Sciencexpress Report

Herpesviral Protein Networks and Their Interaction with the Human Proteome

Peter Uetz,[1*] Yu-An Dong,[2*] Christine Zeretzke,[3] Christine Atzler,[3] Armin Baiker,[3] Bonnie Berger,[2] Seesandra Rajagopala,[1] Maria Roupelieva,[3] Dietlind Rose,[3] Even Fossum,[3] Jürgen Haas3[†]

[1]Institut für Genetik, Forschungszentrum Karlsruhe, Postfach 3640, Karlsruhe, D-76021 Germany. [2]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA. [3]Max-von-Pettenkofer Institut, Ludwig-Maximilians-Universität München, Pettenkoferstrasse 9a, 80336 München, Germany.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: haas@lmb.uni-muenchen.de

The comprehensive yeast-two-hybrid analysis of intraviral protein interactions in two members of the herpesvirus family, Kaposi Sarcoma associated Herpesvirus (KSHV) and Varicella Zoster Virus (VZV), revealed 123 and 173 interactions, respectively. Viral protein interaction networks resemble single, highly coupled modules, whereas cellular networks are organized in separate functional submodules. Predicted and experimentally verified interactions between KSHV and human proteins were used to connect the viral interactome into a prototypical human interactome and to simulate infection. The analysis of the combined system showed that the viral network adopts cellular network features, and that protein networks of herpesviruses and possibly other intracellular pathogens have distinguishing topologies.

Herpesviruses are widely spread throughout vertebrates and possess large double-stranded DNA genomes encoding between approximately 70 and 170 viral proteins, which is only one order of magnitude less than small bacterial genomes. While studies have revealed a significant number of interactions between herpesviral and host proteins, surprisingly little is known about interactions among herpesviral proteins, particularly for those herpesviruses that replicate poorly in cell culture, e.g. Kaposi's sarcoma associated herpesvirus (KSHV) (*1*). We thus generated genome–wide protein interaction maps for two human pathogens, KSHV, which is a member of the γ-herpesvirus subfamily associated with Kaposi's sarcoma and B-cell lymphomas (*2*), and Varicella Zoster Virus (VZV), which is a member of the α-herpesvirus subfamily causing chickenpox and shingles (*3*).

We cloned the open reading frames (ORFeome) of both viruses (currently there are 89 ORFs identified in KSHV and 69 ORFs in VZV) by recombinatorial cloning and generated yeast two-hybrid (Y2H) bait and prey arrays. To circumvent the limitation of the Y2H system for transmembrane proteins, full-length proteins as well as extra- and intracellular domains were cloned separately. In KSHV, we tested more than 12,000 viral protein interactions involving both full-length proteins and protein fragments and identified 123 nonredundant interacting protein pairs (fig. S1 and table S1). To date, only a small number of intraviral protein interactions have been reported for KSHV, of which 71% were captured by our screen (table S2). To further confirm the quality of our Y2H results and generate a set of high-confidence interactions, we tested all positive Y2H interactions in parallel by both β-galactosidase assays (Gal) and coimmunoprecipitations (CoIP) (fig. S2). Approximately 50% of the protein interactions could be confirmed by CoIP (table S1), while many of the remaining ones have orthologous interactions in other herpesviruses that could be confirmed (table S3). While our array-based two-hybrid system is internally controlled, some of the interactions not confirmed by CoIP may be non-physiological and for example caused by autoactivation. A comparison between protein interactions and expression profiles of KSHV-infected cells indicated that protein interactions predominantly occur between proteins expressed at the same time point after infection (fig. S3). In VZV, we detected 173 nonredundant intraviral protein interactions out of ~10,000 tested bait-prey pairs (table S4).

While cellular protein interaction networks exist for several model organisms, none of the studies on viral protein interactions published so far produced a large enough dataset to constitute a protein interaction network. In many cellular protein interaction networks, most nodes have few neighbors while some have many interaction partners (so-called hubs). The degree distributions of cellular protein networks were reported to follow a power-law decay and they have been classified as scale-free (*4*). Like their cellular counterparts, KSHV and VZV have relatively many hubs, a key characteristic of scale-free networks (Fig. 1a; fig. S4). However, in contrast to known cellular protein interaction

networks, in which nodes with a single interaction partner are most abundant, the viral networks have relatively few such "peripheral" nodes lying on the "edge" of the network. In KSHV, for example, the degree distribution peaks at nodes with three neighbors. This unusual characteristic at low node degrees is one of the reasons that viral networks appear as single, highly coupled modules and presumably reflects their incompleteness as stand-alone networks. While cellular protein networks have been shown or assumed to be scale-free, the degree distribution of viral networks does not present such a clear-cut picture (fig. S5). If viral networks are approximated by a power-law distribution, they possess unusually small power coefficients distinctive from known cellular networks and defying current dynamic network evolution models (Fig. 1b; Table 1).

