Molecular BioSystems

Cite this: DOI: 10.1039/c1mb05028a

www.rsc.org/molecularbiosystems

PAPER

A predicted protein–protein interaction network of the filamentous fungus *Neurospora crassa*†

Ting-You Wang,[ab] Fei He,[ab] Qi-Wen Hu[a] and Ziding Zhang*[ab]

Received 25th January 2011, Accepted 17th April 2011
DOI: 10.1039/c1mb05028a

The filamentous fungus *Neurospora crassa* is a leading model organism for circadian clock studies. Computational identification of a protein–protein interaction (PPI) network (also known as an interactome) in *N. crassa* can provide new insights into the cellular functions of proteins. Using two well-established bioinformatics methods (the interolog method and the domain interaction-based method), we predicted 27 588 PPIs among 3006 *N. crassa* proteins. To the best of our knowledge, this is the first identified interactome for *N. crassa*, although it remains problematic because of incomplete interactions and false positives. In particular, the established PPI network has provided clues to further decipher the molecular mechanism of circadian rhythmicity. For instance, we found that clock-controlled genes (ccgs) are more likely to act as bottlenecks in the established PPI network. We also identified an important module related to circadian oscillators, and some functional unknown proteins in this module may serve as potential candidates for new oscillators. Finally, all predicted PPIs were compiled into a user-friendly database server (NCPI), which is freely available at http://protein.cau.edu.cn/ncpi.

Introduction

Neurospora crassa has been an experimental model organism for the fundamental understanding of genome defense systems, DNA repair, circadian rhythms, DNA methylation, and post-transcriptional gene silencing for the latter half of the 20th century.[1] In particular, it has become one of the most durable and dependable model organisms for studying circadian rhythmicity, thereby playing an important role in addressing the central questions of chronobiology.[2] In the circadian system, circadian oscillators receive light and temperature signals, and then transmit them to output pathways. Thus, oscillators can control the rhythmic activity of the clock-controlled genes (ccgs) that function in various aspects of the fungal life cycle.[3] Prior to this study, the core oscillator components and ~295 ccgs had been identified in *N. crassa*.[4–6]

As the main participants in most cellular processes, proteins perform their functions by creating macromolecular assemblies and a large number of protein–protein interactions (PPIs). The availability of PPI networks (also referred to as interactomes) will provide a new way to understand the biological organizations of living organisms from the perspective of systems biology. Interactome data have also been widely used to assign biological functions to uncharacterized proteins. This process is supported by the observation that interacting proteins generally have collaborative or similar functions.[7] The genome sequences of *N. crassa* were released in 2003,[1] and only approximately 40% of the proteins have been functionally annotated in the MIPS *N. crassa* Genome Database (MNCDB).[8] Undoubtedly, a comprehensive *N. crassa* PPI dataset will accelerate the functional annotation of *N. crassa* proteins. Additionally, a PPI network will also be helpful to explore the molecular mechanism of circadian rhythmicity from the network perspective.

Using high-throughput techniques, such as yeast two-hybrid,[9] mass spectrometry[10] and protein chip,[11] proteome-scale interactome data have been experimentally extracted for many model organisms, including *Saccharomyces cerevisiae*,[12] *Caenorhabditis elegans*,[13] *Drosophila melanogaster*,[14] *Helicobacter pylori*[15] and *Homo sapiens*.[16] These interactome data have provided insights into biological complexes, pathways and entire organisms, despite the noise and incompleteness of the experimentally determined PPIs.[17] Unfortunately, none of these high-throughput methods has been applied to the filamentous fungus *N. crassa*, although it is highly demanded. Owing to the time-consuming nature of experiments, a variety of computational prediction methods have been developed to complement experimental approaches. These prediction methods can be roughly classified into genome-scale approaches,[18] sequence-based approaches,[19] structure-based approaches,[20] machine learning-based approaches[21] and network-based

[a] *State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing 100193, China. E-mail: zidingzhang@cau.edu.cn*
[b] *Bioinformatics Center, College of Biological Sciences, China Agricultural University, Beijing 100193, China*
† Electronic supplementary information (ESI) available. See DOI: 10.1039/c1mb05028a

approaches.[22] In general, prediction methods can generate protein pairs with functional associations or physical interactions.[23,24] Two sequence-based approaches (*i.e.*, the interolog method[25] and domain interaction-based method[26]) are widely implemented in practical applications to predict physical interactions.[27] The interolog method can be described as the transfer of interaction annotations between species through comparative genomics,[19] which is qualified by the conservation of proteins. The domain interaction-based method is based on the assumption that protein interaction is deduced from domain–domain interactions. Thus, the domain interaction-based method complements the interolog method because interacting proteins without interacting homologs may contain interacting domains that can be obtained experimentally. Until now, these two well-recognized methods have been employed to construct proteome-wide PPI networks for important organisms, such as *Magnaporthe grisea*,[28] *Fusarium graminearum*,[29] *Mycobacterium tuberculosis*[30] and *Deinococcus radiodurans*.[31]

In this study, we used the interolog method and the domain interaction-based method to construct a predicted PPI network of *N. crassa*. To explore the mechanism of circadian rhythmicity from the network perspective, network topology analysis of CCGs and non-CCGs was carried out and network modules associated with CCGs and circadian oscillators were identified. We also built a user-friendly web-interface *N. crassa* Protein Interactome (NCPI) database.

Materials and methods

Datasets

***N. crassa* sequence data.** We obtained 9830 *N. crassa* protein sequences from the Broad Institute (http://www.broadinstitute.org/annotation/genome/neurospora/MultiDownloads.html, Version 3).

Protein–protein interaction data. We collected PPI data from four high-quality interactome databases, including 60 699 PPIs from the Database of Interacting Proteins (DIP; http://dip.doe-mbi.ucla.edu/, 12-30-2009 release),[32] 103 628 PPIs from the Molecular INTeraction database (MINT; http://mint.bio.uniroma2.it/mint/, 07-29-2009 release),[33] 38 788 PPIs from the Human Protein Reference Database (HPRD; http://www.hprd.org, 07-06-2009 release 8),[34] and 193 422 PPIs from The IntAct molecular interaction database (IntAct; http://www.ebi.ac.uk/intact/, 08-25-2009 release).[35]

Domain interaction data. We obtained protein domain interactions from iPfam[36] and 3did,[37] which were compiled from the known domain–domain interaction data in PDB.[38] A total of 5785 non-redundant interacting Pfam-A domain pairs were collected from the iPfam database (http://pfam.sanger.ac.uk/, version 24) and the 3did database (http://3did.irbbarcelona.org/, 10-18-2009 release).

