



# Interactomes: Experimental and In Silico Approaches

# 6

Luíza Lane de Barros Dantas  
and Marcelo Mendes Brandão

## Abstract

Any part of the Central Dogma of Molecular Biology (DNA replication, transcription, and translation) is based on an intricate protein–protein interaction. On this chapter, we will navigate over the techniques that enable us to construct or fulfill the gaps on an interactome study, directly using assessment of the biochemical and/or molecular machinery that allow two proteins to interact with each other; or rely on computational biology techniques to gather information on PPI from public available databases and evaluate this interaction.

## Keywords

Protein–Protein interactions · Interactome · Proteome · Bioinformatics · Databases

L. L. de Barros Dantas (✉)  
John Innes Centre, Norwich Research Park,  
Norwich, UK

M. M. Brandão  
Center for Molecular Biology and Genetic  
Engineering, State University of Campinas,  
Campinas, SP, Brazil  
e-mail: [brandaom@unicamp.br](mailto:brandaom@unicamp.br)

## 6.1 Introduction

The intricate machinery that sustains all living forms is built upon a large well lubricated network of interaction among different biochemical entities, specially, relying on the protein–protein interactions (PPI). Proteins on their course of action, hardly ever act as a lone wolf since their functions tend to be regulated by other proteins to properly achieve its goal.

Protein–protein interactions are the central controller to all biological processes and its revelation provide the basis to comprehend biology as an integrated system. Michael Cusick, on his 2005 manuscript entitled “Interactome: gateway into systems biology,” states that “the full interactome network is the complete collection of all physical protein–protein interactions that can take place within a cell.”

The interactome is the next big step for System Biology, after massive worldwide effort for DNA, RNAs, and proteins sequencing and subsequent gene annotation for many model and non-model organisms. The PPI from a specific cell or organ unravel the roles of each interactor on a signal transduction pathway, improving the discover, quantification and new biochemical targets for biotechnology.

## 6.2 Molecular Technologies for Protein–Protein Interactions (PPI) Identification

The plant cell requires a tight coordination of protein expression, assembly, modification, aggregation into complexes and subcellular localization, in order to properly function. Therefore, it is important to know how proteins work to fully understand how a plant cell works. In addition, as proteins mostly act gathered in complexes rather than isolated, it is critical to understand how proteins interact inside these complexes. What keep proteins together in the macromolecular complexes are protein–protein interactions (PPI). Such interactions are crucial for the maintenance of the cell as a working unity in every plant tissue. The PPI study also helps to elucidate protein cellular localization, which is also relevant to understand protein function. That is why the PPI study provides insights about cell physiology.

PPI can be investigated through many different technologies, used to discover, to confirm or to characterize PPIs and analyze protein proximity on a molecular level in plants. Some techniques are tailored to investigate protein interactions on a binary level or on a multicomplex level with high accuracy. Others allow PPI investigation by imaging living cells or protein complexes, using organisms, purified proteins and cell lysates. There are techniques better suited for PPI screening, while other methods are convenient for confirming PPIs. Before starting a full set of experiments to analyze a specific PPI, a few issues should be considered to avoid both false positive and false negative interactions. Meticulously experiment planning is critical and it is advised to combine at least two different independent molecular methods in PPI analysis (Braun et al. 2013; Hayes et al. 2016). Besides, for already known PPIs, information regarding the binding affinity of the proteins involved in an interaction is useful (Perkins et al. 2010). It also helps on the experiment design when there is prior knowledge about binding domains of the interacting proteins (Keskin et al. 2016) and

about subcellular location where proteins interact (Hayes et al. 2016). To help designing PPI analysis in plants, a brief description of the most well-established molecular technologies used to study protein interactions in plants is shown thereupon, as well as some examples of these technologies applied on plant PPIs analysis.

### 6.2.1 Yeast Two Hybrid (Y2H)

Y2H might be the most popular technique to investigate PPI and for many scientists this is the starting point for PPI studies. This *in vivo* method is based on the direct interaction between two proteins fused to halves of a transcription factor inside yeast nucleus, which reconstructs a transcription factor that expresses a reporter gene. This reporter gene is in charge of yeast survival on selective media (Fields and Song 1989). There are many versions of this technique (Bruckner et al. 2009; White 1996), but Y2H general principle is quite simple. A transcription factor split in two halves: one half is a DNA-binding domain (DB), that allows DNA binding (called bait), and the second half is a transcriptional activation domain (AD), that activates the gene reporter expression (called prey). The transcription of the reporter gene allows yeast to grow in a selective media (Ito et al. 2001) only when a given pair of proteins fused to bait and prey halves physically interact. Y2H is often used as a screening method to start searching for PPI. This method is suitable because it is easy to operate and inexpensive, ideal to start screening PPIs (Braun et al. 2013). Y2H has several limitations, like generate false positive PPIs. Because the candidate interacting proteins should be expressed in yeast nucleus, Y2H might detect interactions between proteins unnaturally co-localized. Also, Y2H might fail to identify interactions involving proteins requiring post-translational modifications, proteins with transient interactions or proteins expressed and or active on the membrane (Braun et al. 2009). That is why Y2H is a technique to be applied in combination with other PPI detection technologies. In plants, there are many examples of PPI analysis done using Y2H. The first plant interac-

tome, the *Arabidopsis* Interactome 1 (AI-1), was completed using Y2H and shows around 6200 interactions among 2700 proteins, approximately (*Arabidopsis* Interactome Mapping 2011). In tomato, Y2H was used to examine the interactions of ABA signaling core components (Chen et al. 2016). In tobacco, Y2H assays showed the role of 14-3-3 isoforms in plant signaling by mapping the interaction between protein 14-3-3 and enzyme sucrose-6-phosphate synthase (SPS) (Bornke 2005; Ferro and Trabalzini 2013).

