Published in final edited form as:
Proteomics. 2009 December ; 9(23): 5296. doi:10.1002/pmic.200900282.

Benchmarking yeast two-hybrid systems using the interactions of bacterial motility proteins

Seesandra Venkatappa Rajagopala[1], **Kelly T. Hughes**[2], and **Peter Uetz**[1]

[1] J Craig Venter Institute (JCVI), 9704 Medical Center Drive, Rockville, MD 20850, USA, uetz@jcvi.org, raja@jcvi.org

[2] Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA, kelly.hughes@utah.edu

Abstract

Yeast two-hybrid (Y2H) screens often produce vastly non-overlapping interaction data when the screens are conducted in different laboratories, or use different vectors, strains, or reporter genes. Here we investigate the underlying reasons for such inconsistencies and compare the effect of 7 different vectors and their Y2H interactions. Genome-wide array screens with 49 motility-related baits from *Treponema pallidum* yielded 77 and 165 interactions with bait vectors pLP-GBKT7 and pAS1-LP, respectively, including 21 overlapping interactions. In addition, 90 motility-related proteins from *E. coli* were tested in all pairwise combinations and yielded 140 interactions when tested with pGBKT7g/pGADT7g vectors but only 47 when tested with pDEST32/pDEST22. We discuss the factors that determine these effects, including copy number, the nature of the fusion protein, and species-specific differences that explain non-conserved interactions among species. The pDEST22/pDEST32 vectors produce a higher fraction of interactions that are conserved and that are biologically relevant when compared to the pGBKT7/pGADT7-related vectors, but the latter appear to be more sensitive and thus detect more interactions.

Keywords

protein-protein interactions; yeast two-hybrid vectors; bacterial motility; flagella

1 Introduction

The yeast two-hybrid (Y2H) system has been among the most powerful methods to identify protein-protein interactions (PPIs). However, it has also been criticized for generating large numbers of false positives and false negatives (e.g. [1]). The lack of reproducibility is even more obvious when Y2H data is compared to other datasets, such as those derived from affinity purification/mass spectrometry (AP/MS) experiments [2]. Even when Y2H data are compared to other Y2H data, the overlaps are usually small. For example, two systematic screens of protein-protein interactions among proteins of Kaposi Sarcoma-associated Herpesvirus (KSHV) yielded only few overlapping interactions [3],[4].

Recently Braun et al. [5] compared Y2H systems to completely different methodologies such as LUMIER [6] and NAPPA [7] and found dramatic differences. In addition, they also used two slightly different Y2H systems and showed that there are differences between CEN vectors

Correspondence to: Peter Uetz.

The authors do not declare any conflicts of interest.

and 2μ plasmids as well as different reporter genes. However, these authors tested only 92 defined "gold-standard" interactions without providing details on their vectors.

Here we investigate this phenomenon by systematically screening an array of 1,000 preys for protein-protein interactions with 49 bait proteins involved in bacterial motility using multiple different Y2H vector systems, including pLP-GBKT7/pLP-GADT7 (Clontech), and pAS1-LP [8]. In addition, we tested 90 ×90 = 8,100 identical protein pairs with two different bait and prey vector pairs, namely pGBKT7g/pGADT7g [9] and pDEST22/pDEST32 (Invitrogen). Clearly, differences of vectors appear to affect the number and nature of the resulting interactions, even when all other parameters are kept constant.

The goal of this study was to answer three questions: (1) What are the differences between several Y2H vector systems when identical protein pairs or libraries are tested? (2) Can multiple Y2H systems be used for quality control purposes? (3) Which system provides the more reliable and biologically significant interaction data?

2 Materials and methods

2.1 Vectors and strains

Bait and prey open reading frames (ORFs) were cloned from entry vectors pUniD [10] or pDONR/zeo (Invitrogen, Carlsbad, CA) into Y2H vectors pGBKT7g/pGADT7g [9], pLP-GBKT7/pLP-GADT7 (Clontech, Mountain View, CA), pAS1-LP (M. McKevitt, pers. comm. and [8]), and pDEST22/pDEST32 (Invitrogen). See Table 1 for a comparison of these vectors. Prey clones in pLP-GADT7, pDEST22 and pGADT7g vectors were transformed into Y187 (MAT**a**) yeast strains; similarly, all the bait clones in pLP-GBKT7, pAS1-LP, pDEST32, and pGBKT7g vectors were transformed into yeast strain AH109 (MATα) [11,12] by a standard LiAc protocol [9].

2.2 Proteins sources, libraries, bait and prey cloning, and Y2H screening

We used three different prey collections in this study, the first two of which were studied in the context of an array of bacterial motility proteins while the third was studied in the context of a whole-genome array.

2.2.1 The bacterial motility array—We selected 90 *E. coli* proteins assigned to the bacterial motility function (Supplementary Table S1), cloned as ORFs into the Gateway entry vector pDONR/zeo (Invitrogen). The 90 entry clones were cloned into four Y2H vectors, namely pDEST22, pDEST32, pGADT7g, and pGBKT7g (Table 1). The prey clones in pDEST22 vector and pGADT7g vector were arrayed separately as quadruplicates onto 384-well formatted Omnitray-agar plates (Nunc). Each pDEST32 bait was individually tested for two-hybrid interactions against the pDEST22 prey array. Similarly, each pGBKT7g bait was individually tested for two-hybrid interactions with the pGADT7g prey array. Thus, the two screens involved a total of $2 \times 90 \times 90 \approx 16{,}200$ individual tests, each of which was carried out in quadruplicate to ensure reproducibility. The interaction testing was done as described in [13].

