

OPEN

# Genome-wide inference of the *Camponotus floridanus* protein-protein interaction network using homologous mapping and interacting domain profile pairs

Shishir K. Gupta<sup>1,2,4</sup>, Mugdha Srivastava<sup>1,4</sup>, Özge Osmanoglu<sup>1</sup> & Thomas Dandekar<sup>1,3\*</sup>

Apart from some model organisms, the interactome of most organisms is largely unidentified. High-throughput experimental techniques to determine protein-protein interactions (PPIs) are resource intensive and highly susceptible to noise. Computational methods of PPI determination can accelerate biological discovery by identifying the most promising interacting pairs of proteins and by assessing the reliability of identified PPIs. Here we present a first in-depth study describing a global view of the ant *Camponotus floridanus* interactome. Although several ant genomes have been sequenced in the last eight years, studies exploring and investigating PPIs in ants are lacking. Our study attempts to fill this gap and the presented interactome will also serve as a template for determining PPIs in other ants in future. Our *C. floridanus* interactome covers 51,866 non-redundant PPIs among 6,274 proteins, including 20,544 interactions supported by domain-domain interactions (DDIs), 13,640 interactions supported by DDIs and subcellular localization, and 10,834 high confidence interactions mediated by 3,289 proteins. These interactions involve and cover 30.6% of the entire *C. floridanus* proteome.

In terms of biodiversity and biomass, insects are the most successful animals on earth. They provide major beneficial impacts such as pollination, food source and soil improvement. On the contrary, some insects damage crops and spread deadly diseases as vectors. One of the harmful pests is the ant *Camponotus floridanus* which is widely distributed throughout Florida and the neighboring states<sup>1</sup>. They hollow the wood softened by moisture and damage the structural integrity of houses by affecting the wood work with their strong mandibles. Besides this, this ant species serves as a good model system to understand host-endosymbiont relationships regarding its bacterial endosymbiont *Blochmannia*<sup>2</sup>.

The complete genome sequences of *C. floridanus* has revealed the composition of proteins, based mainly on theoretical predictions utilizing their corresponding DNA sequence. We analyzed the transcriptome level evidence of protein existence and re-annotated the *C. floridanus* gene models and proteins<sup>3</sup>. How these proteins interact is not yet explored, in part due to limited genetic studies in this organism, high cost, time-consuming and labor-intensive nature of experimental methods. Protein interactions are at the core of nearly all biological processes, and knowledge about protein-protein interactions (PPIs) is vital for understanding biological systems. Despite advances in high-throughput experimental methods for detecting PPIs, the interaction networks for even the well-studied experimental model organisms are far from complete<sup>4</sup>. Nevertheless, high throughput assays typically include false positives PPIs<sup>5</sup> which stipulate an enduring need for efficient computational methods to complement existing experimental approaches. In this context, combining the interolog method<sup>6</sup> with adding domain information<sup>7</sup>, gene ontology (GO) annotation<sup>8</sup> and cellular localization<sup>9,10</sup> yields a graphical representation of the interaction networks, a robust and well-established approach to provide an intuitive vision and useful insights to help and analyze complex relations therein, as indicated by several previous studies in the reconstruction and

<sup>1</sup>Functional Genomics and Systems Biology Group, Department of Bioinformatics, Biocenter, Am Hubland, D-97074, Würzburg, Germany. <sup>2</sup>Department of Microbiology, Biocenter, Am Hubland, D-97074, Würzburg, Germany. <sup>3</sup>EMBL Heidelberg, BioComputing Unit, Meyerhofstraße 1, 69117, Heidelberg, Germany. <sup>4</sup>These authors contributed equally: Shishir K. Gupta and Mugdha Srivastava. \*email: [dandekar@biozentrum.uni-wuerzburg.de](mailto:dandekar@biozentrum.uni-wuerzburg.de)

understanding of PPIs in various organisms<sup>11–14</sup>. Here we used domain information, subcellular localization and isoform information to filter the preliminary global PPI network of *C. floridanus* reconstructed on stringent interolog based criteria. We focus on interactions predicted with high confidence to reduce noise. This conservative approach rejects 79.1% of the preliminary predicted interactions. We then explored the topologically important and evolutionary conserved proteins by analyzing the reconstructed interactome regarding cellular functions.

## Results and Discussion

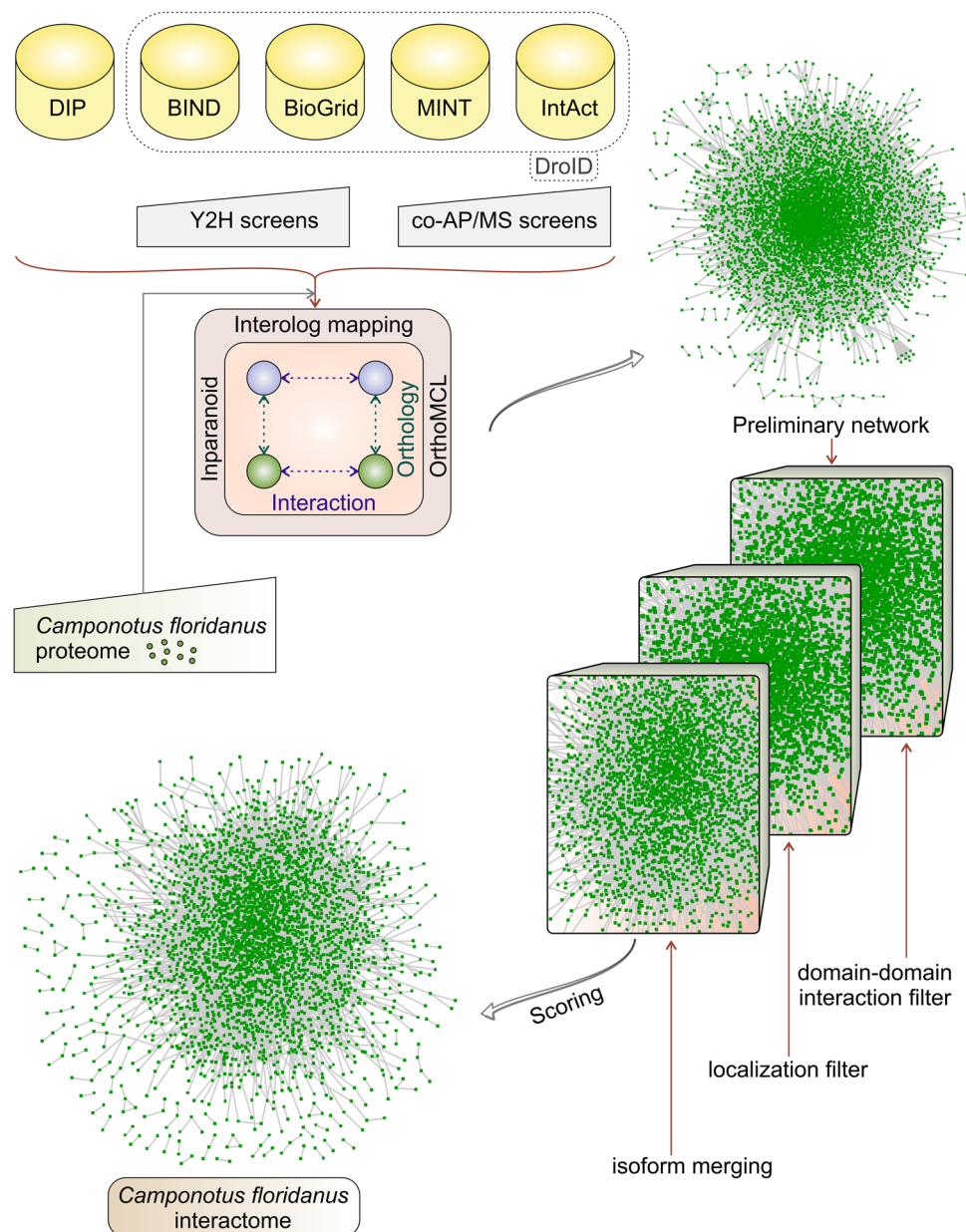
**Generating the interactome of ant *C. floridanus*.** PPIs are typically mediated by interactions between domains that are often evolutionary conserved across species<sup>15</sup> and form stable interactions<sup>16,17</sup>. PPI (protein-protein interaction) maps from experiments on *D. melanogaster* were collected and augmented by PPI data from the DIP database (Database of Interacting Proteins). This provided a basis for interaction predictions according to interologs from *C. floridanus*: conserved proteins compared to *Drosophila* should also be conserved in their interactions<sup>6,18</sup> (see Materials and methods for details).