### 6.2.2 Pull-Down

Pull-down assays are widely used for PPI detection and/or confirmation. This in vitro technique is based on affinity purification, similarly to co-immunoprecipitation (Co-IP). The difference between them is, while Co-IP uses antibodies fused to known proteins, pull-down uses tags fused to known proteins. In pull-down experiments, a known protein is expressed in cells with a tag (called bait). This fused protein is immobilized to an affinity matrix specifically compatible with this tag. The interacting candidate proteins (called preys) are trapped in a protein complex attached to matrix. After a few purification steps, this protein complex is eluted and ready to analysis on SDS-PAGE and western blot or mass spectrometry (Louche et al. 2017). The bait proteins can come from various sources, such as cell lysate, expression systems or purified. That is also true for the prey proteins, depending on the purpose of the pull-down assay, which could be PPI identification or characterization of a known PPI. A crucial step in a pull-down assay preparation is the choice of a tag. Since the tag is going to act as the link between the specific affinity matrix and the protein complex, aspects such as size and polarity of a tag before expressing the bait fused protein must be considered. The glutathione S-transferase (GST) tag has affinity for glutathione-based matrixes. GST tags are significantly large (26 kDa), expensive, and can interact in a nonspecific fashion. An extensively used tag is the histidine (His) tag. This tag is made of six histidine amino acid residues and has a high

affinity for nickel-based resins, such Ni-NTA agarose. This is a small tag (1.1 kDa), unlikely to affect the bait protein folding and it is inexpensive. Even though pull-down is a good method to study PPI in complexes, this might not be the best approach to investigate transient PPIs. An example of pull-down assays use in plants comes from rice RING UB E3 ligase (OsSIRP2), whose gene is upregulated under abiotic stress conditions (i.e., salinity stress). E3 ligase was shown to interact with TRANSKETOLASE 1 (OsTKL1) under salinity conditions and to increase OsTKL1 degradation (Chapagain et al. 2017). Pull-down experiments were also done to confirm interactions between JASMONATE ZIM DOMAIN (JAZ) protein and NOVEL INTERACTOR OF JAZ (NINJA) transcriptional repressor in jasmonate responses (Pauwels et al. 2010).

### 6.2.3 Co-immunoprecipitation (Co-IP)

This technique is another in vitro method based on affinity purification for PPI analysis in a larger scale on protein complexes. Co-IP is generally used for PPI confirmation and/or characterization (Dwane and Kiely 2011; Hayes et al. 2016; Rao et al. 2014). Similarly to the pull-down mechanism, Co-IP assays are based on a known protein (called bait), with which other proteins in a complex (called prey) interact. The complex is isolated due to the connection between the bait protein and a specific antibody. For Co-IP, whole cell lysates can be used as a starting point, as well as purified proteins. The protein complex detected due to the antibody specific connection can be immobilized in a matrix, isolated, eluted and analyzed by western blot or mass spectrometry. Because this method allows the use of cell lysate, Co-IP is a suitable approach for proteins bearing post-translational modifications and is also indicated to analyze endogenous proteins (Rao et al. 2014). Besides, Co-IP can evaluate proteins PPIs in their native conformation and it is relatively inexpensive. A great disadvantage of Co-IP for plant studies is the fact that it is a technique based on the use of antibodies, since there is little vari-

ety of antibodies for plant proteins (Braun et al. 2013). Also, Co-IP produces background and false positives, requiring careful planning and use of negative controls (Braun et al. 2013; Ransone 1995). Transient PPI are challenging to be detected using Co-IP. In *Arabidopsis*, Co-IP assays were used to expose the interactions of EFR receptor kinases triggered by innate immunity responses (Roux et al. 2011). Recently, the interaction between PROTEIN TARGETING TO STARCH (PTST) PTST2 and PTST3 with STARCH SYNTHASE4 (SS4) was shown to be related to starch granule initiation regulation in *Arabidopsis* leaves (Seung et al. 2017).

#### 6.2.4 Tandem Affinity Purification: Mass Spectrometry (TAP-MS)

This is a high throughput method for PPIs identification, designed to investigate them in the cell standard conditions (Rigaut et al. 1999). TAP-MS employs a tag fused to the C- or N-terminus of a known protein, called bait (Kaiser et al. 2008). As the tag used in TAP-MS assays is built as a double tag, with two proteins connected by a protease, this method requires a two-step purification using two immobilized matrixes with affinity for each part of the double tag (Gunzl and Schimanski 2009). The protein complex that interacts with the bait protein is isolated from the initial cell lysate or purified protein solution and subsequently analyzed by mass spectrometry. There are several types of double tags used in TAP-MS. One of them is the combination of a double-protein-A domain connected by a tobacco etch virus (TEV) protease cleavage site to a calmodulin-binding peptide (Rigaut et al. 1999). Another tag is the GS tag, which has a double-protein-G domain and a streptavidin-binding-peptide connected by a protease from TEV or rhinovirus 3C (Braun et al. 2013; Van Leene et al. 2008). TAP-MS is a very efficient method able to detect both transient and stable PPI (Yates et al. 2009). However, due to the necessity of specific equipment, it can be expensive. The fused tag might interfere in the bait protein expression and folding, and the two-round purification might