2.2.2 The *Treponema pallidum* proteome array—We cloned all protein encoding genes of the spirochete *T. pallidum* into the prey vector pLP-GADT7 (Clontech), as described previously [8]. However, while we did not distinguish between different bait vectors in our previous study, here we present the results of genome-wide screens of two different bait vectors. In brief, 49 *T. pallidum* ORFs, which are part of our KEGG motility collection, were cloned into bait vectors pAS1-LP and pLP-GBKT7 (Clontech) by Cre-loxP mediated recombination

of pUni entry clones [10,14]. Each bait strain was individually screened for protein interactions against the whole *T. pallidum* prey array (in pLP-GADT7) as previously described [13].

3 Results

3.1 Differences between bait and prey vectors

The Gal4 transcription factor has been commonly used for Y2H assays since its description by Fields and Song [15]. Here we compare 3 different prey and 4 different bait vectors that are based on the Gal4 system (Table 1,Figure 1). While these vectors are all broadly similar, there are several notable differences. For example, five vectors use the 2-micron (2μ) origin which results in 50–100 copies per cell [16] whereas the CEN origin of the other other vectors results in only 1-2 copies per cell. While the level of fusion protein expression is driven by the **ADH1 promotor** in all vectors used here [17], two of the four bait vectors carry the full length *ADH1* promoter (pDEST32 and pAS1-LP) and the other two use a truncated version (pGBKT7g and pGBKT7-LP). Since others have reported differences in expression levels resulting from these promoters [18], we have not measured the expression levels of our vectors. Finally, the nature of the **fusion protein** differed in our assays. The fusion proteins differ in the linker amino acid sequence encoded between the Gal4 DNA-binding or activation domain and the ORFs. Because the recombined ORFs do not contain an endogenous stop codon, some vectors encode additional peptide sequences at the C-terminal ends of the ORFs, which may affect their interaction patterns.

3.2 Different baits but identical genome-wide libraries: pLP-GBKT7 and pAS1-LP

In our global survey of protein-protein interactions in *T. pallidum* [8,19], we screened all possible pairwise combinations. We also used the same genome-wide array of all preys for screens with all known or suspected motility proteins [19]. Here we report the screening of 49 motility proteins as baits, cloned into two different bait vectors, pLP-GBKT7 and pAS1-LP, against the aforementioned prey array (in pLP-GADT7, Figure 2A). Surprisingly, despite their similarity, these two vectors produced strikingly different results (Figures 2A and Figures 3A,C,E; Table S4). On average, the number of interactions was higher with the pAS1-LP bait compared to the pLP-GBKT7 bait (Figure 3A). Overall, pLP-GBKT7 and pAS1-LP yielded 77 and 165 interactions, of which 21 were identical with both vectors (Figure 3C). While 17 of the 49 baits (35%) produced interactions with both vectors, 7 (14%) yielded interactions exclusively with pLP-GBKT7 and 10 (20%) only with pAS1-LP (Figure 3E). When both datasets were combined, we found interactions with 29 (59%) of all baits, i.e. either with pLP-GBKT7 or pAS1-LP.

3.3 Pairwise array screens: pDEST22/pDEST32 vs. pGBKT7/pGADT7

Library screens as described above are more complex and involve more variables than individual tests. For example, two series of 49 (i.e. 98) screens cannot be done in one day, which causes day-to-day experimental variations. Thus, we designed a series of individual tests using two different vector pairs that can be done in parallel under exactly the same conditions, such that the results are highly standardized and hence comparable. For these individual tests, known flagellar proteins and interactions manually curated from the literature served as gold-standard data [19]. We cloned all 90 known motility-related ORFs from *E. coli* into two vector pairs (pDEST32/pDEST22 and pGBKT7g/pGADT7g) and tested them in pairwise combinations (Figure 2B). Surprisingly, there was a striking difference in the number of interactions found with the two systems (Figures 3B,D,F): the pGBKT7/pGADT7 system yielded a total of 138 interactions while pDEST32/22 yielded a total of 47, of which only 6 were found in both systems (Figure 3D). All of these interactions were reproducible when retested. When pGKBT7g was used as bait vector, 37 out of 90 bait proteins (41%) yielded interactions as opposed to 26 (29%) when pDEST32 was used (Figure 3F, Tables S1,S2). Only

12 baits produced interactions in both vectors. However, when both datasets were combined, 57% of all baits yielded interactions.

3.4 Assay Sensitivity

Assay sensitivity is the fraction of all the biological interactions that can be identified by an assay conducted under a specific set of experimental conditions. Like other protein interaction detection systems, the Y2H system is not able to detect all protein-protein interactions that occur *in vivo* [5]. Using a literature-curated protein-protein interaction reference set, Venkatesan *et al.* [20] showed that Y2H assays typically capture around 20% of such "true" positives. As was done by Venkatesan *et al*. [20], we compiled all previously published *E. coli* PPIs from the MPIDB database [21] as a "true positive" reference dataset, including interactions of the 90 *E. coli* motility-related proteins we used in this study (Table S3). Based on the reference dataset of 24 PPIs, we estimate the assay sensitivity for the pDEST22/pDEST32 vector system to be 25% (6 out of 24) and for the pGADT7g/pGBKT7g vector system to be 17% (4/24). The combination of two vector systems increased the assay sensitivity to a combined value of 33% (8/24).