Prediction of PPIs based on the interolog approach

Using the interolog method, we identified potential *N. crassa* PPIs based on the known PPIs in DIP, MINT, HPRD and IntAct (Fig. 1). First, the PPIs in these four interactome databases were preprocessed. Briefly, the sequence identifiers in different interactome databases were treated with a standardization pipeline. This process allowed us to avoid any possible confusion caused by the fact that one protein may have different identifiers. All sequence identifiers were mapped to NCBI RefSeq or EBI Uniprot identifiers. The unmatched proteins were considered to be unverified and were discarded. In addition, six known *N. crassa* PPIs from these four interactome databases were also discarded at this step because these PPIs were later used to assess the prediction results. Next, the sequences of all the remaining PPIs were obtained from EBI Uniprot if they were unavailable in the corresponding interactome database. For each PPI in the interactome databases, the corresponding two protein sequences were used as queries and BLASTed against the whole proteome of *N. crassa* to identify homologs with e-value, sequence identity and aligned sequence length coverage cut-offs of 1.0×10^{-10}, 40% and 40%, respectively. The aligned sequence length coverage was defined as the aligned sequence length of the query (without gaps) divided by the whole sequence length of the query. The corresponding homolog pairs identified in *N. crassa* were predicted to interact with each other. In general, PPIs inferred from more than one experimentally verified PPI have higher confidence. Therefore, PPIs predicted from only one PPI in the interactome datasets were removed.

Prediction of PPIs based on domain interaction-based approach

The key idea of the domain interaction-based approach in this study was to infer whether *N. crassa* protein pairs can interact based on the domain–domain interaction information from the iPfam and 3did databases (Fig. 1). Briefly, *N. crassa* proteins were mapped to Pfam-A domains from the Pfam database with the e-value and aligned sequence length coverage cut-offs of 0.001 and 80%, respectively. The aligned sequence length coverage was computed as the aligned sequence length of the query (without gaps) divided by the total sequence length of the corresponding Pfam-A domain. If the *N. crassa* protein pair contained an interacting Pfam-A domain pair, the proteins were expected to interact with each other. Compared with the interolog method, the domain interaction-based method generally shows lower accuracy.[39] To remove potentially false positives, two filtration procedures were conducted. We first filtered the predictions based on interacting Pfam-A domain pairs that were able to infer more than 1% of the total predictions.[40] The interactions among multiple domain proteins may be mediated by multidomain interactions or domain–motif interactions, rather than single domain pairs.[29] Therefore, the predicted interacting protein pair was further filtered if any protein in the pair contained multiple domains.

Analysis of network topology

A PPI network can be represented as an undirected graph, $G(V, E)$, that consists of a set of vertices, V, and a set of edges, E. Each vertex (*i.e.*, node) represents a singular protein, while each edge represents an interaction between two proteins. The degree refers to the number of interacting partners of a vertex.

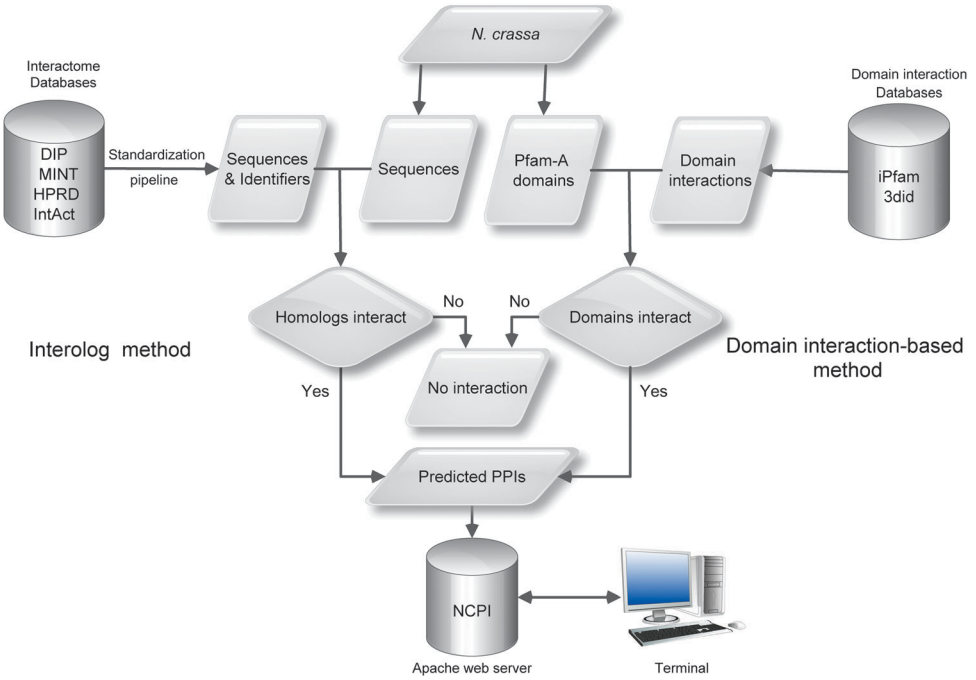

Fig. 1 Flowchart of the interolog and domain interaction-based approaches in predicting *N. crassa* protein–protein interactions. This picture was prepared by using ConceptDraw PRO 7(http://www.conceptdraw.com).

The betweenness is measured by the total number of shortest paths through a certain vertex. More shortest paths crossing through a given vertex results in a higher score of betweenness for the vertex. The closeness depicts how close a vertex is to all other vertices in the graph, which is defined as the average distance from a vertex to any other vertex. The clustering coefficient measures the network cohesiveness, which reflects the density of connected neighborhoods of a vertex. We used NetworkAnalyzer[41] (http://med.bioinf.mpi-inf.mpg.de/netanalyzer/), a Java plugin for Cytoscape,[42] to calculate these parameters. In addition, hubs were defined as the top 20% high degree proteins. Bottlenecks were defined as the top 20% high betweenness proteins.[43]

Assessment of the reliability of predicted PPIs

First, we collected experimentally-verified *N. crassa* protein interactions from public interactome databases and previous studies to provide a direct validation of the predicted *N. crassa* PPIs. We also utilized three different methods to assess the reliability of the predicted PPIs indirectly: the protein localization (PL) method,[44] the expression profile reliability (EPR) method,[45] and the annotation similarity (AS) method. The reliability of the predicted PPIs was tested indirectly because there is no comprehensive experimental *N. crassa* PPI dataset available. The general strategy of these three methods is to compare the predicted network with a power law-preserving randomized network.[46] In this study, the randomized network was constructed by keeping the degree distribution consistent with that of the predicted network and randomizing the degree of proteins. We used "randomized network" hereafter to describe a power law–preserving randomized network, unless stated otherwise.

