BMC Evolutionary Biology

Aligning functional network constraint to evolutionary outcomes

--Manuscript Draft--

Manuscript Number:	EVOB-D-18-00353R3
Full Title:	Aligning functional network constraint to evolutionary outcomes
Article Type:	Research article
Section/Category:	Genome Evolution and Evolutionary Systems Biology
Funding Information:	
Abstract:	Background. Functional constraint through genomic architecture is suggested to be an important dimension of genome evolution, but quantitative evidence for this idea is rare. In this contribution, existing evidence and discussions on genomic architecture as constraint for convergent evolution, rapid adaptation, and genic adaptation are summarized into alternative, testable hypotheses. Network architecture statistics from protein-protein interaction networks are then used to calculate differences in evolutionary outcomes on the example of genomic evolution among yeast, and the results are used to evaluate statistical support for these longstanding hypotheses. Results. A discriminant function analysis lent statistical support to classifying the yeast interactome into hub, intermediate and peripheral nodes based on network neighborhood connectivity, betweenness centrality, and average shortest path length. Quantitative support for the existence of genomic architecture as mechanistic basis for evolutionary constraint is then revealed through utilizing these statistical parameters of the protein-protein interaction network in combination with estimators of protein evolution. Conclusions. As functional genetic networks are becoming increasingly available, it will now be possible to evaluate functional genetic network constraint against variables describing complex phenotypes and environments, for better understanding of commonly observed deterministic patterns of evolution in non-model organisms. The hypothesis framework and methodological approach outlined herein may help to quantify the extrinsic versus intrinsic dimensions of evolutionary constraint, and result in a better understanding of how fast, effectively, or deterministically organisms adapt.
Corresponding Author:	Katharina Wollenberg Valero University of Hull UNITED KINGDOM
Corresponding Author E-Mail:	k.wollenberg-valero@hull.ac.uk
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Hull
Corresponding Author's Secondary Institution:	
First Author:	Katharina Wollenberg Valero
First Author Secondary Information:	
Order of Authors:	Katharina Wollenberg Valero
Order of Authors Secondary Information:	
Response to Reviewers:	- the email on my profile has been updated to match the manuscript - the response to reviews has been removed from files - the manuscript was uploaded as a clean version
Additional Information:	
Question	**Response**
Has this manuscript been submitted before to this journal or another journal in the <a	Yes

href="https://www.biomedcentral.com/p/the-bmc-series-journals#journallist" target="_blank" >BMC series</ a>?	
Please provide the manuscript identification number from your previous submission. If you no longer have the identification number, please specify this in the text box below. as follow-up to "Has this manuscript been submitted before to this journal or another journal in the <a href="https://www.biomedcentral.com/p/the-bmc-series-journals#journallist" target="_blank" >BMC series</ a>?"	GBIO-D-18-01393

Aligning functional network constraint to evolutionary outcomes

Katharina C. Wollenberg Valero

University of Hull, Department of Biological and Marine Sciences, Cottingham Road HU6 7RX, Kingston-Upon-Hull, United Kingdom
Contact: k.wollenberg-valero@hull.ac.uk

Abstract

Background. Functional constraint through genomic architecture is suggested to be an important dimension of genome evolution, but quantitative evidence for this idea is rare. In this contribution, existing evidence and discussions on genomic architecture as constraint for convergent evolution, rapid adaptation, and genic adaptation are summarized into alternative, testable hypotheses. Network architecture statistics from protein-protein interaction networks are then used to calculate differences in evolutionary outcomes on the example of genomic evolution among yeast, and the results are used to evaluate statistical support for these longstanding hypotheses.

Results. A discriminant function analysis lent statistical support to classifying the yeast interactome into hub, intermediate and peripheral nodes based on network neighborhood connectivity, betweenness centrality, and average shortest path length. Quantitative support for the existence of genomic architecture as mechanistic basis for evolutionary constraint is then revealed through utilizing these statistical parameters of the protein-protein interaction network in combination with estimators of protein evolution.

Conclusions. As functional genetic networks are becoming increasingly available, it will now be possible to evaluate functional genetic network constraint against variables describing complex phenotypes and environments, for better understanding of commonly observed deterministic patterns of evolution in non-model organisms. The hypothesis framework and methodological approach outlined herein may help to quantify the extrinsic versus intrinsic dimensions of evolutionary constraint, and result in a better understanding of how fast, effectively, or deterministically organisms adapt.

Keywords: evolution, constraint, adaptation, systems biology

Background

Genetic constraint and evolutionary outcomes

Understanding the nature of the genetic dimension of evolutionary constraint is crucial to understanding adaptation and other repeatedly observed outcomes of evolution such as convergent phenotypes, rapid adaptation, or genic evolution (see Glossary in Table 1). For example, divergent genetic populations of the well-studied Caribbean lizard *Anolis cybotes* [1] have nonetheless evolved convergent phenotypic, ecological, reproductive, and physiological adaptations to high elevations on three separate mountain chains, which is mirrored by genomic adaptations in a subset of genes [2–5]. These observations made in natural populations suggest that the variants available to mutation and selection may be constrained at the genomic level, enabling faster adaptations and higher rates of convergent evolution than it were possible without constraint.

Many studies have shown the non-independence of genes from one another, be it through physical linkage, phylogenetic relationship (e.g., in the case of whole genome duplications), or functional interaction (Figure 1). Futuyma [6] cited Schluter [7], noting that correlations between genes could reduce the degrees of freedom on which selection can operate. Mayr [8] stated that "coadapted" genes are a result of natural selection, being brought together to form a "balanced system", but ruled out that such gene complexes would be of any interest to evolutionary biology, as ultimately only the complete phenotype is selected ([8], p.184ff). Nonetheless, evolutionary trajectories of complex phenotypes have been extensively studied through the concept of the genetic variance-covariance or G-matrix [9]. Some mechanistic properties of the genome leading to these constraints that can be expressed as a G-matrix are trait polygeny, trait pleiotropy, and linkage (Figure 1, [10]), but the evolutionary constraints of correlated traits implied from a G-matrix can be rapidly overcome in only a few generations [11], hinting at additional genomic properties influencing evolutionary constraint. A more detailed look into the mechanistic basis of constrained phenotypic evolution at the molecular level is therefore necessary, and now made possible through the rapid accumulation of genomic and other molecular -omics data sets in the public domain.

Genetic constraint through gene functional importance

The functional importance of a gene has long been thought to cause such evolutionary constraint at the molecular level: protein-coding genes that are indispensable for the organism should be highly conserved and thus, be constrained through evolution, as most nonsynonymous mutations would be detrimental to protein function and thus would most likely result in non-adaptive phenotypes. Consequently, these genes should have a lower rate of molecular evolution. Such genes have formerly been identified through their “dispensability”. This term describes how essential genes are for organismal function within a certain environmental context, which can be estimated through knockout experiments (Table 1).

Zhang and Yang [12] reviewed evidence from empirical studies, but found that essential genes are not evolving more slowly than nonessential genes. Instead, highly expressed genes seem to have lower rates of protein evolution (dubbed the “E-R anticorrelation” [13, 14], which some authors relate to translational selection on amino acids with different metabolic cost [13]. Many studies have ascribed an important role to gene expression levels in constraining evolutionary rate of proteins [15–17]. But perhaps, functional importance needs to be defined differently than via gene essentiality or dispensability, and expression level may be a correlative variable linked to another cause. In *Saccharomyces cerevisiae* (in following: yeast), which was used for many studies on protein evolutionary rate and functional importance, essential genes are required for organismal growth and performance under optimal environmental conditions. A gene that renders an organism nonfunctional may thus predominantly be active in genetic pathways related to development and growth. However, in a multicellular organism such as a vertebrate, also the genes that are essential for organismal viability and reproduction are of high functional importance, and potentially could be under evolutionary constraint [18], such as in the example of genes coding for eye color determining mating success in *Drosophila melanogaster* [19]. A high proportion of the human genome has also been found to be under selective constraint in other mammals, indicating that gene dispensability is not a binary variable [20]. As reviewed by Zhang and Yang [12], Wilson et al. [21] suggested that evolutionary rate might be determined by both functional importance and functional constraint [12, 21]. If functional importance measured as (negative) gene dispensability does not predict variations and constraints of evolutionary rate, perhaps functionally more constrained genes are the ones evolving slowly. Prior studies have attempted to identify functional constraint in terms of which sites within a protein are essential for performing its function, called protein functional site constraint in Figure 1. The Neutral Theory [22] already identified codon constraint where nonsynonymous mutations are of larger consequence than synonymous ones as being important for evolution (Figure 1).

Genetic constraint through gene pleiotropy and network architecture

During the recent decades, network thinking has emerged as a powerful approach for better understanding biological realities [23]. The network concept might also have deep implications in evolutionary biology. Gene interaction networks were found to evolve either faster or slower than comparable genes functioning without being connected to others [24–27], and gene regulatory circuits convergently evolve in the absence of shared ancestry [28]. The overall network architecture or hierarchy of genes within the network is likely to contribute to the speed and mode of evolution and the phenotype components associated with it, regulated through functional constraint of nodes within the network [23]. For example, a study by Jeong and colleagues [29] found that genes with many functional interaction partners are also likely to be essential, which, however, does not provide enough evidence to extrapolate directly from functional constraint to evolutionary outcome.

Functional genetic network structure has been shown to affect evolutionary outcomes through "gene pleiotropy" in yeast: gene products that interact with many others are thought to be involved in many cellular pathways and by that means, to have multiple (pleiotropic) effects on the cellular function [30, 31]. Fitness effects of mutations in pleiotropic genes could be partitioned across several phenotypic components, increasing the likelihood of maladaptive effects, which means that they should be more conserved through evolution and evolve more slowly [30]. However, it is important to note, that despite that a connection between pleiotropic gene and pleiotropic phenotype was implied in these studies [30, 31], gene pleiotropy or the number of functional connections a gene has with others should be regarded as a distinct concept from phenotypic pleiotropy, unless such a relationship to pleiotropic phenotype has been demonstrated.