interfere in the final PPI yielding in an initial protein material. In plants, TAP-MS was used to elucidate the TCP4 complex components, helping to regulate the expression of CONSTANS (CO) at the right time of the day (Kubota et al. 2017). A classic example of TAP-MS in plants is the platform for *Arabidopsis* cell suspension cultures created to analyze protein complexes (Van Leene et al. 2011).

#### 6.2.5 Förster Resonance Energy Transfer

The resonance energy transfer methods are proximity-dependent techniques that use recombinant fused proteins to analyze proteins pairs within a distance of 10 nm or less from each other (Kerppola 2006; Piston and Kremers 2007). The interacting proteins pairs are fused to donor-acceptor molecules pairs, either fluorescent or bioluminescent, and the energy of an excited donor molecule is transferred to the acceptor molecule, which emits energy as photons (Lonn and Landegren 2017; Wiens and Campbell 2018). There are two different methods based on the principle of resonance energy transfer, according to the molecular nature of the donor-acceptor pair: Fluorescent Resonance Energy Transfer (FRET) (Piston and Kremers 2007) and Bioluminescent Resonance Energy Transfer (BRET) (Pfleger and Eidne 2006). FRET is based on fluorophores donor-acceptors pairs. An example of widely used donor-acceptor pairs in FRET assays are Cyan Fluorescent Protein (CFP), used as the donor fluorophore, and Yellow Fluorescent Protein (YFP), used as the acceptor fluorophore. Each one of these fluorescent proteins is fused to an interacting protein from a PPI pair and, in case both interacting proteins are brought together in a distance of 10 nm or less, light emission can be imaged using standard confocal microscopy or wide-field microscope, for example (Lonn and Landegren 2017). BRET depends upon an enzyme-catalyzed luminescence reaction. The oxidation reaction of a compatible substrate, such as coelenterazine, by luciferase enzyme causes emission of bioluminescence. In BRET, the lucif-

erase acts as the donor that excites the acceptor fluorophore, if the acceptor-donor pair is within a radius of 10 nm or less. The bioluminescence emission can be captured using a cooled-CCD camera (Lonn and Landegren 2017; Xu et al. 2007, 1999). In both FRET and BRET, PPI can be imaged *in situ* and *in planta*. Nonetheless, both techniques require expensive equipment for analysis. BRET assays were efficiently used to image tobacco and *Arabidopsis* tissues (Xu et al. 2007), and also to show the role of interaction between enzymes SUCROSE PHOSPHATE SYNTHASE (SPS) and SUCROSE PHOSPHATE PHOSPHATASE (SPP) in *Arabidopsis* growth (Maloney et al. 2015). FRET assays were applied on experiments to identify interactions between VACUOLAR SORTING RECEPTORS (VSRs) and vacuole-targeted proteins, crucial to target proteins for degradation in the vacuole (Kunzl et al. 2016).

### 6.2.6 Bimolecular Fluorescence Complementation (BiFC)

This molecular *in vivo* method for PPI analysis is an established form of protein complementation assay (PCA), based on protein-fragment complementation. In BiFC assays, a fluorescent protein, like GFP or YFP, is split in half and each of these parts is fused to the N- or C-terminal end of a candidate interacting proteins pair. Note that those fluorescent protein parts alone are unfunctional. If the recombinant protein pair interacts, both fluorescent protein halves are linked and the fluorescent protein is restored to its full folded version (Ghosh et al. 2000; Lonn and Landegren 2017). The resultant fluorescence emission can be imaged using live microscopy or confocal microscopy. In plants, BiFC experiments are mostly performed prior to transient protein expression in either *Nicotiana* or *Arabidopsis* (Bracha-Drori et al. 2004; Braun et al. 2013; Citovsky et al. 2008). Even though BiFC is a suitable method for identifying the subcellular cell location where PPI occurs, the recombinant fluorescent fused half-protein might affect protein conformation and location.

Another limitation is that BiFC assays might give high background fluorescence because of the fluorescent protein parts spontaneous self-assembling. The spontaneous self-assembling might also generate false positives and, therefore, BiFC experiments need a very careful planning and rigorous control. As an alternative to BiFC, but using the same PCA principle, there is the Bimolecular Luminescent Complementation (BiLC). BiLC uses luciferases from different sources instead of fluorescent proteins for complementation (Buntru et al. 2016; Wiens and Campbell 2018). In plants, BiFC assays were performed to prove the homodimerization of transcription factors LATERAL ORGAN BOUNDARIES DOMAIN/ASYMMETRIC LEAVES2-LIKEs (LBD) LBD16 and LBD18, required for activating lateral root formation in *Arabidopsis* (Lee et al. 2017). In rice, experiments showed the relationship between flowering time and phosphorus homeostasis with help of BiFC experiments confirming the interaction between proteins UBIQUITIN-CONJUGATING E2 ENZYME (OsPHO2) and GIGANTEA (OsGI) (Li et al. 2017). BiLC performed in *Nicotiana* showed PPIs in the Golgi apparatus relevant to xyloglucan biosynthesis (Lund et al. 2015).

