OPEN ACCESS Freely available online

PLoS one

Non-Uniform Survival Rate of Heterodimerization Links in the Evolution of the Yeast Protein-Protein Interaction Network

Takeshi Hase[1], Yoshihito Niimura[1]*, Tsuguchika Kaminuma[1], Hiroshi Tanaka[1,2]

1 Department of Bioinformatics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan, **2** Department of Bioinformatics, School of Biomedical Science, Tokyo Medical and Dental University, Tokyo, Japan

Abstract

Protein-protein interaction networks (PINs) are scale-free networks with a small-world property. In a small-world network, the average cluster coefficient ($<C>$) is much higher than in a random network, but the average shortest path length ($<L>$) is similar between the two networks. To understand the evolutionary mechanisms shaping the structure of PINs, simulation studies using various network growth models have been performed. It has been reported that the heterodimerization (HD) model, in which a new link is added between duplicated nodes with a uniform probability, could reproduce scale-freeness and a high $<C>$. In this paper, however, we show that the HD model is unsatisfactory, because (i) to reproduce the high $<C>$ in the yeast PIN, a much larger number (n_{HI}) of HD links (links between duplicated nodes) are required than the estimated number of n_{HI} in the yeast PIN and (ii) the spatial distribution of triangles in the yeast PIN is highly skewed but the HD model cannot reproduce the skewed distribution. To resolve these discrepancies, we here propose a new model named the non-uniform heterodimerization (NHD) model. In this model, an HD link is preferentially attached between duplicated nodes when they share many common neighbors. Simulation studies demonstrated that the NHD model can successfully reproduce the high $<C>$, the low n_{HI}, and the skewed distribution of triangles in the yeast PIN. These results suggest that the survival rate of HD links is not uniform in the evolution of PINs, and that an HD link between high-degree nodes tends to be evolutionarily conservative. The non-uniform survival rate of HD links can be explained by assuming a low mutation rate for a high-degree node, and thus this model appears to be biologically plausible.

Citation: Hase T, Niimura Y, Kaminuma T, Tanaka H (2008) Non-Uniform Survival Rate of Heterodimerization Links in the Evolution of the Yeast Protein-Protein Interaction Network. PLoS ONE 3(2): e1667. doi:10.1371/journal.pone.0001667

Editor: Matthew W. Hahn, Indiana University, United States of America

Received November 7, 2007; **Accepted** January 22, 2008; **Published** February 27, 2008

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: niimura@bioinfo.tmd.ac.jp

Introduction

The information of protein-protein interaction networks (PINs) at the whole-genome level is now available from several organisms, including *Saccharomyces cerevisiae* [1–3], *Caenorhabditis elegans* [4], and *Drosophila melanogaster* [5]. These data were provided by using high-throughput experimental techniques such as yeast two-hybrid screens [1,2]. The structure of PINs is represented as nodes (proteins) and links (interactions between proteins). Studies of PIN structures have revealed that PINs exhibit the following interesting properties [6].

First, PINs are scale-free networks [7,8]. The number of links connected to a node is called a degree. The degree distribution $P(k)$ gives the probability that a node has k links (*i.e.*, degree k). In a scale-free network, $P(k)$ decays as a power law, following $P(k) \sim k^{-\gamma}$ [9]. (In the case of PINs, it is known that $P(k)$ better fits a power law with an exponential cut-off, *i.e.*, $P(k) \sim (k_0+k)^{-\gamma}\ e^{-k/kc}$ [7,10].) Therefore, a scale-free network is highly heterogeneous and is characterized by the presence of a large number of nodes having only a few links and a small number of nodes (hubs) that have numerous links. A scale-free network is known to be tolerant to random removal of nodes, but it is very fragile against selective removal of hubs [7,11]. Second, PINs are small-world networks [4,5,8,10]. A small-world network is highly clustered like regular lattices, but it has small path lengths like a random network [12]. A small-world property is quantified by two statistics of a network, the average cluster coefficient $<C>$ and the average shortest path length $<L>$. The cluster coefficient of node i is defined as $C_i = 2e_i/k_i(k_i-1)$, where k_i is the degree of node i and e_i is the number of links connecting k_i neighbors of node i to one another [12]. (When k_i is zero or one, C_i is defined to be zero.) In other words, e_i is the number of triangles that pass through node i. C_i is equal to one when all neighbors of node i are fully connected to one another, while C_i is zero when none of the neighbors are connected to one another. A small-world network is characterized by a $<C>$ that is larger, and an $<L>$ that is similar, to those of a random network [12]. (In a random network, $<C> = <k>/N$ and $<L> \sim \log N/\log<k>$ [13], where $<k>$ is the average degree and N is the number of nodes.) Scale-free and small-world properties are commonly observed in various complex networks such as the Internet [9], coauthorship of scientific papers [14], metabolic pathways [15], and functional connections in the human brain [16]. Third, PINs show a hierarchical structure. In a network showing a hierarchical structure, $<C(k)>$, the average cluster coefficient of k-degree nodes, decays as a power law $<C(k)> \sim k^{-\mu}$ [17,18]. This indicates that a node with a small number of links

has a high C and belongs to a small subnetwork in which all nodes are densely connected, while a hub has a low C and links different subnetworks. Fourth, PINs show a disassortative structure, in which $<K_{nn}(k)>$ (''nn'' represents ''nearest neighbors''), the average degree among the neighbors of all k-degree nodes, follows $<K_{nn}(k)>\sim k^{-\nu}$ [19–21]. Therefore, the connections between a hub and a low-degree node are favored, while those between hubs and those between low-degree nodes are suppressed [19–22].