Optimally, for such predictions several methods are combined<sup>19</sup> (Fig. 1). We combined the orthology prediction methods InParanoid<sup>20</sup> and OrthoMCL<sup>21</sup>. This did yield a first estimate of the *C. floridanus* interactome with 6274 nodes and 51866 edges<sup>22</sup>. However, the preliminary ant PPI network could have several false positive interactions acquired from the interologs of template data as shown previously in similar other studies<sup>5,10,23,24</sup>, including transfer to curated databases<sup>25</sup>. To reduce false predictions, we counter-checked all our data by domain-domain interactions (DDI). DDI are often used as an approach independent from sequence homology-based methods to predict protein-protein interaction networks and thus strongly reduce the number of false positives<sup>7,26,27</sup>. Generally, some of the PPIs are achieved via interactions between short motifs that are often transient interactions<sup>28</sup>. On the other hand, conserved interactions are mediated by conserved interaction domains across species<sup>6</sup>. Moreover, many signals and processes in the cell rely on conserved interacting protein domains<sup>16,29</sup>. There were 51866 conserved proteins (interologs) and 20544 ant protein-protein interactions that also were associated with DDI pairs, yielding a curated *C. floridanus* interactome with 4589 nodes and 20544 edges. For final curation of the interactome we used the subcellular localization of ant proteins: interacting proteins have to share the same subcellular localization (summarized in Table 1), predicted interactions between proteins not in the same location were removed. This led to a consolidated ant interactome consisting of 3914 nodes and 13640 edges. The highest proportion of interactions were identified in the cytoplasm followed by nucleus and plasma membrane respectively. A closer inspection of the interactions that were enriched across subcellular compartments (such as Golgi apparatus-cytoplasm) showed that in numerous cases at least one of the interacting proteins was alternatively localized to a compartment other than its major site of localization and thus the interacting proteins did indeed share a common compartment. For instance, in 482 interaction pairs (Table 1) at least one protein showed both the Golgi apparatus localization and cytoplasmic localization. It should be noted that these interaction partners are multiple localized proteins and may also appear in other cellular compartments. This is not an uncommon situation, as > 50% of proteins of our final interactome network annotated with predicted subcellular localization information are, in fact, localized at two or more compartments.

As a final step of network reduction, isoforms of proteins are shown as a single node. These steps of successive filtering ultimately reduce the complexity of the network and increase the confidence of the *C. floridanus* interactome. Figure 1 summarizes the *C. floridanus* protein-protein interaction databases, our workflow, pruning steps and resulting ant network. It consists of 3289 nodes and 10834 edges (more details in<sup>22</sup>). The complete four networks are provided in the Datasheets 1–4 in Supplementary Material. We also identified several novel interactions predicted to be present in *C. floridanus*. For instance, an interaction was observed between S-phase kinase-associated protein 1 (SkpA, Cflo\_N\_g10272) and immune receptor peptidoglycan-recognition protein LC (PGRP-LC, Cflo\_N\_g10272). As an important component of ubiquitin-proteasome pathway SkpA is involved in Immune Deficiency (IMD) pathway regulation in *D. melanogaster*<sup>30</sup>. Since PGRP-LC is also a regulator of the ant IMD pathway<sup>3</sup>, the interaction we identified suggests that SkpA can modulate the IMD pathway by the interaction with PGRP-LC. Not only the interaction between protein complexes such as laminin subunit beta-1 (Cflo\_N\_g14102) and laminin subunit gamma-1 (Cflo\_N\_g9869) but also the interaction between Cflo\_N\_g14102 and C-type lectin precursor (Cflo\_N\_g765) was resolved (see Datasheets 3 in Supplementary Material for all the interactions).

To further supplement the proposed ant interactome, we performed a topology-based scoring of the network. The method CAPPIC<sup>31</sup> used the intrinsic modularity of PPI network for assessing the confidence of individual interactions. 88.5% of the total interactions are high confidence (Fig. 2) while 9.65% were assigned to medium confidence and 1.8% to low confidence.

We applied the Mann-Whitney test to compare the average confidence scores of all four PPI networks and observed significant increase of confidence score for the first three steps from the preliminary network through DDI mediated filtering and localization-based filtering (Supplementary Fig. 1). The mean confidence score of the final interactome, after the isoform merging, did not change much. This is because the merging of this last step also eliminated some high confidence PPIs mediated by the isoforms. Nevertheless, the comparison of the proportions of high-confidence PPIs in the preliminary interactome and the final ant interactome indicates that it has a significantly increased number of high confidence interactions (in the preliminary network these are 78%, in the final 89%; Fisher's exact test p-value < 2.2e-16). Note that the applied filtering steps also eliminated most of the low confidence PPIs (see low confidence zone in Supplementary Fig. 1). To further confirm the elimination after successive filtering steps, we compared the low confidence PPIs proportions in all four interactomes in a pairwise way with Fisher's exact test and show a significant decrease in the number of the low confidence PPIs between the preliminary, DDI-filtered and localization-filtered interactomes (in preliminary 4.5%, in DDI-filtered 2.2%, in localization-filtered 1.6% with maximum p-value < 3.4e-05). These analyses clearly demonstrate the improvement of network quality after filtering steps.



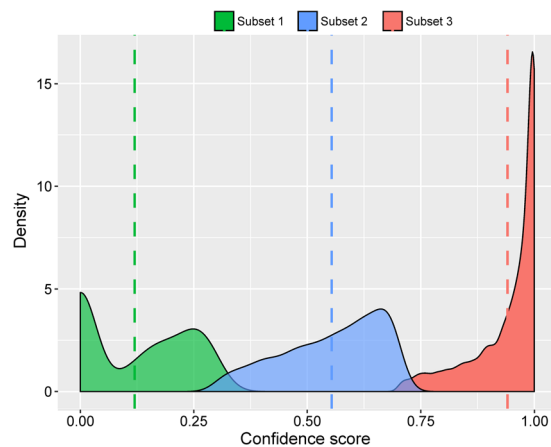
**Figure 1.** Pipeline for reconstruction of *C. floridanus* interactome. Top: Five databases, yeast two hybrid screens and coAP/MS screens and the *C. floridanus* interactome provide data for calculating conserved proteins (interologs) and protein-protein interactions. The preliminary protein-protein interaction network (top, right) is then filtered (right) by demanding in addition support by domain-domain interactions and shared subcellular localizations (right, bottom). It is then simplified merging similar isoforms (bottom, left).

**Network analysis of *C. floridanus* interactome and accuracy assessment.** The resulting PPI summarizes the whole network and reveals central connecting nodes. The final high confidence ant interactome showed a clustering coefficient of 0.094 with a mean shortest path length of 4.359, network diameter 14 and an average degree of 6.970. As a typical biological network<sup>32–34</sup> it shows small-world connectivity and scale-free topology.