Proteins with many interactants may be constrained in the evolution of their functional sites to instances of co-evolution with the interactant genes, in order to maintain their functionality [32]. A corresponding model of evolutionary constraint on evolution through gene pleiotropy that was explicitly based on functional network node hierarchy within an interactome was proposed by Pavlicev and Wagner [33]. They argued that for genetic adaptation in a target gene to happen, selection has to overcome the inertia generated through stabilizing selection of the genes functionally connected with the target [33]. The premise of this model is that any change in genotype-phenotype interaction represents a change in a developmental pathway and, due the position of a gene within a network, will have pleiotropic effects on the phenotype [33]. For example, most pleiotropic genes with many interaction partners only had a small pleiotropic effect on the phenotype, but some genes with large phenotypic effect were also more pleiotropic [18]. High gene pleiotropy is assumed to have a cost for adaptation, which was explained as nodes central to a network with many interaction partners evolving slower [18, 34]. This idea, dubbed the “cost of complexity” [35] would lead to faster evolution of organisms with less complex genomic architecture due to this constraint being relaxed [18], and to adaptive selection on standing genetic variation preferentially to occur in genes with low pleiotropic effects [36]. Concerning evolutionary outcomes, gene pleiotropy was suggested to limit events of genomic co-evolution [32], genomic adaptation [36], and convergent evolution [36] in nodes central to a network. Consequently, the properties of nodes within a functional genetic network may be informative to understand their evolutionary constraint. However, gene pleiotropy was defined by most of these authors [30–33, 35] as synonymy with number of interactants and with centrality in a network - but looking at more recently generated interactomes, nodes topologically central to the network are usually not the nodes with the highest number of connections. Instead, nodes with most connections are located in intermediate positions within a network [37]. The number of edges of a node consequently, may not be sufficient to disentangle the effects of network structure on evolutionary constraint since it only measures one of a network’s many properties. The concept of variable genomic networks existing within populations was first explored by Wagner [38] and was represented through a hypothetical “genotype space” of similar phenotypes that might correspond to the concept of “phenotypic optima”. Selection can cause a population to modify their genotype networks in a way that renders them more robust to changes in the fitness landscape.

While the concept of network architecture influencing evolutionary outcomes is known from the studies outlined above and from others, in many cases this concept has not been sufficiently transformed into testable hypotheses yet, and correspondingly, no straightforward methodology exists for biologists to test them empirically. **The first aim** of this paper is to deconstruct the abstract concept of gene pleiotropy by setting genomic network architecture in relation to the three evolutionary outcomes: 1) genomic re-use generating convergent phenotypes, 2) the simultaneous occurrence of convergence and divergence within a genome, resulting in genic adaptation, and 3) the speed of adaptation. For this purpose, I propose three categories of nodes with different putative evolutionary trajectories. I set these categories in relation to previously defined hypotheses and expectations aligning functional network constraint to evolutionary outcomes. **The second aim** is to demonstrate how these hypotheses can be quantitatively tested. First, the nodes of the yeast interaction network are transformed into categories based on network statistical parameters and discriminant function analysis. Secondly, these are then tested for differences of proxy measures of evolutionary outcomes using data from yeast evolution. For this I use published data, which aligns this study to others with similar questions [13, 39, 40] but utilizes a novel approach.

Results and Discussion

Aim 1: Novel node classification scheme based on network statistics

Figure 2 outlines a testable, hypothetical scenario of how functional genetic network architecture could influence evolutionary outcomes: When selection acts upon a population (for example, through a sudden change in climate), advantageous mutations will be selected from standing genetic variation (allele frequencies). A population will have standing genetic variation in different nodes of the network, which is dependent on the topological position of the nodes. Organisms possess a finite subset of biochemical pathways (underlying functional genetic networks) such as those related to temperature homeostasis [41–43], and that align to a finite amount of selected phenotype components. The population must adapt to this newly arising selective pressure through selection of advantageous mutations in one of these subnetworks, but not by selecting mutations in any other subnetwork, as these are unrelated to the stimulus or organismal fitness in response to it, and would therefore not result in adaptation. This does not mean that other subnetworks are not under any selection, or under stabilizing selection for other causes, or that selection on one subnetworks does not influence others, but is a simplification here used for the purpose of classifying nodes. This constrains the number of mutations in the genome that selection will operate on, and thus determines the evolutionary response through genetic constraint. Second, and of high importance for the new classification scheme proposed here, node hierarchy within these subnetworks poses an additional level of constraint: and this additional level reduces the "evolutionary search space" for potential beneficial variants.

This can be illustrated through the following hypothetical construct, which reduces network structure to distinct types of nodes. Network nodes, which are functionally important for the operation of the network (hub nodes central to the network - in following **H-nodes**), are less likely to harbor significant genetic variation in first- or second-codon positions or regulatory regions because of their high functional constraint. Consequently, genetic variation, as well as adaptation to an environmental selective pressure, should both be more likely to occur within non-hub nodes within the subnetwork. Nodes with highest number of edges are intermediately positioned within a network (intermediate nodes, in following **I-nodes**) and were shown to have weaker selective constraint [44, 45] than centrally positioned nodes, as they have lower functional constraint than H-nodes. Consequently, they should evolve faster. This assumption differs from the gene pleiotropy hypothesis, which places the highest functional constraint on these nodes. However, because of their high cost of complexity, adaptation in I-nodes should be highly constrained in terms of which genes can adapt (depending on the nature of their functional interactions) and how (through changing the wiring pattern with other nodes). Because of gene pleiotropy, adaptations that do evolve in these nodes should have a larger phenotypic effect, which combined with the reduced possibilities for adaptation, increases the likelihood for convergent evolution in them. Genes peripheral in the network (peripheral nodes, in following, **P-nodes**) have higher degrees of freedom due to the lowest degree of gene pleiotropy and should be able to accumulate genetic variation with least cost. Therefore, the population should already harbor more genetic variation within these peripheral genes on which selection can operate. Change in such nodes however, due to lower gene pleiotropic interactions, would result in less phenotypic effect and thus they are less likely to promote large evolutionary changes. In such nodes, divergence is more likely to accumulate than convergence. The expectation is thus that different node types will differ in standing genetic variation due to the different genetic constraints acting upon them. **H-nodes** will be very strongly constrained and only can accumulate little standing genetic variation, resulting in a low potential for selection to operate on. I-nodes will harbor sufficient standing genetic variation but be under high functional constraint, so that selection can only operate on a limited amount of variants that all have multiple phenotypic effects. In different organisms, the same variants can be selected quickly due to this reduced search space, which leads to parallel genomic evolution resulting in convergent phenotypes. **P-nodes** will be least constrained, allowing a lot of variation but less sweeping phenotypic effects due to lower gene pleiotropy. Selection can operate on multiple variants in these; selective advantages are more likely due to the lower gene pleiotropy in more genes, so selection will less likely lead to convergence. All three evolutionary outcomes can (among other factors such as gene expression levels) be explained with this mechanism of constraint through functional genetic network structure.

As Fraser [46] pointed out, objective classification of nodes into any such type of categories is important, and they should reflect true topological properties of an interactome. Therefore, values for three network statistical parameters were obtained from the yeast interactome whose definition corresponds to the above outlined node types. Those parameters were average shortest path length (maximal in peripheral nodes), neighborhood connectivity (maximal in nodes intermediate to the network), and betweenness centrality (maximal in nodes connecting subnetworks). Maximal values for each statistical parameter were used to bin nodes into P, I and H nodes, and a Discriminant Function Analysis yielded significant support for the allocation of network statistical parameters to these **P, I, and H** node categories (Figure 3, Table 3). To explore the network position of nodes that have undergone convergent adaptation, yeast ORF IDs that were demonstrated experimentally to show convergent genomic adaptation in independent experiments, strains, or species of yeasts were identified from the literature (66, 74-78), Table 4). These nodes were classified as (**C-nodes**).

All network statistical parameters significantly differed between node categories, as shown with Kruskal-Wallis tests: Average shortest path length: KW-H (3,2208)=1220.590, $p<0.0001$; neighborhood connectivity, KW-H(3,2204)=926.571, $p<0.0001$; betweenness centrality KW-H(3,2208)=293.849, $p<0.0001$ (Figure 4). Fraser's [46] first exploration of the influence of modules within genomes and their hub nodes (called "modularity") found that the rate of protein evolution is faster in intramodule hubs (nodes that link genes with high co-expression in response to a stimulus) compared to intermodule hubs (linking low genes with low co-expression in response to a stimulus, defined after [47]. The node classification scheme of Fraser [46] was based on gene co-expression and not on node topology, and gene co-expression in response to a stimulus followed a bimodal distribution. In this contribution, **I nodes** instead have the highest number of edges and connect sub-networks, but are not defined with respect to their expression.

Aim 2: Genetic constraint and network architecture influencing evolution

A recent paper published by Schoenrock et al. [39] uses a data set of 4,179 protein-coding genes (sourced from [13, 40] to investigate the involvement of network structure in protein evolution. This data set was generated for five species of yeasts (*Saccharomyces cerevisiae*, *S. paradoxus*, *S. bayanus*, *S. kudriavzevii*, and *S. mikatae*). The study compared a quantitative variable related to network structure (computationally predicted re-wiring of nodes through evolution γ), with an estimator of protein evolutionary rate on nodes (substitution rate ω, measured as dN/dS). The authors find that the degree of rewiring of nodes across the phylogeny is only poorly associated with evolutionary sequence divergence, but nodes with very low evolutionary rate had high variability of rewiring scores, which indicates that changing gene interactions is an important mechanism how functionally constrained genes may evolve. While the study remained somewhat inconclusive about the influence of network structure and node rewiring on protein evolution, the data contained within it, combined with additional data, allow demonstrating a test of the hypotheses outlined above using the new node classification scheme (The results are compared to previous studies in Table 2).

To test the influence of functional network constraint on the evolutionary outcomes rapid adaptation and convergent evolution, as well as the important factor of gene expression, I rearranged and expanded on this [39] data set (see Methods). I then tested, how network statistical parameters relate to estimators of evolutionary parameters ω, γ, and CAI. The amount of mRNA produced by each gene in regular somatic cells can be estimated by CAI (Codon Adaptation Index) which is derived from codon use bias in yeast that correlates with mRNA levels (based on [48, 49]). First, a general linear model was run with evolutionary parameters as dependent variables, and network parameters as predictor variables. All three network statistical parameters were found to significantly predict estimators for evolutionary outcomes (Table 5). All node categories have significantly different values for ω (KW-H(3,2204) = 20.1345, $p = 0.0002$), CAI (KW-H(3,2195) = 26.1472, $p = 0.00001$) and γ (KW-H(3,2195) = 36.7936, $p = 0.00000$), as shown by Kruskal-Wallis tests (Figure 5).