It has been reported that the emergence of scale-free networks can be explained by the mechanisms of network growth and preferential attachment, in which a new node is preferentially attached to a node that already has many links [23]. In the PIN evolution, gene duplication is thought to be responsible for preferential attachment, because gene duplication creates a new node having the same interacting patterns as the original node, and a high-degree node is more likely to gain a new link by the duplication of a randomly selected node than a low-degree node [6]. To account for the properties of PINs mentioned above, several network growth models have been proposed. These models are generally based on gene duplication and divergence. In a divergence process, some of the links created by duplication are removed and some new links are added to a network. Sole et al. [10] proposed a model in which a divergence process includes two mechanisms, random removal of links from one of the duplicated nodes and random attachment of new links between a duplicated node and another node. Both simulation and analytical studies have shown that this model can generate scale-free and small-world properties [10,24–27]. However, studies have reported that a network generated by this model having the same number of nodes and links as those in the yeast and fly PINs showed a much smaller $<C>$ than these PINs [10,28,29]. To overcome this difficulty, Vazquez et al. [30] and Ispolatov et al. [29] proposed the heterodimerization (HD) model. In this model, gene duplication is followed by divergence and HD; in the divergence process links are removed from duplicated nodes with a uniform probability α, and in the HD process a new link is established between two duplicated nodes with another probability β, forming a heterodimer [29,30]. When a self-interacting protein is duplicated, the duplicated proteins will interact to each other. Therefore, β in the HD model represents the probability that a randomly selected protein is self-interacting and the link between two duplicated proteins survives after divergence. Simulation and analytical studies have shown that the HD model could reproduce a similar $<C>$ to the yeast and fly PINs as well as a scale-free property [28–30]. The reason for the successful reproduction of a large $<C>$ is that an HD process creates a triangle, and a network containing a large number of triangles shows a large $<C>$. Middendorf et al. [28] reported that the HD model could best reproduce the fly PIN among seven network growth models using a technique from machine learning.

In this paper, we examine the yeast PIN, since it constitutes the most reliable PIN data currently available at the whole genome level [31]. We first show that the HD model is unsatisfactory as an evolutionary model of the yeast PIN. We then propose a new model named the non-uniform heterodimerization (NHD) model, in which an HD link is preferentially attached between two duplicated nodes that share many common neighbors. The NHD model can successfully reproduce various features of the yeast PIN that cannot be explained by the HD model.

Results

In this study, we examined two models, the heterodimerization (HD) model and the non-uniform heterodimerization (NHD) model (see Materials and Methods for details). In the HD model, there are two parameters, the probability that a link is removed from one of the duplicated nodes (α) and the probability that a new link is attached between two duplicated nodes (β) (Figure 1A), which represents the probability that a duplicated protein is self-interacting and the interaction between two duplicated proteins survives after the divergence process. These parameters were determined to let $<k>$ and $<C>$ in a generated network be the same as those in the yeast PIN. To compare the number of HD links in a generated network with that in the yeast PIN, we defined an evolutionary distance (Figure 1B). Two nodes are defined to be homologous when the evolutionary distance between these nodes is lower than or equal to a given threshold value d_T. The statistics of the networks generated by the HD model are shown in Table 1. The number of homologous pairs in the yeast PIN ($n_H = 6{,}544$) is between n_H for $d_T = 3$ (5,309) and that for $d_T = 4$ (8,337). However, the number of interactions between homologous nodes ($n_{HI} = 395$ and 514 for $d_T = 3$ and 4, respectively) is much larger than that in the yeast PIN (175). This observation is consistent with the investigation of the fly PIN by Ispolatov et al. [29], in which it was reported that the HD model requires a much larger number of HD links (270) than the actual number in the fly PIN (142) [32] to generate the 1,405 triangles present in the fly PIN.

As was mentioned in the Introduction, the HD model can generate a network with a high $<C>$, because an HD link produces triangles. When two duplicated nodes share n_N common neighbors, n_N new triangles are created by an HD link between them (Figure 1C). Therefore, if a new link is attached between duplicated nodes more preferentially when a larger number of neighbors are shared between them, it is expected that HD links fewer than those required by the HD model can reproduce the high $<C>$ in the PIN. For this reason, we examined the NHD model, in which the probability that a new link is added between duplicated nodes is proportional to the number of neighbors shared by these nodes. The probability of removing a link (α) and the proportionality constant to add a new link (β) were adjusted to let $<k>$ and $<C>$ in a generated network be the same as those in the yeast PIN. The results of simulations by the NHD model are shown in Table 1. Both n_H (6,544) and n_{HI} (175) in the yeast PIN are between the values for the NHD model with $d_T = 3$ and 4 ($n_H = 5{,}315$ and 8,351, respectively, and $n_{HI} = 157$ and 208, respectively). Moreover, the values of n_{HI}/n_H for the NHD model with $d_T = 3$ and 4 are very close to that in the yeast PIN. Therefore, both a high $<C>$ and a low n_{HI} were well reproduced by the NHD model. Table 1 also shows that $<L>$ in both HD and NHD networks are similar to that in a random network, indicating that they are small-world networks. However, interestingly, $<L>$ in the yeast PIN is much lower than that in a random network (see Discussion).