We further tested whether the proposed interactome aligns with the properties of a real biological network. To assess this, we derived three independent datasets and compared their topological properties with the proposed network. The average z-statistic value (Datashet 5 in Supplementary Material) clearly indicates comparatively less variation of the ant interactome from the ‘Barabási-Albert scale free model’ (z-statistic = 23.06, –5.28) in terms of clustering coefficient and mean shortest path. However, the differences were high while comparing that of with ‘Watts–Strogatz small world graph model’ (z-statistic = –30.95, –58.49) and ‘Erdős-Rényi flat-random’ network model (z-statistic = 171.03, –52.72). Scale-free networks have been often observed in biological systems such as PPI and gene regulatory networks<sup>35</sup>, therefore the bias towards such a network is an indicator of the equality of the reconstructed ant interactome. To test another factor, the degree distribution of the ant interactome was

Cytoplasm (2176)	6871										
Golgi apparatus (243)	482	720									
Peroxisome (21)	11	3	4								
Other (1) †	1	0	0	1							
Nucleus (1866)	4686	227	3	0	6596						
Cytoskeleton (427)	1580	134	0	1	926	387					
Lysosome (50)	106	17	0	0	53	26	11				
Mitochondrion (319)	443	39	8	0	297	61	12	211			
ER (270)	368	59	0	1	179	79	19	24	155		
Extracellular (123)	120	16	4	0	87	36	1	19	22	51	
Plasma membrane (1522)	2365	383	10	1	1187	507	79	245	416	83	2636
	Cytoplasm (2176)	Golgi apparatus (243)	Peroxisome (21)	Other (1) †	Nucleus (1866)	Cytoskeleton (427)	Lysosome (50)	Mitochondrion (319)	ER (270)	Extracellular (123)	Plasma membrane (1522)

**Table 1.** Numbers of PPIs, by subcellular localization of interacting proteins in localization supported network. The numbers in the parentheses indicate the total number of proteins in the interactome that are present in each subcellular compartment (based on localization prediction, including multiple localized proteins). ER, endoplasmic reticulum. †The number given here in parentheses is the number of proteins that do not have subcellular localization in any of the other compartments in the table.



**Figure 2.** Density plot of the confidence scores for interactions in the *C. floridanus* interactome. How sure are the predicted protein-protein interactions? The distribution of the different confidence levels were computed with CAPPIC<sup>31</sup>. Score distributions were separated into low, medium and high confidence category and the density for each category was plotted. In the three subsets scores range between 0 and 0.3 for subset 1 (green, low confidence), 0.3 and 0.7 for subset 2 (blue, medium confidence) and 0.7 and 1 for subset 3 (red, high confidence). Stippled coloured lines indicate median values for the different categories.

much closer to Watts–Strogatz model ( $z$ -statistic = 0.49), although the differences with Barabási–Albert model was not too high ( $z$ -statistic = 2.45). The nodes in the network obey a power-law distribution indicating a typical, biological small-world and scale-free network.

**Gene ontology (GO) enrichment analysis.** The molecular function GO term over-representation analysis indicates enriched protein functions in the ant networks (FDR < 0.05; Table 2 and Datasheet 6 in Supplementary Material). Over-represented functional categories include the term ‘binding’ as to be expected from the PPI construction and a validation criterion. Out of 2804 proteins annotated as GO term GO:0005488 ‘binding’ in *C. floridanus* proteome, 46.11% proteins are present in the final interactome. In total, 64 binding-related GO terms were identified constituting 34.97% of all over-represented GO terms. We only found the under-representation of two GO terms: GO:0003964 ‘RNA-directed DNA polymerase activity’ and GO:0034061 ‘DNA polymerase activity’. This indicates during the filtering we did not lose most of the functional proteins that are involved in molecular binding.

We further compared the semantic similarity scores of the interacting pairs with the random networks of non-interacting proteins. We first assigned the level-4 GO annotations (for molecular function) to all the proteins coded by the ant genome using Blast2GO<sup>36</sup>. Next, we used the GOGO algorithm<sup>37</sup> to measure the semantic similarity scores of the high confidence interacting pairs in the proposed ant interactome. We further generated 30

GO ID	GO molecular function term	FDR
GO:0005488	Binding	3.27E-130
GO:0003824	Catalytic activity	8.17E-92
GO:0097159	Organic cyclic compound binding	9.44E-88
GO:1901363	Heterocyclic compound binding	9.44E-88
GO:0036094	Small molecule binding	3.59E-80
GO:1901265	Nucleoside phosphate binding	7.83E-78
GO:0000166	Nucleotide binding	7.83E-78
GO:0043168	Anion binding	2.45E-66
GO:0017076	Purine nucleotide binding	5.15E-59
GO:0032553	Ribonucleotide binding	7.09E-59
GO:0032555	Purine ribonucleotide binding	1.76E-58
GO:0035639	Purine ribonucleoside triphosphate binding	2.85E-58
GO:0097367	Carbohydrate derivative binding	2.68E-55
GO:0043167	Ion binding	3.31E-52
GO:0016817	Hydrolase activity, acting on acid anhydrides	2.18E-39
GO:0005515	Protein binding	2.18E-39
GO:0016818	Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	9.29E-39
GO:0016462	Pyrophosphatase activity	6.95E-38
GO:0017111	Nucleoside-triphosphatase activity	3.90E-37
GO:0030554	Adenyl nucleotide binding	3.90E-37

**Table 2.** Top 20 over-represented GO molecular function terms in the ant-interactome<sup>a</sup>. <sup>a</sup>See Datasheet 6 in Supplementary material for the complete list of over-represented GO molecular functions terms in ant-interactome.

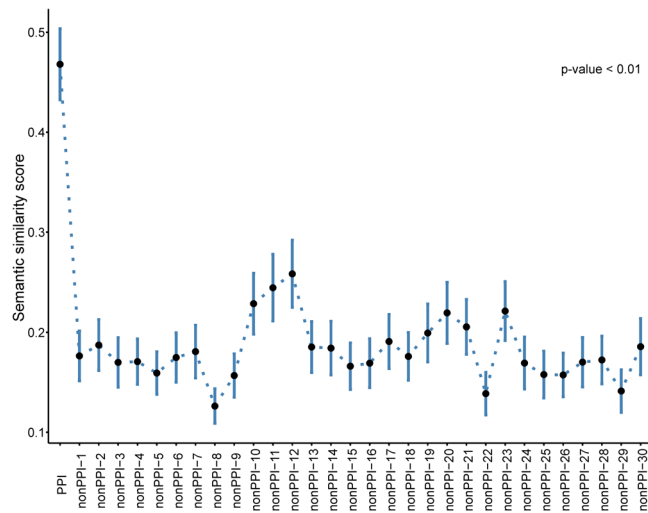
random networks each with 100 random interactions among the proteins that were assigned to level-4 molecular function GO annotations using a custom-made Perl script which can be accessed from the GitHub repository ([https://github.com/ShishirGupta-Wu/ant\\_ppi](https://github.com/ShishirGupta-Wu/ant_ppi)). We made sure the random networks did not contain any proteins pairs apparent in the preliminary interactome. Using the GOGO algorithm<sup>37</sup> semantic similarity scores were also assigned to the random networks (non-PPIs) and these scores were further compared with the interacting proteins in a pairwise way using the Mann-Whitney U test. We observed that the interacting protein set had not only the highest average score of 0.47, this was also well separated and significantly higher than the average score in all the 30 non-PPI sets (Fig. 3). This comparison demonstrates the interactions in our calculated ant interactome are functionally relevant and clearly different from random networks.

