), the closer v is to other nodes.	$C(v) = \frac{n-1}{\sum_{u=1}^{n-1} d(u,v)}$, where *n* is the number of nodes and $d(u,v)$ is the shortest path distance between u and v.
Degree Centrality (**DEG**), $C_D(v)$	Degree centrality of node v is the fraction of nodes in the network that node v is connected to.	$C_D(v) = \frac{deg^+(v)+deg^-(v)}{n-1}$, where *n* is the number of nodes in the network.
Core Number (**CORE**)	A *k*-core is the largest subgraph that contains nodes of at least degree *k*. The core number of node v is the largest value *k* of a *k*-core containing node v.	Calculated using the algorithm of Batagelj and Zaversnik (2011).

Table 1. Definitions of Network Parameters, following the documentation of NetworkX, v.1.11 (Hagberg et al. 2008). The abbreviations of the network parameters used in Table 2 follow the parameter name in parentheses in bold type.

	BTW	CLO	DEG	IN°	OUT°	CORE	ΔM	|ΔM|	I_M	I_E	h_M^2
BTW		0.60	0.84	0.86	0.66	0.79	-0.009 (0.96)	-0.055 (0.77)	-0.007 (0.97)	-0.122 (0.52)	0.128 (0.51)
CLO			0.54	0.51	0.47	0.54	0.012 (0.94)	0.297 (0.11)	0.119 (0.53)	0.034 (0.86)	0.089 (0.64)
DEG				0.88	0.92	0.83	0.038 (0.84)	-0.078 (0.68)	0.178 (0.35)	-0.062 (0.74)	0.218 (0.25)
IN°					0.65	0.85	0.099 (0.60)	0.043 (0.82)	0.188 (0.32)	0.007 (0.97)	0.277 (0.14)
OUT°						0.68	0.031 (0.87)	-0.200 (0.29)	0.133 (0.49)	-0.096 (0.62)	0.139 (0.47)
CORE							0.245 (0.20)	0.104 (0.59)	0.298 (0.11)	0.025 (0.89)	0.481 (0.008)