Figure 2A shows the degree distribution of the networks generated by the HD and NHD models and that of the yeast PIN. Although there is a discrepancy between the model networks and the yeast PIN for a large k (see Discussion), the results showed that both models can reproduce the degree distribution of the yeast PIN that follows a power law with an exponential cut-off. Figure 2B shows that $<C(k)>$ in the networks generated by the models follows a power law, indicating that these networks exhibit a hierarchical structure. In the case of the yeast PIN, however, $<C(k)>$ decreases following a power law as k increases only for a non-small k ($k>10$). This relationship was also observed in the previous studies [18,19]. As shown in Figure 2C, both the HD and NHD networks display a disassortative structure $<K_{nn}(k)>\sim k^{-\nu}$, but the values of ν are smaller than that in the yeast PIN (see Discussion).

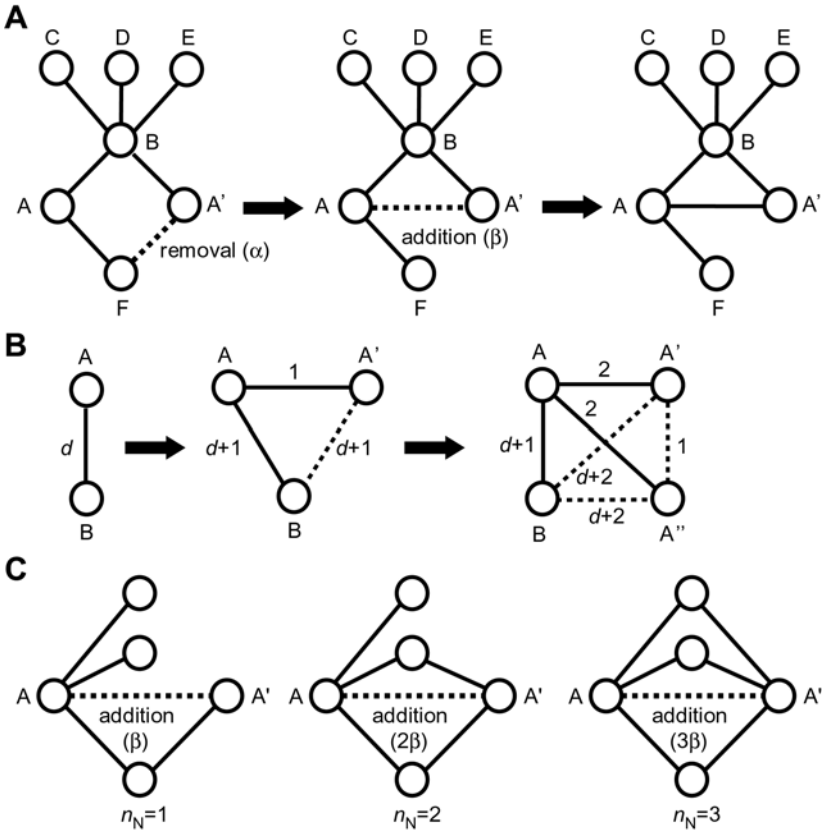

Figure 1. Simulation. (A) HD model. Node A is duplicated to generate node A′. Each of the links to node A′ is removed with a uniform probability α (left). Note that this method is based on completely asymmetric divergence [44], in which only one (A′) of the duplicated nodes is the target of removal of links. An HD link between node A and node A′ is attached with a uniform probability β (middle). (B) Evolutionary distance. When a node is duplicated, the evolutionary distance between each of the duplicated nodes and each of the other nodes in a network is assumed to increase by one due to mutations occurring in the duplicated nodes during the divergence process. Suppose that the evolutionary distance between node A and node B is *d* (left). After the duplication of node A to generate node A′ and the divergence of them, the evolutionary distance between nodes A and B, and that between nodes A′ and B become *d*+1 whether a link between nodes A and B and that between A′ and B are present or not (middle). (A dashed line indicates absence of a link.) The evolutionary distance between nodes A and A′ is defined to be 1 regardless of the presence of a link between them. After that, if node A′ is duplicated to create node A″, the evolutionary distance between nodes A and B continues to be *d*+1, while the evolutionary distances between nodes A and A′, A and A″, B and A′, and B and A″ become 2, 2, *d*+2, and *d*+2, respectively (right). (C) NHD model. In this model, the probability that a link is added between A and A′ is proportional to the number (n_N) of common neighbors shared by these nodes.
doi:10.1371/journal.pone.0001667.g001

Figure 2D shows the probability $P_T(n_T)$ that a given link is contained in n_T triangles in a network. For example, $n_T = 2$ for the link between nodes A and A′ (dashed line) in the middle of Figure 1C. The probability distribution $P_T(n_T)$ is a statistic describing a spatial distribution of triangles in a network. In a network generated by the HD model, the spatial distribution of triangles can be regarded to be random, because addition of a new HD link occurs randomly. As shown in this figure, the distribution of $P_T(n_T)$ in the yeast PIN is quite different from that in the network by the HD model, suggesting that the spatial distribution of triangles in the yeast PIN is highly skewed. In other words, in the yeast PIN, the extent of overlapping of triangles is larger than the expectation from a random distribution. On the other hand, the $P_T(n_T)$ distribution for the NHD model is close to that in the yeast PIN. Therefore, the structure of a network generated by the NHD model is more similar to the PIN than that by the HD model.