With regards to **rapid adaptation,** Figure 5 shows that the highest values of ω are found both in P and I-nodes with almost identical median values (0.93 vs. 0.91), and the lowest values were found in H nodes. This shows that nodes located less centrally in the network evolve faster than other nodes, but does not identify peripheral nodes as adapting particularly fast. CAI increases towards the center of the network, with mRNA expression level being highest in hub nodes. Network node hierarchy may therefore be able to explain the E-R anticorrelation (gene expression levels being negatively correlated with evolutionary rate [12]. H-nodes connect various subnetworks with one another, and thus are likely to be involved in more diverse functions (which might be partitioned across different tissues, processes or life history phases), than nodes more peripheral in a network (Figure 5, [47]). Such common functions may require a high amount of product, which may translate into high levels of mRNA expression in these nodes. γ is highest in P and I-nodes, indicating that evolutionary rewiring events are more common in less central parts of the networks.

An interesting subject for further study may be to compare explicit topologies of nodes that underwent re-wiring through evolution, in order to determine whether they can additionally move between I and P node categories over time. I-nodes harbor the majority of edges within a network - genetically re-wiring these nodes could lead to rapid adaptation [50]. Centrality of H-nodes seems to reduce their adaptability while peripheral and intermediate nodes are less constrained to adapt, and this process may involve rewiring within the network. This demonstrates how functional constraint can explain evolutionary outcomes better than estimators for gene dispensability can. Rapid genomic adaptation within diversifying populations has been shown to occur as a rapid response to selection such as anthropogenic pollution [51]. Such rapid adaptation often occurs in the presence of gene flow [51, 52]. This means that adaptation is constrained to specific genes under selection, which can interrupt their gene flow between populations, while alleles of other genes show uninterrupted gene flow. The speed of such adaptation related to divergence of a subset of genes within the same genome has been dubbed the "genic theory of speciation" [52]. Such genic evolution was shown to occur in *Timema* stick insects [53]. Future studies could test whether such rapidly adapting loci are preferentially located in P- and I-nodes, and whether this leads to a change in wiring patterns.

With respect to **genic adaptation**, Schoenrock et al. [39] could show that some functionally similar nodes experienced lower than expected levels of protein evolution, indicating purifying selection. Nodes that were evolving through fewer re-wiring events than expected, included functions related to phosphorylation, mitochondrial translation, response to pheromone, small GTPase mediated signal transduction, and transport. Nodes that were evolving among the five yeast species with higher than expected degrees of re-wiring, included the functions metabolic process, and various gene ontologies related to transcription and its regulation, as well as the regulation of transposition regulation. As indicated in Figure 2, these results prove that evolutionary outcomes are different for functionally different subnetworks within an interactome. It might be worth noting that, as outlined above, none of these functions is particularly related to growth but rather to maintaining organismal function, which is why they would be overlooked if conserved genes were only classified by the criterion of dispensability for colony growth. Gresham et al. [54] similarly showed that evolutionary constraint in experimentally evolved yeast populations over 200 generations is dependent on the type of selection (limiting Glucose or Phosphate vs. Sulphur), with convergence being an outcome of the system level organization of the respective metabolic pathway. Additionally, the same differences in evolutionary estimators between node categories that could promote rapid adaptation, would allow genes in different node categories to evolve with differential speed, which allows for genic adaptation. Traditionally, **convergence** has been studied in non-model organisms, and with a focus on adaptive modification of the phenotype (e.g., [55]. More recently, phenotypic convergence has been traced back to in some instances resulting from identical genotypic variants (called "genomic re-use", reviewed in [36]). These can arise either as new parallel mutations or from parallel selection of the same alleles from standing genetic variation [36] such as in the independent selection of body armor in the ectodysplasin locus of stickleback fish [56]. Other examples have recently been uncovered in skin toxin transport in poison frogs, [57] or in functional genomic adaptation to cold in a range of extant and extinct mammals including the mammoth [58]. Such genomic re-use causing convergence in distantly related lineages may indicate that constraint at the genomic level is important to generate convergent evolution and has been identified to drive speciation [59]. However, convergent phenotypic adaptations can alternatively be produced by different genes. They may also be exaptations, where a similar allele evolved due to ancestrally different selective pressures with a subsequent change of function [60]. In this study, I could test whether or not a small number of convergently adapting genes are preferentially located within I-nodes, as described above. Figures 4 and 5 support this idea. C-nodes have network statistical parameters most similar to I-nodes, showing that convergently adapting genes have similar evolutionary rates, expression levels, and degrees, as nodes that are located intermediately in the interactome (cf. inset network in Figure 4). This supports the notion that nodes with highest number of edges and intermediate network position are constrained to adapt and thus increase the likelihood for convergent evolution. Gresham et al. [54], from which five C-nodes were obtained, also showed that convergent evolution is related to system level organization of the respective metabolic pathway. In summary, these results clearly demonstrate a relationship between network architecture and convergence, and if additional genes are becoming known to evolve convergently in yeast, this hypothesis can be further tested.

In previous studies, gene expression level estimates (using CAI as proxy) were identified as the predominant explanatory variable related to functional genetic constraint influencing evolutionary outcomes. Therefore, the effect size of network statistical parameters as predictors for variables estimating evolutionary outcomes relative to CAI was determined. When CAI was incorporated into the analysis to predict values of ω and γ, average shortest path length (P-node classifier) was the predictor with highest power (0.99), followed by its interaction term with CAI (0.97), then CAI itself (0.92), followed by neighborhood connectivity (the I-node classifier, 0.83). Only betweenness centrality (the H-node classifier) was not significantly contributing to this model (Supplementary Tables 1 and 2). As mentioned previously, Figure 5 shows that the estimator of gene expression levels is highest in H-nodes, which might explain why CAI was seen as a better predictor for ω and γ than betweenness centrality. However, AIC based model selection revealed that a global model of all four variables including CAI and network parameters explains ω and γ better than CAI itself (Supplementary Tables 3 and 4). For ω, the only model with higher likelihood than the global model is that excluding betweenness centrality, whilst the CAI-only model ranks 8th. The rewiring score γ is best explained by the global model and the CAI-only model ranks 5th. These results confirm that whilst gene expression levels are an important element of genetic constraint, the position of highly expressed nodes as hub nodes in the network, together with the other network topology parameters, yield better explanatory power for two estimators of evolutionary outcomes. These results further support network topology as an important agent of evolutionary constraint.

Conclusions

Metagenomic resequencing of every 500 generations within a 60,000 generation *E. coli* long term evolution experiment [61] revealed that certain genes accumulated beneficial mutations through selection significantly more often than expected by chance, and were very often affected by parallel adaptation [61].

These results, together with the incidences of recurrent genomic adaptations reviewed herein, demonstrate that the above-described relationship between network structure and convergent evolution may be expandable to organisms other than yeasts [36]. Apart from the quick assessment performed in this contribution, the influence of network structure in shaping evolutionary outcomes in more complex organisms than yeast such as vertebrates still needs to be comprehensively tested. Additionally, statistics computed on edge distributions may change over time as more experimental evidence on interactions becomes available, and evolutionary constraint might differ by the type of interactions studied.

As demonstrated above in the yeast example, the impending advent of large-scale functional genomic networks for many new species makes it possible to convert functional genomic network structure of related species into variables describing hierarchical node position within the network. Future tests relating evolution to genomic constraint could include node architecture, and revolve around (1) Comparing standing genetic variation to network node position (while considering the effect of demography, selective sweeps, genetic drift, bottlenecks, and other levels of extrinsic constraint); (2) Testing whether similar subnetworks/node hierarchies adapt to same selection pressure in different organisms. (3) Comparing the speed of realized adaptation to a mutation/selection expectation, without considering network constraint. The potential benefits of better understanding genetic constraint leading to deterministic evolution may be wide ranging-- in humans, the use of functional interaction networks is omnipresent in genomic and transcriptomic study of cancer data, and recently, calls have been made for evolutionary methods to be applied to cancer problems [62]. A recent study demonstrates how the early progression of pancreatic cancer is defined through evolutionary constraints resulting from following one of three tumor suppressive pathways, and thus may be predictable [63]. Recognizing network constraint as evolutionary force, rather than disregarding evolution through natural selection [64], would allow quantifying “background genetic constraint” through functional network structure. The remaining variance could be better allocated to mutation and selection in directing rapid, convergent, and genic phenotypic evolution.

Methods

To assess evolutionary outcomes rapid adaptation and convergent evolution, as well as to address the important factor of gene expression in shaping protein-coding gene evolution, the data set of Schoenrock et al. including yeast ORF ID, computationally predicted evolutionary PPI re-wiring score (γ), and substitution rate (ω)[39] was downloaded. The re-wiring score was obtained from comparing networks across five species of yeast [39] and was used here to assess whether nodes that change wiring patterns are linked to specific positions within the network. The dataset was then rearranged and integrated with data downloaded from Wall et al. [13] including ORF ID, and CAI (Codon Adaptation Index, a measure of RNA expression levels, based on [48]. When analyzing networks, it is important to do so on exhaustive data sets [65] to avoid experimental bias [66]. Such an exhaustive interactome for yeast generated from the BIOGRID database [67] was obtained from CYTOSCAPE v.3.6.0 [68], which contained 6,508 nodes and 340,000 edges, with data curated from 5,500 studies. With the goal to calculate a classifier that will aid in describing hierarchical node position within networks, common network statistical parameters were calculated from this exhaustive yeast interactome in CYTOSCAPE v.3.6.0 [68] using the Network Analyzer function. Data for the matching node ORFs were appended to the data set, and variables with non-normal distribution were BoxCox transformed. The final data set contained 2209 ORFs with only a few missing data points per variable. The network statistical parameters obtained from the yeast interactome were average shortest path length (maximal in peripheral nodes), neighborhood connectivity (maximal in nodes intermediate to the network), and betweenness centrality (maximal in nodes connecting subnetworks). Nodes with maximum values for each one of these three statistical parameters, and that were not overlapping with each other (1081 nodes, Figure 2), were each assigned to a category: P (peripheral nodes), I (intermediate nodes) and H (hub nodes). To assign node categories to the remaining nodes in the network that may be harder to allocate visually, a discriminant function analysis (DFA) was employed in STATISTICA (V13, Tibco). All remaining nodes with significant statistical support could be associated to one of these three categories (Table 3). To explore the network position of nodes that have undergone convergent adaptation, ORF IDs that were demonstrated experimentally to show convergent genomic adaptation in independent experiments, strains, or species of yeasts (C-nodes) were identified from the literature (66, 74-78), Table 4). Out of the 26 obtained C-nodes, 21 nodes were allocated by DFA to the I-category, and five were allocated to the P-category. It was then tested, how network statistical parameters relate to the evolutionary parameters ω, γ, and CAI. First, a general linear model was run with evolutionary parameters as dependent variables, and network parameters as predictor variables. Differences in, respectively, network statistical parameters or estimators for evolutionary parameters, and node categories were tested with Kruskal-Wallis tests. Because previous studies [15–17] have ascribed gene expression (here measured as CAI) an important role for constraining evolution, it is possible that whilst network statistical parameters do explain evolutionary parameters well, this effect could disappear once CAI itself is considered as a predictor for ω and γ. This assumption was therefore tested through (i) comparing power of predictors in another linear model, including network statistical parameters, CAI, as well as interaction terms as predictors and (ii) comparing Akaike information criteria of models generated from these variables and their interaction terms.