**C. floridanus interactome protein conservation compared with seven organisms.** Proteins that perform essential functions are expected to be evolutionary conserved. We further investigated the evolutionary conservation of ant interactome proteins. Higher degree proteins are generally evolutionary better conserved<sup>38</sup>, some caveats are discussed in<sup>39,40</sup>. To analyze this, node degree and the fraction of proteins present in the ant interactome that are conserved in different model organisms were compared. It turns out that in general the interactions are conserved and supported by most species tested and not just by one (Fig. 4). For exact quantification we did not check the possible restricted conservation of the binary ant PPIs, but more general the conservation of proteins that are present in the ant interactome and have orthologs in seven other species. For instance, in the ant interactome there are 535 proteins of degree 2. Out of these 535 proteins 451 have an ortholog in *Anopheles*, 209 have an ortholog in *Arabidopsis*, 298 have an ortholog in *C. elegans*, 404 have an ortholog in mouse, 82 have an ortholog in plasmodium, 151 have an ortholog in yeast, and 402 have a human ortholog.

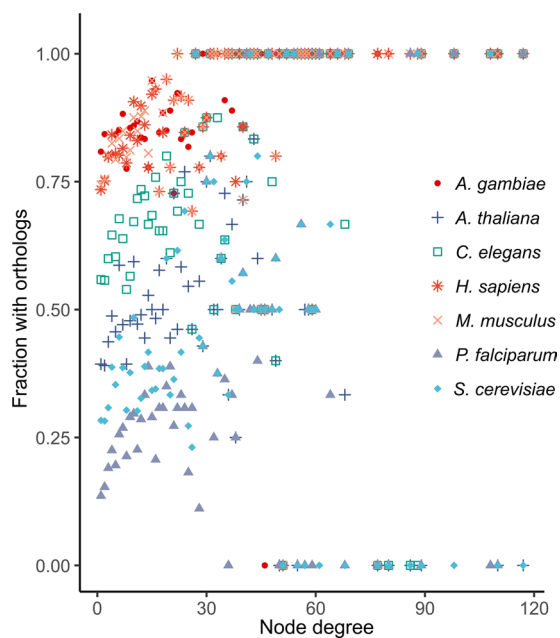
There was a positive correlation between degree and conservation in the evolutionary closest analyzed species *A. gambiae* (Spearman's rank  $r = 0.62$ ,  $p\text{-value} = 3.5\text{e-}09$ ). Similar correlations are observed between ant and human ( $r = 0.60$ ), and mouse ( $r = 0.51$ ). Between ant and worm the correlation was weak ( $r = 0.33$ ), while no significant correlation is observed between ant and *A. thaliana*, *P. falciparum*, and yeast. An ortholog table is provided in Datasheet 7 in Supplementary Material.

**Overall conservation and infection induced hubs and bottlenecks in the ant interactome.** We also evaluated the overall conservation of all the ant proteins with the other seven model organisms and compared the relatedness of the ant interactome proteins using the chi-square test. The analysis indicated the relatedness of corresponding proportions with  $p\text{-value} < 0.05$  in each case. The differences in the number of orthologs can be clearly visualized (Fig. 5a) in case of ant comparison with protozoan parasite, yeast and plant.

Due to the large phylogenetic distance to these three organisms there are less orthologs but these are well conserved (chi-square test).

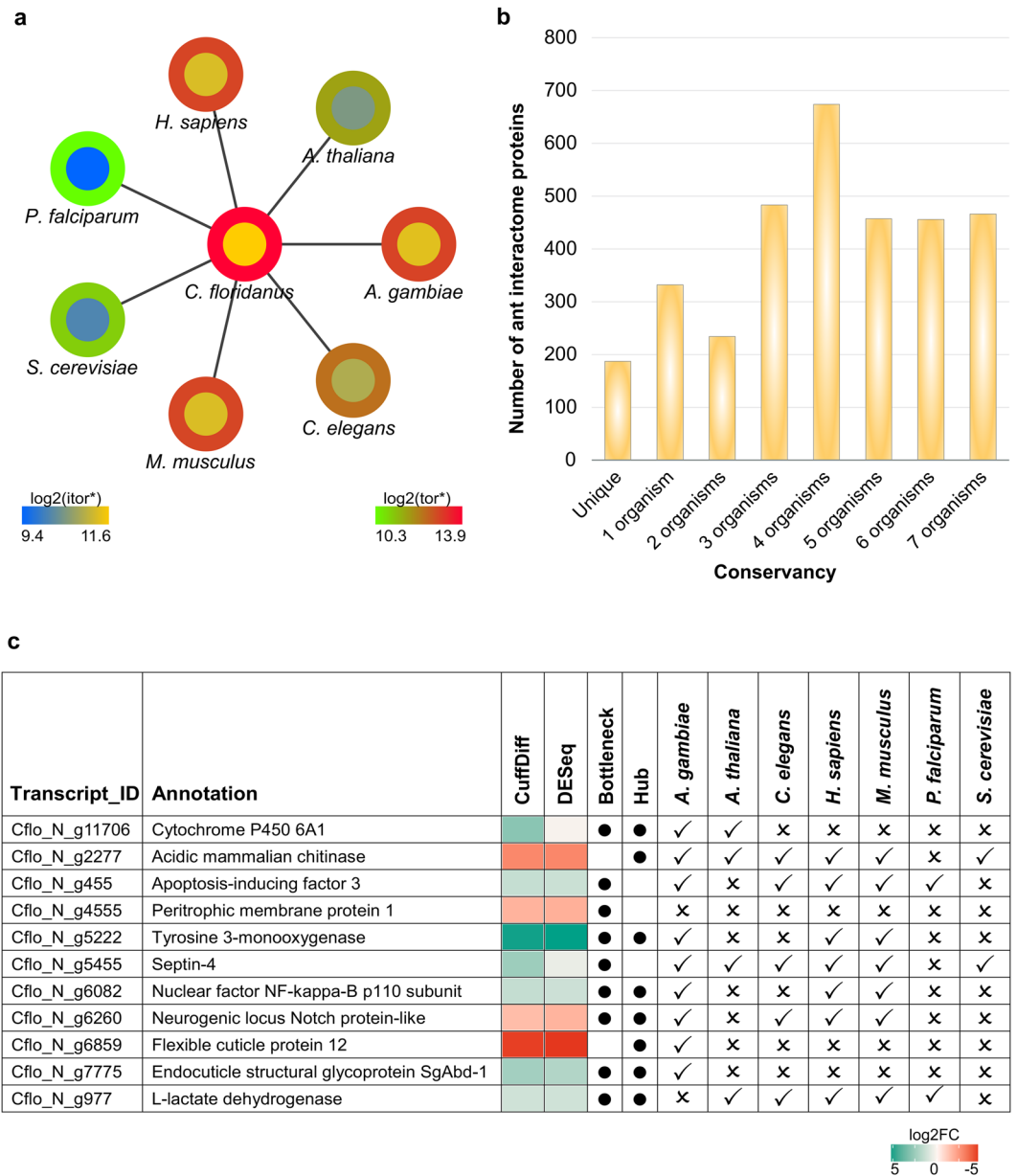


**Figure 3.** Pairwise Gene Ontology (GO) similarity of the ant interactome compared against non-interacting proteins. The high average similarity score of the ant interactome (PPI, left) stands out against the low similarity scores of the non-interacting PPIs. Semantic similarity score between interacting (PPI) and non-interacting (nonPPI) protein pairs were compared in a pairwise fashion using the Mann-Whitney U test. The Average scores for semantic similarity in molecular function level-4 GO annotations of interacting proteins and 30 random networks of non-interacting proteins are shown.



**Figure 4.** Conservation of interacting proteins by degree in different model species. The conservation level of ant interactome proteins with varying degrees are shown in each analyzed species and in general the interactions are well confirmed by several species (symbols listed on the right, middle). Higher connected proteins (right) become rare and so if present in the compared model organism, they are fully confirmed (1.00, top) or nothing is found in some other model species (0.00, bottom). Each protein in the *C. floridanus* interactome was examined for orthologous proteins in the seven organisms, and binned according to degree. The proportion of each bin with orthologous proteins is shown. A trendline is not shown on the graph since the data is analyzed with Spearman's rank correlation and a trendline could be misleading.