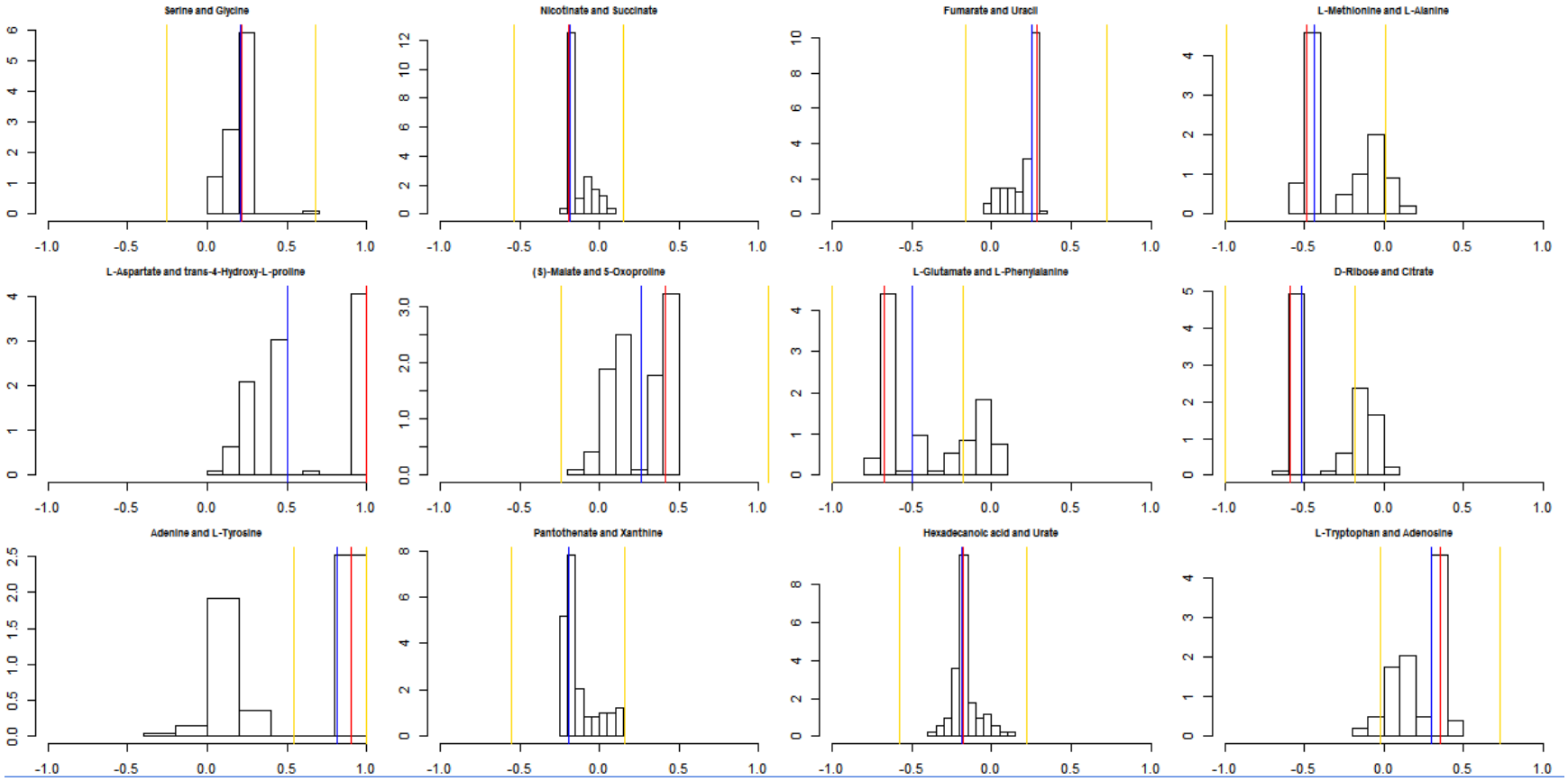
Table 2. Spearman's rank correlation ρ between network parameters (rows/first five columns) and between network parameters and mutational parameters (rows/last four columns). Abbreviations of network parameters are: BTW, betweenness centrality; CLO, closeness centrality; DEG, degree centrality; IN°, in-degree, OUT°, out-degree; CORE, core number. Network parameters are defined mathematically and heuristically in Table 1. Abbreviations of mutational parameters are: ΔM, per-generation change in the trait mean; |ΔM|, absolute value of ΔM; I_M, squared mutational coefficient of variation; I_E, squared residual coefficient of variation; h_M^2, mutational heritability. See text and Supplementary Table S1 for details of mutational parameters. Uncorrected P-values of mutational parameters are given in parentheses.

Supplementary Figure S1.



Supplementary Figure S1. Schematic depiction of the *k*-core(s) of a graph. The *k*-core of a graph is the largest subgraph that contains nodes of degree at least *k*. The colored balls represent nodes in a network and the black lines represent connecting edges. Each red ball in the darkest gray area has core number *k*=3; note that each node with *k*=3 is connected to <u>at least</u> three other nodes. From Batagelj and Zaveršnik (2011).

