EffectorK, a comprehensive resource to mine for pathogen effector targets in the Arabidopsis proteome

Manuel González-Fuente[1¶], Sébastien Carrère[1¶], Dario Monachello[2,3], Benjamin G. Marsella[4], Anne-Claire Cazalé[1], Claudine Zischek[1], Raka M. Mitra[5], Nathalie Rezé[2,3] ,Ludovic Cottret[1], M. Shahid Mukhtar[4], Claire Lurin[2,3], Laurent D. Noël[1*] and Nemo Peeters[1*]

[1] Laboratoire des Interactions Plantes Micro-organismes (LIPM), INRAE, CNRS, Université de Toulouse, F-31326 Castanet-Tolosan, France

[2] Institut des Sciences des Plantes de Paris Saclay (IPS2), UEVE, INRAE, CNRS, Université Paris Sud, Université Paris-Saclay, F-91192 Gif-sur-Yvette, France

[3] Université de Paris, IPS2, F-91192 Gif-sur-Yvette, France

[4] Department of Biology, University of Alabama at Birmingham, 1300 University Blvd., Birmingham, AL 35294, USA

[5] Department of Biology, Carleton College, One North College Street, Northfield, MN 55057, USA

* Corresponding authors

E-mail: nemo.peeters@inra.fr (NP) and laurent.noel@inra.fr (LDN)

¶These authors contributed equally to this work.

Abstract

Pathogens deploy effector proteins that interact with host proteins to manipulate the host physiology to the pathogen's own benefit. However, effectors can also be recognized by host immune proteins leading to the activation of defense responses. Effectors are thus essential components in determining the outcome of plant-pathogen interactions. Despite major efforts to decipher effector functions, our current knowledge on effector biology is scattered and often limited. In this study, we conducted two systematic large-scale yeast two-hybrid screenings to detect interactions between *Arabidopsis thaliana* proteins and effectors from two vascular bacterial pathogens: *Ralstonia pseudosolanacearum* and *Xanthomonas campestris*. We then constructed an interactomic network focused on Arabidopsis and effector proteins from a wide variety of bacterial, oomycete, fungal and animal pathogens. This network contains our experimental data and protein-protein interactions from 2,035 peer-reviewed publications (48,200 Arabidopsis-Arabidopsis and 1,300 Arabidopsis-effector protein interactions). Our results show that effectors from different species interact with both common and specific Arabidopsis targets suggesting dual roles as modulators of generic and adaptive host processes. Network analyses revealed that effector targets, particularly effector hubs and bacterial core effector targets, occupy important positions for network organization as shown by their larger number of protein interactions and centrality. These interactomic data were incorporated in EffectorK, a new graph-oriented knowledge database that allows users to navigate the network, search for homology or find possible paths between host and/or effector proteins. EffectorK is available at www.effectork.org and allows users to submit their own interactomic data.

Author summary

Plant pests and diseases caused by bacteria, oomycetes, fungi or animals are threatening food security worldwide. Understanding how these pathogens infect and manipulate the host is key to develop sustainable crop resistance in the long term. Effector proteins are secreted by pathogens to subvert the host immune responses. The roles of several effector proteins have been described; however, it is yet poorly understood how effectors interact with host proteins at a global level. To address this issue, we have generated EffectorK, an interactive database focused on the model plant species *Arabidopsis thaliana*. This database contains manually curated Arabidopsis-effector protein interactions from the available literature on a wide variety of pathogens. It also contains new experimental data on effectors from two vascular pathogens: *Ralstonia pseudosolanacearum* and *Xanthomonas campestris*. This work integrates all the gathered knowledge over the last decades and allows to identify general patterns of how effectors interact with the host proteome. This knowledge is easily accessible and searchable at www.effectork.org.

Introduction

Plants are continuously confronted with a wide variety of pathogens including bacteria, oomycetes, fungi, nematodes and insects. To prevent their proliferation, plants have evolved a complex multilayered immune system [1]. The first layer of defense corresponds to constitutive physical and chemical barriers such as the cuticle, cell wall or secondary metabolites [2,3]. Plants are also able to recognize highly conserved pathogen-associated molecular patterns (PAMPs) through pattern-recognition receptors triggering induced defense responses collectively known as

60 'PAMP-triggered immunity' (PTI) [4]. These responses are usually enough to prevent most
61 potential invaders; however, some pathogens secrete effector proteins to subvert the defense
62 responses and alter diverse cellular processes to ease their proliferation [5]. Plants, on the other
63 hand, have evolved several intracellular nucleotide-binding site-leucine-rich repeat (NBS-LRR)
64 receptors recognizing these effectors and activating potent defense responses collectively known
65 as 'effector-triggered immunity' (ETI) [6].

66 Although the targets and molecular functions of some effectors have been well
67 characterized [7–10], for most effectors they are still unknown. The main factors complicating the
68 large-scale identification and characterization of effector-host protein interactions are: the wide
69 diversity of pathosystems, the difficulty to identify *bona fide* effector genes, the collective
70 contribution of effector proteins, the complexity of the host responses and the lack of robust high
71 throughput techniques. For the model species *Arabidopsis thaliana* (*Ath*), to our knowledge, there
72 are only two studies in which systematic effector-host protein interactions at the effectome-scale
73 have been identified [11,12]. In these studies plant targets of effector proteins from *Pseudomonas*
74 *syringae* (*Psy*, bacterium), *Hyaloperonospora arabidopsidis* (*Hpa*, oomycete) and *Glovinomyces*
75 *orontii* (*Gor*, fungus) were identified by yeast two-hybrid (Y2H). They reported that the effectors
76 of these three species converged onto a limited set of *Ath* proteins. These studies also demonstrated
77 that many effector targets are important for plant immunity and showed that their importance
78 correlates with the level of effector convergence.

79 Bacterial wilt, caused by *Ralstonia pseudosolanacearum* (*Ralstonia solanacearum*
80 phylotype I, *Rps*), and black rot, caused by *Xanthomonas campestris* pathovar *campestris* (*Xcc*)

81 are listed among the top five plant bacterial diseases in the world [13]. Both *Rps* and *Xcc* are
82 xylem-colonizing pathogens and rely on their type III secretion systems for full virulence [14,15].
83 This 'molecular syringe' allows the pathogen to deliver type III effector proteins (T3Es) directly
84 into the host cell in order to promote disease. The roles of several of their T3Es have been
85 characterized [16,17], but most knowledge on T3E functions comes from the study of *Psy*, which
86 resides on leaf surfaces and in the leaf apoplast [7,18]. Focusing mainly on a few species offers a
87 partial view of effector biology. It is therefore crucial to expand our studies to other species to
88 grasp most of the existing diversity of effector proteins and pathogen lifestyles.

89 To obtain a deeper understanding of the global *Ath*-effector protein interactome, we
90 conducted two systematic large-scale screenings with T3Es from *Rps* and *Xcc*, the first vascular
91 pathogens screened in this manner. Additionally, we conducted an extensive literature survey to
92 gather published *Ath* targets of effector proteins from pathogens from four different kingdoms of
93 life: Bacteria, Chromista, Fungi and Animalia. Combining all these data allowed us to identify 100
94 new 'effector hubs' (i.e., *Ath* proteins interacting with two or more effectors). Together with *Ath*-
95 *Ath* protein interactions retrieved from public databases, we generated a comprehensive *Ath*-
96 effector protein network that captures the wide diversity of *Ath* pathogens. This network allowed
97 us to detect general trends of effector interference with the host proteome. We have created a
98 publicly available interactive knowledge database called EffectorK (for Effector Knowledge)
99 which allows users to access and augment this network.

Results

Systematic identification of Arabidopsis targets of *R. pseudosolanacearum* and *X. campestris* effectors.

Multiple Y2H screenings were performed to identify *Ath* targets of *Rps* and *Xcc* effector proteins. In a first screening, we identified 42 *Ath* targets for 21 out of 56 T3Es from *Rps* strain GMI100 screened against a library of more than 8,000 full-length *Ath* cDNAs (8K space). In the second and third screenings, we identified 176 *Ath* targets for 32 out of 48 T3Es from *Rps* strain GMI1000 and 52 *Ath* targets for 18 out of 25 T3Es from *Xcc* strain 8004 screened against an extended version of the previous library containing more than 12,000 *Ath* full-length cDNAs (12K space) (S1 Fig and S1 Table). On average, T3Es from *Rps* interacted with 10.7 *Ath* proteins while T3Es from *Xcc* interacted with 5.3 *Ath* proteins. These *Ath* cDNA libraries had been previously used to test interactions with effector proteins from *Hpa*, *Psy* (8K space) and *Gor* (12K space) [11,12]. The subset of interactions of effectors from *Rps*, *Xcc* and *Gor* in the 8K space was used to compare with previously published *Hpa* and *Psy* data (Fig 1). In general, *Rps* effectors interacted on average with more *Ath* proteins than the other screened species; however, this difference is only statistically significant when compared to *Gor* effectors (one-tailed Wilcoxon signed-rank test p-value = 0.0005). These data show that effector proteins from these five different species, on average, tend to interact with a similar number of *Ath* proteins regardless the kingdom, lifestyle or effectome size.

Effectors converge onto a limited set of Arabidopsis proteins

We compared the *Rps* and *Xcc* effector targets identified in our screenings with the targets previously identified for *Hpa*, *Psy* and *Gor* effector proteins [11,12]. To avoid bias related to the size of the screened library, we considered only the subset of effector targets present in the 8K space (S2 Fig). At the kingdom level, the highest target specificity was found in Bacteria with 158 exclusive out of a total of 217 targets (72.8%) followed by Chromista, with 31 out of 117 (51.7%), and Fungi, with 16 out of 45 (35.6%). In total, 235 out of 299 effector targets (78.6%) were kingdom-specific. At the species level, when comparing all five pathogens, the percentage of exclusive targets was 58.9% for *Psy*, 58.7% for *Rps*, 51.7% for *Hpa*, 48.8% for *Xcc* and 35.6% for *Gor*. The total number of species-specific effector targets was 221 out of 299 (73.9%). These data show that most effector targets are kingdom- and species-specific.

To evaluate whether *Rps* and *Xcc* effectors interact randomly or converge onto a common set of *Ath* protein we performed simulations rewiring effector-*Ath* protein interactions. In these simulations, each effector was assigned randomly as many *Ath* proteins as it had interacted with in our screenings. Then, the number of targets found on all simulations was plotted and compared with the experimental data (Fig 2A). The number of effector targets observed in our screenings was significantly lower than the numbers obtained in the random simulations for both *Rps* and *Xcc*. Similar results had been reported for effectors from *Hpa*, *Psy* and *Gor* [11,12]. This shows that, similarly to other species, both *Rps* and *Xcc* effectors also interact with a common subset of *Ath* proteins (i.e., intraspecific convergence).