The PL method assumes that interacting proteins are localized in the same cellular compartment. To perform the PL method, the protein sub-cellular localization information of *N. crassa* proteins was predicted using WoLFPSORT with default parameters.[47] Because the overall prediction accuracy of WoLFPSORT is reasonably good (*e.g.*, >80%), the PL method can verify the reliability of predicted PPIs.

The EPR method is based on the assumption that interacting proteins tend to be co-expressed. The level of gene co-expression of an interacting protein pair is measured by applying a Pearson correlation coefficient (PCC) between the corresponding gene expression profiles, which were downloaded from the filamentous fungal microarray database[48] (http://www.yale.edu/townsend/Links/ffdatabase/introduction.htm, accession number 13).

The AS method is based on the finding that 70%–80% of interacting protein pairs share similar functions.[7] The Gene Ontology (GO) annotations of *N. crassa* were obtained using InterProScan with default parameters[49] because the GO annotations for *N. crassa* were not available in the GO Consortium. The GO annotations of 4473 *N. crassa* proteins were assigned, which covered 2533 proteins in the predicted PPI network. Similar to Zhao *et al.* (2009),[29] the Jaccard index was employed to measure the functional similarity between two interacting proteins. The Jaccard index is defined as the size of the intersection of two proteins' GO terms divided by the size of the union of the corresponding GO terms. The Jaccard index ranges from 0 to 1, and higher values represent higher functional similarity. For comparison, we also conducted the AS assessment using the functional annotations from the MIPS FunCat system,[50] which is based on manual curation and may be more reliable.[51] 4073 *N. crassa* proteins were annotated in FunCat, which covered 2315 proteins in the predicted PPI network.

Results and discussion

Predicted protein–protein interaction network of *N. crassa*

In this study, 9803 interactions among 1163 proteins were predicted based on the interolog approach, and 18 437 interactions among 2617 proteins were inferred from the domain interaction-based approach. By integrating predictions from the interolog and domain interaction-based approaches and then removing redundancy, we obtained a PPI dataset of 27 588 interactions among 3006 proteins. Each protein had an average of 18.4 partners. In general, the overlap between the PPIs predicted by these two different methods was negligible (Fig. 2A), suggesting that these two methods are complementary.

Scale-free network topological property is explicitly founded on the predicted PPI network (Fig. 2B), which has been frequently observed in the experimentally determined protein interaction networks of model organisms.[52] In a scale-free network, the probability, $P(k)$, of nodes having k edges decays as a power law $P(k) \approx k^{-\gamma}$. Our predicted network can be approximately characterized by a power law distribution, where $P(k) \approx k^{-1.38}$ ($R^2 = 0.827$). In the predicted PPI network, 85 out of 3006 proteins had a degree higher than 100. These proteins were ranked as the top 15% high-degree hubs. These hub proteins may perform important cellular functions involved in different biological processes. For instance, the top three hubs in the predicted PPI network [*i.e.*, *hsp70-1* (NCU09602), *hsp70-5* (NCU08693) and *grp78* (NCU03982)] belong to the heat shock-induced protein HSP70 family.

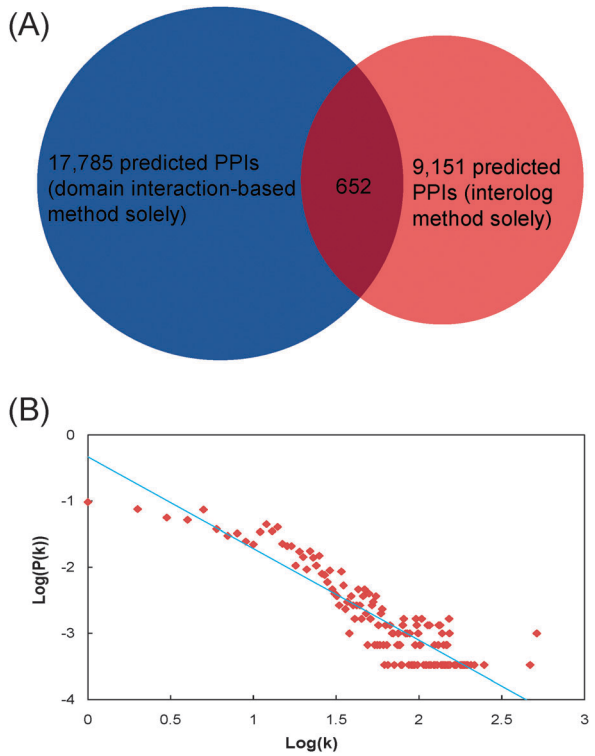

Fig. 2 The predicted *N. crassa* PPIs. (A) Area-proportional Venn diagram of predicted PPIs based on the two methods. (B) Degree distribution of the predicted network, with both axes plotted on logarithmic scales.

Functioning as chaperones that stabilize the conformation of a variety of proteins, HSP70 proteins are ubiquitously and abundantly synthesized to protect cells from thermal or oxidative stress.[53]

Assessment of the reliability of the predicted PPIs

Up to now, proteome-scale interactome data are still not available for the fungus *N. crassa*, which made the assessment of the predicted PPIs based on a large dataset of experimentally-determined PPIs impossible. Despite the lack of sufficient gold standard positives, we manually curated protein interactions from public interactome databases and previous studies to validate the predicted *N. crassa* PPIs. Of the six PPIs deposited in public interactome databases and the twelve PPIs collected from the literature, ten PPIs were successfully predicted in the established PPI network (Table 1), which demonstrates that the overall performance of the predicted PPIs was reasonably high. For instance, two *N. crassa* proteins (NCU01227 and NCU08471) were experimentally characterized to be two interacting subunits of *N. crassa* succinyl-CoA ligase and were successfully predicted to have a physical interaction in this study. This particular PPI was deduced from the interolog method, and the corresponding homologs in other model organisms include human (P53597 and Q96199), *Escherichia coli* (P0AGE9 and P0A836), yeast (P53598 and P53312) and *Campylobacter jejuni* (Q0PAY2 and Q9PHY1).