The remaining set of the other four organisms including insect, human, mouse and worm together consists of contains higher number of orthologs in comparison to the ant proteins (Fig. 5b and Datasheet 7 in Supplementary Material). 187 proteins of the ant interactome are ant-specific in this comparison: they do not have orthologs in any of the analysed organisms (Fig. 5b). The analysis of central topological properties of a PPI network helps to identify key multifunctional components of the network<sup>41</sup>. Infection induced proteins of *C. floridanus* are conserved in related organisms including key interactions. The degree of the node<sup>42</sup> and the betweenness centrality<sup>43</sup>



**Figure 5.** Conservation of *C. floricidanus* proteins. (a) Network showing the similarity of *C. floricidanus* proteins (central node) with other organisms proteins (peripheral nodes) quantitatively by color codes. Color gradients represent the number of proteins in each species that have orthologs in either *C. floricidanus* interactome (itor, inner circle) or in the whole *C. floricidanus* proteome (tor, outer circle). A similar color to the ant node shows close orthology relation while nodes of more distant species have different colors. A correlation between the colors of two circles is expected to show that the interactome successfully represent the orthology relationships between ant and other species. For instance, if both circles of a species show a similar color to ant, it depicts that the close orthology to ant can be also observed in the interactome. Pearson’s Chi-squared test with Yates’ continuity correction showed relationship ( $p$ -value  $< 2.2e-16$ ) between the total ant proteins orthologs and interactome proteins orthologs for each corresponding organisms. (b) Conservation of ant interactome proteins in set of any analyzed organisms. Number of ant interactome proteins that are unique to ant and that are orthologous to proteins in different numbers of analyzed species is shown. (c) Differentially expressed top hubs and bottlenecks (represented by black dots) and presence/absence of their orthologs in other seven organisms. Differential expression values ( $\log_2\text{FC}$ ) of hubs and bottlenecks in bacterial infection of *C. floricidanus* are represented by a color gradient.

represent the most important properties in the PPI network because of their role in maintaining the functional integrity and connectivity of the network. Proteins with high degree are termed as hubs while the proteins with high betweenness centrality are termed bottlenecks.

We applied Fisher's exact test to compare the proportion of multi-localized proteins in hubs and bottlenecks to non-hubs and non-bottlenecks, respectively. Supplementary Fig. 2 shows differences between the localization of bottlenecks and hubs of the ant interactome. For bottleneck proteins, 70% were found to be multi-localized (versus 56% for non-bottleneck proteins; significant difference;  $p = 9.6 \times 10^{-10}$ ). On the other hand, 62% of the hub proteins had multiple localizations (versus 56% for non-hub proteins; significant difference,  $p = 0.001575$ ).

Integration of the RNASeq data<sup>3</sup> with the ant interactome revealed differentially expressed infection-induced hubs and bottlenecks during the bacterial infection of *C. floridanus* (Fig. 5c). These include also well-known key proteins involved in *C. floridanus* immune response such as nuclear factor NF-kappa-B p110 (Relish, Cflo\_N\_g6082), acidic mammalian chitinase (Cflo\_N\_g2277), as well as stress-related protein cytochromes P450 6A1 (Cflo\_N\_g11706)<sup>3</sup>. Given the high importance of hubs and bottlenecks in PPI networks and their differential expression during bacterial infection, all the identified proteins are expected to participate in the defense against bacterial pathogen, and hence can also be examined for decoding immune mechanisms. The insect peritrophic membrane (PM) imposes protective physical barriers over the midgut epithelium<sup>44</sup>. The PM related proteins have shown their potential as targets for pest control<sup>45,46</sup>. Therefore, the important ant peritrophic membrane protein 1 (Cflo\_N\_g4555) (Fig. 5c) with no human homology could be further tested as a potential pest target. However, differential expression does not guarantee a protein to be the best target<sup>47,48</sup> and therefore, other topologically important proteins in the network without human homology (Datasheet 7 in Supplementary Material) should also be considered as potential pest targets in future.

## Conclusions

Our curated ant interactome is the first large-scale PPI network of an ant. It allows besides numerous analysis of network biology to study how different cellular processes connect to each other including hub proteins and different types of crosstalk, for instance in immunity.

Similarly, the PPI maps of other sequenced ants can be reliably predicted using the interologs of the reconstructed high-confidence *C. floridanus* interactome. Moreover, detailed cross-validation, comparison with random networks, GO annotation, and conservation analysis support the high quality of the resulting ant interactome and its construction steps. The network analysis including evolutionary conserved network proteins further suggest that topologically important proteins could also be exploited as future pest targets. For instance, cytochrome P450 6A1 (Cflo\_N\_g11706), peritrophic membrane protein 1 (Cflo\_N\_g4555), flexible cuticle protein 12 (Cflo\_N\_g6859), endocuticle structural glycoprotein SgAbd-1 (Cflo\_N\_g7775) were identified as topologically important differentially expressed proteins with no human orthologs. Nevertheless, specific interactions highlighted from our global analysis will need individual follow up by detailed investigations.

## Materials and Methods

**Reconstructing protein-protein interaction map of *C. floridanus*.** We compiled the list of experimentally verified high-confidence PPIs available in Database of interacting proteins (DIP)<sup>49</sup>, *D. melanogaster* PPIs from DroID<sup>50</sup> database which includes data from different studies including interactions from high throughput Gal4 proteome-wide yeast two-hybrid (Y2H) screens<sup>32</sup>, LexA Y2H system screens<sup>51–53</sup>, PPIs from fly protein interaction map<sup>54</sup>, interactions determined in large-scale co-affinity purification (co-AP)/MS screens<sup>55,56</sup>, interactions from BIND<sup>57</sup>, BioGRID<sup>58</sup>, MINT<sup>59</sup>, IntAct<sup>60</sup>, and databases available in DroID v2014\_10.

The *C. floridanus* interologs of the entire template PPIs were determined using orthology predictions from the software InParanoid<sup>20,61</sup> and OrthoMCL<sup>21</sup>. These were further customized using own perl and bash scripts. For DIP interactors we used the default parameters of InParanoid. For the fly data orthology was determined using the stricter Blosum80 matrix. For the OrthoMCL based interologs mapping a Blast e-value of  $1e-05$  was used and the MCL inflation index set to 1.5. InParanoid distinguished seed orthologs with co-orthologs and left fewer possibilities of mixing outparalogs in orthologous clusters. Consensus predictions of InParanoid and OrthoMCL were added to InParanoid seed orthologs to create a set of interologs.

**Pruning PPIs with domain-domain interactions.** The amino acid sequences of non-redundant preliminary PPIs were extracted and domains were assigned to them using Pfam version 27.0<sup>62</sup>. The list of non-redundant domain-domain interactions was prepared from the meta-databases Domine<sup>63</sup>, DIMA 3.0<sup>64</sup> and IDDI database<sup>65</sup>. These use complexes available in the Protein Data Bank (PDB)<sup>66</sup> to identify by interacting domains the Pfam families containing these domains. These Pfam families are then predicted to be interacting. This list was used to parse the template PPIs. All interactions were categorized whether they are supported (good interactions, used for further filtering steps) or not by domain-domain interactions (DDIs).

**Subcellular localization filtering.** The subcellular localization of *C. floridanus* proteins was determined with orthology to Swiss-Prot proteins and the extended version of KnowPredsite<sup>67</sup> available at UniLoc server (bioapp.iis.sinica.edu.tw/UniLoc/), a knowledge-based classifier for protein subcellular localization. If in a binary interaction, both proteins do not share the same localization or at least one compartment in multiple localized proteins, the interaction was ruled out as probable not occurring.