3.5 Validating interactions through interologs in multiple species

Protein-protein interactions are often conserved in evolution and this conservation can be used to validate Y2H data, if an interaction can be identified in multiple species. We used the Microbial Protein Interaction Database (MPIDB) [21] as a reference dataset to benchmark the fraction of conserved PPIs ("interologs") obtained from the pDEST22/32 and pGADT7g/pGBKT7g motility array screens (Figure 4A). The pDEST22/32 vector system detected 47 interactions, of which 21 (45%) have interologs in MPIDB. By contrast, the pGADT7g/pGBKT7g screens detected 138 interactions, of which only 5 (3.6%) had interologs in MPIDB (Table S6). The genome-wide screenings of *T. pallidum* flagellum baits with pLP-GBKT7 vector detected 77 interactions, of which 9 (13%) have interologs in MPIDB. The pAS1-LP vector detected 165 interactions of which 8 (5%) have interologs in MPIDB (Table S5). While these numbers indicate that pDEST22/pDEST32 detect more interolog-validated interactions, it remains unknown which of the PPIs without interologs are "true" interactions.

3.6 Biological validation

In addition to interologs, we evaluated the biological value of the interactions from the motility array. We classified these interaction into one of three groups: known or very plausible ("known"), plausible, or unknown/unclear. The assessment was made based on known physical or genetic interactions; plausibility was estimated based on the location of proteins in the flagellar machinery or other evidence (Table S2). This analysis suggested that the pDEST22/32 vector pair yields a much larger fraction of "known" or "plausible" interactions (57%) than the pGBKT7/pGADT7 vector pair (34%, Figure 4B). However, since the pDEST22/32 vectors are more stringent, it is more likely that these vectors miss relevant interactions that the pGBKT7/pGADT7 vectors may detect.

A similar analysis of the interactions obtained in the *Treponema* library screens did not yield any conclusive difference between the pAS1-LP and pLP-GBKT7 bait vectors. However, this analysis was hampered by the fact that the biological significance is difficult to evaluate in *Treponema* and thus only interologs in other species could be used. Although pAS1-LP yielded more than twice as many interactions than pLP-GBKT7 none of them appears to yield a higher fraction of "known" and/or plausible interactions (Table S4).

4 Discussion

Yeast two-hybrid screens often produce dramatically different interactions even when the same baits are used. Here we attempt to identify some of the conditions and features that may cause these differences.

4.1 Plasmid copy number and expression level

Several studies (e.g. [5]) have pointed out that the origin of replication and thus the copy number is a critical component that determines the sensitivity and thus the outcome of a Y2H screen. In fact, while 2μ plasmids are present in 50–100 copies per cell [16] and CEN plasmids in one or two copies, this copy number difference may translate to only a 20–30-fold difference in expression level [16]. Our results confirm that CEN vectors indeed yield fewer interactions than 2μ plasmids, likely due to the fewer transcripts from CEN vectors and correspondingly reduced protein abundance. However, if copy number differences resulted only in differences in Y2H sensitivity, then we would expect that CEN vectors simply produce a subset of 2μ interactions (namely, the “stronger” interactions).

A high plasmid copy number should have a similar effect as a strong promoter from which the fusion proteins are expressed, i.e. more interactions. In fact, the number of interactions found with the pAS1-LP vector, which has the full-length *ADH1* promoter and a high expression level, is greater than that of pLP-GBKT7, a low-expression vector with a truncated ADH1 promoter [17,18] (Table 1). However, the low expression vector does not yield a subset of interactions but rather a different set (Figure 3C). Interestingly, screens with identical bait proteins in different bait vectors (pLP-GBKT7 and pAS1-LP) yielded significant differences despite their identical 2μ origin (Figure 3C). The main difference between the two vectors is the length of the ADH promoter (full-length in pAS1-LP, truncated in pLP-GBKT7) and the linker region of the fusion protein (Figure 1).In fact, two of our bait vectors (pLP-GBKT7 and pGBKT7g) use this truncated promoter, so their reduced promoter activity should somewhat compensate for their high copy number. Again, even if there are significant differences in expression levels, this would not explain the mostly non-overlapping list of interactions we find. Thus, there must be other factors that determine the nature of interactions detected by a Y2H screen.

4.2 Structural differences of bait and prey fusion proteins

While the baits and preys encoded by the various Gal4 vectors used here appear superficially identical, they carry several sequence features that render the fusion proteins different (Figure 1). First, the Gal4 AD is slightly truncated in pDEST22 as opposed to pGADT7g. We do not know what the consequence is of this truncation. Second, the linkers between Gal4 (AD or DBD) and the fused ORF are significantly different between different constructs, ranging from 14 amino acids in pDEST22 to 56 amino acids in pLP-GADT7, with a similar range in the bait vectors (Figure 1). Given the many fewer interactions detected using the pDES22/pDEST32 pair, their shorter linker may reduce the flexibility of the fusion protein and thus result in fewer interactions. However, this hypothesis needs to be tested by increasingly longer linker sequences and additional Y2H assays. Third, fusion proteins encoded by both pDEST as well as the pGBKT7g/pGADT7g vectors generate C-terminal tail sequences of 13 to 29 amino acids appended to bait and prey proteins, which may affect their interactions (Figure 1). Again, it remains unclear if these tails have an effect on interactions but it has been shown previously that C- vs. N-terminal fusions can prevent interactions [22,23].