In addition to the direct assessment, we also employed three different indirect methods to comprehensively evaluate the reliability of the predicted *N. crassa* PPI network. For the PL method, 8278 PPIs in the predicted PPI network were co-localized. The average number of co-localizations in the 1000 randomized networks was 5877 ± 5.66, and the largest number of co-localizations was 6371. Therefore, the predicted network had a significantly higher number of co-localized PPIs than any of the randomized PPI networks (empirical $P < 0.001$). For the EPR method, the distribution of PCCs for PPIs in the predicted PPI dataset was significantly different from that of the randomized network ($P < 2.2 \times 10^{-16}$, Pearson's Chi-squared test, 19 d.f.). As shown in Fig. 3A, the ratio of PPIs with higher PCC in the predicted network was larger than that in the randomized network. In contrast, the proportion of PPIs with lower PCCs in the predicted network was smaller than that in the randomized network. Therefore, the protein interaction pairs in the predicted PPI network were prone to co-expression, implying that the predicted PPI network was more credible than the randomized network. Moreover, we also employed the AS method to verify the predicted PPI network. Fig. 3B depicts the distribution of Jaccard indices of interacting proteins in the predicted network compared with the randomized network based on the GO annotations. In general, the predicted interacting protein pairs tended to have more similar functions than the random pairs. The fractions corresponding to a Jaccard index of 0 in the predicted and randomized networks were 0.62 and 0.90, respectively. When the Jaccard index was equal to 1.0, the corresponding fractions were 0.11 and 0.003, respectively. In addition, we did this assessment based on the MIPS FunCat annotations and similar results were obtained (see ESI†).

Table 1 A collection of curated protein interactions from interactome databases and previous studies

No.	Locus	Locus	Cited databases or references	Be predicted
1	NCU01635	NCU04402	IntAct	No
2	NCU06698	NCU06687	MINT	No
3	NCU04202	NCU08791	MINT	No
4	NCU04202	NCU00355	MINT	No
5	NCU03982	NCU02455	MINT	No
6	NCU00902	NCU02356	MINT, DIP	Yes
7	NCU02247	NCU04017	ref. 65	No
8	NCU09068	NCU08294	ref. 66	No
9	NCU01605	NCU07296	ref. 67	Yes
10	NCU01605	NCU09071	ref. 67	Yes
11	NCU02234	NCU06419	ref. 68	Yes
12	NCU06419	NCU11376	ref. 68	Yes
13	NCU03071	NCU00587	ref. 68	No
14	NCU00587	NCU07024	ref. 68	Yes
15	NCU06182	NCU04612	ref. 68	Yes
16	NCU04612	NCU02393	ref. 68	Yes
17	NCU01227	NCU08471	ref. 53	Yes
18	NCU06605	NCU00272	ref. 69	Yes

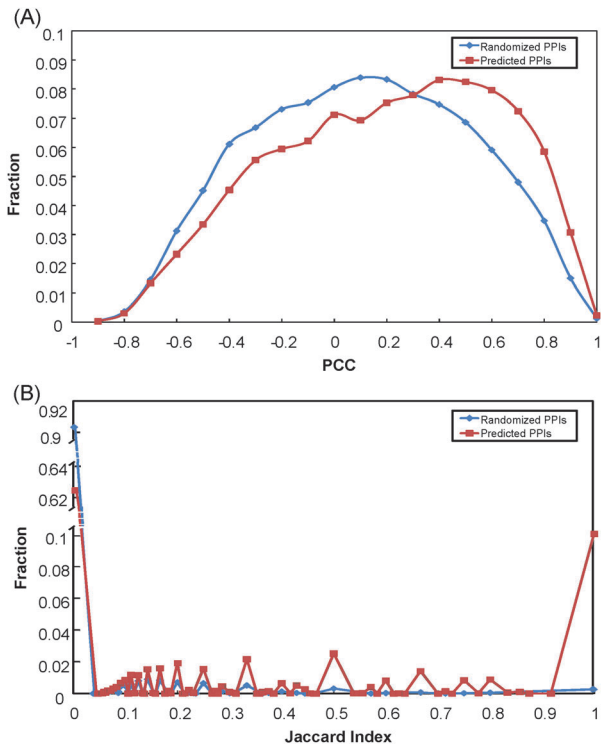

Fig. 3 Validation of the predicted PPI network based on the EPR and AS methods. (A) Distribution of PCCs in the predicted network and randomized networks. (B) Distribution of Jaccard indices for interacting proteins from the predicted and randomized networks based on the GO annotations.

Thus, the AS assessment also shows the relative reliability of the predicted PPI network. Taken together, we have clearly revealed the overall reliability of the predicted PPI network through direct and indirect approaches.

Topological features of clock-controlled proteins in the predicted PPI network

Circadian clocks are endogenous cellular timekeepers that control a great diversity of daily physiological and molecular rhythms in most eukaryotes and some prokaryotes.[54,55] The fungus *N. crassa* is a leading model organism for circadian clock studies, in which circadian oscillators guide ccgs to function in various aspects of the fungal life cycle. Previous studies have identified 20 ccgs[4] through targeted approaches, and 295 ccgs[5] have been recognized by the use of microarray technology. Of those identified CCGs, only 109 clock-controlled proteins have been found in the predicted PPI network.

We compared the topological properties of 109 clock-controlled and 2897 non-clock-controlled proteins in the predicted PPI network. As shown in Table 2, the betweenness centrality of the clock-controlled proteins was significantly higher than that of the non-clock-controlled proteins (average betweenness: 0.0261 *versus* 0.00481, $P = 8.31 \times 10^{-4}$). The clustering coefficient of clock-controlled proteins was significantly lower (average clustering coefficient: 0.471 *versus* 0.567, $P = 3.60 \times 10^{-3}$). However, no significant differences in the degree and closeness were detected between clock-controlled and non-clock-controlled proteins. Moreover, we found that

Table 2 Comparison of the average topological properties between clock-controlled and non-clock-controlled proteins of *N. crassa* in the predicted interactome

	Clock-controlled proteins	Non-clock-controlled proteins	*P*-value
Degree	25.9	18.1	0.240
Betweenness	0.0261	0.00481	8.31×10^{-4}
Closeness	0.413	0.391	0.0320
Clustering coefficient	0.471	0.567	3.60×10^{-3}