Another important characteristic of complex networks is the so-called small-world property (*5*). In a small-world network the average distance between any two nodes is short (short characteristic path length or the *six degrees of separation* phenomenon) and local neighborhoods are more densely connected (high clustering coefficient). Both viruses exhibit a short characteristic path length and a short network diameter (the maximum distance between any two nodes), which also suggests their coupling as single modules (Table 1). To assess the viral levels of local clustering, we generated random networks of the same size and degree distribution. Our results show that the level of local clustering is low in KSHV and VZV, in fact comparable to equivalent random networks, and thus these viral networks cannot be classified as small-world. In contrast, most cellular protein interaction networks are unambiguously small-world, even after the effect on local clustering due to degree distribution and network size, which turns out to be substantial, is filtered out. For example, the clustering coefficient of the *S. cerevisiae* protein interaction network is increased approximately 30-fold over simulated networks (Table 1). In *S. cerevisiae*, Maslov and Sneppen demonstrated that hubs tend to avoid each other while prefering low-connectivity nodes (*6*). As a result, the yeast network has well-separated modules, and errors in one module do not easily propagate to other modules. In the viral networks, there is no such declining degree correlation and hubs do not tend to avoid each other, which offers additional evidence that these viral networks could be viewed as single, highly coupled modules (fig. S6). As a consequence of these unusual topological features, viral networks are more resistant to deliberate attacks than cellular networks, as both network size and characteristic path length remain more stable after the most highly connected nodes are removed (Fig.1, c and d).

While sequence and phylogenetic analyses identify a core set of genes conserved in all herpesviruses, the KSHV interactome allowed us to determine a core set of interactions conserved in all herpesviruses. Using the reciprocal best BLAST hit approach, we determined the orthology relationships among the viruses and predicted 112 orthologous protein interactions in HSV-1, VZV, mCMV and EBV (table S1). Intriguingly, the majority of the predicted interactions (74/112 or 78%; table S3) could be confirmed by CoIP despite the rather low level of sequence similarity between KSHV proteins and their orthologs (only in the 20%-40% range). While in general KSHV proteins with viral interaction partners are not more conserved than those without (fig. S7a), for KSHV proteins with interaction partners there is a significant correlation between the number of protein interactions and homology to the respective EBV ortholog (fig. S7b). When we compared the protein networks of KSHV and VZV, we found that only 9 of the 50 protein interactors in KSHV have orthologous interactors in VZV (Fig. 1a; table S1). The KSHV network predicts 19 interactions between orthologous viral proteins in VZV, of which we could detect 5 in our Y2H screen (26.3%). This low result is not surprising, as the number of predicted interactions in HSV-1, mCMV and EBV that were confirmed by Y2H was also considerably lower than the number confirmed by CoIP, indicating an inherent technical limitation of the Y2H system.

The network analyses of the KSHV and VZV interactomes revealed unique features of viral systems which also manifested themselves on the local level (fig. S8). Since we hypothesized that many of them could be attributed to missing virus-host interactions, we modeled the interplay between viral and human protein networks. To date, only rather small subnets of the human interactome have been reported (*7*, *8*). Unfortunately, only few published human proteins targeted by KSHV lie within these reported human subnets and only few virus-host interactions could be predicted based on homology between KSHV and human proteins. However, using a prototypical human protein interaction network (derived from high-confidence interactions in *S. cerevisiae*, *C. elegans* and *D. melanogaster* (*9–11*) which is considerably larger than the experimental subnets, we were able to efficiently connect the viral and the human networks by predicting interactions between KSHV and human proteins if both proteins have known interacting orthologs in either *S. cerevisiae*, *C. elegans* or *D. melanogaster* (*12*). By this approach, we received 20 predicted interactions between 8 KSHV and 20 human proteins which are connected within the human network. Nineteen of these 20 virus-host interactions were tested by CoIP and a surprisingly large percentage (13/19 or 68.4%) could be confirmed (table S9). While published viral-host interactions tend to involve genes or interactions specific to human or higher eukaryotes (since most human targets have no orthologs or orthologous interactions in the three lower-

eukaryotic model organisms), our predicted viral-host interactions involve genes and interactions conserved from yeast to human and hence might reflect more general host-interacting mechanisms.