4.3 Species-specific differences

In many cases, protein interactions are species-specific. First, there are many pairs of proteins that are only present in one species but not another. For example, of the 52 interactions reported

among flagellar proteins, only 39 have (predicted) interologs in *T. pallidum*, simply because not all *E. coli* proteins are present in *T. pallidum* and vice versa. Second, many interactions do not appear to be conserved, or are at least not detectable in homologous systems. For example, we detected the FliI dimerization in by Y2H system in *E.* coli but not in *T. pallidum* [19] which has a clear homolog of the *E. coli* protein. However, given the fact that we have done our genome-wide screens using *T. pallidum* proteins (in pLP vectors) and the motility array screens using *E. coli* proteins (in pDEST and pGBKT7/pGADT7 vectors), it remains impossible to state which of the detected differences is truly species-specific or merely a difference in vectors. We are planning to carry out systematic interolog testing with multiple bacterial species but identical vectors in the near future, so that this issue can be settled.

4.4 Open questions and future experiments

While this study documents the differences between various vector systems, a mechanistic explanation of these differences requires additional experiments which are beyond the scope of this paper. In particular, it remains unclear how expression levels and fusion proteins affect the quality and quantity of interactions, i.e. why do different systems generate non-overlapping interactions. The following experiments could be done to answer these questions: (1) Swap the CEN origin of pDEST22/pDEST32 and the 2μ origin of pGBKT7g/pGADT7g. (2) Swap full-length and truncated promoters in pDEST and pGBK/GAD vectors (Table 1). (3) Test different lengths of linkers between Gal4 portions and bait/prey ORFs, e.g. linkers of 17, 37, and 57 amino acids. (4) Remove C-terminal tails in both pDEST and pGBK/GAD vectors.

We have done extensive experiments involving permutations of N- and C-terminal fusions of Gal4 AD and DBD domains and found dramatic differences between them while copy number and expression levels of these fusion proteins were held constant (Stellberger & Uetz, unpublished). It is thus clear that expression levels play an important role but the nature of the fusion proteins appears to be equally important.

4.5 Implications for interactome mapping

Given the dramatic differences between different vector systems, future interactome projects may need to employ multiple vectors in parallel to achieve maximum coverage. This will be especially important with the recent demonstration that no single method can detect more than about ¼ of all interactions [5]. While the pDEST32/pDEST22 system seems to be more specific than pGBKT7g/pGADT7g, both detect a number of *bona fide* interactions that the other does not detect. Application of multiple systems also provides a built-in quality control system such that interactions found with two or more vector pairs are much more reliable. In addition to the vectors investigated here, many more variations are possible, e.g. different DBD and AD fusions such as LexA and B42 [24] or different reporter systems such as *ADE2* or *lacZ* [25]. Ideally, these systems should be combined so that the yeast two-hybrid system could become even more powerful and at the same time yield much more reliable and biologically relevant data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank B. Titz for his support with the *T. pallidum* genome-wide screens. *E. coli* ORF clones were kindly provided by Hirotada Mori. We thank J. Goll for his help with MPIDB. Stan Fields made many valuable comments on an earlier draft of this manuscript. This project was supported by The J. Craig Venter Institute and NIH grant RO1GM79710.

Abbreviations

AD	activation domain
DBD	DNA-binding domain
PPI	protein-protein interactions
Y2H	Yeast two-hybrid