These random rewiring simulations also allowed us to determine whether effectors from different species interact randomly or convergently with *Ath* proteins. For this, the number of common interactors of effectors from different species was compared with the experiment data (Fig 2B). When comparing all three kingdoms, the number of common targets observed was significantly higher than expected by random rewiring. We then analyzed all possible binary, ternary, quaternary and quinary combinations of species and in all cases, the number of common targets observed was higher than expected randomly (Fig 2C). These differences were all statistically significant except for the common targets of effectors from *Psy* and *Xcc* (p-value = 0.0579) (S3 Fig). This could indicate that these two species are the most different in terms of effector targeting. However, considering that *Psy* and *Xcc* are precisely the two species with the lowest number of effectors for which targets have been identified (*Psy*: 32 and *Xcc*: 18 effector proteins), it is likely that the high p-value is caused by the limited sample size. This shows that effectors from all these five species interact with a common subset of *Ath* proteins (i.e., interspecific convergence).

Altogether, our data indicate that *Rps* and *Xcc* effectors converge both intra- and interspecifically onto a set of limited *Ath* proteins, behaving similarly to effectors from other previously screened pathogen species. This suggests the existence of a convergent set of effector targets common to evolutionary distant pathogens that might have a predominant role in the general modulation of the host responses.

Manual curation of the literature to gather new Arabidopsis-effector protein interactions

In order to gather more knowledge on *Ath*-effector protein interactions, we conducted an extensive literature search compiling data from a wider spectrum of bacterial, fungal, oomycete and animal effector proteins. We only considered published direct protein-protein interactions that had been confirmed by classic techniques such as Y2H, co-immunoprecipitation, pull-down, protein-fragment complementation, fluorescence resonance energy transfer or mass spectrometry. We compiled 287 interactions found in 80 peer-reviewed publications involving 218 *Ath* proteins and 72 effectors from 22 pathogen species (S2 Table). Among these 22 pathogens, there were nine bacterial species, mostly proteobacteria but also a phytoplasma species; eight animal species including both nematodes and insects; four oomycete and one fungal species. While this collection of species does not represent the full diversity of *Ath* pathogens, it covers the majority of pathogens for which effector targets have been found. We can notice that despite being one of the major pathogen classes, few studies have described fungal effector interactors. This illustrates one of the current gaps in our knowledge of effector targets.

Identification of 100 new effector hubs

To compare experimental and published data, we combined all the interactions curated from the published data together with data from our large-scale Y2H screenings. This resulted in a total of 564 different *Ath* proteins targeted by pathogen effectors. Our screenings on *Rps* and *Xcc* effectors identified 235 targets. Similar published screenings on *Psy*, *Gor* or *Hpa* effectors had identified 200 targets [11,12]. The literature curation allowed us to identify 218 effector targets. From the 235 *Rps* and *Xcc* effectors targets found in our screening, 166 were new which represents 29.4% of the total targets compiled in this study (Fig 3). This highlights the potential of such

systematic and high throughput large-scale screenings in identifying novel effector targets. The average effector degree (i.e., number of effectors interacting with an *Ath* protein) was 2.3 but it was unevenly distributed among the 564 targets with 350 of them interacting with only one effector (62%) and 14 interacting with more than 10 effectors (2.5%) (S4 Fig). The contribution of our experimental data was important in the identification of single targets as we identified 93 out of the 350 (26.6%). More remarkable was our contribution in the identification of "effector hubs", what we defined as *Ath* proteins interacting with two or more effectors (Fig 4). The definition of hub has been debated and it has been traditionally associated with proteins that are highly connected in interactomic networks [19]. Our definition of "effector hub" came from the need to designate the *Ath* proteins that interact with several effectors and is based exclusively on the number of interacting effector proteins. We identified 100 new effector hubs and increased the degree of 42 previously described effector hubs (S3 Table).

To evaluate the potential relevance of the newly identified effector hubs in plant immunity, we conducted a second literature survey to check if the corresponding *Ath* genes had been previously characterized to be involved in plant immunity or pathogen fitness *in planta* (Table 1). 16 out of the 100 new effector hub genes, have already been described for their altered infection or other immunity-related phenotype when mutated, silenced or overexpressed. Additionally, the orthologs of 3 other new hubs in other plant species, also produced altered infection phenotypes when silenced or overexpressed. A total of 19 out of the 100 newly identified effector hubs have already been shown to be involved in biotic stress responses. Considering that many of the remaining newly defined effector hubs have been poorly characterized (e.g., hypothetical proteins

201 or descriptions based on homology or belonging to a protein family), it is likely that the number
202 of effector hubs involved in immunity was underestimated. This constitutes a valuable source of
203 novel candidates for further functional characterization.

204 **Table 1. List of 19 new effector hubs involved in plant immunity.**

Effector hub	Protein name	Effector degree[a]	Description of observed phenotype	Reference
AT1G58100	TCP domain protein 8 (TCP8)	13	Triple *tcp8 tcp14 tcp15* mutant showed enhanced *Pseudomonas syringae* strain DC3000 Δ*avrRps4* growth.	[20]
AT1G71230[b]	COP9-signalosome 5B (CSN5B)	8	Wheat *TaCSN5* mutant showed enhanced disease symptoms caused by *Puccinia triticina*.	[21]
AT3G12920	BOI-related gene 3 (BRG3)	7	*brg3* mutant showed increased *Botrytis cinerea* lesion size.	[22]
AT5G08330[b]	TCP domain protein 21 (TCP21)	7	Rice *OsTCP21* silenced and overexpressing plants showed enhanced and reduced disease symptoms caused by rice rust stunt virus (RRSV) respectively.	[23]
AT5G61010	Exocyst subunit EXO70 family protein E2 (EXO70E2)	6	*exo70e2* mutant showed reduced flg22-induced callose deposition.	[24]

AT4G00270	STOREKEEPER-related 1 (STKR1)	6	*STKR1* overexpressing plants showed reduced *Hyaloperonospora arabidopsidis* spore formation.	[25]
AT3G01670	SIEVE ELEMENT OCCLUSION-related 2 (SEOR2)	4	*Myzus persicae* feeding from *seor2* mutant showed reduced progeny.	[26]
AT5G17490	RGA-like protein 3 (RGL3)	3	*rgl3* mutant showed reduced *P. syringae* growth and increased SA content upon infection.	[27]
AT3G54230	Suppressor of *abi3-5* (SUA)	3	*sua* mutant showed enhanced *P. syringae* growth and reduced chitin-induced ROS production.	[28]
AT3G11410	Protein phosphatase 2CA (PP2CA)	3	*pp2ca* mutant showed reduced *P. syringae* colonization and stomatal aperture. *PP2CA* overexpressor showed enhanced stomatal aperture.	[29]
AT2G17290	Calcium-dependent protein kinase 6 (CPK6)	3	Double *cpk5-cpk6* mutant showed enhanced *P. syringae* growth and reduced flg22-induced ROS production.	[30]
AT5G41410[b]	Homeobox protein BEL1 homolog (BELL1)	3	Rice *OsBIHD1* mutant and overexpressing plants showed increased and reduced *Magnaporthe oryzae* lesion area respectively.	[31]
AT4G26750	LYST-interacting protein 5 (LIP5)	2	*lip5* mutant showed enhanced *P. syringae* growth and disease symptoms and reduced endosomal structure formation upon infection.	[32]

AT4G35090	Catalase-2 (CAT2)	2	*cat2* mutant showed increased ROS accumulation upon infection with incompatible *P. syringae* strain.	[33]
AT3G02870	Inositol-phosphate phosphatase (VTC4)	2	*vtc4* mutant showed reduced *P. syringae* growth.	[34]
AT5G53060	Regulator of CBF gene expression 3 (RCF3)	2	*rcf3* mutant showed reduced percentage of diseased plants and higher percentage of plant survival upon *Fusarium oxysporum* infection.	[35]
AT3G02540	RAD23 family protein C (RAD23C)	2	*rad23BCD* mutant (and not *rad23BD*) did not show *Candidatus Phytoplasma*-induced flower virescence and phyllody.	[36]
AT5G38470	RAD23 family protein D (RAD23D)	2	*rad23D* mutant did not show flower virescence and phyllody upon transgenic expression of *C. Phytoplasma* SAP54 effector.	[36]
AT2G37630	Asymmetric leaves 1 (AS1)	2	*as1* mutant showed reduced lesion size caused by *B. cinerea* and *Alternaria brassicicola* and enhanced *Pseudomonas fluorescens* and *P. syringae* growth.	[37]

[a] Ranked in decreasing order.

[b] Orthologous gene in other plant species, as defined by EnsemblPlants [38], characterized for a role in immunity.

In terms of organism of origin, most of the 564 targets are bacterial effector targets as it could be expected considering that 132 out of the 266 total effectors compiled came from bacteria (S4 Fig). In the case of effector hubs, it is noteworthy that 133 out of the 214 hubs described in this work are targeted by effectors from a single kingdom while there are only 64, 16 and 1 hubs interacting with effectors from 2, 3 or 4 different kingdoms respectively. Although biased by the structure of the data, this could suggest kingdom specificity of effector targeting.

Effector targets tend to occupy key positions for the network organization

We constructed an *Ath*-effector protein interaction network compiling the previously described experimental and literature-compiled data with *Ath*-*Ath* protein interactions from public databases and the literature [39–42]. From the total of 49,500 interactions compiled in this study, 48,597 were grouped into a single connected component constituting what we defined as our *Ath*-effector interactomic network (Table S4). This network was constituted of 47,314 *Ath*-*Ath* and 1,283 *Ath*-effector protein interactions between 8,036 *Ath* proteins and 245 effector proteins. Effectors came from 23 different species including bacteria (128 effectors), oomycetes (61 effectors), fungi (46 effectors) and animals (10 effectors). The uneven distribution of effectors among kingdoms highlights the contribution of the large-scale screenings in the identification of effector targets as 1,002 out of 1,283 *Ath*-effector protein interactions came from either our experimental data or previous screenings of the same library [11,12].