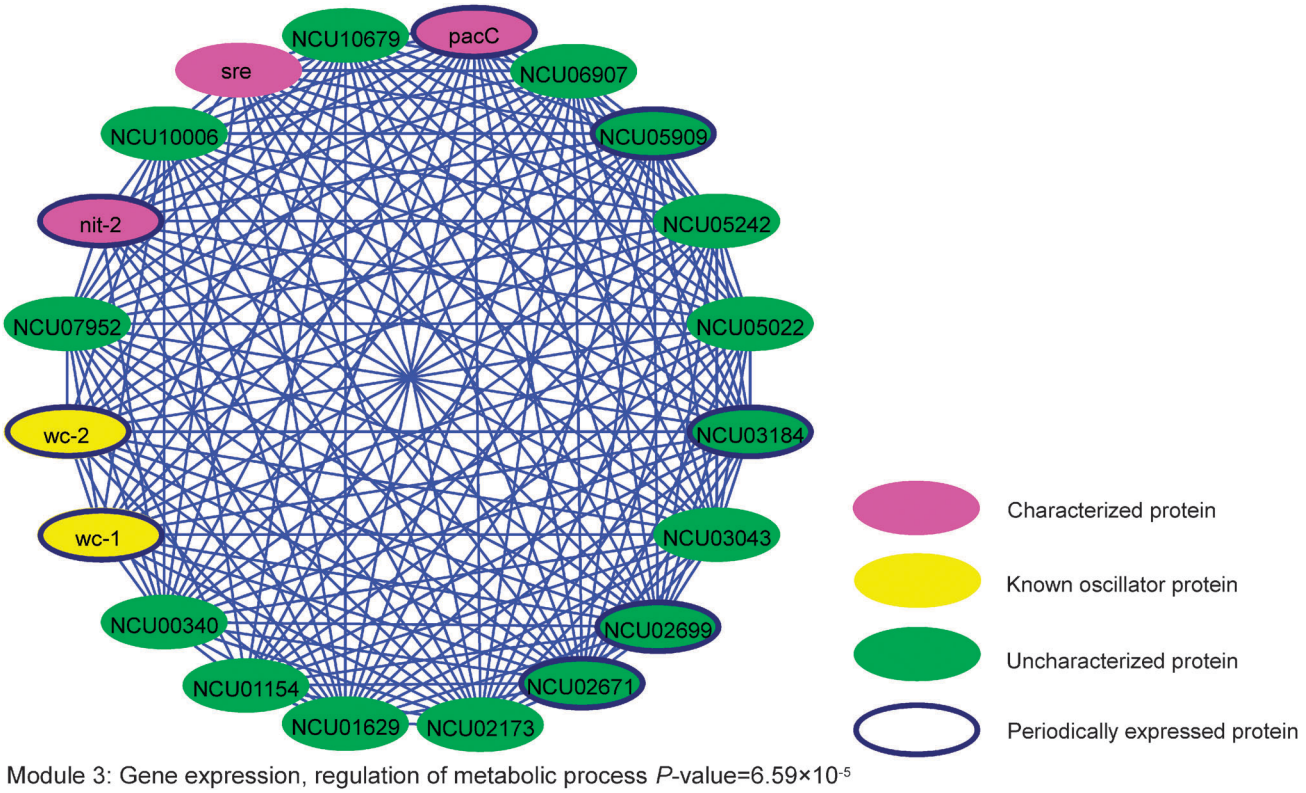

Fig. 4 Network visualization of a key module (Module 3) associated with circadian rhythmicity.

clock-controlled proteins might tend to act as bottleneck nodes. In the predicted network, nearly 40% of the CCGs were defined as bottlenecks, which we termed "bottleneck CCGs" ($P = 6.72 \times 10^{-5}$). Because bottlenecks are likely composed of regulatory proteins,[43] these clock-controlled proteins might be heavily involved in receiving signals from the oscillators and regulating biological pathways. For instance, the bottleneck CCGs NCU04883 and NCU06630 were determined to participate in or regulate cell wall synthesis during the asexual development.[56] In addition, the bottleneck *ccg-7* (NCU01528), which encodes glyceraldehyde 3-phosphate dehydrogenase (GAPDH), is a key enzyme in the pathway of glycolysis and gluconeogenesis.[4] Its yeast homolog, *tdh1*, has also been reported to be a cell-wall associated gene, which is able to respond to stress.[57] A reasonable hypothesis is that *ccg-7* can regulate carbohydrate metabolism based on oscillator signals from the environment. Interestingly, the three bottlenecks mentioned above have been shown to be essential genes.[58] In conclusion, the clock-controlled proteins might facilitate *N. crassa* to generate molecular responses to changes in ambient light and temperature, and play an important regulatory role in pathways that allow *N. crassa* to synchronize with the environment.

Identification of clock-related modules in the predicted network

Most PPI networks contain regions where the proteins are more highly connected to each other than to the rest of the whole network.[59] The densely-connected regions of a PPI network are referred to as clusters or modules. We utilized a graph theoretic clustering algorithm called MCODE[60] to identify ccg- or circadian oscillators-related modules (*i.e.*, clock-related modules) in the predicted network. Ranked by the generated MCODE scores, we list the first six clock-related modules in Figure S2, ESI.† The GO enrichments in these six modules include biological processes of various aspects of the fungal life cycle. The potential biological impact of the identified modules is exemplified in Module 3, where the GO enrichments are DNA binding and gene expression regulation (Fig. 4). Module 3 contains 20 proteins, including five eukaryotic unique GATA transcription factors. The GATA transcription factors have been reported to be responsible for regulating critical biological processes, such as nitrogen utilization and light regulation.[53] Interestingly, these five GATA transcription factors contain two components of the well-established *frq*/*wc*-based circadian oscillator (FWO)[54] [*i.e.*, *wc-1* (NCU02356 and *wc-2* (NCU00902)], although *frq* (NCU02265) does not appear in the predicted PPI network. Moreover, two WC-2 binding genes [*nit-2* (NCU09068) and *sre* (NCU07728)] were also presented in this module,[61] although the corresponding two PPIs were not predicted correctly. Therefore, Module 3 might be an important functional module associated with circadian rhythmicity. In the Pfam database, all remaining proteins of the module were found to contain a C2H2 zinc finger domain (Pfam entry: PF00096), suggesting that these proteins might be C2H2 zinc finger transcription factors. Recently, there has been convincing evidence of the existence of FRQ-less oscillators (FLOs).[62] It has been hypothesized that FLOs, which are fully independent of *frq*, can generate daily rhythm in collaboration with *wc-1* and *wc-2*. In general, the genes encoding oscillator

components should have periodical expression profiles. To detect the periodicity of gene expression profiles in Module 3, the corresponding microarray data were processed by ARSER.[63] In addition to *wc-1* and *wc-2*, six proteins in Module 3 were also identified to be expressed periodically. Therefore, these six proteins might be candidates for distinct oscillators. It is worth mentioning that *wc-1* and *wc-2* are also required for all known blue-light responses. In response to light, they function as transcription factors to regulate the expression of light-responsive genes.[61] Thus, these uncharacterized genes in Module 3 may also be predicted as light-responsive genes.