---

## 6.3 In Silico Approaches for Protein-Protein Interactions (PPI) Identification

### 6.3.1 Databases

Independently on which molecular technique has been used to identify protein interaction, it is necessary to storage this information in a way that useful information might be gathered from the data set, enabling data comparison, exchange and verification. This storage can be done locally using a correlational database manager, such as MySQL or its fork MariaDB, or on spreadsheet software.

The basic database structure and elements for PPI stowage are presented on Table 6.1.

**Table 6.1** Basic PPI database schema elements for storing interaction information

IntA	IntB	Type	Method	Pub	FSW
AT5G47790	AT5G63310	Predicted	Affinity Capture-MS	Mol Syst Biol. 2007;3:89. Epub 2007 Mar 13	0.0410
AT1G09570	AT5G63310	Experimental	Yeast two hybrid	J Biol Chem. 2005 Feb 18;280(7):5740–9. Epub 2004 Nov 23	0.0465

IntA = interactor A; IntB = interactor B; Type = How this interaction was identified; Method = Experimental: This means that the indicated PPI was experimentally demonstrated using the same organism model of study. Predicted: The indicated PPI was proposed based on orthology studies; Pub = indicated reference to the publication of which this interaction was annotated from; FSW = the Functional Similarity Weight. It represents the proportion of interaction partners that two proteins have in common

### 6.3.2 In Silico PPI Reliability Based on Interaction Topology

Two related mathematical approaches, the Czekanowski-Dice distance (CD-distance) (Brun et al. 2003) and Functional Similarity Weight (FSW) (Chua et al. 2006), have been proposed to

assess the reliability of protein interaction data based on the number of common neighbors of two proteins.

The FSW algorithm was originally proposed by Chua et al. (2006) and the functional similarity weight index on a pair of proteins A and B in an interaction graph ( $FSW_{A,B}$ ) is defined as:

$$FSW_{A,B} = \left( \frac{2|N_A \cap N_B|}{|N_A - N_B| + 2|N_A \cap N_B| + \lambda_{A,B}} \right) \times \left( \frac{2|N_A \cap N_B|}{|N_B - N_A| + 2|N_A \cap N_B| + \lambda_{B,A}} \right),$$

where

$N_A$  = set of interaction partners of A;  $N_B$  = set of interaction partners of B;  $\lambda_{A,B}$  is a weight to penalize similarity weights between protein pairs when any of the proteins has few interacting partners and is calculated as:

$$\lambda_{A,B} = \max\left(0, N_{avg} - (|N_A - N_B| + |N_A \cap N_B|)\right),$$

where

$N_{avg}$  = Average of interactions made by each protein on a database.

The Czekanowski-Dice distance between two proteins a and b is given by:

$$D(a,b) = \frac{|N'_a \Delta N'_b|}{|N'_a \cup N'_b| + |N'_a \cap N'_b|}$$

where

$N'_a$  = a set of proteins that contain a and its interaction neighbors;  $a \Delta b$  = symmetric difference between two sets, a and b.

Both algorithms were initially projected to predict protein functions, and lately have been shown to perform well for assessing the reliability of protein interactions (Liu et al. 2009). Wong (2008) has shown that using FSW, which estimates the strength of functional association, to remove unreliable interactions (low FSW) improves the performance of clustering algorithms.

The effectiveness of using FSW as a PPI reliability index was demonstrated using 19,452 interactions in yeast obtained from the GRID database (Breitkreutz et al. 2003). Over 80% of the top 10% of protein interactions ranked by FSW have a common cellular role, and over 90% of them have a common subcellular localization (Chen et al. 2006b, c).

One example of FSW application can be seen on the *Arabidopsis thaliana* protein interaction network database—AtPIN (Brandao et al. 2009). Due to its integrative profile, the reliability index for a reported PPI can be postulated in terms of

interaction partners proportion that two proteins have in common, and these pairs of interacting proteins highly ranked by this method are likely to be true positive interactors. Contrariwise, the proteins pairs lowly ranked are likely to be false positives. With the same benchmarking approach indicated above, the top 10% of protein interactions, ranked by FSW in AtPIN (release 9 of AtPINDB), have indicated that 59% of PPIs share the same subcellular compartment, and 83% have the same function or participate in the same cellular process. A decent FSW value threshold starting point is the top 20%, since Chua et al. (2006) and Chen et al. (2006b) have demonstrated that a protein pair having a high FSW value, above this value, is likely to share a common function.

The most interesting feature of the CD-distance and FSW is that they can rank the reliability of an interaction between a pair of proteins using only the topology of the interactions between that pair and their neighbors within a short radius in a graph network (Chen et al. 2006b, c).

### 6.3.3 PPI Reliability Evaluation Based on Subcellular Localization

An additional reliability checking point for in silico PPI predictions is the Cellular Compartment Classification or  $C^3$ . The  $C^3$  value is represented as classes and is calculated using simple mathematical sum of three parameters:

$$C^3 = A + B + C$$

where

$A$  = type of interaction;  $B$  = co-localization;  $C$  = determination of subcellular localization (experimentally or predicted).