To further investigate the potential impact of effectors on the plant interactome, we evaluated the importance of their targets for the organization of the network. We focused on two main network topology parameters: degree and betweenness centrality (Fig 4). The degree of a

229 protein represents the number of proteins that it interacts with. In this study we differentiated two
230 types of degrees depending on the nature of the interacting proteins: the *Ath* degree of a given
231 effector or *Ath* protein (i.e., number of interacting *Ath* proteins) and effector degree for a given *Ath*
232 protein (i.e., number of interacting effector proteins). The betweenness centrality of a protein is
233 the fraction of all shortest paths connecting two proteins from the network that pass through it.
234 There are two main types of key proteins in a network [43]: 1) proteins important for local network
235 organization, typically showing high degree, and 2) proteins important for the global diffusion of
236 the information through the network, characterized by high betweenness centrality. It had been
237 previously reported in more limited networks that effectors tend to target host proteins with high
238 degree and centrality [43–45]. We then analyzed whether this was the case in our network
239 comparing effector targets with the rest of the *Ath* proteins (Fig 5). The fraction of proteins
240 decreased rapidly as the Ath degree increased. This indicates that most *Ath* proteins present low
241 *Ath* degree and only a few of them show high *Ath* degree values. This tendency was significantly
242 shifted towards higher *Ath* degree values in effector targets compared to the rest of *Ath* proteins.
243 To represent this tendency shift we estimated and compared the area under the curve values of the
244 cumulative distribution of *Ath* degree for effector targets and the rest of *Ath* proteins (Table 2).
245 Effectively, the area under the curve value of effector targets was higher than the value of the rest
246 of *Ath* proteins. This indicates that effector targets present generally higher *Ath* degree than the
247 rest of *Ath* proteins. Similarly, we compared the betweenness centrality of these two groups of
248 proteins (Table 2 and Fig S5). Effector targets also presented significantly higher betweenness
249 centrality values than the rest of *Ath* proteins. Altogether, these results indicate that effectors

250 preferentially interact with *Ath* proteins that are more connected to other *Ath* proteins and that
251 occupy more central positions in the interactomic network.

252 **Table 2. Cumulative *Ath* and effector degrees and betweenness centrality of different groups**
253 **of effector targets**

	Area under the curve[a]		**Figure[b]**	**p-value[c]**
	Effector targets	**Other *Ath* proteins**		
***Ath* degree**	2,737	1,010	5	< 0.0001
Betweenness centrality	0.23	0.033	S5A	< 0.0001
	Multi-pathogen effector targets	**Pathogen-specific effector targets**		
***Ath* degree**	5,344	1,790	S5B	< 0.0001
Betweenness centrality	0.657	0.136	S5C	< 0.0001
	Effector hubs	**Single effector targets**		
***Ath* degree**	4,067	1,810	S5D	< 0.0001
Betweenness centrality	0.407	0.118	S5E	< 0.0001
	Bacterial core T3Es	**Rest of bacterial T3Es**		
***Ath* degree**	656	712	S7A	0.4571
Betweenness centrality	0.072	0.074	S7B	0.9198
	Targets of bacterial core T3Es	**Other bacterial T3Es targets**		
Effector degree	347	123	S7C	< 0.0001
***Ath* degree**	3,610	2,714	S7D	0.0131
Betweenness centrality	0.369	0.239	S7E	0.0007

254 [a] Estimated area under the curve of the cumulative distribution of *Ath* degree, effector degree and
255 betweenness centrality for each group of proteins as represented in figures 5, S5 and S7. Estimation
256 based on numerical integration using Simpson's Rule.

[b] Figure illustrating the cumulative distribution graphic from which the areas under the curve compared were calculated.

[c] One-tailed Wilcoxon signed-rank test p-value of the comparison of the *Ath* degree, effector degree or betweenness centrality values of all proteins from each compared group.

We then wanted to test if the *Ath* degree and betweenness centrality values differed among distinct types of effector targets (Table 2 and Fig S5). First we compared multi-pathogen and pathogen-specific targets as previously described (S2 Fig). Multi-pathogen effector targets presented significantly higher *Ath* degree and betweenness centrality compared to pathogen-specific effector targets. We also compared effector hubs with single effector targets. Similarly, effector hubs also showed higher betweenness centrality and *Ath* degree than single effector interactors. This last observation implies that an *Ath* protein that interacts with several effectors tends to interact with more *Ath* proteins as well. To evaluate whether this is biologically relevant or a bias of the 'stickiness' of a protein, we compared the *Ath* and effector degree values of all targets. Our results showed that these two parameters are not correlated (Pearson correlation coefficient = 0.3221) (S6 Fig). This suggests that effector hubs interact with more *Ath* proteins than single effector targets and not because they might be stickier.

In this work, by compiling our experimental interactomic data on *Xcc* and *Rps* and the literature-curated interactions from a wide variety of other pathogen effectors, we extended the notion that effectors tend to privilege interactions with host proteins with higher *Ath* degree and betweenness centrality [43,45]. Furthermore, we showed that this tendency is stronger among effector hubs compared to single targets and among multi-pathogen effector targets compared to

pathogen-specific targets. This reflects the importance of interfering with these key position proteins in the modulation of host-pathogen interactions.

Bacterial core T3Es interact with more connected and central *Ath* proteins

Our work on *Rps* and *Xcc* together with previous work on *Psy* T3Es [11] provided a large amount of interactomic data on bacterial pathogen species for which other resources have been generated, particularly in terms of abundance and diversity of sequenced genomes and thus curated T3E repertoires [18,46–50]. The most conserved set of T3Es, or 'core effectome', from each of the three bacterial species has been previously defined [47,48,50]. We then tested whether these subsets of T3Es behaved differently from the rest of bacterial T3Es in terms of interaction with host proteins (Table 2 and Fig S7). Our data showed that core and variable T3Es from the three species do not differ in *Ath* degree nor betweenness centrality. We then tested if there were any differences between the network properties of the targets of core T3Es and the other bacterial T3E targets. Core T3Es targets showed higher effector degree, *Ath* degree and betweenness centrality than the rest of targets of bacterial T3Es. This suggests that, although core T3Es in general do not have more targets than the rest of bacterial T3Es, they do interact with more highly connected and central *Ath* proteins. This might imply that core T3Es have a larger potential to interfere with the host interactome what could explain the selective pressure to maintain them in the majority of strains.

296 **EffectorK, an online interactive knowledge database to explore the Arabidopsis-effector**
297 **interactomic data**

298 In order to facilitate the access and exploration of all the data presented in this work, we
299 have generated EffectorK (for 'Effector Knowledge'), an interactive web-based knowledge
300 database freely available at www.effectork.org. The latest version (October 2, 2019) contains
301 49,875 interactions 8,617 proteins coming from 2,035 publications. From these, 1,300 are *Ath*-
302 effector protein interactions. Searches can be done based on a wide range of supported identifiers
303 such as different protein names, NCBI or TAIR accession numbers, PubMed identifiers and
304 InterPro terms. Additionally, users can also query nucleotide or amino acid sequences directly with
305 BLAST or use accession numbers from other model and crop plants to find homologs within the
306 database. All proteins found by query are then listed in tabular format and hyperlinked to the
307 corresponding interactomic data, external resources and amino acid sequences. Interactomic data
308 for a given protein can be then explored and downloaded in graphical or tabular format. The visual
309 interface for the graphical representation of the interactomic data allows users to expand or re-
310 center a local subnetwork based on a given protein, get information and access to external
311 resources linked to either a protein (node) or an interaction (edge) or modify the layout and the
312 position of the elements for optimal visualization. Additionally, EffectorK also allows users to find
313 the shortest paths between two queried proteins in the network.

314 In order to update, expand and further improve EffectorK, we encourage users to submit
315 their own interactomic data by filing in and sending the dedicated template. These data will be
316 verified by the curator team prior to their incorporation in the database. More information about

usage, content and data submission is accessible online, under the tabs 'Help' and 'Contribute' of the database web server. Please contact us if you have any question or suggestions by email at contact@effectork.org.

Discussion

In this study we identified systematically *Ath* targets of effectors from the vascular bacterial pathogens *Rps* and *Xcc*. We combined this information with other *Ath* targets identified in similar experimental setups. Additionally, we conducted an extensive literature review to gather published *Ath* targets of effectors from a wide variety of pathogens including other bacterial species and also oomycete, fungal and animal pathogens. Studying this combined interactomic dataset allowed us to identify new trends of how effectors interfere with the plant proteome and evaluate whether previously described network principles were still supported on a wider scale. We showed that there are no substantial differences in terms of connectivity among the effectomes of five different pathogen species screened systematically (Fig 1). We have reinforced previously described intra- and interspecific convergence of effector targeting with effectors from two new species [11,12], and showed at the same time that most effector targets are pathogen specific (Fig 2 and S2). Our analyses also supported the previously described tendency of effectors to interact with plant proteins better connected and central in the network [43,45], and showed that this tendency is even stronger among effector hubs, multi-pathogen targets and bacterial core T3E targets (Table 2 and Fig S5).

The balance between target specificity and convergence

Our data showed that most effector targets were pathogen-specific (S2 Fig) but at the same time, effectors converge interspecifically onto a small subset of *Ath* proteins (Fig 2B-C). These *a priori* contradictory observations open an interesting question: what is the balance between specificity and convergence of effector targets? At this point, it is impossible to assert whether this specificity is merely caused by the limited number of pathogens screened at the effectome-scale or if it is a reflection of the different and unique ways that each pathogen has evolved to interfere with the host physiology and immunity. This issue can only be addressed by increasing the number of pathogen effectors screened thoroughly and at a large-scale. Comparing large datasets of effector targets of a wider and more diverse set of pathogens would allow evaluating in which sense this balance between specificity and convergence tilts: 1) If the target specificity decreased, it would mean that the effectomes from the different pathogens tend to interact similarly with the host proteome. This was the case when we compared the percentage of species-specific targets of effectors from *Hpa*, *Psy* and *Gor* that passed from being 73.9%, 64.9% and 46.7% in previous works [11,12], to 51.7%, 58.9% and 35.6%, respectively in the present study (S2 Fig). Nevertheless, a total of five screened species is probably not powerful enough to sustain this claim. 2) If, on the contrary, the target specificity increased with the number of screened species, it would mean that the different pathogens have evolved unique ways to modulate the interaction with the host. If this were be the case, deeper analyses comparing related pathogens (e.g., species with similar lifestyle or from the same kingdom) could allow identifying trait-specific targets (e.g., effector targets exclusive among vascular pathogen effectors). In any case, to better understand the

357 similarities and particularities on how effectors modulate host processes, it is essential to increase
358 the number of pathogen species screened for effector targets at the effectome-scale.

359 **Large-scale screenings fill the gap in the identification of effector targets**

360 Including manually curated data from literature has allowed us to broaden significantly the
361 diversity of plant pathogen species compared to similar studies. However, 346 out the 564
362 described *Arabidopsis* effector targets have been identified exclusively through large-scale Y2H
363 screenings against partial libraries of *Ath* cDNAs. As with any other large-scale screening, the
364 technical limitations together with the incompleteness of the library might have probably led to an
365 underestimation of the plant-effector interactome of the five screened species [51]. The relatively
366 small overlap between the large-scale Y2H screenings and manually curated literature datasets
367 might be a consequence of this limitation (Fig 3). This small overlap illustrates the current
368 knowledge gap in the characterization of the full plant interactome of pathogen effectors.
369 Extensive work will be required to characterize further effector-host protein interactions in other
370 pathosystems. As one of the simplest yet powerful high throughput techniques for protein-protein
371 interaction detection, our work, like others before, highlights the potential of such large-scale Y2H
372 screenings in the identification of novel effector targets in an easy, cheap and systematic manner.