**Isoform filtering.** The information on *C. floridanus* protein isoforms and their function was extracted from our previous publication of *C. floridanus* re-annotation and transcriptome sequencing<sup>3,68</sup>. To reduce network complexity and noise, isoforms of any specific protein present in the network were represented as a single node. Although, the data files for all the networks are provided in the Supplementary Tables (1–5) which allow interested readers to analyze the network of their choice further if they wish.



**Assigning the confidence score.** In fact, the preliminary network is filtered successively as mentioned above to reconstruct the final network, in this way the final network is already of high-confidence as many network biologists working on PPI networks have used DDIs and subcellular localization either to increase confidence or validate the interacting pairs. Here additionally we used topology-based method CAPPIC (cluster-based assessment of protein-protein interaction confidence) to assign the interaction confidence score<sup>31,69</sup> in the filtered network. In brief, CAPPIC calculations are based on the assumption that the proteins existing in the same network module are expected to have a higher number of common neighbours (neighbourhood interconnectedness<sup>70</sup>), and a short path length inbetween<sup>71</sup>. For scoring the confidence level, CAPPIC first performs the clustering of the network using a robust clustering algorithm, Markov Cluster (MCL)<sup>72</sup> and then scores the interactions according to their level of compliance with the basic assumptions of topology-based methods. For the clustering we used an MCL inflation value of 1.5. Scores were classified to three subsets; low confidence score between 0 to 0.3, medium confidence score between 0.3 to 0.7, and high confidence score between 0.7 to 1.

**Network analysis and visualization.** The *C. floridanus* interactome was subjected to topological analysis using Network Analyzer plugin version 2.7 of Cytoscape version 2.8.1<sup>73</sup>. The node degree distribution, mean path length, network diameter and betweenness centrality (BC) were determined with graph theoretic analysis implemented with CentiScaPe<sup>74</sup>. For the network  $G(V,E)$ , the BC of node  $n$  is defined as follows

$$BC(n) = \sum_{s \neq n \neq t} \left( \frac{\sigma_{st}(n)}{\sigma_{st}} \right) \quad (1)$$

here  $s$  and  $t$  are network nodes different from node  $n$ ,  $\sigma_{st}$  is the number of shortest paths from  $s$  to  $t$ , and  $\sigma_{st}(n)$  gives the number of shortest paths from  $s$  to  $t$  that goes through node  $n$ .

Hubs and bottlenecks in the network were identified with cytoHubba<sup>75</sup>. Hubs were defined as proteins connecting with  $\geq 5$  proteins. Moreover, top 20% of bottlenecks and hubs were considered for mapping of the RNASeq expression data which was collected from our previous publication<sup>3</sup>.

**Random networks.** We generated random networks following the Erdős-Rényi Model<sup>76</sup>, Barabási-Albert Model<sup>77</sup> and randomized the proposed (final) ant interactome while preserving the total number of interactome nodes using the Network Randomizer plugin<sup>78</sup> of Cytoscape<sup>73</sup>. A total of 1000 random simulation were employed to generate the undirected random graphs. For all three network sets we computed topological parameters, mean shortest path, degree distribution and clustering coefficient and compared their differences to the native ant interactome using the statistical Z-test<sup>79</sup>.

**Functional annotation.** Blast2GO<sup>36</sup> was used to annotate the Gene Ontology (GO) terms of proteins involved in the reconstructed interactome. Over-representation analyses of GO terms was performed using the Gossip package<sup>80</sup> of the Blast2GO suite. A two-tailed Fisher's exact test followed by false discovery rate (FDR) correction for multiple testing<sup>81</sup> was applied to see the functional difference of ant interactome proteins annotations (foreground set) and full *C. floridanus* proteome annotations<sup>3</sup> (background set). Only differences having an adjusted p-value  $< 0.05$  were considered significant.

**Orthology analysis.** InParanoid<sup>20</sup> was used to identify the orthologs of topologically important nodes in seven model organisms: *Anopheles gambiae*, *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Homo sapiens*, *Mus musculus*, *Plasmodium falciparum*, and *Saccharomyces cerevisiae*. Only the ortholog with 100% bootstrap support was considered as true ortholog. As a note of caution, the conservation was calculated rather conservatively demanding double orthology relations. Hence, the absence of an ortholog (Suppl. Datasheet 7) only indicates that the highly restrictive threshold was not met. Generally, a sequence related protein may still be found by less restrictive algorithms (e.g. BLAST).

For exact quantification of the degree of conservation of ant PPIs we did not check the possible restricted conservation of the binary ant PPIs, but more general the conservation of proteins that are present in the ant interactome and have orthologs in seven other species. After calculation of the orthology relationships between ant and other organisms we identified for every degree the occurrence value of the ant interactome and how many orthologs are present in other species. For each organism the fraction of proteins at a particular ant interactome degree is considered as the number of ant protein orthologs at that particular degree and greater divided by the number of proteins in the set.

## Data availability

All data generated and analysed during this study are included in this published article (and its supplementary Information files). The dataset and codes for random network generation are also available at [https://github.com/ShishirGupta-Wu/ant\\_ppi](https://github.com/ShishirGupta-Wu/ant_ppi).

Received: 9 May 2019; Accepted: 22 January 2020;

Published online: 11 February 2020

## References

1. Gadau, J., Heinze, J., Holldobler, B. & Schmid, M. Population and colony structure of the carpenter ant *Camponotus floridanus*. *Mol. Ecol.* **5**, 785–792 (1996).
2. Zientz, E., Beyaert, I., Gross, R. & Feldhaar, H. Relevance of the endosymbiosis of *Blochmannia floridanus* and carpenter ants at different stages of the life cycle of the host. *Appl. Env. Microbiol.* **72**, 6027–6033, <https://doi.org/10.1128/AEM.00933-06> (2006).