References

1. Edwards AM, Kus B, Jansen R, Greenbaum D, et al. Bridging structural biology and genomics: assessing protein interaction data with known complexes. Trends Genet 2002;18:529–536. [PubMed: 12350343]
2. Cornell M, Paton NW, Oliver SG. A critical and integrated view of the yeast interactome. Comp Funct Genom 2004;5:382–402.
3. Uetz P, Dong YA, Zeretzke C, Atzler C, et al. Herpesviral protein networks and their interaction with the human proteome. Science 2006;311:239–242. [PubMed: 16339411]
4. Rozen R, Sathish N, Li Y, Yuan Y. Virion-wide protein interactions of Kaposi's sarcoma-associated herpesvirus. J Virol 2008;82:4742–4750. [PubMed: 18321973]
5. Braun P, Tasan M, Dreze M, Barrios-Rodiles M, et al. An experimentally derived confidence score for binary protein-protein interactions. Nat Methods 2009;6:91–97. [PubMed: 19060903]
6. Barrios-Rodiles M, Brown KR, Ozdamar B, Bose R, et al. High-throughput mapping of a dynamic signaling network in mammalian cells. Science 2005;307:1621–1625. [PubMed: 15761153]
7. Ramachandran N, Raphael JV, Hainsworth E, Demirkan G, et al. Next-generation high-density self-assembling functional protein arrays. Nat Methods 2008;5:535–538. [PubMed: 18469824]
8. Titz B, Rajagopala SV, Goll J, Hauser R, et al. The binary protein interactome of Treponema pallidum--the syphilis spirochete. PLoS ONE 2008;3:e2292. [PubMed: 18509523]
9. Rajagopala, SV.; Titz, B.; Uetz, P. Yeast Gene Analysis. 2. 2007. p. 139-163.
10. McKevitt M, Patel K, Smajs D, Marsh M, et al. Systematic cloning of Treponema pallidum open reading frames for protein expression and antigen discovery. Genome Res 2003;13:1665–1674. [PubMed: 12805273]
11. James P, Halladay J, Craig EA. Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. Genetics 1996;144:1425–1436. [PubMed: 8978031]
12. Harper JW, Adami GR, Wei N, Keyomarsi K, Elledge SJ. The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. Cell 1993;75:805–816. [PubMed: 8242751]
13. Cagney G, Uetz P. High-throughput two-hybrid screening using yeast arrays. Current Protocols in Protein Science 2001;19:1–12. [PubMed: 18429143]
14. Liu Q, Li MZ, Liu D, Elledge SJ. Rapid construction of recombinant DNA by the univector plasmid-fusion system. Methods Enzymol 2000;328:530–549. [PubMed: 11075365]
15. Fields S, Song O. A novel genetic system to detect protein-protein interactions. Nature 1989;340:245–246. [PubMed: 2547163]
16. Criekinge, Wv; Beyaert, R. Yeast Two-Hybrid: State of the Art. Biol Proced Online 1999;2:1–38. [PubMed: 12734586]
17. Legrain P, Dokhelar MC, Transy C. Detection of protein-protein interactions using different vectors in the two-hybrid system. Nucleic Acids Res 1994;22:3241–3242. [PubMed: 8065941]
18. Tornow J, Santangelo GM. Efficient expression of the Saccharomyces cerevisiae glycolytic gene ADH1 is dependent upon a cis-acting regulatory element (UASRPG) found initially in genes encoding ribosomal proteins. Gene 1990;90:79–85. [PubMed: 2199331]
19. Rajagopala SV, Titz B, Goll J, Parrish JR, et al. The protein network of bacterial motility. Mol Syst Biol 2007;3:128. [PubMed: 17667950]
20. Venkatesan K, Rual JF, Vazquez A, Stelzl U, et al. An empirical framework for binary interactome mapping. Nat Methods 2009;6:83–90. [PubMed: 19060904]

21. Goll J, Rajagopala SV, Shiau SC, Wu H, et al. MPIDB: the microbial protein interaction database. Bioinformatics 2008;24:1743–1744. [PubMed: 18556668]

22. Brown MA, MacGillivray RT. Vectors for expressing proteins at the amino-terminus of an activation domain for use in the yeast two-hybrid system. Anal Biochem 1997;247:451–452. [PubMed: 9177714]

23. Millson SH, Truman AW, Piper PW. Vectors for N- or C-terminal positioning of the yeast Gal4p DNA binding or activator domains. Biotechniques 2003;35:60–64. [PubMed: 12866406]

24. Fashena SJ, Serebriiskii IG, Golemis EA. LexA-based two-hybrid systems. Methods Enzymol 2000;328:14–26. [PubMed: 11075335]

25. Fang Y, Macool DJ, Xue Z, Heppard EP, et al. Development of a high-throughput yeast two-hybrid screening system to study protein-protein interactions in plants. Molecular Genetics and Genomics 2002;267:142–153. [PubMed: 11976957]