Database server

We also built a user-friendly web-interface *Neurospora crassa* Protein Interactome (NCPI) database (http://protein.cau.edu.cn/ncpi), which was developed on the open source web platform LAMP (Linux-Apache-MySQL-PHP) and was tested using Internet Explorer (Versions 7 and higher), Firefox, Chrome, Opera and Safari web browsers. Furthermore, the functionality of network visualization was implemented based on VisANT.[64] In addition to the predicted PPIs, eight manually curated protein interactions from public interactome databases and previous studies, which were not successful predicted, were also compiled into NCPI.

Conclusions

Using two well-recognized PPI prediction methods, we constructed a predicted PPI network of *N. crassa*. Experimental PPI information from previous studies or public PPI databases indicated that some predictions were actual interactions. Moreover, the overall reliability of the predicted PPI network was also clearly demonstrated by the results of three distinct assessment methods. Although the established network is far from complete and certainly contains false positives, we hope that the established network can provide new insights into the functions of *N. crassa* proteome at the system level. Based on the network topology analysis of CCGs in the predicted PPI network, we found that CCGs are inclined to act as bottlenecks. Therefore, they receive signals from oscillators and play an important regulatory role in biological pathways. As an example of the application of the established network, we also identified a key module related to circadian rhythmicity, which provides new candidates for circadian oscillators.

Acknowledgements

We thank the anonymous referees whose constructive comments were very helpful in improving the quality of this work. We also thank Chuan Wang, Rendong Yang and Yuan Zhou from the bioinformatics center of the China Agricultural University for helpful discussions. This research was supported by the National Natural Science Foundation (31070259) and the State Key Laboratory of Agrobiotechnology (2010SKLAB05-11).

References

1 J. E. Galagan, S. E. Calvo, K. A. Borkovich, E. U. Selker, N. D. Read, D. Jaffe, W. FitzHugh, L.-J. Ma, S. Smirnov, S. Purcell, B. Rehman, T. Elkins, R. Engels, S. Wang, C. B. Nielsen, J. Butler, M. Endrizzi, D. Qui, P. Ianakiev, D. Bell-Pedersen, M. A. Nelson, M. Werner-Washburne, C. P. Selitrennikoff, J. A. Kinsey, E. L. Braun, A. Zelter, U. Schulte, G. O. Kothe, G. Jedd, W. Mewes, C. Staben, E. Marcotte, D. Greenberg, A. Roy, K. Foley, J. Naylor, N. Stange-Thomann, R. Barrett, S. Gnerre, M. Kamal, M. Kamvysselis, E. Mauceli, C. Bielke, S. Rudd, D. Frishman, S. Krystofova, C. Rasmussen, R. L. Metzenberg, D. D. Perkins, S. Kroken, C. Cogoni, G. Macino, D. Catcheside, W. Li, R. J. Pratt, S. A. Osmani, C. P. C. DeSouza, L. Glass, M. J. Orbach, J. A. Berglund, R. Voelker, O. Yarden, M. Plamann, S. Seiler, J. Dunlap, A. Radford, R. Aramayo, D. O. Natvig, L. A. Alex, G. Mannhaupt, D. J. Ebbole, M. Freitag, I. Paulsen, M. S. Sachs, E. S. Lander, C. Nusbaum and B. Birren, *Nature*, 2003, **422**, 859–868.
2 J. C. Dunlap, *J. Biol. Chem.*, 2006, **281**, 28489–28493.
3 R. M. de Paula, M. W. Vitalini, R. H. Gomer and D. Bell-Pedersen, *Cold Spring Harbor Symp. Quant. Biol.*, 2007, **72**, 345–351.
4 M. W. Vitalini, R. M. de Paula, W. D. Park and D. Bell-Pedersen, *J. Biol. Rhythms*, 2006, **21**, 432–444.
5 W. Dong, X. Tang, Y. Yu, R. Nilsen, R. Kim, J. Griffith, J. Arnold and H. B. Schüttler, *PLoS One*, 2008, **3**, e3105.
6 A. Correa, Z. A. Lewis, A. V. Greene, I. J. March, R. H. Gomer and D. Bell-Pedersen, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 13597–13602.
7 A. Vazquez, A. Flammini, A. Maritan and A. Vespignani, *Nat. Biotechnol.*, 2003, **21**, 697–700.
8 MIPS Neurospora crassa Genome Database http://mips.helmholtz-muenchen.de/genre/proj/ncrassa/.
9 S. Fields and O. Song, *Nature*, 1989, **340**, 245–246.
10 M. Mann, R. C. Hendrickson and A. Pandey, *Annu. Rev. Biochem.*, 2001, **70**, 437–473.
11 G. MacBeath and S. L. Schreiber, *Science*, 2000, **289**, 1760–1763.
12 P. Uetz, L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields and J. M. Rothberg, *Nature*, 2000, **403**, 623–627.
13 S. Li, C. M. Armstrong, N. Bertin, H. Ge, S. Milstein, M. Boxem, P. O. Vidalain, J. D. Han, A. Chesneau, T. Hao, D. S. Goldberg, N. Li, M. Martinez, J. F. Rual, P. Lamesch, L. Xu, M. Tewari, S. L. Wong, L. V. Zhang, G. F. Berriz, L. Jacotot, P. Vaglio, J. Reboul, T. Hirozane-Kishikawa, Q. Li, H. W. Gabel, A. Elewa, B. Baumgartner, D. J. Rose, H. Yu, S. Bosak, R. Sequerra, A. Fraser, S. E. Mango, W. M. Saxton, S. Strome, S. Van Den Heuvel, F. Piano, J. Vandenhaute, C. Sardet, M. Gerstein, L. Doucette-Stamm, K. C. Gunsalus, J. W. Harper, M. E. Cusick, F. P. Roth, D. E. Hill and M. Vidal, *Science*, 2004, **303**, 540–543.
14 L. Giot, J. S. Bader, C. Brouwer, A. Chaudhuri, B. Kuang, Y. Li, Y. L. Hao, C. E. Ooi, B. Godwin, E. Vitols, G. Vijayadamodar, P. Pochart, H. Machineni, M. Welsh, Y. Kong, B. Zerhusen, R. Malcolm, Z. Varrone, A. Collis, M. Minto, S. Burgess, L. McDaniel, E. Stimpson, F. Spriggs, J. Williams, K. Neurath, N. Ioime, M. Agee, E. Voss, K. Furtak, R. Renzulli, N. Aanensen, S. Carrolla, E. Bickelhaupt, Y. Lazovatsky, A. DaSilva, J. Zhong, C. A. Stanyon, R. L. Finley, Jr., K. P. White, M. Braverman, T. Jarvie, S. Gold, M. Leach, J. Knight, R. A. Shimkets, M. P. McKenna, J. Chant and J. M. Rothberg, *Science*, 2003, **302**, 1727–1736.
15 J. C. Rain, L. Selig, H. De Reuse, V. Battaglia, C. Reverdy, S. Simon, G. Lenzen, F. Petel, J. Wojcik, V. Schachter, Y. Chemama, A. Labigne and P. Legrain, *Nature*, 2001, **409**, 211–215.
16 U. Stelzl, U. Worm, M. Lalowski, C. Haenig, F. H. Brembeck, H. Goehler, M. Stroedicke, M. Zenkner, A. Schoenherr, S. Koeppen, J. Timm, S. Mintzlaff, C. Abraham, N. Bock, S. Kietzmann, A. Goedde, E. Toksoz, A. Droege, S. Krobitsch, B. Korn, W. Birchmeier, H. Lehrach and E. E. Wanker, *Cell*, 2005, **122**, 957–968.