3. Gupta, S. K. *et al.* Scrutinizing the immune defence inventory of *Camponotus floridanus* applying total transcriptome sequencing. *BMC Genomics* **16**, 540, <https://doi.org/10.1186/s12864-015-1748-1> (2015).
4. Ben-Hur, A. & Noble, W. S. Choosing negative examples for the prediction of protein-protein interactions. *BMC Bioinforma.* **7**(Suppl 1), S2, <https://doi.org/10.1186/1471-2105-7-S1-S2> (2006).
5. von Mering, C. *et al.* Comparative assessment of large-scale data sets of protein-protein interactions. *Nat.* **417**, 399–403, <https://doi.org/10.1038/nature750> (2002).
6. Yu, H. *et al.* Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. *Genome Res.* **14**, 1107–1118, <https://doi.org/10.1101/gr.1774904> (2004).
7. Zhang, S., Chen, H., Liu, K. & Sun, Z. Inferring protein function by domain context similarities in protein-protein interaction networks. *BMC Bioinforma.* **10**, 395, <https://doi.org/10.1186/1471-2105-10-395> (2009).
8. Mahdavi, M. A. & Lin, Y. H. False positive reduction in protein-protein interaction predictions using gene ontology annotations. *BMC Bioinforma.* **8**, 262, <https://doi.org/10.1186/1471-2105-8-262> (2007).
9. Saito, R., Suzuki, H. & Hayashizaki, Y. Interaction generality, a measurement to assess the reliability of a protein-protein interaction. *Nucleic acids Res.* **30**, 1163–1168 (2002).
10. Sprinzak, E., Sattath, S. & Margalit, H. How reliable are experimental protein-protein interaction data? *J. Mol. Biol.* **327**, 919–923 (2003).
11. Dyer, M. D., Murali, T. M. & Sobral, B. W. Computational prediction of host-pathogen protein-protein interactions. *Bioinforma.* **23**, 1159–1166, <https://doi.org/10.1093/bioinformatics/btm208> (2007).
12. Remmele, C. W. *et al.* Integrated inference and evaluation of host-fungi interaction networks. *Front. microbiology* **6**, 764, <https://doi.org/10.3389/fmicb.2015.00764> (2015).
13. Wang, Y. C. *et al.* Interspecies protein-protein interaction network construction for characterization of host-pathogen interactions: a *Candida albicans*-zebrafish interaction study. *BMC Syst. Biol.* **7**, 79, <https://doi.org/10.1186/1752-0509-7-79> (2013).
14. Zhou, H. *et al.* Stringent homology-based prediction of *H. sapiens*-*M. tuberculosis* H37Rv protein-protein interactions. *Biol. direct* **9**, 5, <https://doi.org/10.1186/1745-6150-9-5> (2014).
15. Itzhaki, Z., Akiva, E., Altuvia, Y. & Margalit, H. Evolutionary conservation of domain-domain interactions. *Genome Biol.* **7**, R125, <https://doi.org/10.1186/gb-2006-7-12-r125> (2006).
16. Pawson, T. & Nash, P. Assembly of cell regulatory systems through protein interaction domains. *Sci.* **300**, 445–452, <https://doi.org/10.1126/science.1083653> (2003).
17. Schuster-Bockler, B. & Bateman, A. Reuse of structural domain-domain interactions in protein networks. *BMC Bioinforma.* **8**, 259, <https://doi.org/10.1186/1471-2105-8-259> (2007).
18. Sharan, R. *et al.* Conserved patterns of protein interaction in multiple species. *Proc. Natl Acad. Sci. United States of America* **102**, 1974–1979, <https://doi.org/10.1073/pnas.0409522102> (2005).
19. Pereira, C., Denise, A. & Lespinet, O. A meta-approach for improving the prediction and the functional annotation of ortholog groups. *BMC genomics* **15**(Suppl 6), S16, <https://doi.org/10.1186/1471-2164-15-S6-S16> (2014).
20. Sonnhammer, E. L. & Ostlund, G. InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic acids Res.* **43**, D234–239, <https://doi.org/10.1093/nar/gku1203> (2015).
21. Li, L., Stoeckert, C. J. Jr. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189, <https://doi.org/10.1101/gr.1224503> (2003).
22. Gupta, S. K. Re-annotation of *Camponotus floridanus* Genome and Characterization of Innate Immunity Transcriptome Responses to Bacterial Infections PhD thesis, Universität Würzburg, (2016).
23. Deane, C. M., Salwinski, L., Xenarios, I. & Eisenberg, D. Protein interactions: two methods for assessment of the reliability of high throughput observations. *Mol. Cell. proteomics: MCP* **1**, 349–356 (2002).
24. Mrowka, R., Patzak, A. & Herzel, H. Is there a bias in proteome research? *Genome Res.* **11**, 1971–1973, <https://doi.org/10.1101/gr.206701> (2001).
25. Cusick, M. E. *et al.* Literature-curated protein interaction datasets. *Nat. methods* **6**, 39–46, <https://doi.org/10.1038/nmeth.1284> (2009).
26. Wojcik, J. & Schachter, V. Protein-protein interaction map inference using interacting domain profile pairs. *Bioinforma.* **17**(Suppl 1), S296–305 (2001).
27. Zhou, H. *et al.* Stringent DDI-based prediction of *H. sapiens*-*M. tuberculosis* H37Rv protein-protein interactions. *BMC Syst. Biol.* **7**(Suppl 6), S6, <https://doi.org/10.1186/1752-0509-7-S6-S6> (2013).
28. Pawson, T., Raina, M. & Nash, P. Interaction domains: from simple binding events to complex cellular behavior. *FEBS Lett.* **513**, 2–10 (2002).
29. Prieto, C. & Rivas, D. L. J. Structural domain-domain interactions: assessment and comparison with protein-protein interaction data to improve the interactome. *Proteins* **78**, 109–117, <https://doi.org/10.1002/prot.22569> (2010).
30. Khush, R. S., Cornwell, W. D., Uram, J. N. & Lemaire, B. A ubiquitin-proteasome pathway represses the *Drosophila* immune deficiency signaling cascade. *Curr. Biol.* **12**, 1728–1737 (2002).
31. Kamburov, A., Grossmann, A., Herwig, R. & Stelzl, U. Cluster-based assessment of protein-protein interaction confidence. *BMC Bioinforma.* **13**, 262, <https://doi.org/10.1186/1471-2105-13-262> (2012).
32. Giot, L. *et al.* A protein interaction map of *Drosophila melanogaster*. *Sci.* **302**, 1727–1736, <https://doi.org/10.1126/science.1090289> (2003).
33. Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nat.* **411**, 41–42, <https://doi.org/10.1038/35075138> (2001).
34. Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Sci.* **296**, 910–913, <https://doi.org/10.1126/science.1065103> (2002).
35. Albert, R. Scale-free networks in cell biology. *J. Cell Sci.* **118**, 4947–4957, <https://doi.org/10.1242/jcs.02714> (2005).
36. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinforma.* **21**, 3674–3676, <https://doi.org/10.1093/bioinformatics/bti610> (2005).
37. Zhao, C. & Wang, Z. GOGO: An improved algorithm to measure the semantic similarity between gene ontology terms. *Sci. Rep.* **8**, 15107, <https://doi.org/10.1038/s41598-018-33219-y> (2018).
38. Fraser, H. B., Hirsh, A. E., Steinmetz, L. M., Scharfe, C. & Feldman, M. W. Evolutionary rate in the protein interaction network. *Sci.* **296**, 750–752, <https://doi.org/10.1126/science.1068696> (2002).
39. Fraser, H. B., Wall, D. P. & Hirsh, A. E. A simple dependence between protein evolution rate and the number of protein-protein interactions. *BMC Evol. Biol.* **3**, 11, <https://doi.org/10.1186/1471-2148-3-11> (2003).
40. Jordan, I. K., Wolf, Y. I. & Koonin, E. V. 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 (2003).
41. Akhooon, B. A. *et al.* *C. elegans* protein interaction network analysis probes RNAi validated pro-longevity effect of nhr-6, a human homolog of tumor suppressor Nr4a1. *Sci. Rep.* **9**, 15711, <https://doi.org/10.1038/s41598-019-51649-0> (2019).
42. Batada, N. N., Hurst, L. D. & Tyers, M. Evolutionary and physiological importance of hub proteins. *PLoS computational Biol.* **2**, e88, <https://doi.org/10.1371/journal.pcbi.0020088> (2006).
43. Yu, H., Kim, P. M., Sprecher, E., Trifonov, V. & Gerstein, M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS computational Biol.* **3**, e59, <https://doi.org/10.1371/journal.pcbi.0030059> (2007).