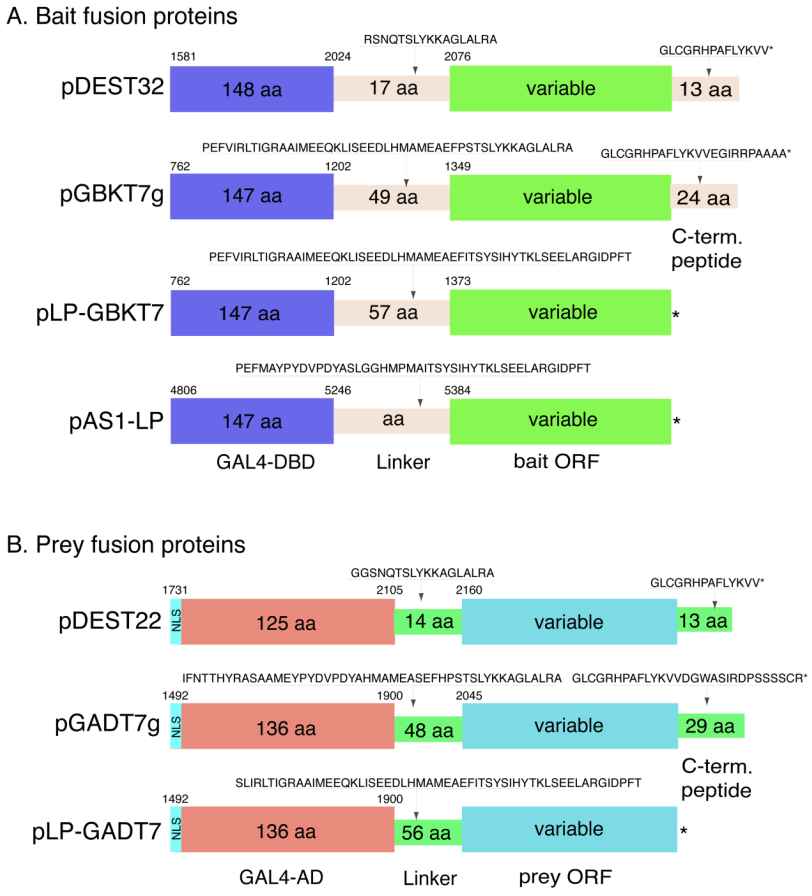

Figure 1. The structure of the bait and prey fusion proteins used in this study

All fusions are based on Gal4 activation domain (AD, preys: **A**) and Gal4 DNA binding domains (DBD, baits: **B**). Prey fusions mainly differ by the size of the linker between the AD and the prey ORF and the C-terminal peptide fused to these ORFs. **NLS** = nuclear localization signal. The amino acid sequence of the linker region between Gal4 AD or DBD and the prey or bait ORF is indicated above each fusion. Similarly, the C-terminal peptide derived from the vector is shown. All loxP-containing vectors (pLP-GADT7, pLPGBKT7, pAS1-LP) were used with ORFs that carried their endogenous stop codons, so they do not contain any C-terminal peptide.

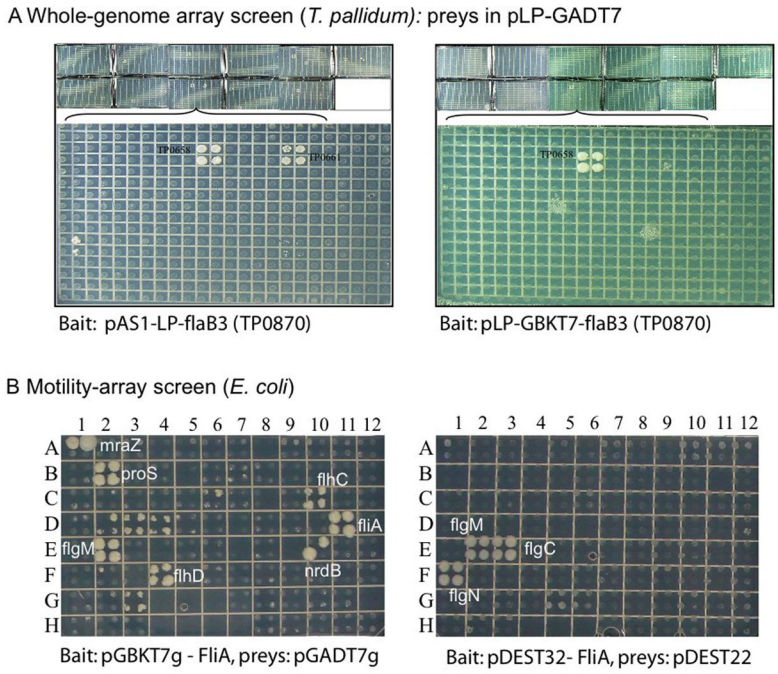

Figure 2. Different vectors yield different results in genome-wide screens and motility-specific array screens

(**A**) *Treponema pallidum* whole-genome array screen as described in [8]. All *T. pallidum* ORFs were expressed in pLP-GADT7g and screened with 49 *T. pallidum* flagellum baits cloned into pLP-GBKT7 and pAS1-LP bait vectors. Two out of 98 screens (2 x 49) are shown in the top bar; a single plate (out of 11) is shown enlarged. Left: pLP-GBKT7-FliC (ORF TP0870), right: pAS1-LP-FliC (ORF TP0870). (**B**) *E. coli* motility-specific prey arrays expressed only 96 known and predicted motility-related prey proteins expressed in the pGADT7g (left) and pDEST32 (right) prey vectors. In contrast to (A), both bait and prey vectors were different although the protein pairs in the left and right panels are exactly the same. That is, except that the left panel has protein pairs expressed in pGBKT7g/pGADT7g vectors and the right panel in pDEST32/pDEST22 vectors. Shown here are screens with *E. coli* FliA, an RNA polymerase sigma factor for flagellar operons, that was screened against all 90 *E. coli* motility preys. The pDEST32/pDEST22 pairs show markedly different and overlapping interactions compared to the pGBKT7g/pGADT7g pairs. Strong and reproducible interactions are labeled, weak and potentially spurious interactions in the pGBKT7g/pGADT7g panel are not labeled