Declarations

Ethics approval and consent to participate

Not applicable

Consent to publish

Not applicable

Availability of data and materials

The datasets generated and/or analysed during the current study are included in this published article and available as Supplementary Table 5

Competing interests

The authors declare that they have no competing interests

Funding

Not applicable

Authors' Contributions

KWV conceived the study, analyzed the data and wrote the manuscript

Acknowledgements

I would like to thank Alexander P. Turner (Department of Computer Science, University of Hull) for helpful discussions

Table 1. Glossary of terms

Evolutionary constraint	[69]**:** the phenomenon of evolution producing a finite number of genomic and associated phenotypic outcomes from a near infinite number of possible genetic variants.
Genetic constraint	The portion of evolutionary constraint, which is determined at the level of genes or their gene products, for example codon constraint or developmental genetic pathways.
Functional network constraint	The portion of network constraint attributed to the structure or architecture of gene interactions that can be expressed in the form of a network. Networks consist of nodes (genes) and edges (functional interactions between these genes).
Genic evolution	The phenomenon of different evolutionary outcomes being the outcome of independent mutation and selection events in different genes. For example, the occurrence of convergent evolution in diverging populations, both of which are caused by evolution in distinct genes.
Rapid adaptation	The phenomenon of adaptive change in allele frequencies of a population to natural selection, taking place within just a few generations.
Convergent evolution / convergence	Traditionally defined as similar phenotypes evolving from similar selective pressure in response to similar environments [70]. May be caused at the genomic level through genomic re-use of the same genes or alleles, which is also called parallel genetic evolution or genomic re-use.
Gene dispensability	A variable to estimate gene essentiality. The less dispensable a gene is for organismal growth and function, the more essential it is. An estimator for the mean fitness effect of all possible mutations of a gene across environments the cell is likely to encounter. In yeast, this is experimentally determined through knockouts.
Pleiotropy and cost of complexity	Traditionally defined as one gene influencing more than one trait. In the papers cited in this study, has been defined as gene products with more than one functional interactions with other gene products, with the link to pleiotropy of phenotypic traits being implied. It is therefore here called "gene pleiotropy".
Gene expression level CAI	The amount of mRNA produced by each gene in regular somatic cells. CAI (Codon Adaptation Index) is used as a substitute variable in this paper, and is derived from codon use bias in yeast that correlates with mRNA levels.
Omega ω	The ratio of nonsynonymous to synonymous substitutions dN/dS. It is assumed that dS remains constant, and dN is used here as an estimator for directional evolution.
Gamma γ	A score developed for estimating events of rewiring functional connections between network nodes over the course of evolution. Developed on the example of five species of yeasts.
Neighborhood connectivity	A network statistic used to describe the structure of a functional genetic network. Describes the number of connections of all neighbors of each node. Highest values are expected in intermediately located nodes within a network.
Betweenness centrality	A network statistic used to describe the structure of a functional genetic network, describing where a node lies within paths between other nodes. Nodes with many paths progressing through them may be important in transmitting information. Highest values are expected in nodes central to a network.
Average shortest path length	A network statistic used to describe the structure of a functional genetic network. Shortest distance between a node and other nodes. Highest values are expected in peripheral nodes of a network.

Table 2. Hypotheses relating network constraint to evolutionary outcomes and results of hypothesis assessment using a node classification scheme in yeast.

Evolutionary outcome	Hypothesis (H)	Alternative Hypothesis (HA)	Results in this paper following assessment with hierarchical node classification scheme.
Speed of evolution	Indispensable or essential genes are more constrained and evolve slowly [71].	Functionally important and thus functionally constrained genes evolve slowly, independent of dispensability [21]. Highly expressed genes evolve slowest [13, 14].	**HA:** Functionally most constrained genes (H-nodes) have the lowest substitution ratios of all categories, and are most highly expressed, but have lower scores of evolutionary rewiring than P and I-nodes.
Speed of evolution	Central nodes have highest number of edges; evolve very slowly because any change will lead to maladaptive pleiotropic effects - causes balancing selection through cost of complexity. [36],[35],[18],[34]	Intermediate nodes evolve fastest as their higher number of edges allows for evolution through rewiring ([44, 45]).	**HA:** Nodes with highest number of edges are intermediate to the network, evolve fast (high ω) and have a high score of rewiring (γ), indicating that the substitution rate of these genes may be associated with evolutionary rewiring events.
Speed of evolution	Nodes with a low number of edges evolve fastest due to higher degrees of freedom, which allows for genetic adaptations minimizing pleiotropic effects. [72], [36]	---	**H:** Peripheral nodes evolve fast (high ω) and have a high score of rewiring (γ), indicating that the substitution rate of these genes may be associated with evolutionary rewiring events.
Convergent evolution	Nodes with a low number of edges should be the prime target of convergent evolution. Pleiotropic negative effects are expected to be low, and mutations in them can maximize adaptation [36].	Peripheral nodes have the highest degrees of freedom and thus divergence is more likely than convergence in them. Convergent evolution should instead be favored in nodes that allow for genetic variance, while having reduced degrees of freedom (I-nodes) (This contribution).	**HA**: 21 out of 26 nodes with convergent evolution demonstrated in yeasts were classified as I-nodes by DFA, and five as P nodes. ω and CAI were similar to I-nodes, but none of these 26 nodes showed evidence of evolutionary rewiring.
Genic evolution	Adaptations can be characterized (either causative or correlative for the speciation process) by any number of divergent genes within the genome, whereas other genes are not associated with adaptation. [52].	Only the complete phenotype is selected, the genic component is less important [8].	**H:** Different clusters of functionally similar nodes experience either higher, lower than expected or neutral rates of evolution across five species of yeast [69]. Causation or correlation to speciation process not testable with data.

Table 3. Discriminant function analysis summary to assign node categories H, I, P to nodes within dataset. Wilks' Lambda: 0.0704 approx. F (6,2152)=992.780 p<0.001.

	Wilks Lambda	**Partial Lambda**	**F-remove 2,1076**	**p-value**	**Toler.**	**1-Toler. (R-sqr.)**
Neighborhood connectivity	0.137	0.514	507.835	<0.001	0.988	0.012
Betweenness centrality	0.105	0.673	261.039	<0.001	0.994	0.006
Average shortest path length	0.133	0.528	480.907	<0.001	0.983	0.017

Table 4. List of yeast genes that were found to adapt to novel environments, and were additionally shown to evolve these adaptations convergently across populations or species of yeast. Node hierarchy categories after discriminant function analysis (DFA) are shown in the first column. P - peripheral nodes, I - intermediate nodes.

DFA estimated Node hierarchy	Gene symbol	ORF ID	Reference
I	STE11	YLR362W	Lang et al., 2013
I	STE12	YHR084W	Lang et al., 2013
I	STE4	YOR212W	Lang et al., 2013
P	KRE6	YPR159W	Lang et al., 2013
I	SFL1	YOR140W	Lang et al., 2013
I	STE5	YDR103W	Lang et al., 2013
P	ANP1	YEL036C	Lang et al., 2013
I	GCN1	YGL195W	Lang et al., 2013
I	ERG5	YMR015C	Gerstein et al., 2012
P	ERG7	YHR072W	Gerstein et al., 2012
I	CNE1	YAL058W	Lang et al., 2013
I	GPB1	YOR371C	Lang et al., 2013
P	KEG1	YFR042W	Lang et al., 2013
I	KRE5	YOR336W	Lang et al., 2013
I	TOH1	YJL171C	Lang et al., 2013
P	SUL4	YBR294W	Gresham et al 2008
I	GAL3	YDR009W	Hittinger et al., 2004; Stern, 2013
I	GIN4	YDR507C	Gresham et al 2008
I	PDR1	YGL013C	Anderson et al. 2003
I	SGF73	YGL066W	Gresham et al 2008
I	SET4	YJL105W	Lang et al., 2013
I	SIR1	YKR101W	Gresham et al 2008
I	ACE2	YLR131C	Lang et al., 2013
I	GAS1	YMR307W	Lang et al., 2013
I	WHI2	YOR043W	Lang et al., 2013
I	CKA2	YOR061W	Gresham et al 2008

Table 5. Multivariate Wilks tests of significance and powers for network parameters to explain protein evolutionary rate (□), gene expression (Codon Adaptation Index CAI), and evolutionary rewiring between species of yeast (□). All predictors were significant.

	Wilks' Lambda	F	Effect df	Error df	p	Observed power (alpha)
Intercept	0.317	1569.597	3	2188	<0.001	1.000
Neighborhood connectivity	0.924	59.892	3	2188	<0.001	1.000
Betweenness centrality	0.995	3.931	3	2188	0.008	0.832
Average shortest path length	0.961	29.553	3	2188	<0.001	1.000

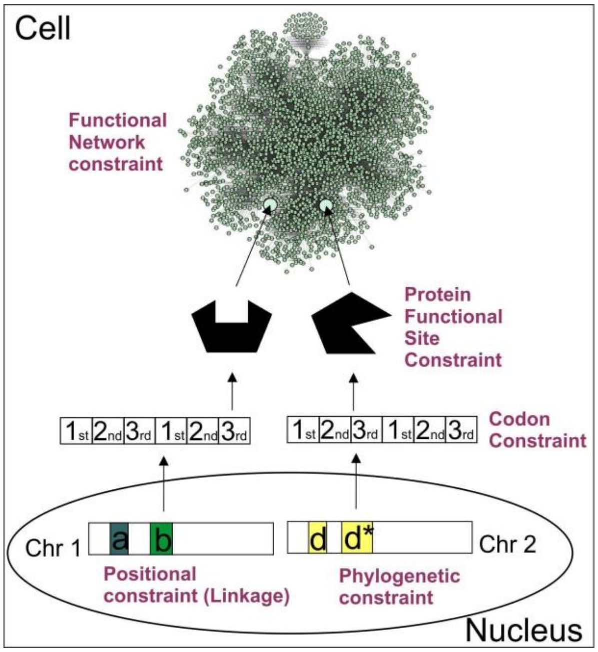


Figure 1. Examples for different levels of genetic constraint. Linkage is a transient constraint, which is broken up through recombination or other chromosome rearrangements. If a gene arises through duplication, phylogenetic constraint means that the function of its gene product may be non-independent with relation to the ancestral gene product. Codon constraint describes the likelihood of the different codon positions to produce beneficial mutations. Protein functional site constraint describes constraint located in genomic regions that code for functional sites of proteins versus other regions of the proteins. This is related to the idea that gene products form a functional genomic network. Within this network, interactions of these gene products also pose an element of constraint on evolution, but this is not well researched.