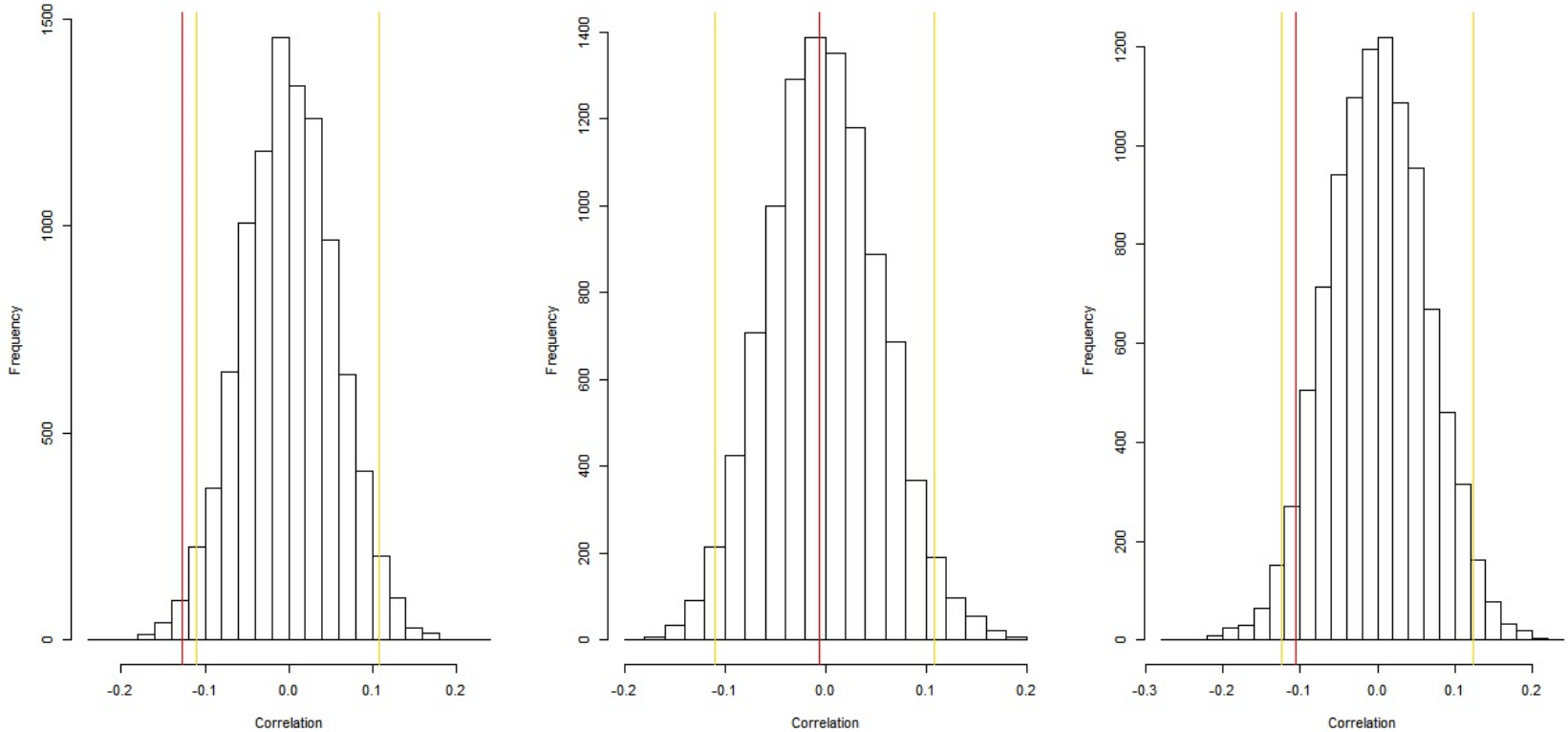
Supplementary Figure S2



Supplementary Figure S2. Bootstrap distributions of mutational correlations (r_M) calculated from the joint REML estimate of the among-line components of covariance from a pair of focal metabolites (listed at the top of each panel) and six other metabolites randomly sampled without replacement. Each distribution is based on 100 resamples from the data. Red lines show the observed

r_M, blue lines show the median of the resampled values, yellow lines show $\pm$ two standard errors of the observed r_M. Details of the bootstrap analysis are given in the Methods.

Supplementary Figure S3



Supplementary Figure S3. Parametric bootstrap distributions of random correlations ρ between (a) r_M and the shortest path length in the directed network, (b) $|r_M|$ and the shortest path length in the directed network, (c) r_M and shortest path length in the undirected network (i.e., the shorter of the two path lengths between metabolites *i* and *j* in the directed network). Red lines show the observed values of ρ, yellow lines show the 95% confidence interval of the distribution of the correlation between the mutational correlation and a random shortest path length drawn from the observed distribution of shortest path lengths. See Methods for details.