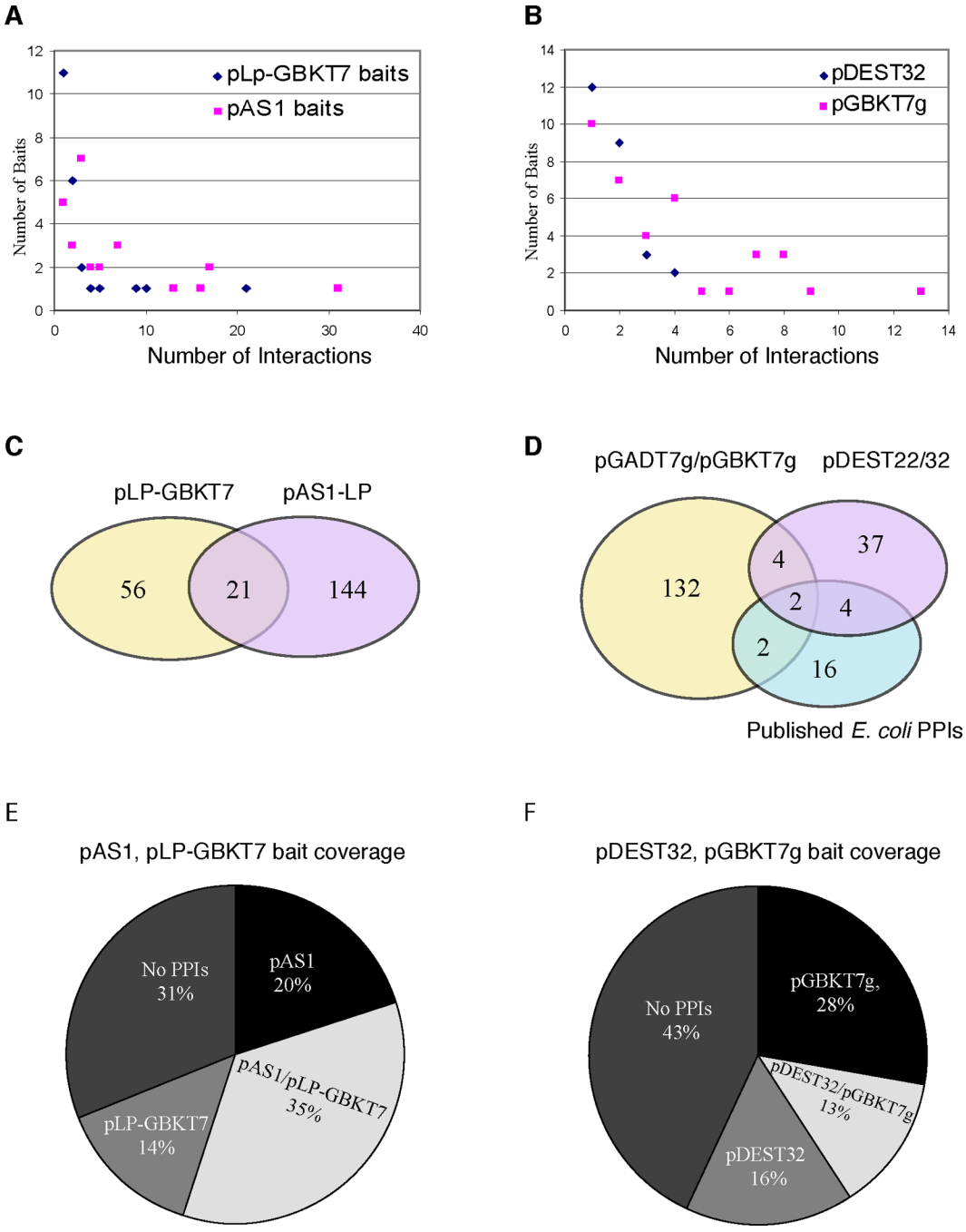

Figure 3. Summary and comparison of screening results with different vectors

(**A**) Number of interactions with individual baits found in *T. pallidum* whole-genome screens with pAS1-LP and pLP-GBKT7 bait constructs. On average, pAS1 baits produce more interactions than pLP-GBKT7 baits. (**B**) Number of interactions in *E. coli* motility array screens, plotted for individual baits. Note that all pDEST32 baits were tested against 90 pDEST22 preys; similarly, all pGBKT7g baits were tested against pGADT7g preys. pGBKT7g baits generally show more interactions than pDEST32. (**C,D**) Overlapping interactions between different datasets. (**C**) Overlap between the total numbers of interactions from 49 screens using motility proteins as baits (in bait vectors pLP-GBKT7 and pAS1-LP) against the whole-genome *T. pallidum* array. (**D**) Overlap between *E. coli* motility array screens using

bait/prey vector pairs pGBKT7g/pGADT7g and pDEST32/pDEST22. Note that exactly the same set of proteins pairs (i.e. *E.coli* flagellum proteins) was tested. 24 published interactions among *E. coli* flagellar proteins (from MPIDB [21]) are included as gold-standard dataset. Note that despite the significant difference in total interactions, the overlap with the gold-standard set is very similar. (**E,F**) Fraction of baits that yielded interaction data in each of the whole genome (**E**) or motility array screens (**F**). For example, in the whole-genome screens in *T. pallidum*, 20% of all baits yielded interactions only as pAS1-LP baits, while 35% of all baits yielded interactions with both pAS1-LP and pLP-GBKT7. Note that the overlap between pAS1-LP and pLP-GBKT7 (**E**) was significantly larger than between the pDEST32/pDEST22 and pGBKT7g/pGADT7g pairs (**F**).