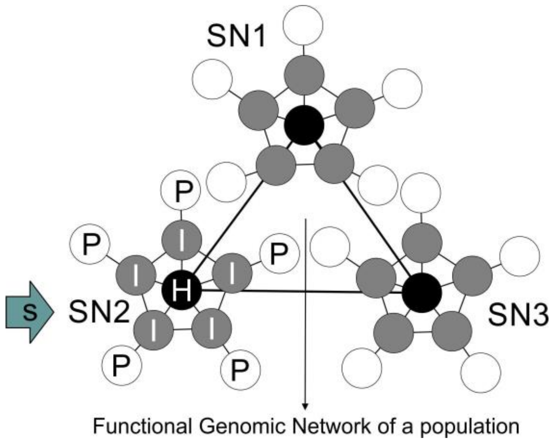


Figure 2. Proposed testable relationship between functional genomic network architecture, network node position, and evolutionary outcomes. SN are subnetworks within the functional genomic network of a population with distinct functions (e.g., metabolic pathways). Standing genetic variation exists within nodes, but depends on their position within the network. Black nodes (H) are essential for organismal function and not likely to accumulate non-synonymous mutations; Grey nodes (I) are functionally connected with many others and constrained in accumulating non-synonymous mutations. White nodes (P) are functionally connected to fewest others and most likely to accumulate non-synonymous mutations. Resulting from this, three evolutionary outcomes can be explained: Rapid adaptation is facilitated in white nodes through their high standing genetic variation. Selection being constrained to operate on these nodes in a specific subnetwork increases the speed of adaptation. Convergent evolution is facilitated through the finite number of networks that are related to specific functions and shared among species through common ancestry. The likelihood of convergent evolution within one subnetwork in response to selection increases through the moderate level of genetic variance, combined with constraint posed by the high number of connections to other nodes. Genic evolution is facilitated through the selection pressure only having an effect in the subnetwork with organismal functions related to it but not in others. Selection is likely to operate on standing genetic variation, which is likely concentrated in white nodes (shown as blue squares). These different processes can explain the coexistence of convergent and divergent (rapid, genic) evolution within the genomes of a population.

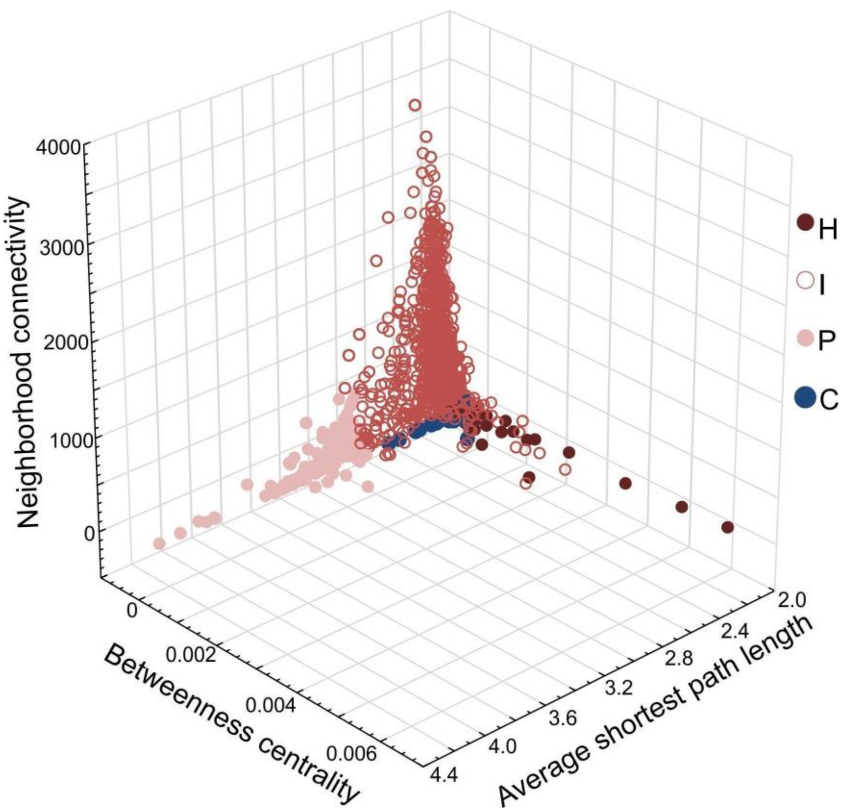


Figure 3. Distribution of yeast interactome nodes within network parameter space (neighborhood connectivity, average shortest path length, and betweenness centrality). The top values for each axis are colored in shades of red (light, filled: P-nodes; light, open: I-nodes; dark, filled: H-nodes). Convergent evolution nodes are indicated in dark blue. These top values for each axis formed the basis to classify the remaining nodes based on discriminant function analysis.

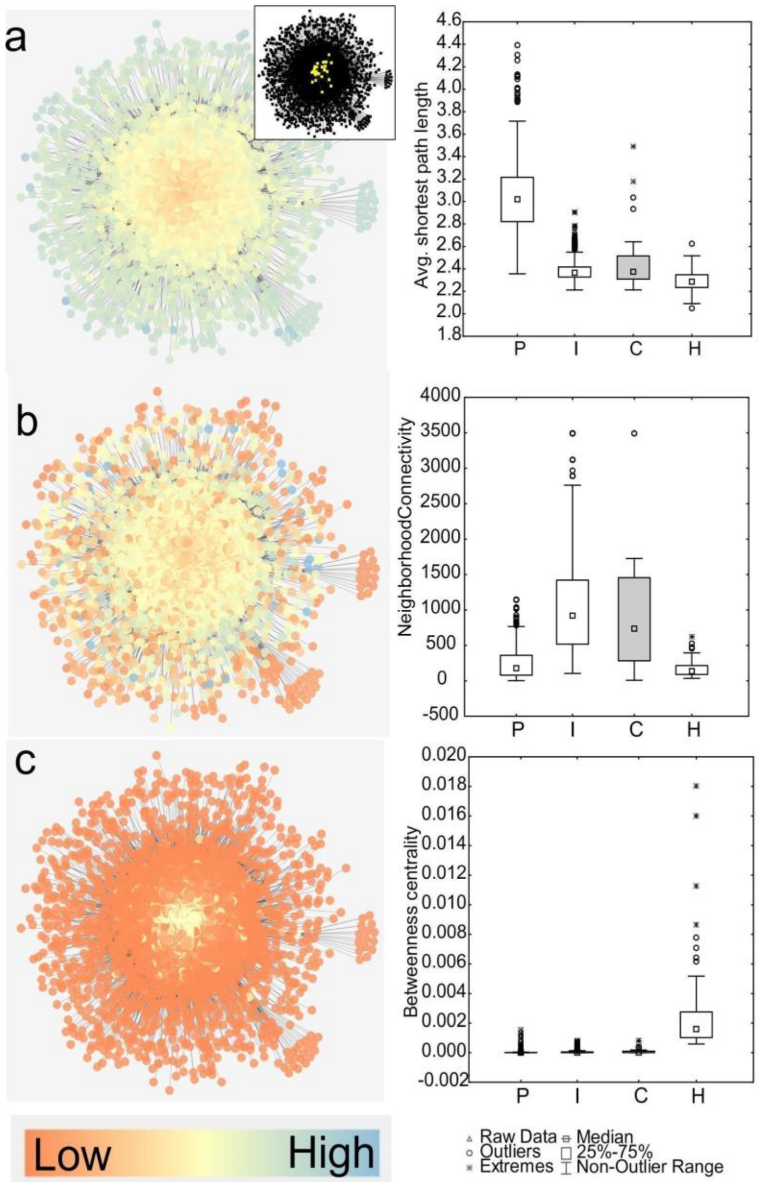


Figure 4. Visualization of node classification scheme in yeast interactome. Values of a) average shortest path length, b) neighborhood connectivity, and c) betweenness centrality within the yeast interactome (left panels), and values for the DFA-derived hierarchical node categories P, I and H, and for nodes known to be under convergent evolution in yeasts (C, N=18). The small inset network shows the location of convergently evolved genes (C-nodes) within the interactome (yellow nodes).

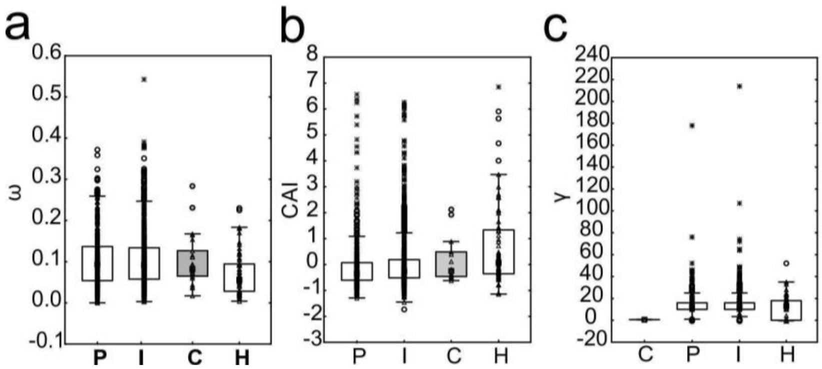


Figure 5. Relationship between hierarchical node structure of yeast interactome and evolutionary parameters. Node types are designated as peripheral (P), intermediate (I), or hub (H) based on discriminant function analysis, and nodes that were found to evolve convergently (C; N=22) in yeasts. Three evolutionary outcomes (a) substitution rate, (b) expression level, approximated through Codon Adaptation Index (CAI), and (c) evolutionary rewiring score significantly differ among node categories (see text). C-node boxes are sorted by Median. Double red line: outliers above median not shown in figure but included in tests. Raw data points - triangles, circles - outliers, stars - extreme values, squares - Medians, boxes - 25-75% data, whiskers - non-outlier range.