x

Metabolite	***ID #***	**BTW**	**CLO**	**DEG**	**IN°**	**OUT°**	**CORE**
trans-4-Hydroxy-L-proline	11	0.004482	0.002761	0.004658	2	1	1
L-3-Amino-isobutanoate	10	0	0	0.001553	1	0	1
Adenine	20	0.003335	0.032205	0.009317	2	4	2
Adenosine	28	0.005516	0.032198	0.009317	4	2	2
L-Alanine	8	0	0	0.001553	1	0	1
L-Asparagine	16	0.002236	0.052579	0.004658	1	2	1
L-Aspartate	9	0.055054	0.056756	0.012422	3	5	2
Citrate	19	0.171567	0.058576	0.012422	4	4	2
Fumarate	6	0.028317	0.053784	0.007764	3	2	2
L-Glutamate	14	0.036659	0.048283	0.020186	5	8	2
Glycine	2	0.02661	0.044377	0.017081	5	6	2
L-Lysine	21	0	0.004658	0.004658	0	3	1
(S) – Malate	12	0.029016	0.057446	0.007764	2	3	2
L-Methionine	7	7.49E-05	0.004969	0.007764	2	3	2
Nicotinate	3	0.00221	0.046122	0.006211	2	2	2
Hexadecanoic acid	25	0.066162	0.019529	0.01087	2	5	2

Metabolite	***ID #***	**BTW**	**CLO**	**DEG**	**In°**	**Out°**	**Core**
Pantothenate	23	0	0.001553	0.001553	0	1	1
L-Phenylalanine	15	0.001137	0.040279	0.009317	1	5	1
Putrescine	18	0.003362	0.002795	0.003106	1	1	1
5-Oxoproline	13	0	0.045387	0.001553	0	1	1
D-Ribose	17	0	0.044346	0.001553	0	1	1
L-Serine	1	0.034582	0.057313	0.017081	5	6	2
Succinate	4	0.02215	0.051067	0.01087	4	3	2
Alpha,alpha-Trehalose	29	0	0.042179	0.001553	0	1	1
L-Tryptophan	27	0	0.004969	0.004658	0	3	1
L-Tyrosine	22	0.004518	0.041547	0.012422	3	5	2
Uracil	5	0.078303	0.044365	0.01087	4	3	2
Urate	26	0	0.029702	0.003106	1	1	1
Xanthine	24	0.004393	0.031003	0.01087	3	4	2

Supplementary Table S1(a), above. Network parameters for the 29 metabolites. Column headings are the abbreviations for the network parameters given in Tables 1 and 2. ID# is the number of the metabolite in the network shown in Figure 1.

Metabolite	***ID #***	***Mean (G0)***	***Mean (MA)***	***ΔM (%)***	***VL***	***VM***	***VE***	I_M	h^2_M ***(x 10³)***
trans-4-Hydroxy-L-proline	11	228.6 (10.4)	149.4 (14.5)	0.14 (0.02)	1058.0 (349.0)	2.12 (0.70)	3577.4 (1052.3)	9.45E-05	0.59 (0.26)
L-3-Amino-isobutanoate	10	20.0 (3.1)	14.5 (0.8)	-0.11 (0.06)	3.58 (4.32)	0.007 (0.009)	95.0 (9.86)	3.43E-05	0.08 (0.09)
Adenine	20	11.3 (1.4)	13.2 (0.8)	0.07 (0.06)	15.4 (4.4)	0.03 (0.009)	28.0 (4.02)	0.000178	1.10 (0.35)
Adenosine	28	2.1 (0.7)	26.0 (3.4)	4.39 (0.62)	407.7 (78.2)	0.82 (0.16)	186.5 (75.9)	0.001261	4.37 (1.97)
L-Alanine	8	5.5 (1.5)	6.1 (0.6)	0.04 (0.12)	5.48 (3.43)	0.011 (0.007)	30.3 (5.0)	0.000278	0.36 (0.23)
L-Asparagine	16	3.7 (1.3)	1.2 (0.1)	-0.27 (0.14)	0.11 (0.15)	0.0002 (0.0003)	2.4 (0.4)	0.000135	0.09 (0.13)
L-Aspartate	9	37.7 (2.7)	19.0 (1.2)	-0.20 (0.03)	48.2 (19.2)	0.096 (0.038)	60.2 (18.2)	0.00026	1.60 (0.80)
Citrate	19	11.4 (2.2)	5.7.1 (0.5)	-0.16 (0.08)	4.56 (1.54)	0.009 (0.003)	19.8 (6.8)	0.000182	0.46 (0.22)