373 **EffectorK, an entry point to explore and make sense of plant-effector interactomics**

374 To conclude, our work also provides valuable resources for the plant-pathogen interaction
375 community. We described 540 new *Ath-Rps* and *Ath-Xcc* effector protein interactions that allowed
376 us to identify 166 new effector targets (S1 Table). We also manually curated several publications
377 to assemble a collection of 287 *Ath*-effector protein interactions from a wide variety of pathogens

(S2 Table). All this, allowed us to identify 100 novel effector hubs (S3 Table). The contribution to plant immunity of these effector hubs has been described for 19 of them, but remains untested for the majority (Table 1). This constitutes a list of promising candidates for further functional characterization. All these data were integrated in EffectorK, a knowledge database where users can have easy access to the *Ath*-effector protein interactions and explore the resulting interactomic network visually and interactively. While major efforts were done to capture the maximal diversity on the pathogen side, we limited our work to the Arabidopsis plant model. Thanks to the built-in homology search tools available, users can also use their own data as query regardless of the species studied. It is therefore feasible to use EffectorK as a starting point to build on and extend to crop plant-effector protein interactomics. In the long term, these data could be exploited to better understand how pathogens interact with these crops with the prospect of selecting breeding candidates for improved tolerance or resistance against pathogens.

Materials and Methods

Cloning of *Rps* and *Xcc* T3E genes

All the cloning of the T3E genes from *Rps* and *Xcc* was performed by BP gateway BP or TOPO cloning (Thermo Fisher Scientific), to generate pENTRY plasmids, which were later transferred into the appropriate Y2H plasmids [11], using the LR gateway reaction (Thermo Fisher Scientific). S5 Table contains all the PCR primers and final plasmid identities describing the collection of plasmids used in this study. Gene sequence information from *Rps* strain GMI1000

397 can be obtained from www.ralsto-T3E.org [47] and from the published genome of *Xcc* strain 8004
398 [52].

399 **Y2H screenings**

400 The Y2H screening was performed in semi-liquid ('8K space' screening) and liquid ('12K
401 space' screening) media as recently reported [53], which is an adaptation of a previously developed
402 Y2H-solid pipeline [54]. In both protocols the same low copy number yeast expression vectors
403 and the two yeast strains, *Saccharomyces cerevisiae* Y8930 and Y8800, were used. The expression
404 of the *GAL1-HIS3* reporter gene was tested with 1 mM 3AT (3-amino-1,2,4-triazole, a competitive
405 inhibitor of the HIS3 gene product), unless described otherwise. Prior to Y2H screening, DB-X
406 strains were tested for auto-activation of the *GAL1-HIS3* reporter gene in the absence of AD-Y
407 plasmid. In case of auto-activation, DB-X were physically removed from the collection of baits
408 and screened against the (DB)-*Ath*-cDNA collections using their AD-X constructs. Briefly, DB-X
409 baits expressing yeasts were individually grown (30°C for 72 hours) into 50-ml polypropylene
410 conical tubes containing 5 ml of fresh selective media (Sc-Leucine, Sc-Leu). Pools were created
411 by mixing a maximum of 72 and 50 individual bait yeast strains for the '8K space' and '12K space'
412 respectively. Subsequently, 120 µl and 50 µl of these individual pools were plated into 96-well
413 and 384-well low profile microplates for *Ath*-cDNA '8K space' and '12K space' collections
414 respectively. Glycerol stocks of the (AD)-*Ath*-cDNA '8K space' and '12K space' collections were
415 thawed, replicated by handpicking or using a colony picker Qpix2 XT into 96-well and 384-well
416 plates filled with 120 µl and 50 µl of fresh selective media (Sc-Tryptophan; Sc-Trp) respectively,
417 and incubated at 30°C for 72 hours. Culture plates corresponding to the DB-baits pools and AD-

418 collection were replicated into mating plates filled with YEPD media and incubated at 30°C for
419 24 hours. In liquid Y2H case ('12K space' screening), mating plates were then replicated into
420 screening plates filled with 50 µl of fresh Sc-Leu-Trp-Histidine + 1 mM 3AT media and incubated
421 at 30°C for 5 days. In order to identify primary positives, the OD_{600} of the 384-well screening
422 plates was measured using a microplate-reader Tecan Infinite M200 PRO. In semi-liquid Y2H
423 case ('8K space' screening), mated yeast were spotted onto Sc-Leu-Trp-Histidine + 1 mM 3AT
424 media agar plates, and incubated at 30 °C for 3 days. Protein pairs were identified by depooling
425 of DB-baits in a similar targeted matricial liquid or semi-liquid assays in which all the DB-baits
426 were individually tested against all the previously identified AD-proteins. Identified pairs were
427 picked and checked by PCR and DNA sequencing.

428 **Database content and manual curation**

429 Binary interactions between *Ath* proteins with each other and with pathogen effector
430 proteins were compiled on tabular form keeping track of the protein names and accessions, species
431 and ecotypes/strains of origin, techniques used to detect the interactions and the reference. *Ath-Ath*
432 protein interactions were compiled from the Arabidopsis Interactome [41,42] and the public
433 databases BioGrid (www.thebiogrid.org [39]; downloaded in September 2019) and IntAct
434 (www.ebi.ac.uk/intact [40]; downloaded in September 2019). We only kept the direct interactions
435 with the evidence codes 'co-crystal structure', 'FRET' (fluorescence resonance energy transfer),
436 'PCA' (protein-fragment complementation assay), 'reconstituted complex' or 'two-hybrid' on
437 BioGrid and 'physical association' on IntAct. *At*-effector protein interactions were gathered from
438 our experimental Y2H data together with the similarly produced data on *Hpa*, *Psy* and *Gor*

effectors [11,12]. In addition, an extensive keyword search on effector-Arabidopsis literature was done to retrieve interactions from 80 published articles. A confidence level was assigned to each interaction depending on the number of independent techniques used in a publication for validation: "1" if the interaction was detected by only one technique and "2" if the interaction was validated by at least a second technique. Some interactions lacked important information but, in order to maximize the extent of our network, several assumptions were taken instead of discarding useful data. First, gene models for *Ath* proteins were rarely mentioned on publications so we assumed the first gene model available on the latest version of the Arabidopsis genome (Araport11 [55]). Secondly, when the ecotype/strain of the organism was not explicitly stated, a generic 'NA' (not available) was assigned.

In silico **analyses**

Computational simulations of random targeting of *Ath* proteins by single pathogen effectors (intraspecific convergence). Significance of the intraspecific convergence was tested comparing our experimental data with random simulations as previously published [12]. Briefly, for each effector of *Xcc* and *Rps* we assigned randomly the same number of *Ath* targets as experimentally observed from the degree-preserved list of 8K proteins. The distribution obtained from 10,000 simulations was plotted and compared to the experimentally obtained data. The p-value of the experimental data was calculated as follows: number of simulations where the number of targets is lower or equal than experimentally observed is divided by the number of simulations. When the number of simulations with less targets than observed was zero, the p-value was set to < 0.001.

$$p - value = \frac{number\ of\ simulations\ where\ the\ number\ of\ interactors \leq experimentally\ observed\ number\ of\ interactors}{number\ of\ simulations} \quad (1)$$

Computational simulations of random targeting of *Ath* proteins by several pathogen effectors (interspecific convergence). Significance of the interspecific convergence was tested comparing our experimental data and previously published data with random simulations as published [11,12]. Briefly, for each effector of all compared pathogens we assigned the same number of *Ath* targets as experimentally observed/published from the list of 8K proteins. The distribution obtained from 10,000 simulations was plotted and compared to experimentally and published data. The p-value of the experimental data was calculated as follows: number of simulations where the number of common targets between species was higher or equal than the experimentally observed is divided by the number of simulations. When the number of simulations with more common targets than observed was zero, the p-value was set to < 0.001.

$$p - value = \frac{number\ of\ simulations\ where\ the\ number\ of\ common\ interactors \geq experimentally\ observed\ number\ of\ common\ interactors}{number\ of\ simulations} \quad (2)$$

Overlap of effector targets. The overlaps of effector targets between the different kingdoms and species were calculated taking into account the targets found in the different large-scale screening and limiting to the 8K space. For representation of the data, Venn diagrams were generated using the Venn Diagrams tool from VIB-UGent Center for Plant Systems Biology (www.bioinformatics.psb.ugent.be/webtools/Venn/). The overlap of effector targets from the different datasets was calculated not limiting to any limited space. For an area-proportional representation of the data, a Venn diagram was generated using BioVenn [56].

Network topology analyses. The topology parameters of the *Ath*-effector interactomic network were calculated on Cytoscape 3.7.2 [57]. Our analyses focused on two key node parameters: degree and betweenness centrality. The degree of a protein is a measure of its connectivity and denotes the number of proteins interacting with it. Throughout this work, we have differentiated two kinds of degrees: 1) effector degree (i.e., number of interacting effector proteins) and 2) *Ath* degree (i.e., number of interacting *Ath* proteins). The betweenness centrality measures the proportion of shortest pathways between two proteins that passes through a given node. These parameters were compared against different subset of data and statistical tests were performed in R language [58]. The cumulative distribution of these parameters among different subset of data was plotted and the area under the curve was estimated using Simpson's rule with the 'Bolstad2' package [59].

Database construction

The database was built using the software architecture recently described [60]. The files submitted by the curator team were automatically checked for typographic mistakes using *ad-hoc* Perl scripts and loaded into a Neo4J database and indexed in an ElasticSearch search engine. Each release was rebuilt from scratch. Data were made accessible through a web interface (see Results and discussion section) built upon Cytoscape.js library [61]. The raw data used for the database setup are available in the 'Data' section of www.effectork.org and the source code is available at https://framagit.org/LIPM-BIOINFO/KGBB.

Acknowledgments

We wish to thank Alberto Macho and Laurent Deslandes for the contribution of RipS3, and RipP1 and RipP2 containing plasmids. We thank colleagues (Corinne Audran, Julien Luneau, Effectome and FNX network members) for their critical opinion on our work.