References

1. Losos JB. Lizards in an Evolutionary Tree: Ecology and Adaptive Radiation of Anoles. 1st edition. University of California Press; 2009.

2. Hertz PE, Huey RB. Compensation for altitudinal changes in the thermal environment by some Anolis lizards on Hispaniola. Ecology. 1981;62:515–21.

3. Wollenberg KC, Wang IJ, Glor RE, Losos JB. Determinism in the diversification of Hispaniolan trunk-ground anoles (Anolis cybotes species complex). Evolution. 2013;67:3175–90.

4. Munoz MM, Stimola MA, Algar AC, Conover A, Rodriguez AJ, Landestoy MA, et al. Evolutionary stasis and lability in thermal physiology in a group of tropical lizards. Proceedings of the Royal Society of London B: Biological Sciences. 2014;281:20132433.

5. Rodríguez A, Rusciano T, Hamilton R, Holmes L, Jordan D, Wollenberg Valero KC. Genomic and phenotypic signatures of climate adaptation in an Anolis lizard. Ecol Evol. 2017;7:6390–403.

6. Futuyma DJ. Evolutionary constraint and ecological consequences. Evolution. 2010;64:1865–84.

7. Schluter D. The Ecology of Adaptive Radiation. OUP Oxford; 2000.

8. Mayr E. Populations, Species, and Evolution: An Abridgment of Animal Species and Evolution. 1st ed. Harvard University Press; 1970.

9. Arnold SJ, Bürger R, Hohenlohe PA, Ajie BC, Jones AG. Understanding the evolution and stability of the G-matrix. Evolution. 2008;62:2451–61.

10. Saltz JB, Hessel FC, Kelly MW. Trait correlations in the Genomics era. Trends Ecol Evol. 2017;32:279–90.

11. Conner JK, Karoly K, Stewart C, Koelling VA, Sahli HF, Shaw FH. Rapid independent trait evolution despite a strong pleiotropic genetic correlation. Am Nat. 2011;178:429–41.

12. Zhang J, Yang J-R. Determinants of the rate of protein sequence evolution. Nat Rev Genet. 2015;16:409–20.

13. Wall DP, Hirsh AE, Fraser HB, Kumm J, Giaever G, Eisen MB, et al. Functional genomic analysis of the rates of protein evolution. Proc Natl Acad Sci U S A. 2005;102:5483–8.

14. Papp B, Pál C, Hurst LD. Dosage sensitivity and the evolution of gene families in yeast. Nature. 2003;424:194–7.

15. Bloom JD, Drummond DA, Arnold FH, Wilke CO. Structural determinants of the rate of protein evolution in yeast. Mol Biol Evol. 2006;23:1751–61.

16. Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. Why highly expressed proteins evolve slowly. Proc Natl Acad Sci U S A. 2005;102:14338–43.

17. Drummond DA, Raval A, Wilke CO. A single determinant dominates the rate of yeast protein evolution. Mol Biol Evol. 2006;23:327–37.

573 18. Wang Z, Liao B-Y, Zhang J. Genomic patterns of pleiotropy and the evolution of complexity. Proc Natl Acad Sci U S A. 2010;107:18034–9.

575 19. Stanić S, Pavković-Lucic S. Mating success of wild type and sepia mutants Drosophila melanogaster in different choice. Riv Biol. 2005;98:513–24.

577 20. Cooper GM, Stone EA, Asimenos G, NISC Comparative Sequencing Program, Green ED, Batzoglou S, et al. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. 2005;15:901–13.

580 21. Wilson AC, Carlson SS, White TJ. Biochemical evolution. Annu Rev Biochem. 1977;46:573–639.

581 22. Kimura M. The Neutral Theory of Molecular Evolution. Cambridge University Press; 1983.

582 23. Proulx SR, Promislow DEL, Phillips PC. Network thinking in ecology and evolution. Trends Ecol Evol. 2005;20:345–53.

584 24. Davidson EH. Emerging properties of animal gene regulatory networks. Nature. 2010;468:911–20.

585 25. Erwin DH, Davidson EH. The evolution of hierarchical gene regulatory networks. Nat Rev Genet. 2009;10:141–8.

587 26. Camps M, Herman A, Loh E, Loeb LA. Genetic constraints on protein evolution. Crit Rev Biochem Mol Biol. 2007;42:313–26.

589 27. Chae L, Lee I, Shin J, Rhee SY. Towards understanding how molecular networks evolve in plants. Curr Opin Plant Biol. 2012;15:177–84.

591 28. Conant GC, Wagner A. Convergent evolution of gene circuits. Nat Genet. 2003;34:264–6.

592 29. Jeong H, Mason SP, Barabási AL, Oltvai ZN. Lethality and centrality in protein networks. Nature. 2001;411:41–2.

594 30. He X, Zhang J. Toward a molecular understanding of pleiotropy. Genetics. 2006;173:1885–91.

595 31. Promislow DEL. Protein networks, pleiotropy and the evolution of senescence. Proc Biol Sci. 2004;271:1225–34.

597 32. Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. Evolutionary rate in the protein interaction network. Science. 2002;296:750–2.

599 33. Pavlicev M, Wagner GP. A model of developmental evolution: selection, pleiotropy and compensation. Trends Ecol Evol. 2012;27:316–22.

601 34. Wagner A. Robustness against mutations in genetic networks of yeast. Nat Genet. 2000;24:355–61.

602 35. Wagner GP, Kenney-Hunt JP, Pavlicev M, Peck JR, Waxman D, Cheverud JM. Pleiotropic scaling of gene effects and the “cost of complexity.” Nature. 2008;452:470.

604 36. Stern DL. The genetic causes of convergent evolution. Nat Rev Genet. 2013;14:751–64.

605 37. Goenawan IH, Bryan K, Lynn DJ. DyNet: visualization and analysis of dynamic molecular interaction networks. Bioinformatics. 2016;32:2713–5.

38. Wagner A. Genotype networks shed light on evolutionary constraints. Trends Ecol Evol. 2011;26:577–84.

39. Schoenrock A, Burnside D, Moteshareie H, Pitre S, Hooshyar M, Green JR, et al. Evolution of protein-protein interaction networks in yeast. PLoS One. 2017;12:e0171920.

40. Scannell DR, Zill OA, Rokas A, Payen C, Dunham MJ, Eisen MB, et al. The awesome power of yeast evolutionary genetics: new Genome sequences and strain resources for the Saccharomyces sensu stricto genus. G3. 2011;1:11–25.

41. Storey KB. Reptile freeze tolerance: metabolism and gene expression. Cryobiology. 2006;52:1–16.

42. Wollenberg Valero KC, Pathak R, Prajapati I, Bankston S, Thompson A, Usher J, et al. A candidate multimodal functional genetic network for thermal adaptation. PeerJ. 2014;2:e578.

43. Porcelli D, Butlin RK, Gaston KJ, Joly D, Snook RR. The environmental genomics of metazoan thermal adaptation. Heredity . 2015;114:502–14.

44. Siegal ML, Promislow DEL, Bergman A. Functional and evolutionary inference in gene networks: does topology matter? Genetica. 2007;129:83–103.

45. Hahn MW, Conant GC, Wagner A. Molecular evolution in large genetic networks: does connectivity equal constraint? J Mol Evol. 2004;58:203–11.

46. Fraser HB. Modularity and evolutionary constraint on proteins. Nat Genet. 2005;37:351–2.

47. Han J-DJ, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, et al. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature. 2004;430:88–93.

48. Coghlan A, Wolfe KH. Relationship of codon bias to mRNA concentration and protein length in Saccharomyces cerevisiae. Yeast. 2000;16:1131–45.

49. Wagner A. Energy constraints on the evolution of gene expression. Mol Biol Evol. 2005;22:1365–74.

50. Roguev A, Bandyopadhyay S, Zofall M, Zhang K, Fischer T, Collins SR, et al. Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. Science. 2008;322:405–10.

51. Reid NM, Proestou DA, Clark BW, Warren WC, Colbourne JK, Shaw JR, et al. The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. Science. 2016;354:1305–8.

52. Wu C-I. The genic view of the process of speciation. J Evol Biol. 2001;14:851–65.

53. Riesch R, Muschick M, Lindtke D, Villoutreix R, Comeault AA, Farkas TE, et al. Transitions between phases of genomic differentiation during stick-insect speciation. Nat Ecol Evol. 2017;1:82.

54. Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, Ward A, et al. The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. PLoS Genet. 2008;4:e1000303.

55. Losos JB, Jackman TR, Larson A, Queiroz K, Rodriguez-Schettino L. Contingency and determinism in replicated adaptive radiations of island lizards. Science. 1998;279:2115–8.

56. Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G Jr, Dickson M, Grimwood J, et al. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. Science. 2005;307:1928–33.

57. Tarvin RD, Santos JC, O'Connell LA, Zakon HH, Cannatella DC. Convergent substitutions in a sodium channel suggest multiple origins of toxin resistance in poison frogs. Mol Biol Evol. 2016;33:1068–81.

58. Yudin NS, Larkin DM, Ignatieva EV. A compendium and functional characterization of mammalian genes involved in adaptation to Arctic or Antarctic environments. BMC Genet. 2017;18:111.

59. Schwenk K, Wagner GP. The Relativism of Constraints on Phenotypic Evolution. In: Pigliucci M, Preston K, editors. Phenotypic Integration: Studying the Ecology and Evolution of Complex Phenotypes. Oxford University Press; 2004. p. 390–408.

60. Losos JB. Convergence, adaptation, and constraint. Evolution. 2011;65:1827–40.

61. Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM. The dynamics of molecular evolution over 60,000 generations. Nature. 2017;551:45–50.

62. von Loga K, Gerlinger M. Cancer (r)evolution. Nat Ecol Evol. 2017;1:1051–2.

63. Mueller S, Engleitner T, Maresch R, Zukowska M, Lange S, Kaltenbacher T, et al. Evolutionary routes and KRAS dosage define pancreatic cancer phenotypes. Nature. 2018;554:62–8.