Metabolite	***ID #***	***Mean (G0)***	***Mean (MA)***	***ΔM (%)***	***VL***	***VM***	***VE***	I_M	h^2_M ***(x 10³)***
Fumarate	6	33.0 (2.1)	25.5 (1.2)	-0.09 (0.03)	35.1 (9.2)	0.070 (0.018)	84.1 (26.1)	0.000109	0.83 (0.34)
L-Glutamate	14	115.2 (24.5)	93.2 (6.3)	-0.08 (0.09)	973.0 (556.8)	1.94 (1.11)	2703.7 (556.9)	0.000214	0.72 (0.44)
Glycine	2	47.1 (5.3)	52.8 (4.5)	0.05 (0.06)	606.7 (411.6)	1.21 (0.82)	939.3 (323.7)	0.000393	1.29 (0.98)
L-Lysine	21	6.7 (2.6)	5.0 (0.6)	-0.10 (0.16)	4.43 (3.99)	0.009 (0.008)	38.4 (7.01)	0.00032	0.23 (0.21)
(S) – Malate	12	44.7 (4.0)	71.9 (4.2)	0.24 (0.05)	397.8 (135.4)	0.80 (0.27)	1179.7 (276.2)	0.000157	0.67 (0.28)
L-Methionine	7	47.9 (4.1)	32.2 (1.7)	-0.13 (0.04)	101.2 (20.5)	0.202 (0.041)	72.4 (13.4)	0.000197	2.79 (0.77)
Nicotinate	3	4.9 (0.6)	33.0 (3.5)	2.29 (0.29)	473.9 (88.1)	0.95 (0.18)	120.3 (21.4)	0.000863	7.88 (2.03)
Hexadecanoic acid	25	238.4 (22.3)	265.1 (15.0)	-0.03 (0.04)	6712.9 (1579.6)	13.43 (3.16)	9579.1 (2985.7)	0.000194	1.40 (0.55)

Metabolite	***ID #***	***Mean (G0)***	***Mean (MA)***	***ΔM (%)***	***VL***	***VM***	***VE***	I_M	h^2_M ***(x 10³)***
Pantothenate	23	22.9 (0.8)	14.4 (0.6)	-0.15 (0.02)	12.8 3.2	0.026 (0.006)	16.0 (3.17)	0.000123	1.60 (0.51)
L-Phenylalanine	15	87.4 (7.9)	94.4 (6.3)	0.03 (0.05)	678.9 (263.6)	1.36 (0.53)	3835.0 (810.3)	0.000155	0.35 (0.16)
Putrescine	18	73.7 (15.3)	57.4 (3.4)	-0.09 (0.08)	70.9 (70.6)	0.14 (0.14)	1672.3 (193.2)	4.15E-05	0.08 (0.08)
5-Oxoproline	13	701.6 (40.0)	528.2 (22.2)	-0.10 (0.03)	7275.6 (3649.0)	14.56 (7.30)	52146.6 (13671.9)	5.24E-05	0.28 (0.16)
D-Ribose	17	5.6 (0.8)	13.3 (1.5)	0.51 (0.11)	76.3 (22.8)	0.15 (0.05)	66.3 (27.2)	0.000841	2.30 (1.17)
L-Serine	1	130.0 (49.8)	85.8 (3.7)	-0.14 (0.15)	373.7 (117.7)	0.75 (0.24)	1221.7 (376.9)	9.87E-05	0.61 (0.27)
Succinate	4	7.3 (0.8)	91.1 (9.7)	4.52 (0.51)	3216.6 (688.3)	6.43 (1.38)	2797.7 (932.6)	0.000778	2.30 (0.91)
Alpha,alpha-Trehalose	29	1772.2 (147.2)	2525.4 (277.6)	0.19 (0.07)	2118803 (105069)	4237.6 (2101.4)	4355039 (1656108)	0.000584	0.97 (0.61)