In the HD and NHD models, self-interactions were not explicitly considered, though it was assumed that an HD link is created only when a self-interacting protein is duplicated. However, in the yeast PIN, the fraction of self-interacting proteins is only 0.049, and $<k>$ increases slightly (3.84) when self-interactions are considered. The effect of self-interactions to other statistical properties of the yeast PIN is negligible (Figure S1). Therefore, it is expected that explicit consideration of self-interactions in a model does not essentially alter the results described above. We should also note that the fraction (0.049) of self-interactions in the yeast PIN is consistent with the value of β (0.028) in the NHD model. The fraction in the yeast PIN is much smaller than that (0.18) in the human transcription factor network, in which the statistical properties are considerably different between the networks with and without self-interactions [33].

We also examined the effect of gene deletions that are caused by mutations. For this purpose, we modified the NHD model by adding the process of random elimination of nodes (NHD+E model). However, the elimination of nodes did not essentially change the results (Table S1 and Figure S2).

Discussion

In this study, we showed that the NHD model can successfully reproduce both a high $<C>$ and a low n_{HI} in the yeast PIN, whereas the HD model cannot regenerate the value of n_{HI}. We also demonstrated that the distribution of triangles in the yeast

Table 1. Statistics of the networks by the HD and NHD models and the yeast PIN

Model	d_T	α[a]	β[a]	n_H[b]	n_{HI}[c]	n_{HI}/n_H	$<k>$[d]	$<C>$[e]	$<L>$[f]
HD model	1	0.725	0.061	1,312 (11)	140 (12)	0.107 (0.009)	3.73 (0.09)	0.066 (0.006)	6.45 (0.14)
	2	–	–	3,031 (27)	269 (19)	0.089 (0.006)	–	–	–
	3	–	–	5,309 (43)	395 (25)	0.074 (0.005)	–	–	–
	4	–	–	8,337 (65)	514 (31)	0.062 (0.004)	–	–	–
	5	–	–	12,363 (92)	628 (42)	0.051 (0.003)	–	–	–
NHD model	1	0.745	0.028	1,308 (11)	52 (6)	0.040 (0.005)	3.74 (0.07)	0.066 (0.006)	6.23 (0.12)
	2	–	–	3,030 (22)	105 (11)	0.035 (0.004)	–	–	–
	3	–	–	5,315 (42)	157 (17)	0.029 (0.003)	–	–	–
	4	–	–	8,351 (61)	208 (21)	0.025 (0.003)	–	–	–
	5	–	–	12,373 (86)	259 (28)	0.021 (0.002)	–	–	–
Yeast PIN[g]				6,544	175	0.027	3.74	0.066	4.85
Random[h]							3.74	0.00096	6.27

The number in parentheses represents the standard deviation calculated from 100 networks generated by simulations. –, the same as above.
[a]Parameters used in the simulations. See Materials and Methods.
[b]The number of homologous pairs. Two nodes are defined to be homologous when the evolutionary distance between the two nodes is d_T or less.
[c]The number of interactions between homologous proteins.
[d]The average degree.
[e]The average cluster coefficient.
[f]The average shortest path length.
[g]The yeast PIN without self-interactions.
[h]A random network that has the same $<k>$ and N as those in the yeast PIN, where N is the number of nodes (3,891) in the yeast PIN. The values of $<C>$ and $<L>$ were calculated using the formulae $<K>/N$ and $\log N/\log<k>$, respectively.
doi:10.1371/journal.pone.0001667.t001

PIN is highly skewed and the skewed distribution can be reproduced by the NHD model but not by the HD model. These results suggest that the NHD model would reflect the actual evolutionary mechanism of PINs.

Is the NHD model biologically realistic? In the PIN evolution, when a self-interacting protein is duplicated, an HD link between duplicated proteins is added to the PIN. Some HD links survive in evolution, but other links disappear because of mutations occurring at interacting sites in one or both of the duplicated proteins. Therefore, in the HD model, an HD link is assumed to survive at a uniform rate. On the other hand, in the NHD model, it is assumed that the survival rate of an HD link is proportional to the number of common neighbors shared by the duplicated nodes. Figure 3A shows the probability $P_{HD}(n_N)$ that two homologous nodes have an HD link when they share n_N common neighbors. This figure indicates that $P_{HD}(n_N)$ is nearly constant regardless of n_N in the networks by the HD model. On the other hand, in the yeast PIN, $P_{HD}(n_N)$ increases in proportion with n_N, which is consistent with the NHD model. These observations suggest that, in the evolution of PINs, the survival rate of HD links is not uniform in terms of n_N. Therefore, the NHD model appears to be realistic. The value of $P_{HD}(n_N)$ in the NHD network is smaller than that in the yeast PIN for $n_N<15$. This appears to happen because, in the NHD network, several protein pairs have very large values of n_N, which is not the case in the yeast PIN. That there are no protein pairs with large n_N in the yeast PIN may be due to the high duplicability of low-degree nodes [34], which was not considered in the NHD model (see below).