References

1. Jones JDG, Dangl JL. The plant immune system. Nature. 2006;444: 323–9. doi:10.1038/nature05286

2. Osbourn AE. Preformed Antimicrobial Compounds and Plant Defense against Fungal Attack. Plant Cell. 1996;8: 1821–1831. doi:10.1105/tpc.8.10.1821

3. Hamann T. Plant cell wall integrity maintenance as an essential component of biotic stress response mechanisms. Front Plant Sci. 2012;3. doi:10.3389/fpls.2012.00077

4. Zipfel C. Plant pattern-recognition receptors. Trends Immunol. 2014;35: 345–51. doi:10.1016/j.it.2014.05.004

5. Ma W, Wang Y, McDowell J. Focus on Effector-Triggered Susceptibility. Mol Plant-Microbe Interact. 2018;31: 5–5. doi:10.1094/MPMI-11-17-0275-LE

6. Cui H, Tsuda K, Parker JE. Effector-Triggered Immunity: From Pathogen Perception to Robust Defense. Annu Rev Plant Biol. 2015;66: 487–511. doi:10.1146/annurev-arplant-050213-040012

7. Büttner D. Behind the lines-actions of bacterial type III effector proteins in plant cells.

FEMS Microbiol Rev. 2016;40: 894–937. doi:10.1093/femsre/fuw026

8. Sharpee WC, Dean RA. Form and function of fungal and oomycete effectors. Fungal Biol Rev. 2016;30: 62–73. doi:10.1016/j.fbr.2016.04.001

9. Giron D, Huguet E, Stone GN, Body M. Insect-induced effects on plants and possible effectors used by galling and leaf-mining insects to manipulate their host-plant. J Insect Physiol. 2016;84: 70–89. doi:10.1016/j.jinsphys.2015.12.009

10. Vieira P, Gleason C. Plant-parasitic nematode effectors — insights into their diversity and new tools for their identification. Curr Opin Plant Biol. 2019;50: 37–43. doi:10.1016/j.pbi.2019.02.007

11. Mukhtar MS, Carvunis A-R, Dreze M, Epple P, Steinbrenner J, Moore J, et al. Independently evolved virulence effectors converge onto hubs in a plant immune system network. Science. 2011;333: 596–601. doi:10.1126/science.1203659

12. Weßling R, Epple P, Altmann S, He Y, Yang L, Henz SR, et al. Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. Cell Host Microbe. 2014;16: 364–375. doi:10.1016/j.chom.2014.08.004

13. Mansfield J, Genin S, Magori S, Citovsky V, Sriariyanum M, Ronald P, et al. Top 10 plant pathogenic bacteria in molecular plant pathology. Mol Plant Pathol. 2012;13: 614–29. doi:10.1111/j.1364-3703.2012.00804.x

14. Genin S, Denny TP. Pathogenomics of the *Ralstonia solanacearum* species complex. Annu Rev Phytopathol. 2012;50: 67–89. doi:10.1146/annurev-phyto-081211-173000

15. Büttner D, Bonas U. Regulation and secretion of *Xanthomonas* virulence factors. FEMS Microbiol Rev. 2010;34: 107–33. doi:10.1111/j.1574-6976.2009.00192.x

16. White FF, Potnis N, Jones JB, Koebnik R. The type III effectors of *Xanthomonas*. Mol Plant Pathol. 2009;10: 749–66. doi:10.1111/j.1364-3703.2009.00590.x

17. Coll NS, Valls M. Current knowledge on the *Ralstonia solanacearum* type III secretion system. Microb Biotechnol. 2013;6: 614–20. doi:10.1111/1751-7915.12056

18. Lindeberg M, Cunnac S, Collmer A. *Pseudomonas syringa*e type III effector repertoires: last words in endless arguments. Trends Microbiol. 2012;20: 199–208. doi:10.1016/j.tim.2012.01.003

19. Vandereyken K, Van Leene J, De Coninck B, Cammue BPA. Hub Protein Controversy: Taking a Closer Look at Plant Stress Response Hubs. Front Plant Sci. 2018;9: 694. doi:10.3389/fpls.2018.00694

20. Kim SH, Son GH, Bhattacharjee S, Kim HJ, Nam JC, Nguyen PDT, et al. The Arabidopsis immune adaptor SRFR1 interacts with TCP transcription factors that redundantly contribute to effector-triggered immunity. Plant J. 2014;78: 978–89. doi:10.1111/tpj.12527

21. Zhang H, Wang X, Giroux MJ, Huang L. A wheat COP9 subunit 5-like gene is negatively involved in host response to leaf rust. Mol Plant Pathol. 2017;18: 125–133. doi:10.1111/mpp.12467

22. Luo H, Laluk K, Lai Z, Veronese P, Song F, Mengiste T. The Arabidopsis Botrytis Susceptible1 Interactor defines a subclass of RING E3 ligases that regulate pathogen and

stress responses. Plant Physiol. 2010;154: 1766–82. doi:10.1104/pp.110.163915

23. Zhang C, Ding Z, Wu K, Yang L, Li Y, Yang Z, et al. Suppression of Jasmonic Acid-Mediated Defense by Viral-Inducible MicroRNA319 Facilitates Virus Infection in Rice. Mol Plant. 2016;9: 1302–1314. doi:10.1016/j.molp.2016.06.014

24. Redditt TJ, Chung E-H, Zand Karimi H, Rodibaugh N, Zhang Y, Trinidad JC, et al. AvrRpm1 Functions as an ADP-Ribosyl Transferase to Modify NOI-domain Containing Proteins, Including Arabidopsis and Soybean RPM1-interacting Protein 4. Plant Cell. 2019; tpc.00020.2019. doi:10.1105/tpc.19.00020

25. Nietzsche M, Guerra T, Alseekh S, Wiermer M, Sonnewald S, Fernie AR, et al. STOREKEEPER RELATED1/G-Element Binding Protein (STKR1) Interacts with Protein Kinase SnRK1. Plant Physiol. 2018;176: 1773–1792. doi:10.1104/pp.17.01461

26. Anstead JA, Froelich DR, Knoblauch M, Thompson GA. Arabidopsis P-protein filament formation requires both AtSEOR1 and AtSEOR2. Plant Cell Physiol. 2012;53: 1033–42. doi:10.1093/pcp/pcs046

27. Li Y, Yang Y, Hu Y, Liu H, He M, Yang Z, et al. DELLA and EDS1 Form a Feedback Regulatory Module to Fine-Tune Plant Growth-Defense Tradeoff in Arabidopsis. Mol Plant. 2019;12: 1485–1498. doi:10.1016/j.molp.2019.07.006

28. Zhang Z, Liu Y, Ding P, Li Y, Kong Q, Zhang Y. Splicing of receptor-like kinase-encoding SNC4 and CERK1 is regulated by two conserved splicing factors that are required for plant immunity. Mol Plant. 2014;7: 1766–75. doi:10.1093/mp/ssu103

575 29. Lim CW, Luan S, Lee SC. A prominent role for RCAR3-mediated ABA signaling in
576 response to *Pseudomonas syringae* pv. tomato DC3000 infection in Arabidopsis. Plant Cell
577 Physiol. 2014;55: 1691–703. doi:10.1093/pcp/pcu100

578 30. Boudsocq M, Willmann MR, McCormack M, Lee H, Shan L, He P, et al. Differential innate
579 immune signalling via Ca(2+) sensor protein kinases. Nature. 2010;464: 418–22.
580 doi:10.1038/nature08794

581 31. Liu H, Dong S, Gu F, Liu W, Yang G, Huang M, et al. NBS-LRR Protein Pik-H4 Interacts
582 with OsBIHD1 to Balance Rice Blast Resistance and Growth by Coordinating Ethylene-
583 Brassinosteroid Pathway. Front Plant Sci. 2017;8: 127. doi:10.3389/fpls.2017.00127

584 32. Wang F, Shang Y, Fan B, Yu J-Q, Chen Z. Arabidopsis LIP5, a positive regulator of
585 multivesicular body biogenesis, is a critical target of pathogen-responsive MAPK cascade
586 in plant basal defense. Li X, editor. PLoS Pathog. 2014;10: e1004243.
587 doi:10.1371/journal.ppat.1004243

588 33. Simon C, Langlois-Meurinne M, Bellvert F, Garmier M, Didierlaurent L, Massoud K, et al.
589 The differential spatial distribution of secondary metabolites in Arabidopsis leaves reacting
590 hypersensitively to *Pseudomonas syring*ae pv. tomato is dependent on the oxidative burst.
591 J Exp Bot. 2010;61: 3355–70. doi:10.1093/jxb/erq157

592 34. Mukherjee M, Larrimore KE, Ahmed NJ, Bedick TS, Barghouthi NT, Traw MB, et al.
593 Ascorbic acid deficiency in arabidopsis induces constitutive priming that is dependent on
594 hydrogen peroxide, salicylic acid, and the NPR1 gene. Mol Plant Microbe Interact. 2010;23:

340–51. doi:10.1094/MPMI-23-3-0340

35. Dagdas YF, Beihaj K, Maqbool A, Chaparro-Garcia A, Pandey P, Petre B, et al. An effector of the irish potato famine pathogen antagonizes a host autophagy cargo receptor. Elife. 2016;5: 1–23. doi:10.7554/eLife.10856

36. MacLean AM, Orlovskis Z, Kowitwanich K, Zdziarska AM, Angenent GC, Immink RGH, et al. Phytoplasma effector SAP54 hijacks plant reproduction by degrading MADS-box proteins and promotes insect colonization in a RAD23-dependent manner. Wagner D, editor. PLoS Biol. 2014;12: e1001835. doi:10.1371/journal.pbio.1001835

37. Nurmberg PL, Knox KA, Yun B-W, Morris PC, Shafiei R, Hudson A, et al. The developmental selector AS1 is an evolutionarily conserved regulator of the plant immune response. Proc Natl Acad Sci U S A. 2007;104: 18795–800. doi:10.1073/pnas.0705586104

38. Kersey PJ, Allen JE, Allot A, Barba M, Boddu S, Bolt BJ, et al. Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. Nucleic Acids Res. 2018;46: D802–D808. doi:10.1093/nar/gkx1011

39. Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 2006;34: D535-9. doi:10.1093/nar/gkj109

40. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, et al. The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res. 2014;42: D358-63. doi:10.1093/nar/gkt1115

615 41. Dreze M, Carvunis A-R, Charloteaux B, Galli M, Pevzner SJ, Tasan M, et al. Evidence for
616 Network Evolution in an Arabidopsis Interactome Map. Science. 2011;333: 601–607.
617 doi:10.1126/science.1203877

618 42. Smakowska-Luzan E, Mott GA, Parys K, Stegmann M, Howton TC, Layeghifard M, et al.
619 An extracellular network of Arabidopsis leucine-rich repeat receptor kinases. Nature.
620 2018;553: 342–346. doi:10.1038/nature25184

621 43. Li H, Zhou Y, Zhang Z. Network analysis reveals a common host–pathogen interaction
622 pattern in Arabidopsis immune responses. Front Plant Sci. 2017;8: 893.
623 doi:10.3389/fpls.2017.00893

624 44. Memišević V, Zavaljevski N, Rajagopala S V, Kwon K, Pieper R, DeShazer D, et al. Mining
625 host-pathogen protein interactions to characterize *Burkholderia mallei* infectivity
626 mechanisms. PLoS Comput Biol. 2015;11: e1004088. doi:10.1371/journal.pcbi.1004088

627 45. Ahmed H, Howton TC, Sun Y, Weinberger N, Belkhadir Y, Mukhtar MS. Network biology
628 discovers pathogen contact points in host protein-protein interactomes. Nat Commun.
629 2018;9: 2312. doi:10.1038/s41467-018-04632-8

630 46. Peeters N, Carrère S, Anisimova M, Plener L, Cazalé A-C, Genin S. Repertoire, unified
631 nomenclature and evolution of the Type III effector gene set in the *Ralstonia solanacearum*
632 species complex. BMC Genomics. 2013;14: 859. doi:10.1186/1471-2164-14-859

633 47. Sabbagh CRR, Carrere S, Lonjon F, Vailleau F, Macho AP, Genin S, et al. Pangenomic
634 type III effector database of the plant pathogenic *Ralstonia* spp. PeerJ. 2019;7: e7346.