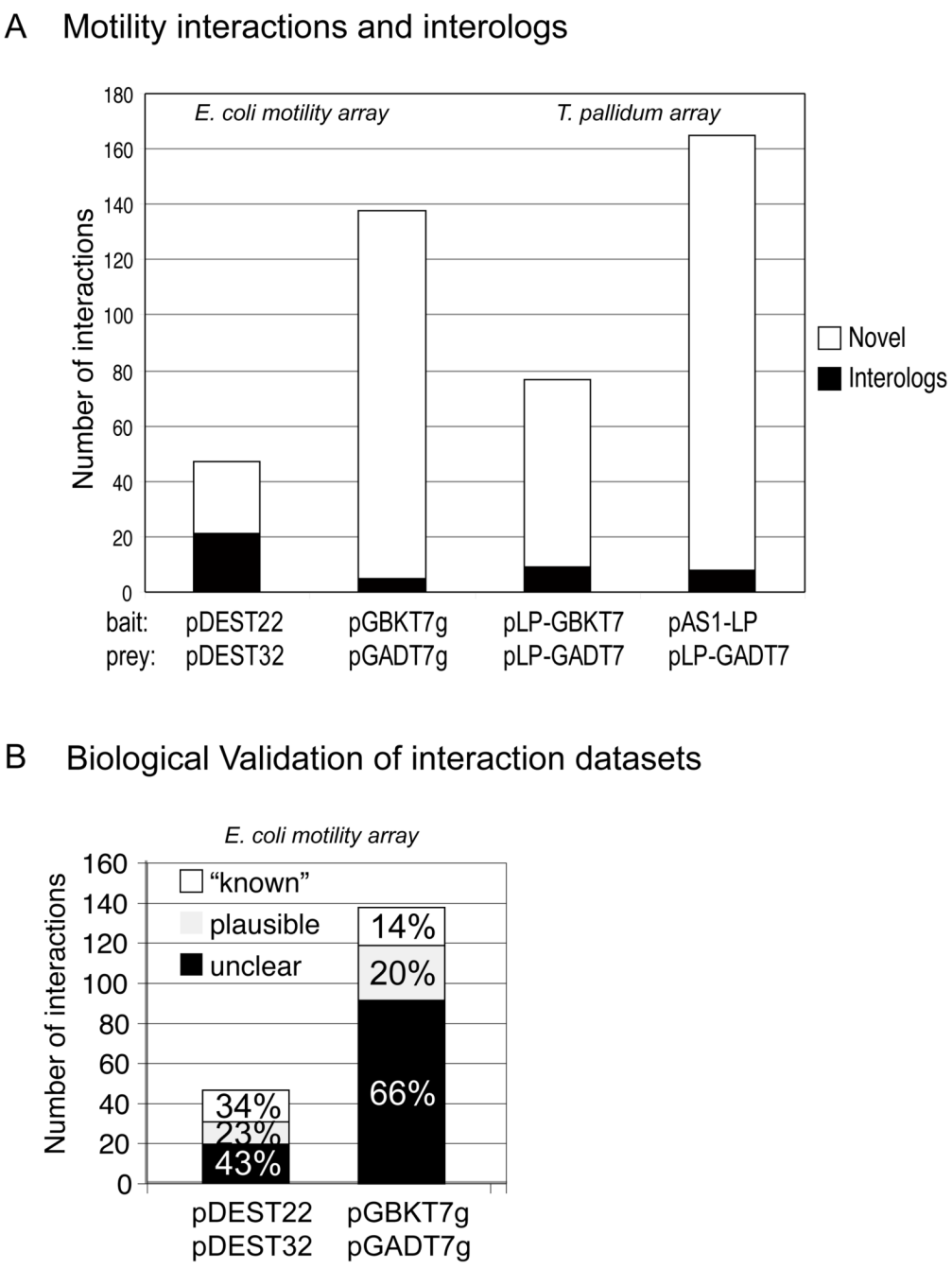

Figure 4. Validation of two-hybrid interactions by interologs and biological evaluation

(**A**) Interaction data are from motility arrays (*E. coli*) or genome-wide screens (in *Treponema pallidum*) as validated by homologous interactions ("interologs") in other species (black). Interactions in open segments do not have interologs in other species. All interologs have been shown experimentally and have been derived from MPIDB [21]. (**B**) Interactions data from motility array screens were classified into one of three classes: "known", plausible, and unclear (unknown). Most interactions (34% + 23% = 57%) detected with pDEST22/pDEST32 were either known or plausible while only 34% (14+20%) of the interactions detected with pGBKT7/pGADT7 were assigned to this class (see Table S2).

Table 1

Y2H vectors

Baits contain DNA-binding domains (DBD) and preys contain activation domains (AD).

		Gal4-Fusion		Selection			
Vector	**Promoter**	**DBD**	**AD**	**yeast**	**bacterial**	**ori**	**Source**
pDEST22	fl-ADH1	-	X	Trp1	Amp	CEN	Invitrogen
pDEST32*	fl-ADH1	X	-	Leu2	Gentam.	CEN	Invitrogen
pGBKT7g	t-ADH1	X	-	Trp1	Kan	2μ	[9]
pGADT7g	fl-ADH1	-	X	Leu2	Amp	2μ	[9]
pAS1-LP*	fl-ADH1	X	-	Trp1	Amp	2μ	[8]
pLP-GADT7	fl-ADH1	-	X	Leu2	Amp	2μ	Clontech
pLP-GBKT7	t-ADH1	X	-	Trp1	Kan	2μ	Clontech

* also encodes CYH2; fl-, t-ADH1 = full length and truncated ADH1 promotors. The bacterial origin in all cases is from pUC (also called ColE1). The pDEST, pGBKT7g, and pGADT7g vectors are Gateway-compatible (as indicate by the “g”) while “LP” indicates loxP sites for recombinational insertion of bait and prey ORFs.