Why, then, is the survival rate of HD links not uniform, but rather proportional to the number of common neighbors? One possible explanation is as follows. It has been reported that the degree of proteins in the yeast PIN is negatively correlated with their evolutionary rates [35–37], though this assertion is controversial [38]. Not surprisingly, proteins connected by an HD link that has a large n_N tend to have a high degree (Figure 3B). Therefore, the evolutionary rates of proteins in an HD link with a large n_N are expected to be low. If this is the case, the possibility of the occurrence of mutations at the binding sites would also be low, and thus the survival rates of HD links having a large n_N are thought to be higher than those of HD links having a small n_N.

Although the degree distribution $P(k)$ of the NHD network is generally in good agreement with that of the yeast PIN, the number of nodes with $k>50$ in the former is much smaller than that in the latter (Figure 2A). The average of the maximum degrees among the NHD networks is 75.2, while the maximum degree in the yeast PIN is 286. Moreover, though the NHD network exhibits a disassortative structure $<K_{nn}(k)>\sim k^{-\nu}$, the value of ν is considerably smaller than that in the yeast PIN (Figure 2C). These discrepancies might be resolved by introducing a mechanism wherein low-degree nodes duplicate more frequently than high-degree nodes. Prachumwat and Li [34] reported a negative correlation between the degree of proteins and their duplicability. Due to the disassortative structure (Figure 2C), low-degree nodes have more links to high-degree nodes than to low-degree nodes. Therefore, as a result of frequent duplication of low-degree nodes, links between a high-degree node and a low-degree node are preferentially generated, and a high-degree node tends to gain new links. For this reason, with the mechanism of high duplicability of low-degree nodes, the degrees of hubs and the value of ν in Figure 2C are expected to become larger than the current values. Moreover, the lack of HD links with high n_N in the yeast PIN (Figure 3A) would also be explained with this mechanism, because HD links with high n_N should be rare if the duplicability of high-degree nodes is low.

Although PINs are generally considered to be small-world networks [4,8,10,24], the $<L>$ in the yeast PIN is much lower

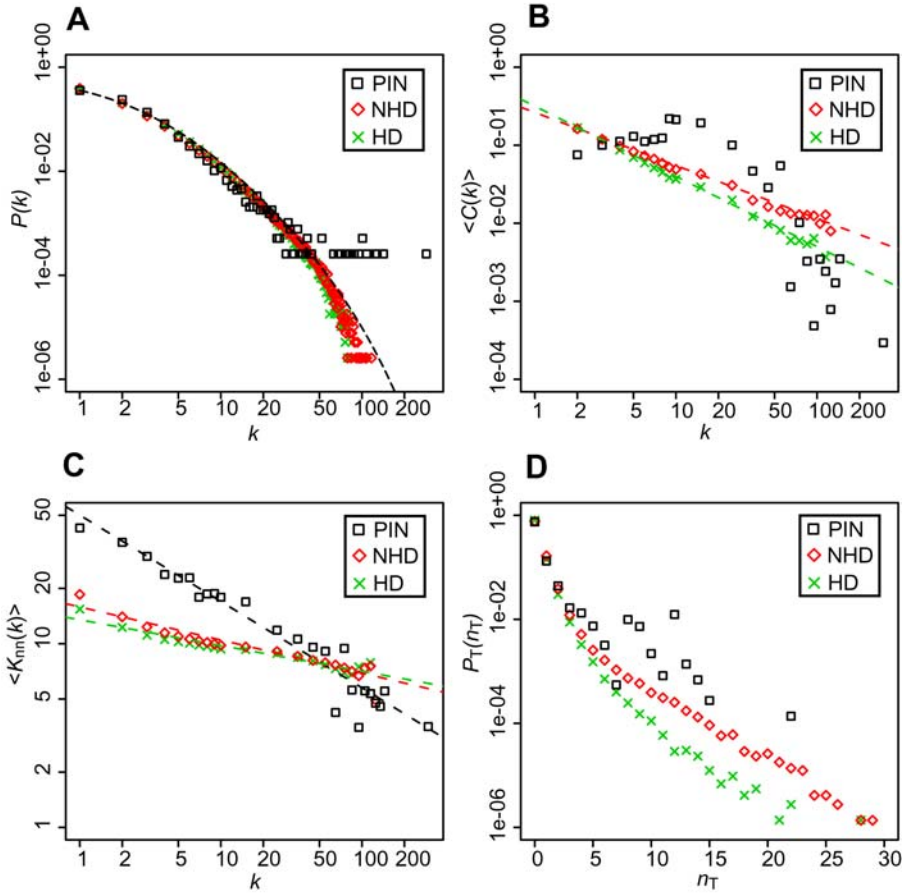

Figure 2. Properties in the networks by the HD and NHD models. Black squares, red diamonds, and green crosses show the values for the yeast PIN, the network generated by the NHD model, and the network by the HD model, respectively. The results for the HD and NHD models were obtained by taking the average among 100 networks generated by simulations. (A) Degree distribution $P(k)$. The dashed line represents $(k_0+k)^{-\gamma}e^{-k/kc}$ with $\gamma=2.7$, $k_0=3.4$, and $k_c=50$. (B) Distribution of the average cluster coefficient $<C(k)>$. Dashed lines in red and green indicate $k^{-0.68}$ and $k^{-0.90}$, respectively. (C) Distribution of $<K_{nn}(k)>$ indicating a disassortative structure. Dashed lines in black, red, and green represent $k^{-0.47}$, $k^{-0.18}$, and $k^{-0.14}$, respectively. (D) Distribution of $P_T(n_T)$, the probability that a given link is contained in n_T triangles.
doi:10.1371/journal.pone.0001667.g002