17 C. von Mering, R. Krause, B. Snel, M. Cornell, S. G. Oliver, S. Fields and P. Bork, *Nature*, 2002, **417**, 399–403.
18 A. J. Enright, I. Iliopoulos, N. C. Kyrpides and C. A. Ouzounis, *Nature*, 1999, **402**, 86–90.
19 H. Yu, N. M. Luscombe, H. X. Lu, X. Zhu, Y. Xia, J. D. Han, N. Bertin, S. Chung, M. Vidal and M. Gerstein, *Genome Res.*, 2004, **14**, 1107–1118.
20 P. Aloy, B. Bottcher, H. Ceulemans, C. Leutwein, C. Mellwig, S. Fischer, A.-C. Gavin, P. Bork, G. Superti-Furga, L. Serrano and R. B. Russell, *Science*, 2004, **303**, 2026–2029.
21 R. Jansen, H. Yu, D. Greenbaum, Y. Kluger, N. J. Krogan, S. Chung, A. Emili, M. Snyder, J. F. Greenblatt and M. Gerstein, *Science*, 2003, **302**, 449–453.
22 J. Chen, W. Hsu, M. L. Lee and S.-K. Ng, *Bioinformatics*, 2006, **22**, 1998–2004.
23 R. Overbeek, M. Fonstein, M. D'Souza, G. D. Pusch and N. Maltsev, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 2896–2901.
24 I. Ulitsky and R. Shamir, *BMC Syst. Biol.*, 2007, **1**, 8.
25 L. R. Matthews, P. Vaglio, J. Reboul, H. Ge, B. P. Davis, J. Garrels, S. Vincent and M. Vidal, *Genome Res.*, 2001, **11**, 2120–2126.
26 M. Deng, S. Mehta, F. Sun and T. Chen, *Genome Res.*, 2002, **12**, 1540–1548.
27 P.-Y. Chen, C. M. Deane and G. Reinert, *PLoS Comput. Biol.*, 2008, **4**, e1000118.
28 F. He, Y. Zhang, H. Chen, Z. Zhang and Y.-L. Peng, *BMC Genomics*, 2008, **9**, 519.
29 X.-M. Zhao, X.-W. Zhang, W.-H. Tang and L. Chen, *J. Proteome Res.*, 2009, **8**, 4714–4721.
30 T. Cui, L. Zhang, X. Wang and Z.-G. He, *BMC Genomics*, 2009, **10**, 118.
31 K. Mezhoud, H. Sghaier and I. Barkallah, *Biol. Direct*, 2009, **4**, 12.
32 L. Salwinski, C. S. Miller, A. J. Smith, F. K. Pettit, J. U. Bowie and D. Eisenberg, *Nucleic Acids Res.*, 2004, **32**, D449–451.
33 A. Ceol, A. Chatr Aryamontri, L. Licata, D. Peluso, L. Briganti, L. Perfetto, L. Castagnoli and G. Cesareni, *Nucleic Acids Res.*, 2010, **38**, D532–539.
34 T. S. Keshava Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D. S. Somanathan, A. Sebastian, S. Rani, S. Ray, C. J. Harrys Kishore, S. Kanth, M. Ahmed, M. K. Kashyap, R. Mohmood, Y. L. Ramachandra, V. Krishna, B. A. Rahiman, S. Mohan, P. Ranganathan, S. Ramabadran, R. Chaerkady and A. Pandey, *Nucleic Acids Res.*, 2009, **37**, D767–772.
35 B. Aranda, P. Achuthan, Y. Alam-Faruque, I. Armean, A. Bridge, C. Derow, M. Feuermann, A. T. Ghanbarian, S. Kerrien, J. Khadake, J. Kerssemakers, C. Leroy, M. Menden, M. Michaut, L. Montecchi-Palazzi, S. N. Neuhauser, S. Orchard, V. Perreau, B. Roechert, K. van Eijk and H. Hermjakob, *Nucleic Acids Res.*, 2010, **38**, D525–531.
36 R. D. Finn, M. Marshall and A. Bateman, *Bioinformatics*, 2005, **21**, 410–412.
37 A. Stein, A. Panjkovich and P. Aloy, *Nucleic Acids Res.*, 2009, **37**, D300–304.
38 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235–242.
39 D. R. Rhodes, S. A. Tomlins, S. Varambally, V. Mahavisno, T. Barrette, S. Kalyana-Sundaram, D. Ghosh, A. Pandey and A. M. Chinnaiyan, *Nat. Biotechnol.*, 2005, **23**, 951–959.
40 F. P. Davis, D. T. Barkan, N. Eswar, J. H. McKerrow and A. Sali, *Protein Sci.*, 2007, **16**, 2585–2596.
41 Y. Assenov, F. Ramirez, S. E. Schelhorn, T. Lengauer and M. Albrecht, *Bioinformatics*, 2008, **24**, 282–284.
42 P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, *Genome Res.*, 2003, **13**, 2498–2504.
43 H. Yu, P. M. Kim, E. Sprecher, V. Trifonov and M. Gerstein, *PLoS Comput. Biol.*, 2007, **3**, e59.
44 E. Sprinzak, S. Sattath and H. Margalit, *J. Mol. Biol.*, 2003, **327**, 919–923.
45 C. M. Deane, L. Salwinski, I. Xenarios and D. Eisenberg, *Mol. Cell. Proteomics*, 2002, **1**, 349–356.