635 doi:10.7717/peerj.7346

636 48. Guy E, Genissel A, Hajri A, Chabannes M, David P, Carrere S, et al. Natural Genetic
637 Variation of *Xanthomonas campestris* pv. *campestris* Pathogenicity on Arabidopsis
638 Revealed by Association and Reverse Genetics. Guttman D, Ausubel FM, editors. MBio.
639 2013;4. doi:10.1128/mBio.00538-12

640 49. Roux B, Bolot S, Guy E, Denancé N, Lautier M, Jardinaud MF, et al. Genomics and
641 transcriptomics of *Xanthomonas campestris* species challenge the concept of core type III
642 effectome. BMC Genomics. 2015;16. doi:10.1186/s12864-015-2190-0

643 50. Dillon MM, Almeida RND, Laflamme B, Martel A, Weir BS, Desveaux D, et al. Molecular
644 Evolution of *Pseudomonas syringae* Type III Secreted Effector Proteins. Front Plant Sci.
645 2019;10. doi:10.3389/fpls.2019.00418

646 51. Brückner A, Polge C, Lentze N, Auerbach D, Schlattner U. Yeast two-hybrid, a powerful
647 tool for systems biology. Int J Mol Sci. 2009;10: 2763–88. doi:10.3390/ijms10062763

648 52. Qian W, Jia Y, Ren S-X, He Y-Q, Feng J-X, Lu L-F, et al. Comparative and functional
649 genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestri*s pv.
650 *campestris*. Genome Res. 2005;15: 757–67. doi:10.1101/gr.3378705

651 53. Monachello D, Guillaumot D, Lurin C. A pipeline for systematic yeast 2-hybrid matricial
652 screening in liquid culture. 2019. doi:10.21203/rs.2.9948/v1

653 54. Dreze M, Monachello D, Lurin C, Cusick ME, Hill DE, Vidal M, et al. High-Quality Binary
654 Interactome Mapping. 2010. pp. 281–315. doi:10.1016/S0076-6879(10)70012-4

655 55. Cheng C-Y, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD.
656 Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. Plant J.
657 2017;89: 789–804. doi:10.1111/tpj.13415

658 56. Hulsen T, de Vlieg J, Alkema W. BioVenn - a web application for the comparison and
659 visualization of biological lists using area-proportional Venn diagrams. BMC Genomics.
660 2008;9: 488. doi:10.1186/1471-2164-9-488

661 57. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a
662 software environment for integrated models of biomolecular interaction networks. Genome
663 Res. 2003;13: 2498–504. doi:10.1101/gr.1239303

664 58. R Core Team. R: A language and environment for statistical computing Authors R Core
665 Team. In: Vienna, Austria [Internet]. 2019. Available: https://www.r-project.org/

666 59. Bolstad WM. Understanding Computational Bayesian Statistics. Hoboken, NJ, USA: John
667 Wiley & Sons, Inc.; 2009. doi:10.1002/9780470567371

668 60. Carrère S, Verdenaud M, Gough C, Gouzy J, Gamas P. LeGOO: An Expertized Knowledge
669 Database for the Model Legume *Medicago truncatula*. Plant Cell Physiol. 2019.
670 doi:10.1093/pcp/pcz177

671 61. Franz M, Lopes CT, Huck G, Dong Y, Sumer O, Bader GD. Cytoscape.js: a graph theory
672 library for visualisation and analysis. Bioinformatics. 2016;32: 309–11.
673 doi:10.1093/bioinformatics/btv557

674 62. Solé M, Popa C, Mith O, Sohn KH, Jones JDG, Deslandes L, et al. The awr Gene Family
37

Encodes a Novel Class of *Ralstonia solanacearum* Type III Effectors Displaying Virulence and Avirulence Activities. Mol Plant-Microbe Interact. 2012;25: 941–953. doi:10.1094/MPMI-12-11-0321

63. Lonjon F, Lohou D, Cazalé A-C, Büttner D, Ribeiro BG, Péanne C, et al. HpaB-Dependent Secretion of Type III Effectors in the Plant Pathogens *Ralstonia solanacearum* and *Xanthomonas campestris* pv. *vesicatoria*. Sci Rep. 2017;7: 4879. doi:10.1038/s41598-017-04853-9

64. Remigi P, Anisimova M, Guidot A, Genin S, Peeters N. Functional diversification of the GALA type III effector family contributes to *Ralstonia solanacearum* adaptation on different plant hosts. New Phytol. 2011;192: 976–987. doi:10.1111/j.1469-8137.2011.03854.x

65. Wang K, Remigi P, Anisimova M, Lonjon F, Kars I, Kajava A, et al. Functional assignment to positively selected sites in the core type III effector RipG7 from *Ralstonia solanacearum*. Mol Plant Pathol. 2016;17: 553–64. doi:10.1111/mpp.12302

66. Deslandes L, Olivier J, Peeters N, Feng DX, Khounlotham M, Boucher C, et al. Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. Proc Natl Acad Sci U S A. 2003;100: 8024–8029. doi:10.1073/pnas.1230660100

Supporting information

S1 Fig. *Ath* degree of T3E proteins from *Rps* strain GMI1000 and *Xcc* strain 8004.

695 *Ath* degree (i.e., number of *Ath* targets per effector) in the in the 12,000 (12K space, light blue)
696 and 8,000 *Ath* cDNA collections (8K space, dark blue) of T3E proteins from *Rps* strain GMI1000
697 (A) and *Xcc* strain 8004 (B). For *Rps* strain GMI1000: in the first screening RipA3, RipAA,
698 RipAB, RipAC, RipAG, RipAL, RipAM, RipAN, RipAO, RipAP, RipAQ, RipAR, RipAZ1,
699 RipB, RipBA, RipG3, RipG4, RipG6, RipG7, RipH2, RipH3, RipI, RipK, RipM, RipN, RipO1,
700 RipP1, RipQ, RipR, RipS2, RipS6, RipT, RipTPS, RipX and RipZ were screened but no targets
701 were found. In the second screening RipAB, RipAC, RipAI, RipAX1, RipAY, RipBM, RipC1,
702 RipE1, RipH1, RipN, RipR, RipS4, RipU, RipX and RipZ were screened but no targets were
703 found, and RipAN and RipM could not be screened because of recalcitrant problems with yeast
704 transformation. For *Xcc* strain 8004: AvrXccA2, HpaA, HrpW, XopAN, XopN and XopQ were
705 screened but no targets were found, and AvrBs2, XopAH, XopAL2, XopD and XopE2 could not
706 be screened because they showed autoactivation in yeast.

707 **S2 Fig. Overlap of *Ath* targets of effector proteins from *Hpa*, *Psy*, *Gor*, *Rps* and *Xcc*.**
708 Venn diagrams showing the overlap among *Ath* targets found in the 8,000-*Ath*-cDNA collection
709 (8K space) of effector proteins from *Hpa*, *Psy*, *Gor*, *Rps* and *Xcc* at the kingdom (A) and species
710 level (B). The total number of effector targets for each kingdom/species is indicated in brackets.

711 **S3 Fig. Interspecific convergence of *Psy* and *Xcc* effector proteins.**
712 Number of *Ath* targets in the 8K space of effectors from *Psy* and *Xcc* and *Rps* strain found in
713 10,000 degree-preserving simulations (grey) versus the observed number (red arrow).

714 **S4 Fig. Effector degree distribution for *Ath* effector targets.**

715 Effector degree (i.e., number of effectors that interact with an *Ath* protein) distribution among the
716 564 identified *Ath* effector targets (A), according to the origin the data: published large-scale
717 screenings in light green, manual curation of literature in mid-green and this study in dark grey or
718 (B), according to the kingdom of the corresponding effector pathogen: Bacteria in light blue,
719 Chromista in dark blue, Fungi in light orange and Animalia in dark orange.

720 **S5 Fig. *Ath* degree and betweenness centrality of different groups of *Ath* effector targets.**
721 Cumulative distribution of *Ath* degree (B and D) and betweenness centrality (A, C and E) for *Ath*
722 proteins targeted (orange) or not (purple) by effectors (B), multi-pathogen (green) and pathogen-
723 specific (pink) effector targets (B and C) and effector hubs (red) and single effector targets (blue)
724 (D and E). The significance of the differences were evaluated by one-tailed Wilcoxon signed-rank
725 test. The illustration in the upper right corner of each graph represents each compared group:
726 effectors are represented by squares, *Ath* proteins by circles, numbers represent different pathogens
727 species and the color code matches the respective cumulative distribution graph. The estimation
728 of the area under the curve of each distribution is compiled in Table 2.

729 **S6 Fig. *Ath* and effector degree of effector targets.**
730 (A) Scatterplot of *Ath* degree versus effector degree of all *Ath* effector targets. Squared in a grey
731 dashed line is the close-up area represented in (B).

732 **S7 Fig. Degrees and betweenness centrality of bacterial core and non-core T3Es and their**
733 **targets.**
734 Cumulative distribution of *Ath* degree (A and D), effector degree (C) and betweenness centrality
735 (B and D) for bacterial core T3Es (yellow) and other bacterial T3Es (cyan) (A and B) and their

736 targets (brown and blue respectively) (C-E). The significance of the differences were evaluated by
737 one-tailed Wilcoxon signed-rank test. The illustration in the upper right corner of each graph
738 represents each compared group: bacterial T3Es are represented by squares, *Ath* proteins by circles
739 and stars represents bacterial core T3Es. The estimation of the area under the curve of each
740 distribution is compiled in Table 2.

741 **S1 Table. List of *Rps* and *Xcc* effector-*Ath* protein interactions detected experimentally in**
742 **this study and composition of the *Ath*-cDNA screening libraries.**

743 **S2 Table. List of manually curated *Ath*-effector protein interactions from the literature.**

744 **S3 Table. List of effector hubs and single effector targets identified.**

745 **S4 Table. List of protein interactions constituting the *Ath*-effector interactomic network.**

746 **S5 Table. List of pENTRY for T3E genes from *Rps* and *Xcc*.**

747 ## Figure captions

748 **Fig 1. *Ath* degree of effector proteins from *Gor, Hpa*, *Psy*, *Xcc* and *Rps*.**
749 Comparison of the *Ath* degree (i.e., number of *Ath* targets per effector) of effector proteins from
750 *Gor*, *Hpa*, *Psy*, *Xcc* and *Rps* found in the 8,000-*Ath*-cDNA collection (8K space). Horizontal black
751 bars represent the median. Colors represent the kingdom (orange: Fungi, yellow: Chromista and
752 blue: Bacteria).