44. Kuraishi, T., Binggeli, O., Opoa, O., Buchon, N. & Lemaitre, B. Genetic evidence for a protective role of the peritrophic matrix against intestinal bacterial infection in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. U S Am.* **108**, 15966–15971, <https://doi.org/10.1073/pnas.1105994108> (2011).
45. Sajjadian, M. & Hosseininaveh, V. Destruction of peritrophic membrane and its effect on biological characteristics and activity of digestive enzymes in larvae of the Indian meal moth, *Plodia interpunctella* (Lepidoptera: Pyralidae). *Eur. J. Entomology* **112**, 245–250, <https://doi.org/10.14411/eje.2015.046> (2015).
46. Zhang, X. & Guo, W. Isolation and identification of insect intestinal mucin HaIM86—the new target for *Helicoverpa armigera* biocontrol. *Int. J. Biol. Sci.* **7**, 286–296 (2011).
47. Gupta, S. K., Gross, R. & Dandekar, T. An antibiotic target ranking and prioritization pipeline combining sequence, structure and network-based approaches exemplified for *Serratia marcescens*. *Gene* **591**, 268–278, <https://doi.org/10.1016/j.gene.2016.07.030> (2016).
48. Kaldorf, M. *et al.* Systematic Identification of Anti-Fungal Drug Targets by a Metabolic Network Approach. *Front. Mol. Biosci.* **3**, 22, <https://doi.org/10.3389/fmolb.2016.00022> (2016).
49. Salwinski, L. *et al.* The Database of Interacting Proteins: 2004 update. *Nucleic acids Res.* **32**, D449–451, <https://doi.org/10.1093/nar/gkh086> (2004).
50. Murali, T. *et al.* DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for *Drosophila*. *Nucleic acids Res.* **39**, D736–743, <https://doi.org/10.1093/nar/gkq1092> (2011).
51. Schwartz, A. S., Yu, J., Gardenour, K. R., Finley, R. L. Jr. & Ideker, T. Cost-effective strategies for completing the interactome. *Nat. methods* **6**, 55–61, <https://doi.org/10.1038/nmeth.1283> (2009).
52. Stanyon, C. A. *et al.* A *Drosophila* protein-interaction map centered on cell-cycle regulators. *Genome Biol.* **5**, R96, <https://doi.org/10.1186/gb-2004-5-12-r96> (2004).
53. Zhong, J., Zhang, H., Stanyon, C. A., Tromp, G. & Finley, R. L. Jr. A strategy for constructing large protein interaction maps using the yeast two-hybrid system: regulated expression arrays and two-phase mating. *Genome Res.* **13**, 2691–2699, <https://doi.org/10.1101/gr.1134603> (2003).
54. Formstecher, E. *et al.* Protein interaction mapping: a *Drosophila* case study. *Genome Res.* **15**, 376–384, <https://doi.org/10.1101/gr.2659105> (2005).
55. Friedman, A. A. *et al.* Proteomic and functional genomic landscape of receptor tyrosine kinase and ras to extracellular signal-regulated kinase signaling. *Sci. Signal.* **4**, rs10, <https://doi.org/10.1126/scisignal.2002029> (2011).
56. Guruharsha, K. G. *et al.* A protein complex network of *Drosophila melanogaster*. *Cell* **147**, 690–703, <https://doi.org/10.1016/j.cell.2011.08.047> (2011).
57. Bader, G. D., Betel, D. & Hogue, C. W. BIND: the Biomolecular Interaction Network Database. *Nucleic acids research* **31**, (248–250 (2003)).
58. Stark, C. *et al.* The BioGRID Interaction Database: 2011 update. *Nucleic acids Res.* **39**, D698–704, <https://doi.org/10.1093/nar/gkq1116> (2011).
59. Licata, L. *et al.* MINT, the molecular interaction database: 2012 update. *Nucleic acids Res.* **40**, D857–861, <https://doi.org/10.1093/nar/gkr930> (2012).
60. Orchard, S. *et al.* The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic acids Res.* **42**, D358–363, <https://doi.org/10.1093/nar/gkt1115> (2014).
61. Remm, M., Storm, C. E. & Sonnhammer, E. L. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* **314**, 1041–1052, <https://doi.org/10.1006/jmbi.2000.5197> (2001).
62. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic acids Res.* **42**, D222–230, <https://doi.org/10.1093/nar/gkt1223> (2014).
63. Yellaboina, S., Tasneem, A., Zaykin, D. V., Raghavachari, B. & Jothi, R. DOMINE: a comprehensive collection of known and predicted domain-domain interactions. *Nucleic acids Res.* **39**, D730–735, <https://doi.org/10.1093/nar/gkq1229> (2011).
64. Luo, Q., Pagel, P., Vilne, B. & Frishman, D. DIMA 3.0: Domain Interaction Map. *Nucleic acids Res.* **39**, D724–729, <https://doi.org/10.1093/nar/gkq1200> (2011).
65. Kim, Y., Min, B. & Yi, G. S. IDDI: integrated domain-domain interaction and protein interaction analysis system. *Proteome Sci.* **10**(Suppl 1), S9, <https://doi.org/10.1186/1477-5956-10-S1-S9> (2012).
66. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic acids Res.* **28**, 235–242 (2000).
67. Lin, H. N., Chen, C. T., Sung, T. Y., Ho, S. Y. & Hsu, W. L. Protein subcellular localization prediction of eukaryotes using a knowledge-based approach. *BMC Bioinforma.* **10**(Suppl 15), S8, <https://doi.org/10.1186/1471-2105-10-S15-S8> (2009).
68. Gupta, S. K. *et al.* in *Big Data Analytics in Genomics* (ed Ka-Chun Wong) 171–195 (Springer International Publishing, (2016)).
69. Kamburov, A., Stelzl, U. & Herwig, R. IntScore: a web tool for confidence scoring of biological interactions. *Nucleic acids Res.* **40**, W140–146, <https://doi.org/10.1093/nar/gks492> (2012).
70. Goldberg, D. S. & Roth, F. P. Assessing experimentally derived interactions in a small world. *Proc. Natl Acad. Sci. United States of America* **100**, 4372–4376, <https://doi.org/10.1073/pnas.0735871100> (2003).
71. Kuchaiev, O., Rasajski, M., Higham, D. J. & Przulj, N. Geometric de-noising of protein-protein interaction networks. *PLoS computational Biol.* **5**, e1000454, <https://doi.org/10.1371/journal.pcbi.1000454> (2009).
72. Vlasblom, J. & Wodak, S. J. Markov clustering versus affinity propagation for the partitioning of protein interaction graphs. *BMC Bioinforma.* **10**, 99, <https://doi.org/10.1186/1471-2105-10-99> (2009).
73. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504, <https://doi.org/10.1101/gr.1239303> (2003).
74. Scardoni, G., Petterlini, M. & Laudanna, C. Analyzing biological network parameters with CentiScaPe. *Bioinforma.* **25**, 2857–2859, <https://doi.org/10.1093/bioinformatics/btp517> (2009).
75. Chin, C. H. *et al.* cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* **8**(Suppl 4), S11, <https://doi.org/10.1186/1752-0509-8-S4-S11> (2014).
76. Erdős, P. & Rényi, A. On Random Graphs I. *Publicationes Mathematicae* **6**, 290–297 (1959).
77. Barabasi, A. L. & Albert, R. Emergence of scaling in random networks. *Sci.* **286**, 509–512 (1999).
78. Tosadori, G., Bestvina, I., Spoto, F., Laudanna, C. & Scardoni, G. Creating, generating and comparing random network models with NetworkRandomizer. *F1000Res* **5**, 2524, <https://doi.org/10.12688/f1000research.9203.3> (2016).
79. Kreyszig, E. *Applied Mathematics*, fourth ed., (Hoboken, NJ: John Wiley & Sons, (1979)).
80. Bluthgen, N. *et al.* Biological profiling of gene groups utilizing Gene Ontology. *Genome Inf.* **16**, 106–115 (2005).
81. Benjamini, Y., Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289–300 (1995).

## Acknowledgements

We gratefully acknowledge funding of this work by the German Research Foundation (DFG-GR1243/8-1; and T.D. by project number 210879364 – TRR 124/B-1). S.K.G. gratefully acknowledges that the present work is a partial output of his PhD thesis<sup>22</sup> and fruitful discussions with GRK 2157 3D tissue infect (project number 270563345). M.S. would like to thank for the financial support from Frauenbeauftragten Büro, University of Wuerzburg, Germany. S.K.G. and Ö.O. would like to thank Dr. Atanas Kamburov (Harvard Medical School,

Massachusetts, USA) for help and useful discussions on the CAPPIC calculations. This publication was funded by the German Research Foundation (DFG) and the University of Wuerzburg in the funding programme Open Access Publishing.

### Author contributions

S.K.G. and T.D. conceived and designed the study. S.K.G., Ö.O., and M.S. performed the procedure and analyzed the data. All authors contributed to biological interpretation of the results. M.S. and T.D. wrote the manuscript. All authors read and agreed to the final version of the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-59344-1>.

**Correspondence** and requests for materials should be addressed to T.D.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020