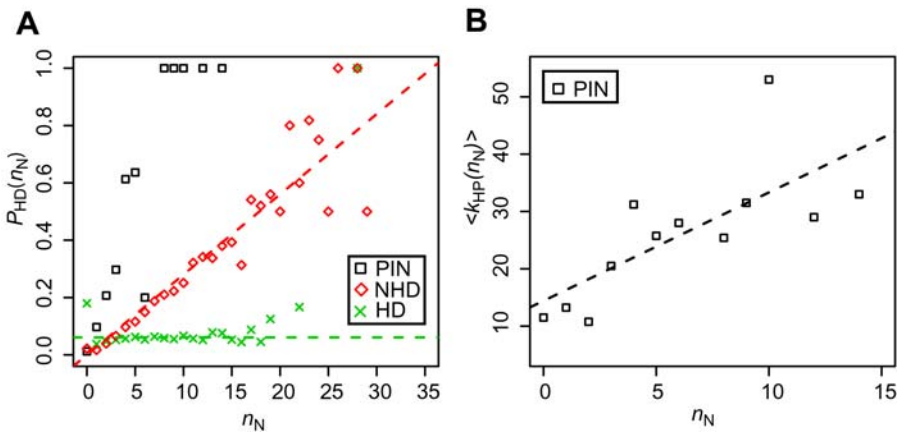

Figure 3. HD links in the yeast PIN and in the networks by simulations. Black squares, red diamonds, and green crosses show the values for the yeast PIN, the network generated by the NHD model, and the network by the HD model, respectively. (A) Distribution of $P_{HD}(n_N)$, the probability that an HD link exists between two homologous proteins when they share n_N common neighbors (for $d_T=3$). The slopes of the dashed lines are 0.028 (red) and 0 (green). The result for $d_T=4$ is nearly identical to this result (data not shown). (B) Distribution of $<k_{HP}(n_N)>$, the average degree of proteins that are connected by HD links and share n_N common neighbors with their homologous proteins. The dashed line is a regression line ($r=0.73$).
doi:10.1371/journal.pone.0001667.g003

than that in a random network (Table 1). In fact, this observation is consistent with the previous result by Pastor-Satorras et al. [24], in which it was reported that $<L>$ in the random network and that in their model network were 8.0 and 6.8, respectively. (However, they mentioned that these two values are "comparable".) Therefore, the yeast PIN is an "ultra-small" network, in which $<L>$ is lower than that in a random network. It is known that a scale-free random network is ultra-small [39,40]. The yeast PIN can be randomized without changing the distribution of $P(k)$ by using the random rewiring method [21]. In this method, two links in a network were chosen randomly, and these links were rewired by exchanging their connecting partners. After randomization of the yeast PIN, $<L>$ (4.49) is similar, but $<C>$ (0.010) becomes much lower than that in the yeast PIN. Therefore, the yeast PIN is far from a scale-free random network. Nevertheless, interestingly, the yeast PIN is an ultra-small network.

The difference in the value of $<L>$ between the yeast PIN and the NHD network may be explained in the following way. It was reported that the removal of hubs drastically increases the value of $<L>$ in the yeast PIN [41]. Therefore, the low $<L>$ in the yeast PIN might be due to the fact that the number of hubs in the yeast PIN is larger than that in the NHD network (Figure 2A). (There are 17 nodes with $k>50$ in the yeast PIN, while the average number of nodes with $k>50$ among the NHD networks is 5.8.) In fact, if we eliminate all nodes with $k>50$ and all links connected to them from the yeast PIN and the NHD network, both $<L>$ and $<C>$ become similar between the two networks ($<L> = 6.13$ and 6.51, and $<C> = 0.063$ and 0.060 for the yeast PIN and the NHD network, respectively). It therefore appears that the presence of a large number of hubs in the yeast PIN would be the reason for a very low $<L>$.

The above discussion would indicate that the NHD model is merely a rough approximation of the actual mechanism of the PIN evolution. However, we should note that although our new model contains only two free parameters, it could well capture various aspects of the structure of the yeast PIN. The availability of high-quality interaction data from other species will thus help to clarify the architecture and evolution of PINs in greater detail.