Table 6.2 presents a summary of the possible entering values to calculate  $C^3$ .

Considering all possibilities, it is possible to divide the PPIs in a dataset into four classes:

- **Class A** ( $C^3 = 7$ ): The PPI and subcellular location have seemed to be experimentally

demonstrated and both proteins are co-localized.

- **Class B** ( $C^3 = 5$ ): The PPI and subcellular location have been experimentally shown; however, the proteins were localized to different subcellular compartments.
- **Class C** ( $C^3 = 3$ ): Same as Class A, but the PPI is based on prediction analyses.
- **Class D** ( $C^3 = 6$ ): Same as Class A, but subcellular location is based on prediction analyses.

### 6.3.4 Publicly Available Databases

We are living on the Big Data ages, and several available databases with proteins interactions have arisen over the past decades. Zahiri et al. (2013) present on their manuscript a comprehensive list of the most popular PPI repositories for model organisms. To integrate the major public interaction data providers in a mutual agreement to share data, to develop a distinct set of curation rules for collect data from directly deposited PPI data and/or from peer-reviewed publications, the IMEx was created, acronym for International Molecular Exchange Consortium (Orchard et al. 2012).

IMEx aims to make these interactions material available in an intuitive browsing and search interface on a single website. One of the key points of this concatenated and curated dataset is to provide all the information in standard format, facilitating the usage and incorporation of this data on a variety of bio computational applications.

This sharing standardization is mandatory since each database provider might storage its PPI datasets on a particular format. Currently, the most used standard format for molecular interaction data exchange is the PSI-MI XML (Kerrien et al. 2007), proposed by the Proteomics Standards Initiative, maintained by the Human Proteome Organization (HUPO). Another very popular exchange format is the PSI-mitab, differently from PSI-MI XML, all the molecular interactions are presented on tab-delimited format with up to 42 fields of information. Both formats

**Table 6.2** Numeric values for each parameter for  $C^3$  calculation

Parameter	Value	Description
Type of interaction	4	Based on experimental data
	0	No experimental data available (predicted)
Co-localization	2	Same cellular compartment
	0	Different cellular compartment
Determination of subcellular localization	1	Based on experimental analyses
	0	If one or both are predicted

**Table 6.3** Largest IMEx partners caretakers of publicly available data

Database	URL	References
Biological General Repository for Interaction Datasets (BioGRID)	<a href="http://www.thebiogrid.org">http://www.thebiogrid.org</a>	Stark et al. (2006)
Database of interacting proteins (DIP)	<a href="http://dip.doe-mbi.ucla.edu">http://dip.doe-mbi.ucla.edu</a>	Salwinski et al. (2004), Xenarios et al. (2002)
IntAct	<a href="http://www.ebi.ac.uk/intact/">http://www.ebi.ac.uk/intact/</a>	Orchard et al. (2014)
Molecular Interaction Database (MINT DB)	<a href="http://mint.bio.uniroma2.it/mint">http://mint.bio.uniroma2.it/mint</a>	Licata et al. (2012)

previously cited, and a few other molecular interactions exchange layouts can be found at HUPO GitHub address at <https://github.com/HUPO-PSI> or at the HUPO-PSI web site (<http://www.psidev.info/>).

The four most active and cited IMEx partners' datasets are presented on Table 6.3. All of them are focused on model organisms PPIs and shareable information on standards formats previously discussed.

### 6.3.5 In Silico Predictions

All the members within a protein family are homologous and can be further separated into orthologs, which are genes of different species that evolved from a common ancestral gene by speciation. Generally, orthologs retain the same molecular function during evolution. Researchers rely on these characteristics to predict possible interactions for a non-model organism from a well curated and annotated PPI dataset.

Studies using multiple sequences alignment, from different organisms, have demonstrated that when average amino acid identity is over 50% of correctly aligned residues, we assume that the involved proteins will present the same ancestry,

and, therefore, might be considered orthologs (Ogden and Rosenberg 2007; Thompson et al. 1999). There are many ways to detect orthologs genes/proteins;

#### 6.3.5.1 Reciprocal BLAST

The simplest is the reciprocal BLAST analyses. Having two datasets named Species A and Species B, two separated blast datasets so-called A\_DB and B\_DB are created. Basically, the appropriate BLAST program is run querying Species A on B\_DB and Species B on A\_DB, using arbitrary threshold for sequence similarity over 80%. A second parameter to evaluate is the *e*-value but keeping in mind that *e*-value depends on the database size, so, the larger the database, the smaller the *e*-value can be. A good cutoff starting point is something between  $10^{-5}$  and  $10^{-20}$ . Moreno-Hagelsieb has summarized few hints on how to choose the best BLAST parameters values for reciprocal BLAST ortholog identification approach (Moreno-Hagelsieb and Latimer 2008). To identify the best reciprocal hit among all sequences on the datasets, the BackBlast Reciprocal Blast script (<https://github.com/LeeBergstrand/>) can be used to automate the analyses process. The BackBlast algorithm will identify those best reciprocal hits and return a fil-

tered list of most plausible orthologs among Species A and Species B. Now it is possible to transfer PPI information from one dataset to another.