753 **Fig 2. Effectors converge intra- and interspecifically onto a common set of *Ath* proteins.**

754 (A) Left: random and intraspecific convergent interactions of effectors (purple squares) with *Ath*
755 proteins (green circles) can be distinguished by random network rewiring and simulation. Adapted
756 from Weßling et al. [12]. Middle and right: number of *Ath* targets in the 8K space of effectors
757 from *Xcc* strain 8004 and *Rps* strain GMI1000 found in 10,000 degree-preserving simulations
758 (grey) versus the observed number (red arrow). (B) Left: random and interspecific convergent
759 interactions of effectors from different species (purple and orange squares) with *Ath* proteins
760 (green circles) can be distinguished by random network rewiring and simulation. Right: number
761 of common *Ath* targets in the 8K space of effectors from Chromista, Bacteria and Fungi found in
762 10,000 simulations (grey) versus the observed number (red arrow). (C) Scatterplot of observed
763 versus simulated number of common *Ath* targets between all binary, ternary, quaternary and
764 quinary combinations of species. x=y regression is represented with a dashed grey line.

765 **Fig 3. Overlap among effector targets depending on the origin of the dataset.**
766 Area-proportional Venn diagram showing the overlap among effector targets identified in the
767 large-scale Y2H screenings performed in this study, in similar large-scale Y2H already published
768 [11,12] and in the manual curation of the literature. The total number of effector targets coming
769 from each dataset is indicated in brackets.

770 **Fig 4. Network topology parameters.**
771 Example of a simple interactomic network of three effector proteins (purple squares) and eight *Ath*
772 proteins (green circles) to illustrate our definition of "effector hub" (i.e., *Ath* protein interacting
773 with two or more effectors; highlighted in red) and the three network topology parameters analyzed
774 in this study. 1) Effector degree: number of effectors that interact with a given *Ath* protein. 2) *Ath*

775 degree: number of *Ath* proteins that interact with a given effector or *Ath* protein. 3) Betweenness
776 centrality: fraction of all shortest paths connecting two proteins from the network that pass through
777 a given protein.

778 **Fig 5. *Ath* degree of Ath proteins targeted or not by effectors.**

779 Cumulative distribution of *Ath* degree of *Ath* proteins targeted (orange) or not (purple) by effectors.
780 The significance of the difference was validated by one-tailed Wilcoxon signed-rank test. The
781 illustration in the upper right corner represents each compared group. Effectors are represented by
782 squares, *Ath* proteins by circles and the color code matches the cumulative distribution graph.