Materials and Methods

Data

Human-curated interaction data of the yeast PIN were downloaded from the MIPS (Munich Information Center for Protein Sequences) database (http://mips.gsf.de) (18 May 2006) [3]. The interaction data are separated into several components that are not connected to each other; we used the largest component containing 3,891 proteins and 7,270 non-redundant interactions. Among these proteins, 191 proteins are self-interacting. The amino acid sequences of 6,736 yeast proteins were also obtained from the MIPS database. In order to estimate the number of interactions between homologous proteins in the yeast PIN, we identified homologous gene pairs. Self-against-self homologous searches were conducted for the 6,736 sequences by using the BLASTP program [42] with the cut-off E-value of 1e-5. We identified 6,544 homologous pairs (n_H) and 175 interactions between these pairs of proteins (n_{HI}) in the yeast PIN (see Table 1). The value of n_{HI}/n_H did not essentially change when a more stringent cut-off E-value was used (n_{HI}/n_H was 0.027 and 0.032 for the E-values of 1e-5 and 1e-10, respectively).

Simulation

In this study, we used the "minimal genome" containing 113 proteins as the initial network, because the first living organism is assumed to have had at least 113 proteins [43]. We generated a single component random network containing 113 nodes with $<k> = 3.74$, which is the average degree of the yeast PIN. The evolutionary distance between two nodes present in the initial network was assumed to be infinity. We obtained very similar results when we started a simulation from the initial network containing only two nodes.

At each time step of simulation in the HD model, a new node is added to the network according to the following rules (Figure 1A). (1) A node is randomly selected (A) and is duplicated to generate a new node (A′), having the same interacting pattern as node A. (2) Each of the links to node A′ is removed with a probability α (completely asymmetric divergence [44]). (3) A link between node A and node A′ is created with a probability β. If node A′ does not have any links after these processes (all links to node A′ were removed and no links were created), node A′ is not added to the network. These processes were repeated until the number of nodes became 3,891, which is the number of nodes contained in the yeast PIN. In the NHD model, the probability that a new link is added between two duplicated nodes (A′ and A) is defined to be βn_N (when $\beta n_N \leq 1$), where n_N is the number of common neighbors shared by these two nodes (Figure 1C). The probability is defined to be one when $\beta n_N > 1$. (However, there were no such cases in the simulations.) We performed simulations using various values of α and β. For a given α and β, we conducted simulations 100 times and computed the average of $<k>$ and the average of $<C>$ from the 100 networks. The values of α and β that could reproduce $<k>$ (3.74) and $<C>$ (0.066) in the yeast PIN were used (Table 1).

In the NHD+E model, the following process was added after the addition of new links at each step of the NHD model. A node in a network is randomly selected, and the selected node is eliminated from the network with a probability δ together with all interactions connecting to the selected node. If the selected node is connected to one-degree nodes, all of these one-degree nodes are also removed. We changed the value δ from 0.001 to 0.1 (see Table S1). The values of α and β were determined in the same way as in the NHD model.

Supporting Information

Table S1

Found at: doi:10.1371/journal.pone.0001667.s001 (0.04 MB DOC)

Figure S1 Properties in the yeast PIN with and without self-interactions. Red triangles and black squares show the values for the yeast PINs with and without self-interactions, respectively. (A) Degree distribution $P(k)$. (B) Distribution of the average cluster coefficient $<C(k)>$. (C) Distribution of $<K_{nn}(k)>$. (D) Distribution of $P_T(n_T)$.

Found at: doi:10.1371/journal.pone.0001667.s002 (0.78 MB TIF)

Figure S2 Properties in the networks by the NHD and NHD+E models. Black squares, red diamonds, and blue crosses show the values for the yeast PIN, the network generated by the NHD model, and the network by the NHD+E model with $\delta = 0.1$, respectively. The results for the NHD and NHD+E models were obtained by taking the average among 100 networks generated by simulations. (A) Degree distribution $P(k)$. The dashed line represents $(k_0+k)^{-\gamma}e^{-k/k_c}$ with $\gamma = 2.7$, $k_0 = 3.4$, and $k_c = 50$. (B) Distribution of the average cluster coefficient $<C(k)>$. Dashed line in red indicates $k^{-0.68}$. (C) Distribution of $<K_{nn}(k)>$. Dashed lines in black and red represent $k^{-0.47}$ and $k^{-0.18}$, respectively. (D) Distribution of $P_T(n_T)$.

Found at: doi:10.1371/journal.pone.0001667.s003 (0.82 MB TIF)

Acknowledgments

The authors thank T. Masuda, Y. Fukuoka, T. Takai, F. Ren, M. Koeda, S. Ogishima, E. Campos, M. Morioka, and S. Nakagawa for their useful comments and discussion.

Author Contributions

Conceived and designed the experiments: YN HT TH TK. Performed the experiments: TH. Analyzed the data: TH. Contributed reagents/materials/analysis tools: TH. Wrote the paper: YN TH.