### 6.3.5.2 OrthoMCL

Its algorithm firstly identify sequence similarities by reciprocal best BLAST, and then, joins proteins into ortholog groups based on normalized BLAST scores between proteins using Markov clustering (Enright et al. 2002; Li et al. 2003). It is also available in the orthoMCL-DB website (Chen et al. 2006a), which contains ortholog groups for most completely sequenced and annotated eukaryotes and for a number of completely sequenced and annotated prokaryotes (<http://orthomcl.org/>). There is an ample tutorial written by Fischer et al. (2011), encompassing all the steps needed to identify the most plausible orthologs.

### 6.3.5.3 InParanoid

This program uses the pairwise similarity scores between two datasets, calculated using BLASTP, for assembling orthology groups. These orthology groups are initially composed of two so-called seed orthologs found by reciprocal best hits between two datasets. On second step, more sequences are added to the group if the sequences in the two datasets are closer to the corresponding seed ortholog than to any sequence not present into the ortholog group in question. The orthology group participants are now called inparalogs, and, a confidence value is provided for each of them, representing how closely related it is to its seed ortholog (O'Brien et al. 2005). The Inparanoid DB (Sonnhammer and Ostlund 2015) is an online database for ortholog groups with inparalogs (<http://inparanoid.sbc.su.se/>).

## References

- Arabidopsis Interactome Mapping C (2011) Evidence for network evolution in an arabidopsis interactome map. *Science* 333:601–608. <https://doi.org/10.1126/science.1203877>
- Bornke F (2005) The variable C-terminus of 14-3-3 proteins mediates isoform-specific interaction with sucrose-phosphate synthase in the yeast two-hybrid system. *J Plant Physiol* 162:161–168. <https://doi.org/10.1016/j.jplph.2004.09.006>
- Bracha-Drori K, Shichrur K, Katz A, Oliva M, Angelovici R, Yalovsky S, Ohad N (2004) Detection of protein-protein interactions in plants using bimolecular fluorescence complementation. *Plant J* 40:419–427. <https://doi.org/10.1111/j.1365-3113X.2004.02206.x>
- Brandao MM, Dantas LL, Silva-Filho MC (2009) AtPIN: Arabidopsis thaliana protein interaction network. *BMC Bioinformatics* 10:454. <https://doi.org/10.1186/1471-2105-10-454>
- Braun P et al (2009) An experimentally derived confidence score for binary protein-protein interactions. *Nat Methods* 6:91–97. <https://doi.org/10.1038/nmeth.1281>
- Braun P, Aubourg S, Van Leene J, De Jaeger G, Lurin C (2013) Plant protein interactomes. *Annu Rev Plant Biol* 64:161–187. <https://doi.org/10.1146/annurev-arplant-050312-120140>
- Breitkreutz BJ, Stark C, Tyers M (2003) The GRID: the general repository for interaction datasets. *Genome Biol* 4:R23
- Bruckner A, Polge C, Lentze N, Auerbach D, Schlattner U (2009) Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci* 10:2763–2788. <https://doi.org/10.3390/ijms10062763>
- Brun C, Chevenet F, Martin D, Wojcik J, Guenoche A, Jacq B (2003) Functional classification of proteins for the prediction of cellular function from a protein-protein interaction network. *Genome Biol* 5:R6
- Buntru A, Trepte P, Klockmeier K, Schnoegl S, Wanker EE (2016) Current approaches toward quantitative mapping of the interactome. *Front Genet* 7:74. <https://doi.org/10.3389/fgene.2016.00074>
- Chapagain S, Park YC, Kim JH, Jang CS (2017) Oryza sativa salt-induced RING E3 ligase 2 (OsSIRP2) acts as a positive regulator of transketolase in plant response to salinity and osmotic stress. *Planta* 247:925. <https://doi.org/10.1007/s00425-017-2838-x>
- Chen F, Mackey AJ, Stoeckert CJ Jr, Roos DS (2006a) OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res* 34:D363–D368. <https://doi.org/10.1093/nar/gkj123>
- Chen J et al (2006b) Increasing confidence of protein-protein interactomes. In: 17th International Conference on Genome Informatics, Yokohama, Japan, pp 284–297
- Chen J, Hsu W, Lee ML, Ng SK (2006c) Increasing confidence of protein interactomes using network topological metrics. *Bioinformatics* 22:1998–2004
- Chen Pet al (2016) Interactions of ABA signaling core components (SIPYLs, SIPP2Cs, and S1SnRK2s) in tomato (*Solanum lycopersicon*). *J Plant Physiol* 205:67–74. <https://doi.org/10.1016/j.jplph.2016.07.016>
- Chua HN, Sung WK, Wong L (2006) Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions. *Bioinformatics* 22:1623–1630