Metabolite	***ID #***	***Mean (G0)***	***Mean (MA)***	***ΔM (%)***	***VL***	***VM***	***VE***	I_M	h^2_M ***(x 10³)***
L-Tryptophan	27	107.7 (14.3)	92.2 (2.8)	-0.06 (0.05)	205.7 (87.3)	0.411 (0.174)	496.9 (64.0)	4.93E-05	0.83 (0.37)
L-Tyrosine	22	74.7 (9.3)	47.9 (2.9)	-0.14 (0.05)	197.2 (70.9)	0.394 (0.142)	643.3 (99.7)	0.00017	0.61 (0.24)
Uracil	5	9.5 (1.0)	8.8 (0.4)	-0.03 (0.05)	4.67 (1.59)	0.009 (0.003)	8.3 (1.4)	0.000123	1.13 (0.43)
Urate	26	20.7 (3.4)	11.5 (1.2)	-0.18 (0.07)	45.3 (20.1)	0.091 (0.040)	45.3 (4.75)	0.000654	2.00 (0.91)
Xanthine	24	0.4 (0.3)	6.7 (1.0)	6.64 (1.09)	36.5 (8.04)	0.073 (0.016)	15.5 (2.8)	0.001711	4.70 (1.34)
Mean			0.73* (0.30)				1.46 (0.31)		
Median			0.14				0.83		

Supplementary Table S1(b). Mutational statistics of the 29 metabolites, reprised from Supplementary Table S1 of Davies et al. (2016). Column Headings are: *ID#*, the number of the metabolite in the network shown in Figure 1; *Mean (G0)*, mean metabolite pool of the G0 ancestor; *Mean (MA)*, mean metabolite pool of the MA lines; *ΔM (%)*, percent change per-generation in mean trait

value; *VL*, among-line component of variance, *VM*, mutational variance; $CV_{M,G0}$, mutational coefficient of variation standardized by the G0 mean; $CV_{M,MA}$, mutational coefficient of variation standardized by the MA mean; *VE*, environmental (= within-line) component of variance; h_M^2, mutational heritability.

* - Mean *ΔM (%)* is the mean absolute value.

Reaction Number	Reaction
R01579	D-Glutamine + H2O = D-Glutamate + NH3
R01887	gamma-Amino-gamma-cyanobutanoate + 2 H2O = DL-Glutamate + NH3
R04936	Se-Adenosylselenohomocysteine + H2O = Adenosine +Selenohomocysteine
R00891	L-Serine + Hydrogen sulfide = L-Cysteine + H2O
R09099	L-Serine + 5,6,7,8-Tetrahydromethanopterin = 5,10-Methylenetetrahydromethanopterin + Glycine + H2O
R02853	D-O-Phosphoserine + H2O = D-Serine + Orthophosphate
R00904	3-Aminopropanal + NAD+ + H2O = beta-Alanine + NADH + H+
R03542	alpha-Aminopropiononitrile + 2 H2O = Alanine + NH3
R01324	Citrate = Isocitrate
R00483	ATP + L-Aspartate + NH3 = AMP + Diphosphate + L-Asparagine
R01221	Glycine + Tetrahydrofolate + NAD+ = 5,10-Methylenetetrahydrofolate+ NH3 + CO2 + NADH + H+
R02078	3,4-Dihydroxy-L-phenylalanine + L-Tyrosine + Oxygen = Dopaquinone+ 3,4-Dihydroxy-L-phenylalanine + H2O
R01706	Hexadecanoyl-[acp] + H2O = Acyl-carrier protein + Hexadecanoicacid
R04666	3-Ureidoisobutyrate + H2O = 3-Aminoisobutyric acid + CO2 + NH3

Supplementary Table S2. Discrepancies between the metabolic networks constructed using the MZ and YW methods. All reactions listed here are in the Ma and Zeng KEGG database (http://www.ibiodesign.net/kneva/) but not in the Wormflux database (http://wormflux.umassmed.edu/) and were used in the generation of the metabolic network. There is a total of 1203 reactions in the network, these represent about 1% of all reactions.