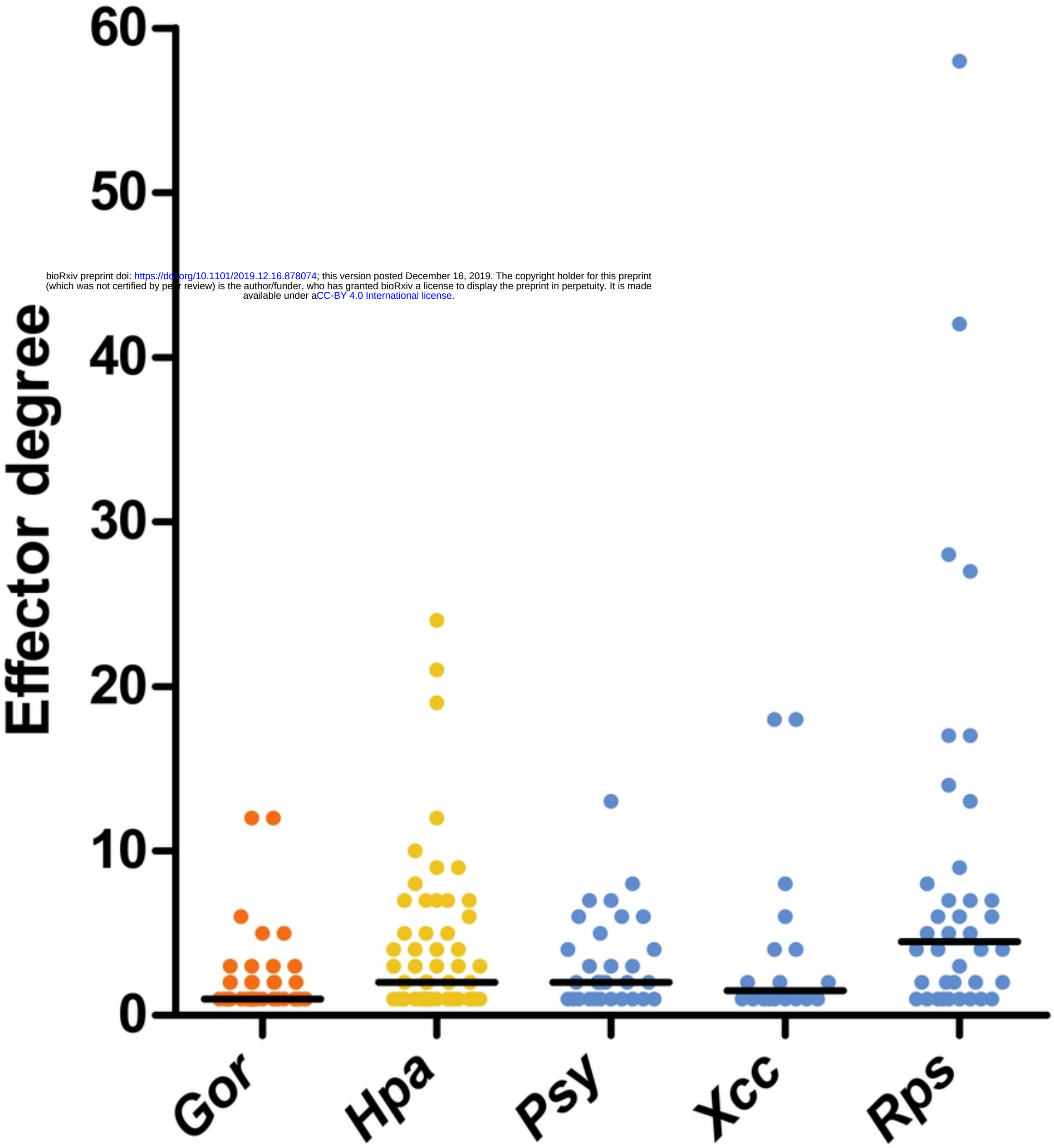


Figure1

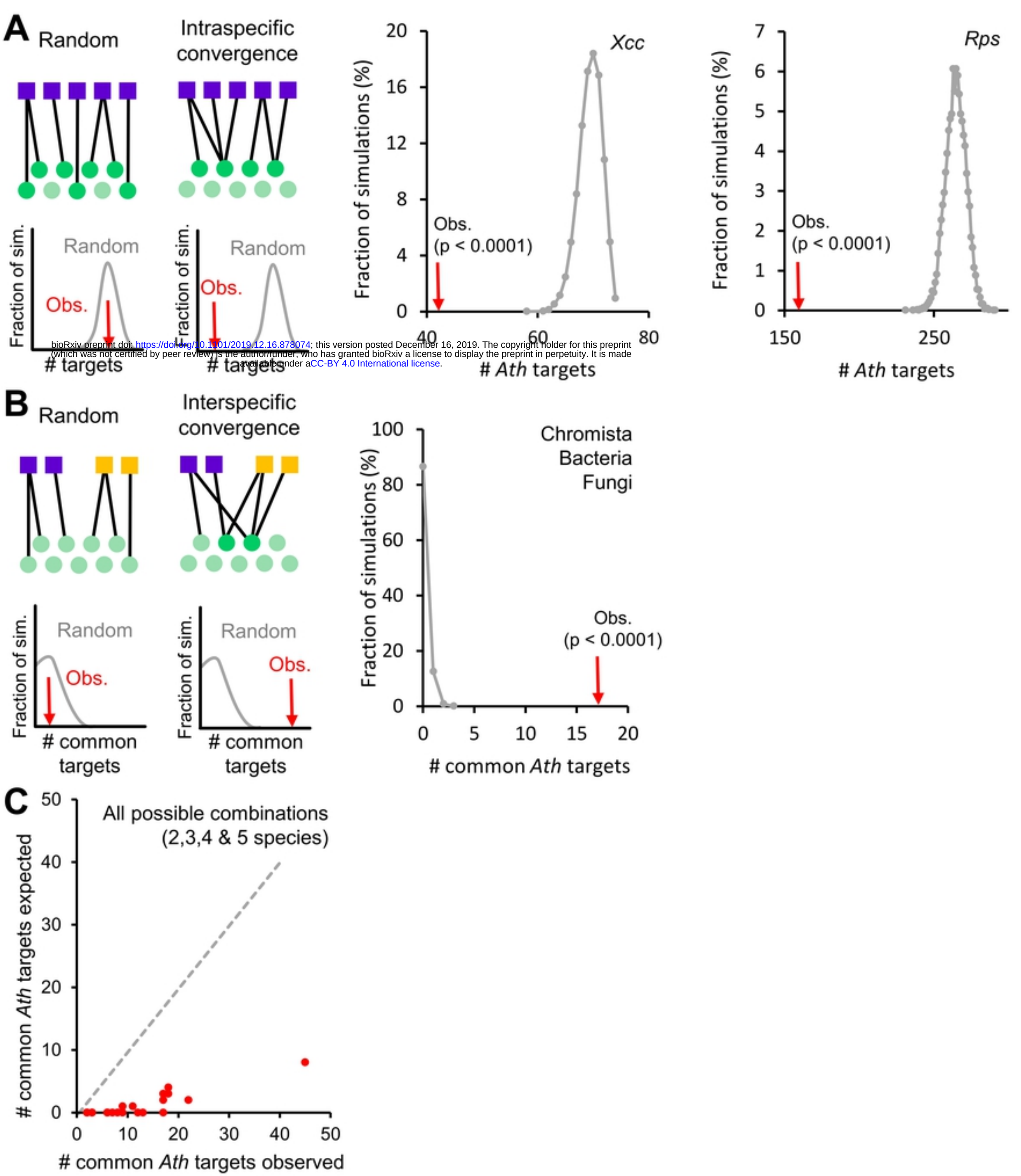


Figure2

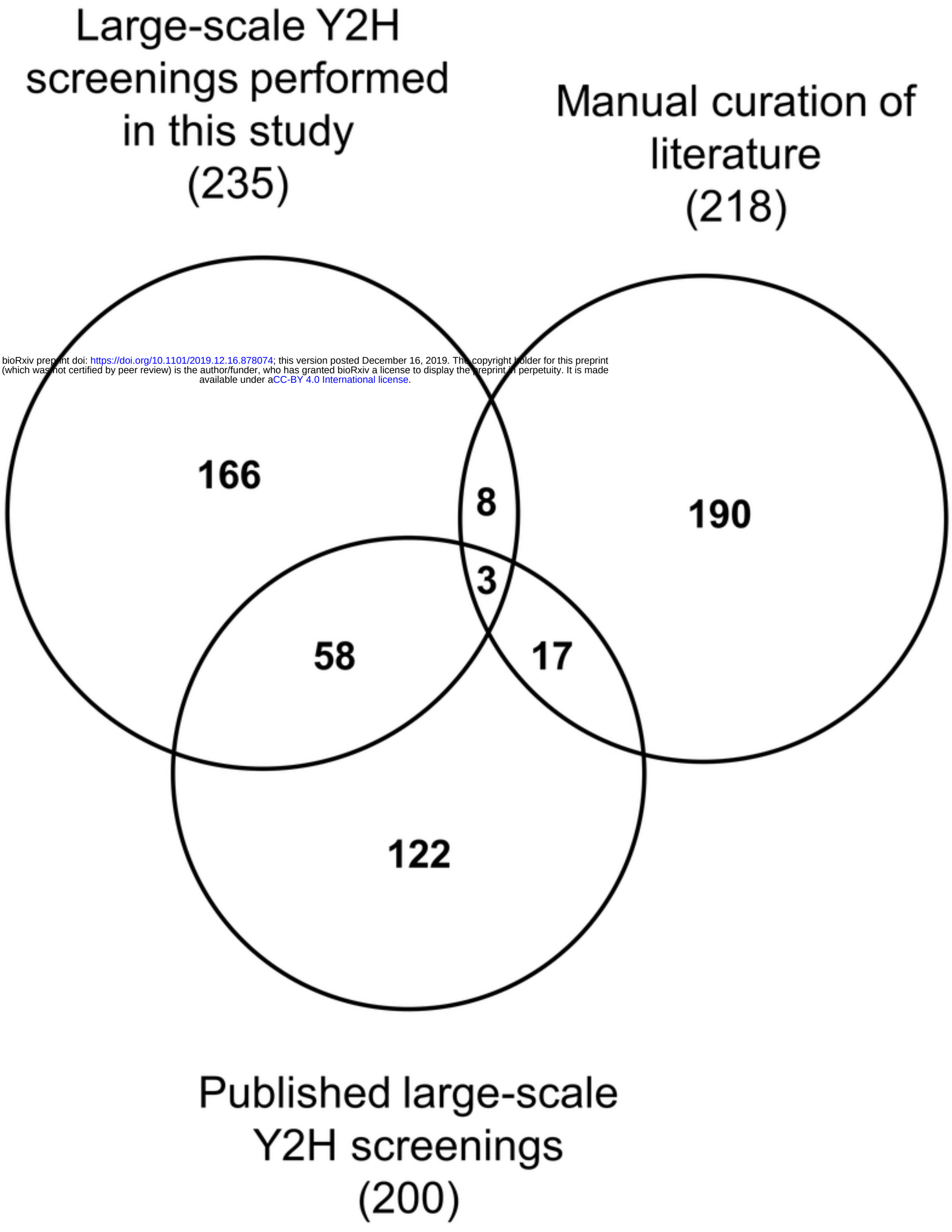
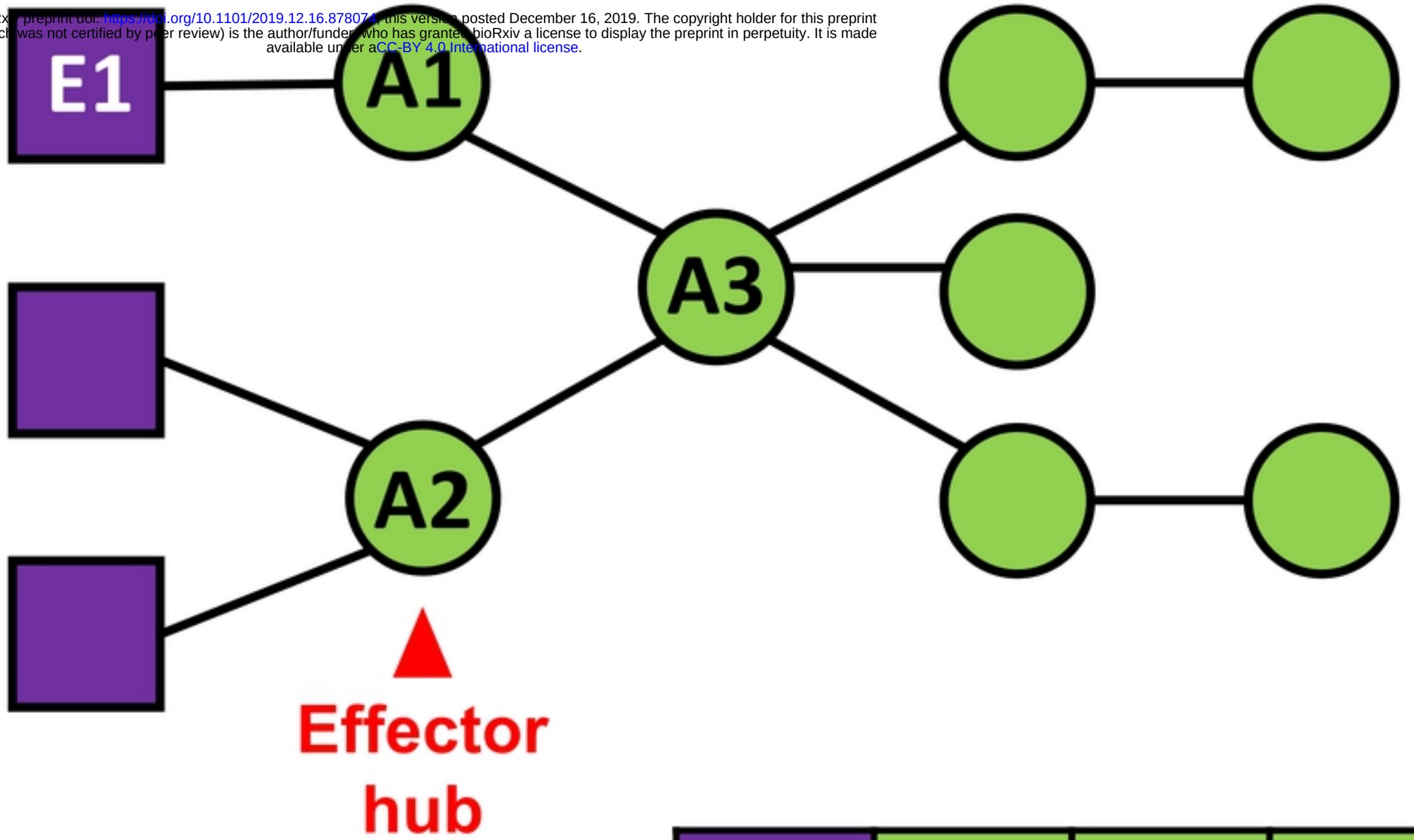


Figure3



	E1	A1	A2	A3
Effector degree	-	1	1	0
***Ath* degree**	1	1	1	5
Betweenness centrality	0	0.2	0.4	0.9

Figure4

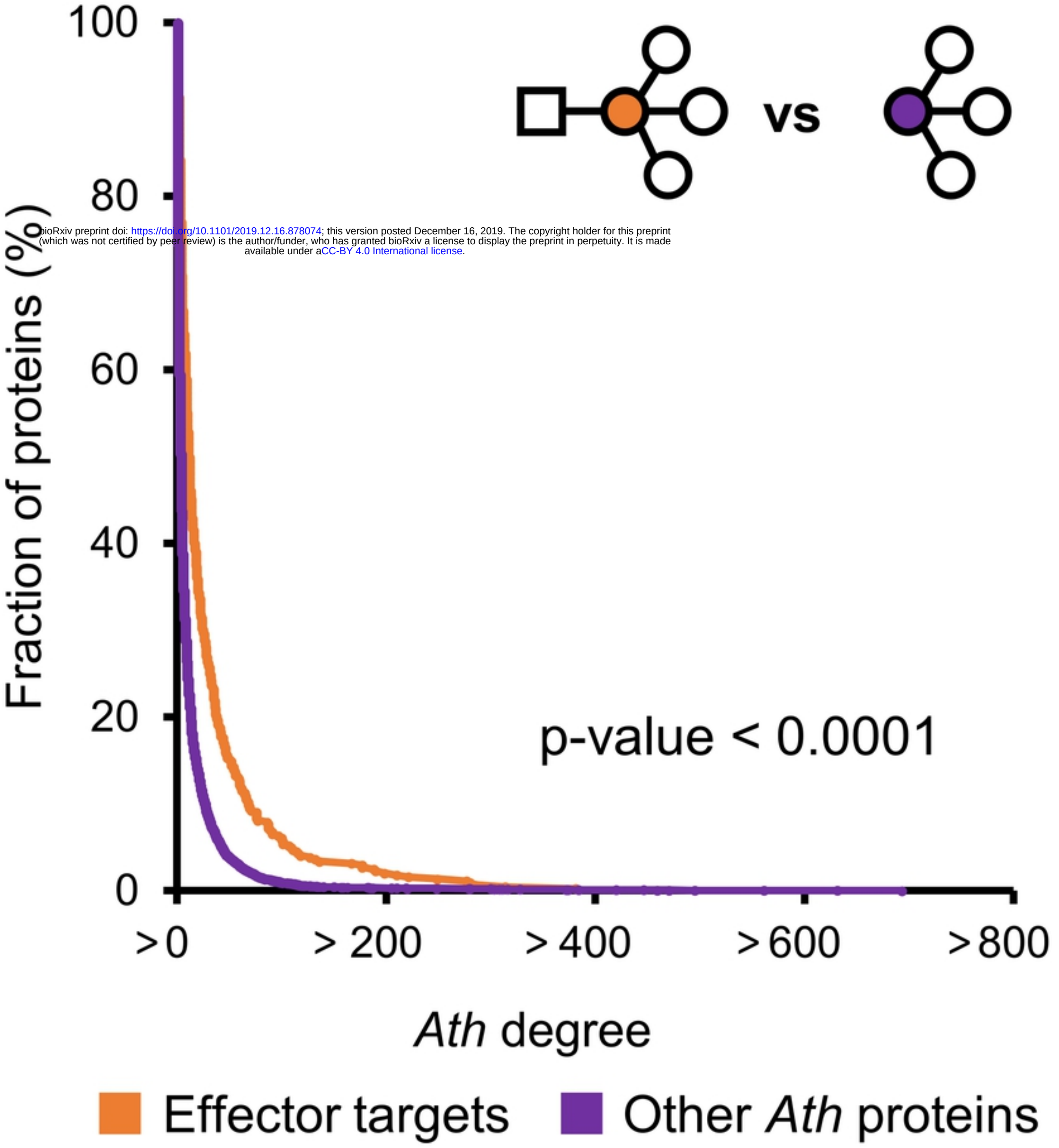


Figure5