- Citovsky V, Gafni Y, Tzfira T (2008) Localizing protein-protein interactions by bimolecular fluorescence complementation in planta. *Methods* 45:196–206. <https://doi.org/10.1016/j.ymeth.2008.06.007>
- Dwane S, Kiely PA (2011) Tools used to study how protein complexes are assembled in signaling cascades. *Bioeng Bugs* 2:247–259. <https://doi.org/10.4161/bbug.2.5.17844>
- Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30:1575–1584
- Ferro E, Trabalzini L (2013) The yeast two-hybrid and related methods as powerful tools to study plant cell signalling. *Plant Mol Biol* 83:287–301. <https://doi.org/10.1007/s11103-013-0094-4>
- Fields S, Song O-K (1989) A novel genetic system to detect protein-protein interactions. *Nature* 340:245–246. <https://doi.org/10.1038/340245a0>
- Fischer S et al (2011) Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Curr Protoc Bioinformatics Chapter 6:Unit 6.12.11–Unit 6.12.19*. <https://doi.org/10.1002/0471250953.bi0612s35>
- Ghosh I, Hamilton AD, Regan L (2000) Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein. *J Am Chem Soc* 122:5658–5659. <https://doi.org/10.1021/ja994421w>
- Gunzl A, Schimanski B (2009) Tandem affinity purification of proteins. *Curr Protoc Prot Sci Chapter 19:Unit.19.19*. <https://doi.org/10.1002/0471140864.ps1919s55>
- Hayes S, Malacrida B, Kiely M, Kiely PA (2016) Studying protein-protein interactions: progress, pitfalls and solutions. *Biochem Soc Trans* 44:994–1004. <https://doi.org/10.1042/BST20160092>
- Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98:4569–4574. <https://doi.org/10.1073/pnas.061034498>
- Kaiser P, Meierhofer D, Wang X, Huang L (2008) Tandem affinity purification combined with mass spectrometry to identify components of protein complexes. In: Starkey M, Elaszwarapu R (eds) *Methods in molecular biology*. Humana Press, Totowa, NJ, pp 309–326. [https://doi.org/10.1007/978-1-59745-188-8\\_21](https://doi.org/10.1007/978-1-59745-188-8_21)
- Kerppola TK (2006) Visualization of molecular interactions by fluorescence complementation. *Nat Rev Mol Cell Biol* 7:449–456. <https://doi.org/10.1038/nrm1929>
- Kerrien S et al (2007) Broadening the horizon—level 2.5 of the HUPO-PSI format for molecular interactions. *BMC Biol* 5:44. <https://doi.org/10.1186/1741-7007-5-44>
- Keskin O, Tuncbag N, Gursoy A (2016) Predicting protein-protein interactions from the molecular to the proteome level. *Chem Rev* 116:4884–4909. <https://doi.org/10.1021/acs.chemrev.5b00683>
- Kubota A et al (2017) TCP4-dependent induction of CONSTANS transcription requires GIGANTEA in photoperiodic flowering in Arabidopsis. *PLoS Genet* 13:e1006856
- Kunzl F, Fruholz S, Fassler F, Li B, Pimpl P (2016) Receptor-mediated sorting of soluble vacuolar proteins ends at the trans-Golgi network/early endosome. *Nat Plants* 2:16017. <https://doi.org/10.1038/nplants.2016.17>
- Lee HW, Kang NY, Pandey SK, Cho C, Lee SH, Kim J (2017) Dimerization in LBD16 and LBD18 transcription factors is critical for lateral root formation. *Plant Physiol* 174:301–311. <https://doi.org/10.1104/pp.17.00013>
- Li L, Stoeckert CJ Jr, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13:2178–2189. <https://doi.org/10.1101/gr.1224503>
- Li S, Ying Y, Secco D, Wang C, Narsai R, Whelan J, Shou H (2017) Molecular interaction between PHO2 and GIGANTEA reveals a new crosstalk between flowering time and phosphate homeostasis in *Oryza sativa*. *Plant Cell Environ* 40:1487–1499. <https://doi.org/10.1111/pce.12945>
- Licata L et al (2012) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res* 40:D857–D861. <https://doi.org/10.1093/nar/gkr930>
- Liu G, Wong L, Chua HN (2009) Complex discovery from weighted PPI networks. *Bioinformatics* 25:1891–1897
- Lonn P, Landegren U (2017) Close encounters - probing proximal proteins in live or fixed cells. *Trends Biochem Sci* 42:504–515. <https://doi.org/10.1016/j.tibs.2017.05.003>
- Louche A, Salcedo SP, Bigot S (2017) Protein-Protein Interactions: Pull-Down Assays. In: Journet L, Cascales E (eds) *Bacterial protein secretion systems*. *Methods in Molecular Biology*, vol 1615. Humana Press, New York, NY. [https://doi.org/10.1007/978-1-4939-7033-9\\_20](https://doi.org/10.1007/978-1-4939-7033-9_20)
- Lund CH, Bromley JR, Stenbæk A, Rasmussen RE, Scheller HV (2015) A reversible Renilla luciferase protein complementation assay for rapid identification of protein – protein interactions reveals the existence of an interaction network involved in xyloglucan biosynthesis in the plant Golgi apparatus. *J Exp Bot* 66:85–97. <https://doi.org/10.1093/jxb/eru401>
- Maloney VJ, Park JY, Unda F, Mansfield SD (2015) Sucrose phosphate synthase and sucrose phosphate phosphatase interact in planta and promote plant growth and biomass accumulation. *J Exp Bot* 66:4383–4394. <https://doi.org/10.1093/jxb/erv101>
- Moreno-Hagelsieb G, Latimer K (2008) Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* 24:319–324. <https://doi.org/10.1093/bioinformatics/btm585>
- O'Brien KP, Remm M, Sonnhammer EL (2005) Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res* 33:D476–D480. <https://doi.org/10.1093/nar/gki107>
- Ogden TH, Rosenberg MS (2007) How should gaps be treated in parsimony? A comparison of approaches using simulation. *Mol Phylogenet Evol* 42:817–826
- Orchard S et al (2012) Protein interaction data curation: the International Molecular Exchange (IMEx)