64. Coyne JA. Comment on "Gene regulatory networks and the evolution of animal body plans." Science. 2006;313:761; author reply 761.

65. Fraser HB, Wall DP, Hirsh AE. A simple dependence between protein evolution rate and the number of protein-protein interactions. BMC Evol Biol. 2003;3:11.

66. Bloom JD, Adami C. Apparent dependence of protein evolutionary rate on number of interactions is linked to biases in protein-protein interactions data sets. BMC Evol Biol. 2003;3:21.

67. Chatr-Aryamontri A, Breitkreutz B-J, Heinicke S, Boucher L, Winter A, Stark C, et al. The BioGRID interaction database: 2013 update. Nucleic Acids Res. 2013;41 Database issue:D816–23.

68. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.

69. Antonovics J, van Tienderen PH. Ontoecogenophyloconstraints? The chaos of constraint terminology. Trends Ecol Evol. 1991;6:166–8.

70. Conway Morris S. The runes of evolution: how the universe became self-aware. 1st.ed. edition. Templeton Foundation Press; 2015.

71. Hurst LD, Smith NG. Do essential genes evolve slowly? Curr Biol. 1999;9:747–50.

72. Davidson EH, Erwin DH. Gene regulatory networks and the evolution of animal body plans. Science. 2006;311:796–800.

Figure 1

Click here to access/download;Figure;Valero_NetworkEvo_Figure_1.jpg

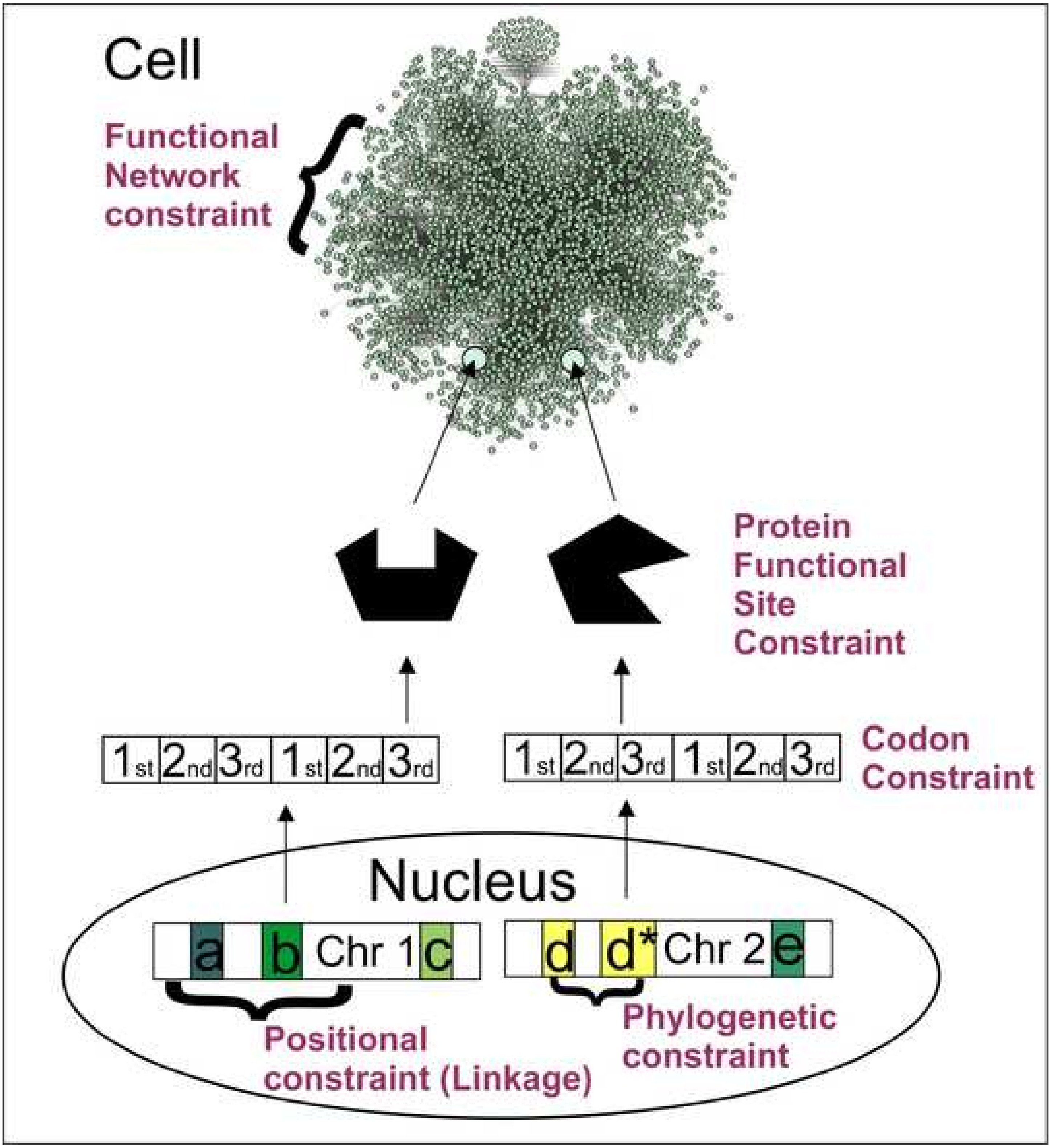


Figure 2

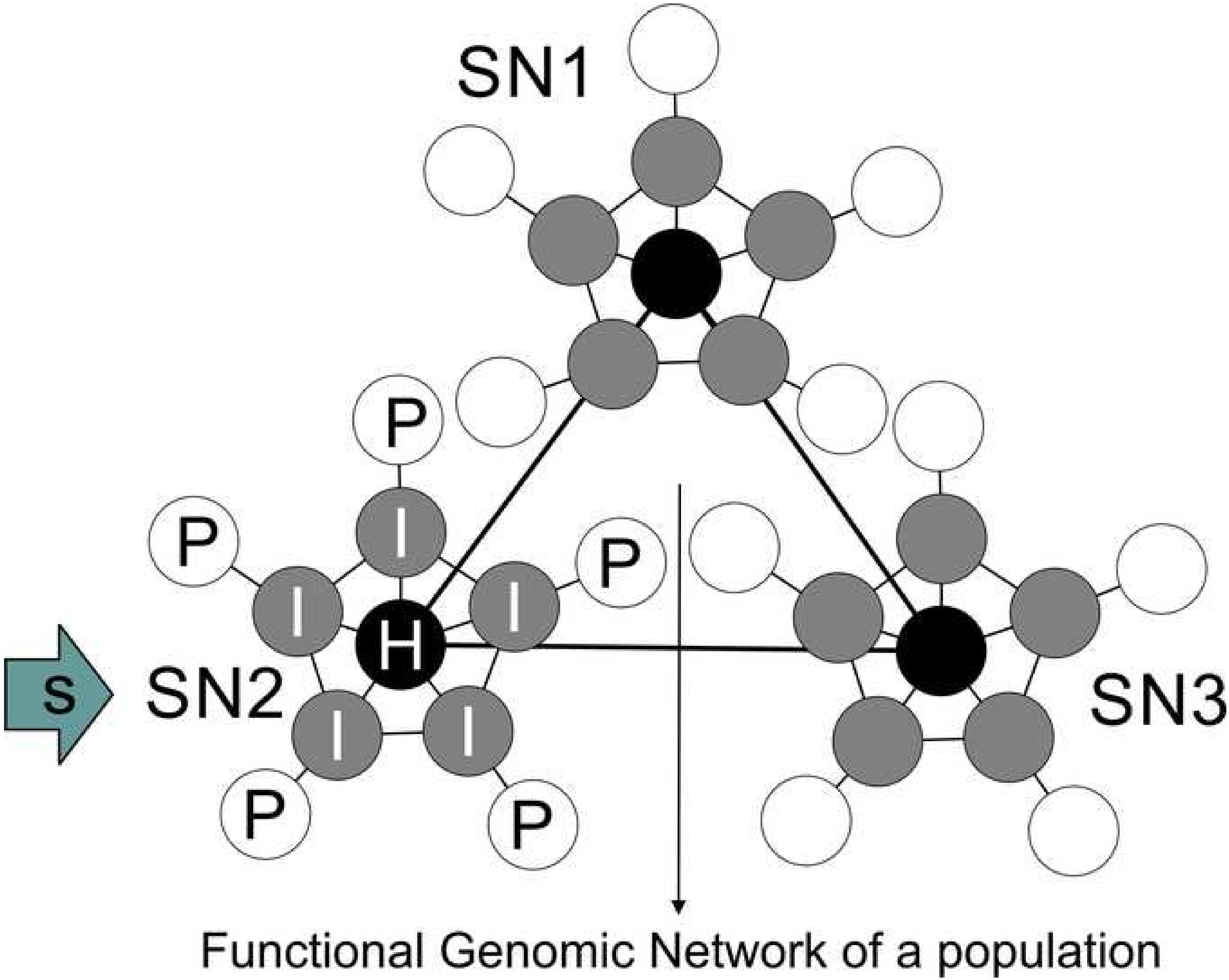

SN1
SN2
SN3
S
H
I
I
I
I
I
P
P
P
P
P
Functional Genomic Network of a population

Figure 3

Click here to access/download;Figure;Valero_NetworkEvo_Figure_3.jpg

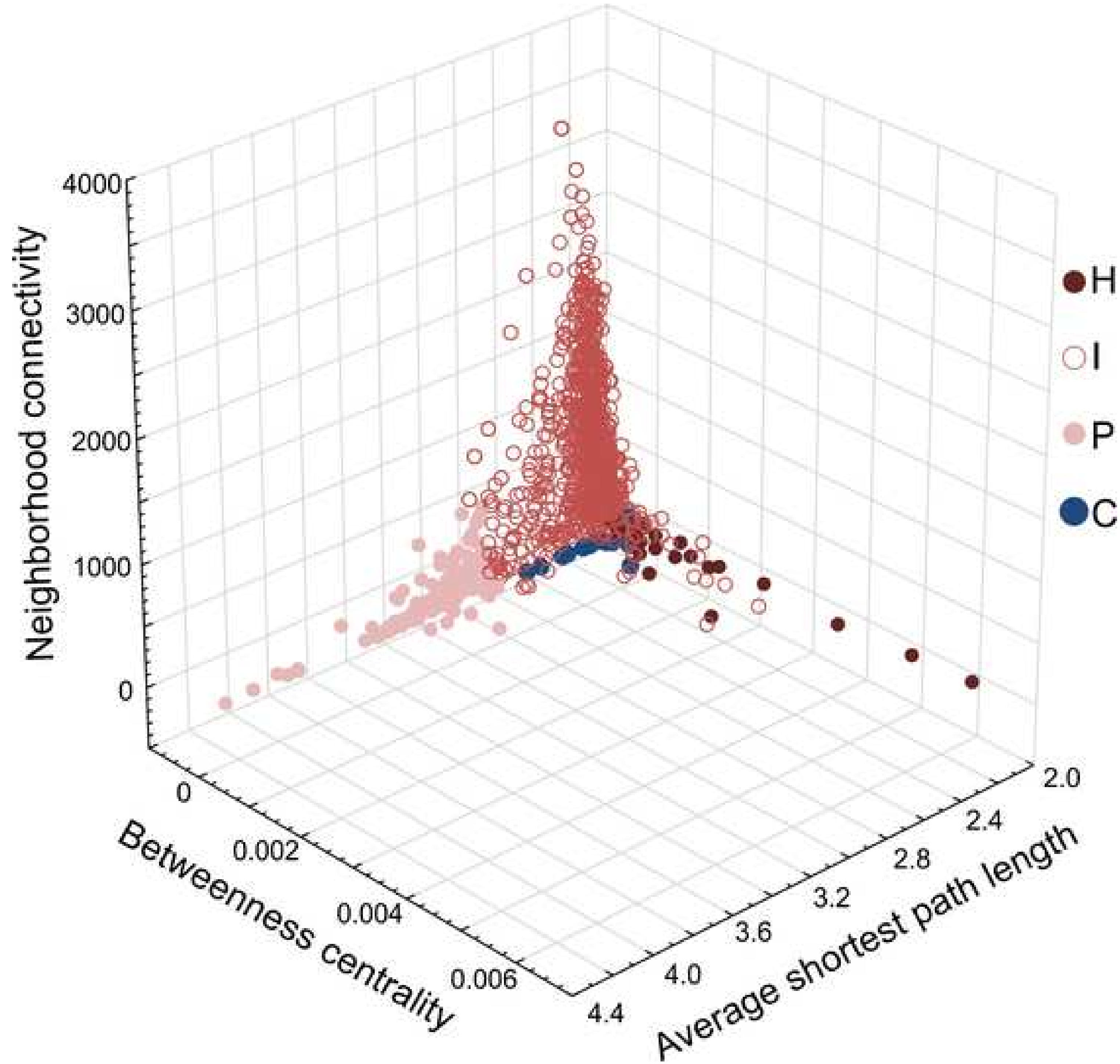
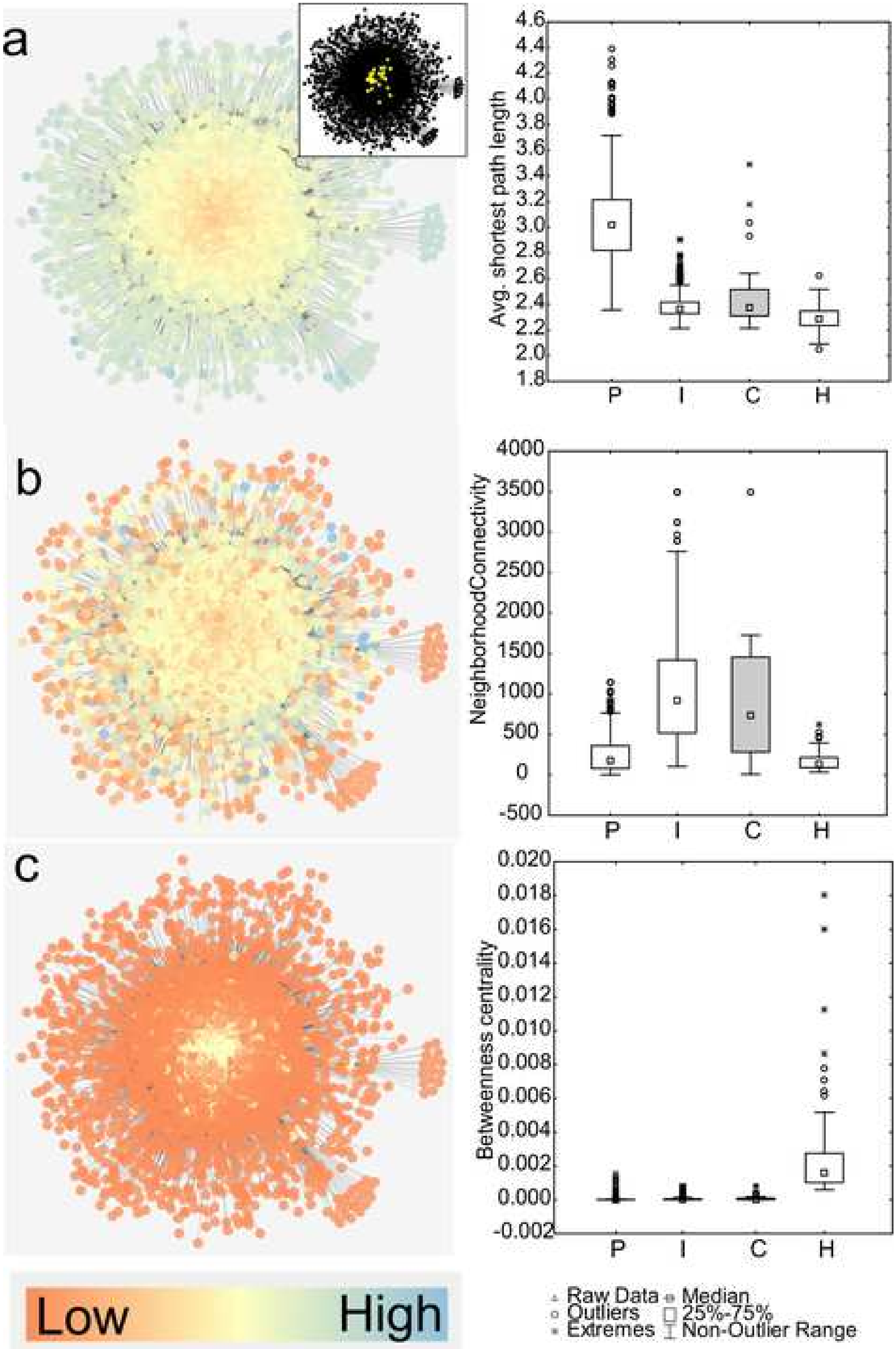
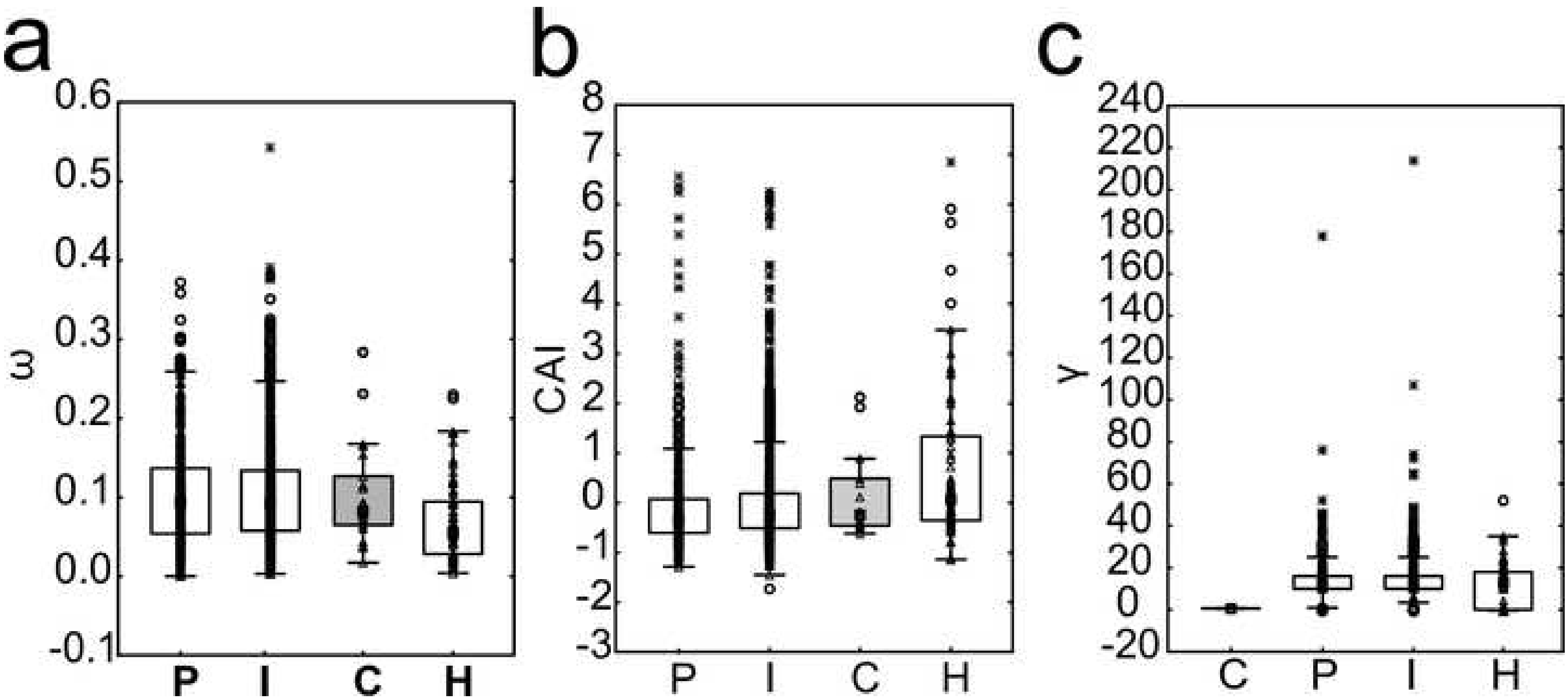


Figure 4

Click here to access/download;Figure;Valero_NetworkEvo_Figure_4.jpg




a
b
c
ω
CAI
γ
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
8
7
6
5
4
3
2
1
0
-1
-2
-3
240
220
200
180
160
140
120
100
80
60
40
20
0
-20
P
I
C
H
P
I
C
H
C
P
I
H

Supplementary Table 5

Click here to access/download
Supplementary Material
Valero_NetworkEvo_Supplementary Table 5.csv