References

1. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, et al. (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature 403: 623–627.
2. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, et al. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci USA 98: 4569–4574.
3. Guldener U, Munsterkotter M, Oesterheld M, Pagel P, Ruepp A, et al. (2006) Mpact: The MIPS protein interaction resource on yeast. Nucleic Acids Res 34: D436–441.
4. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, et al. (2004) A map of the interactome network of the metazoan C. elegans. Science 303: 540–543.
5. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, et al. (2003) A protein interaction map of drosophila melanogaster. Science 302: 1727–1736.
6. Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5: 101–113.
7. Jeong H, Mason SP, Barabasi AL, Oltvai ZN (2001) Lethality and centrality in protein networks. Nature 411: 41–42.
8. Wagner A (2001) The yeast protein interaction network evolves rapidly and contains few redundant genes. Mol Biol Evol 18: 1283–1292.
9. Albert R, Jeong H, Barabasi AL (1999) Diameter of the World-Wide Web. Nature 401: 130–131.
10. Sole RV, Pastor-Satorras R, Smith ED, Kepler T (2002) A model of large-scale proteome evolution. Adv Comp Syst 5: 43–54.
11. Albert R, Jeong H, Barabasi AL (2000) Error and attack tolerance of complex networks. Nature 406: 378–382.
12. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393: 440–442.
13. Albert R, Barabasi AL (2002) Statistical mechanics of complex networks. Reviews of Modern Physics 74: 47–97.
14. Newman MEJ (2001) The structure of scientific collaboration networks. Proc Natl Acad Sci USA 98: 404–409.
15. Jeong H, Tombor B, Albert R, Oltvai ZN Barabasi AL (2000) The large-scale organization of metabolic networks. Nature 407: 651–654.
16. Eguiluz VM, Chialvo DR, Cecchi GA, Baliki M, Apkarian AV (2005) Scale-free brain functional networks. Phys Rev Lett 94: 018102.
17. Williams RJ, Martinez ND, Berlow EL, Dunne JA, Barabasi AL (2002) Hierarchical organization of modularity in metabolic networks. Science 297: 1551–1555.
18. Yook SH, Oltvai ZN, Barabasi AL (2004) Functional and topological characterization of protein interaction networks. Proteomics 4: 929–942.
19. Vazquez A (2003) Growing networks with local rules: preferential attachment, clustering hierarchy and degree correlations. Phys Rev E 67: 056104.
20. Costa LF, Rodrigues FA, Travieso G, Boas V (2007) Characterization of complex networks: A survey of measurements. ADV PHYS 56: 167–242.
21. Maslov S, Sneppen K (2002) Specificity and stability in topology of protein networks. Science 296: 910–913.
22. Pastor-Satorras R, Vazquez A, Vespignani A (2001) Dynamical and correlation properties of the internet. Phys Rev Lett 87: 258701.
23. Barabasi AL, Albert R (1999) Emergence of scaling in random networks. Science 286: 509–512.
24. Pastor-Satorras R, Smith E, Sole RV (2003) Evolving protein interaction networks through gene duplication. J Theor Biol 222: 199–210.
25. Kim J, Krapivsky PL, Kahng B, Render S (2002) Infinite-order percolation and giant fluctuations in a protein interaction network. Phys Rev E 66: 055101.
26. Chung F, Lu L, Dewey TG, Galas DJ (2003) Duplication models for biological networks. J Comput Biol 10: 677–87.
27. Raval A (2003) Some asymptotic properties of duplication graph. Phys Rev E 68: 066119.
28. Middendorf M, Ziv E, Wiggins CH (2005) Inferring network mechanisms: The Drosophila melanogaster protein interaction network. Proc Natl Acad Sci USA 102: 3192–3197.
29. Ispolatov I, Krapivsky PL, Mazo I, Yuryev A (2005) Cliques and duplication-divergence network growth. New J Phys 7: 145.
30. Vazquez A, Flammini A, Maritan A, Vespignani A (2003) Modeling of protein interaction networks. ComPlexUs 1: 38–44.
31. Patil A, Nakamura H (2005) Filtering high-throughput protein-protein interaction data using a combination of genomic features. BMC Bioinformatics 6: 100.
32. Ispolatov I, Yuryev A, Mazo I, Maslov S (2005) Binding properties and evolution of homodimers in protein-protein interaction networks. Nucleic Acids Res 33: 3629–3635.
33. Rodriguez-Caso C, Medina MA, Sole RV (2005) Topology, tinkering and evolution of the human transcription factor networks. FEBS J 272: 6423–34.
34. Prachumwat A, Li WH (2006) Protein function, connectivity, and duplicability in yeast. Mol Biol Evol 23: 30–39.
35. Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW (2002) Evolutionary rate in the protein interaction network. Science 296: 750–752.
36. Fraser HB (2003) A simple dependence between protein evolution rate and the number of protein-protein interactions. BMC Evol Biol 3: 11.
37. Fraser HB (2005) Modularity and evolutionary constraint on proteins. Nature Genetics 37: 351–352.
38. Jordan IK, Wolf YI, Koonin EV (2003) No simple dependence between protein evolution rate and the number of protein-protein interactions: only the most prolific interactors tend to evolve slowly. BMC Evol Biol 3: 1.
39. Chung F, Lu L (2002) The average distances in random graphs with given expected degrees. Proc Natl Acad Sci USA 99: 15879–82.
40. Cohen R, Havlin S (2003) Scale-free networks are ultrasmall. Phys Rev Lett 90: 058701.
41. Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, et al. (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 430: 88–93.
42. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.
43. Forster AC, Church GM (2006) Towards synthesis of a minimal cell. Mol Syst Biol 2: 45.
44. Ispolatov I, Krapivsky PL, Yuryev A (2005) Duplication-divergence model of protein interaction network. Phys Rev E 71: 061911.