- consortium. *Nat Methods* 9:345–350. <https://doi.org/10.1038/nmeth.1931>
- Orchard S et al (2014) The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res* 42:D358–D363. <https://doi.org/10.1093/nar/gkt1115>
- Pauwels L et al (2010) NINJA connects the co-repressor TOPLESS to jasmonate signalling. *Nature* 464:788–791. <https://doi.org/10.1038/nature08854>
- Perkins JR, Diboun I, Dessailly BH, Lees JG, Orengo C (2010) Review transient protein-protein interactions: structural, functional, and network properties. *Struct Fold Des* 18:1233–1243. <https://doi.org/10.1016/j.str.2010.08.007>
- Pflegler KDG, Eidne KA (2006) Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET). *Nat Methods* 3:165–174. <https://doi.org/10.1038/nmeth841>
- Piston DW, Kremers G-J (2007) Fluorescent protein FRET: the good, the bad and the ugly. *Trends Biochem Sci* 32:407–414. <https://doi.org/10.1016/j.tibs.2007.08.003>
- Ransone LJ (1995) Detection of protein-protein interactions by coimmunoprecipitation and dimerization. In: *Methods in enzymology*, vol 254. Academic Press, New York, NY, pp 491–497. [https://doi.org/10.1016/0076-6879\(95\)54034-2](https://doi.org/10.1016/0076-6879(95)54034-2)
- Rao VS, Srinivas K, Sujini GN, Kumar GNS (2014) Protein-protein interaction detection: methods and analysis. *Int J Proteom* 2014:147648. <https://doi.org/10.1155/2014/147648>
- Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Séraphin B (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* 17:1030–1032
- Roux M et al (2011) The Arabidopsis leucine-rich repeat receptor – like kinases BAK1/SERK3 and BKK1/SERK4 are required for innate immunity to hemibiotrophic and biotrophic pathogens. *Plant Cell* 23:2440–2455. <https://doi.org/10.1105/tpc.111.084301>
- Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D (2004) The database of interacting proteins: 2004 update. *Nucleic Acids Res* 32:D449–D451
- Seung D et al (2017) Homologs of PROTEIN TARGETING TO STARCH control starch granule initiation in arabidopsis leaves. *Plant Cell* 29:1657–1677. <https://doi.org/10.1105/tpc.17.00222>
- Sonnhammer EL, Ostlund G (2015) InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Res* 43:D234–D239. <https://doi.org/10.1093/nar/gku1203>
- Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* 34:D535–D539
- Thompson JD, Plewniak F, Poch O (1999) A comprehensive comparison of multiple sequence alignment programs. *Nucleic Acids Res* 27:2682–2690
- Van Leene J, Witters E, Inze D, De Jaeger G (2008) Boosting tandem affinity purification of plant protein complexes. *Trends Plant Sci* 13:517–520. <https://doi.org/10.1093/dnares/dsn008>
- Van Leene J et al (2011) Isolation of transcription factor complexes from arabidopsis cell suspension cultures by tandem affinity purification. In: Yuan L, Perry SE (eds) *Methods in molecular biology*. Humana Press, Totowa, NJ, pp 195–218. [https://doi.org/10.1007/978-1-61779-154-3\\_11](https://doi.org/10.1007/978-1-61779-154-3_11)
- White MA (1996) The yeast two-hybrid system: forward and reverse. *Proc Natl Acad Sci U S A* 93:10001–10003
- Wiens MD, Campbell RE (2018) Surveying the landscape of optogenetic methods for detection of protein-protein interactions. *Wiley Interdiscip Rev Syst Biol Med* 10:e1415. <https://doi.org/10.1002/wsbm.1415>
- Wong L (2008) Constructing more reliable protein-protein interaction maps. In: *International Symposium on Computational Biology & Bioinformatics*, University of Kerala, 17–19 January 2008, pp 284–297
- Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D (2002) DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res* 30:303–305
- Xu Y, Piston DW, Johnson CH (1999) A bioluminescence resonance energy transfer (BRET) system: application to interacting circadian clock proteins. *Proc Natl Acad Sci U S A* 96:151–156
- Xu X, Soutto M, Xie Q, Servick S, Subramanian C, von Arnim AG, Johnson CH (2007) Imaging protein interactions with bioluminescence resonance energy transfer (BRET) in plant and mammalian cells and tissues. *Proc Natl Acad Sci U S A* 104:10264–10269. <https://doi.org/10.1073/pnas.0701987104>
- Yates JR, Ruse CI, Nakorchevsky A (2009) Proteomics by mass spectrometry: approaches, advances, and applications. *Annu Rev Biomed Eng* 11:49–79. <https://doi.org/10.1146/annurev-bioeng-061008-124934>
- Zahiri J, Bozorgmehr JH, Masoudi-Nejad A (2013) Computational prediction of protein-protein interaction networks: algorithms and resources. *Curr Genom* 14:397–414. <https://doi.org/10.2174/1389202911314060004>