46 H. Li and S. Liang, *PLoS One*, 2009, **4**, e6410.
47 P. Horton, K.-J. Park, T. Obayashi, N. Fujita, H. Harada, C. J. Adams-Collier and K. Nakai, *Nucleic Acids Res.*, 2007, **35**, W585–587.
48 J. C. Dunlap, K. A. Borkovich, M. R. Henn, G. E. Turner, M. S. Sachs, N. L. Glass, K. McCluskey, M. Plamann, J. E. Galagan, B. W. Birren, R. L. Weiss, J. P. Townsend, J. J. Loros, M. A. Nelson, R. Lambreghts, H. V. Colot, G. Park, P. Collopy, C. Ringelberg, C. Crew, L. Litvinkova, D. DeCaprio, H. M. Hood, S. Curilla, M. Shi, M. Crawford, M. Koerhsen, P. Montgomery, L. Larson, M. Pearson, T. Kasuga, C. Tian, M. Basturkmen, L. Altamirano and J. Xu, *Adv. Genet.*, 2007, **57**, 49–96.
49 E. M. Zdobnov and R. Apweiler, *Bioinformatics*, 2001, **17**, 847–848.
50 A. Ruepp, A. Zollner, D. Maier, K. Albermann, J. Hani, M. Mokrejs, I. Tetko, U. Guldener, G. Mannhaupt, M. Munsterkotter and H. W. Mewes, *Nucleic Acids Res.*, 2004, **32**, 5539–5545.
51 G. Mannhaupt, C. Montrone, D. Haase, H. W. Mewes, V. Aign, J. D. Hoheisel, B. Fartmann, G. Nyakatura, F. Kempken, J. Maier and U. Schulte, *Nucleic Acids Res.*, 2003, **31**, 1944–1954.
52 J.-D. J. Han, D. Dupuy, N. Bertin, M. E. Cusick and M. Vidal, *Nat. Biotechnol.*, 2005, **23**, 839–844.
53 K. A. Borkovich, L. A. Alex, O. Yarden, M. Freitag, G. E. Turner, N. D. Read, S. Seiler, D. Bell-Pedersen, J. Paietta, N. Plesofsky, M. Plamann, M. Goodrich-Tanrikulu, U. Schulte, G. Mannhaupt, F. E. Nargang, A. Radford, C. Selitrennikoff, J. E. Galagan, J. C. Dunlap, J. J. Loros, D. Catcheside, H. Inoue, R. Aramayo, M. Polymenis, E. U. Selker, M. S. Sachs, G. A. Marzluf, I. Paulsen, R. Davis, D. J. Ebbole, A. Zelter, E. R. Kalkman, R. O'Rourke, F. Bowring, J. Yeadon, C. Ishii, K. Suzuki, W. Sakai and R. Pratt, *Microbiol. Mol. Biol. Rev.*, 2004, **68**, 1–108.
54 Y. Liu and D. Bell-Pedersen, *Eukaryotic Cell*, 2006, **5**, 1184–1193.
55 S. Krystofova and K. A. Borkovich, *Eukaryotic Cell*, 2005, **4**, 365–378.
56 T. Zeke, E. Kókai, B. Szöor, E. Yatzkan, O. Yarden, K. Szirák, Z. Fehér, P. Bagossi, P. Gergely and V. Dombrádi, *Comp. Biochem. Physiol., Part B: Biochem. Mol. Biol.*, 2003, **134**, 161–170.
57 M. L. Delgado, J. E. O'Connor, I. Azorin, J. Renau-Piqueras, M. L. Gil and D. Gozalbo, *Microbiology*, 2001, **147**, 411–417.
58 H. V. Colot, G. Park, G. E. Turner, C. Ringelberg, C. M. Crew, L. Litvinkova, R. L. Weiss, K. A. Borkovich and J. C. Dunlap, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 10352–10357.
59 G. Palla, I. Derenyi, I. Farkas and T. Vicsek, *Nature*, 2005, **435**, 814–818.
60 G. Bader and C. Hogue, *BMC Bioinformatics*, 2003, **4**, 2.
61 K. M. Smith, G. Sancar, R. Dekhang, C. M. Sullivan, S. Li, A. G. Tag, C. Sancar, E. L. Bredeweg, H. D. Priest, R. F. McCormick, T. L. Thomas, J. C. Carrington, J. E. Stajich, D. Bell-Pedersen, M. Brunner and M. Freitag, *Eukaryotic Cell*, 2010, **9**, 1549–1556.
62 J. C. Dunlap and J. J. Loros, *Curr. Opin. Microbiol.*, 2006, **9**, 579–587.
63 R. Yang and Z. Su, *Bioinformatics*, 2010, **26**, i168–i174.
64 Z. Hu, J.-H. Hung, Y. Wang, Y.-C. Chang, C.-L. Huang, M. Huyck and C. DeLisi, *Nucleic Acids Res.*, 2009, **37**, W115–121.
65 S. Honda and E. U. Selker, *Mol. Cell. Biol.*, 2008, **28**, 6044–6055.
66 B. Feng and G. A. Marzluf, *Mol. Cell. Biol.*, 1998, **18**, 3983–3990.
67 S. Maerz, A. Dettmann, C. Ziv, Y. Liu, O. Valerius, O. Yarden and S. Seiler, *Mol. Microbiol.*, 2009, **74**, 707–723.
68 S. Maerz, C. Ziv, N. Vogt, K. Helmstaedt, N. Cohen, R. Gorovits, O. Yarden and S. Seiler, *Genetics*, 2008, **179**, 1313–1325.
69 H. Xu, J. Wang, Q. Hu, Y. Quan, H. Chen, Y. Cao, C. Li, Y. Wang and Q. He, *PLoS Genet.*, 2010, **6**, e1001132.