Using the predicted KSHV-human interactions, we were able to dock the two interactomes at each other (Fig. 2, a and b). Strikingly, the topology of the KSHV network changes completely from a highly coupled module to a more typical scale-free network of interacting submodules once it is connected to its host. To rigorously assess the impact of the two systems upon each other, we performed a combined viral-host network analysis. Starting with the KSHV network (level 0), we first added in their direct human targets (level 1), subsequently we added in those human targets' own cellular interactions partners (level 2), and so on, until the viral network is completely assimilated into the host network. To evaluate the topology of the combined virus-host network, we reasoned that a correctly combined system should be different from randomly combined networks. To generate an ensemble of equivalent random viral-host networks, we adopted the following simulation strategy: the identity and degree of KSHV interactors are fixed, while their human targets are randomly chosen from the host network so that each random target has the same degree as the predicted target that it replaces. Since degree distribution reflects global topology, with the KSHV and human protein networks differing significantly in this respect, it offers an ideal measure to assay the coupled system. Indeed, at level 2, the predicted viral-host network not only exhibits a better power-law fit, but the power coefficient is also bigger, both crucial features of the human network (empirical p-value < 0.01 in 1000 simulations) (Fig. 2c). Thus, at the level carrying most biological relevance (KSHV's human targets and their direct interaction partners) and suffering from minimal noise (level 3 already includes a sizable fraction of the human network and many of the interactions are conceivably no longer relevant to the viral-host context), the combined virus-host network significantly assimilates human network properties (fig. S9). When we predicted VZV-human interactions and modelled the combined system, we received very similar results demonstrating the general utility of our approach (fig. S10; table S10).

While we have shown that virus and host interactomes possess distinct network topologies, their interplay may lead to emergent new system properties representing specific features of the viral pathogenesis. Obviously, numerous biological hypotheses resulting from our study remain to be investigated in detail. The availability of protein interaction networks in other herpesviruses and large-scale virus-host interaction data in the near future will boost our knowledge on the function of many still poorly characterised viral proteins and the phylogeny of herpesviruses. It will eventually lead to a considerably improved understanding of viral pathogenesis and evoke novel therapeutic strategies.

References and Notes

1. P. Uetz, S. V. Rajagopala, Y. Dong, J. Haas, *Genome Res.* **14**, 2029 (2004).
2. Y. Chang *et al*., *Science* **266**, 1865 (1994).
3. A. J. Davison and J. E. Scott, *J.Gen.Virol.* **67**, 1759 (1986).
4. A. L. Barabasi and R. Albert, *Science* **286**, 509 (1999).
5. D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998).
6. S. Maslov and K. Sneppen, *Science* **296**, 910 (2002).
7. U. Stelzl *et al*., *Cell* **122**, 957 (2005).
8. J. F. Rual *et al*., *Nature* **437**, 1173 (2005).
9. P. Uetz *et al*., *Nature* **403**, 623 (2000).
10. L. Giot *et al*., *Science* **302**, 1727 (2003).
11. S. Li *et al*., *Science* **303**, 540-543 (2004).
12. B. Lehner and A. G. Fraser, *Genome Biol.* **5**, R63 (2004).
13. S. McCraith, T. Holtzmann, B. Moss, S. Fields, *Proc Natl Acad Sci U S A* **97**, 4879 (2000).
14. G. Cagney, P. Uetz, S. Fields, *Methods Enzymol.* **328**, 3 (2000).

15. This work was supported by grants provided by BayGene (Bayerisches Staatsministerium für Wissenschaft, Forschung und Kunst), DFG (SFB 576), Sander-Stiftung (#1999.096.1), Helmholtz Gemeinschaft (P.U. and Y.-A.D.) and LMU München (FoeFoLe). We thank Ann Arvin, Yuan Chang, Don Ganem, Shou-Jiang Gao, Wolfgang Hammerschmidt, Yasushi Kawaguchi, Ulrich Koszinowski, Patrick Moore, Bernard Moss, Frank Neipel, Paula Pitha-Rowe, Thomas Schulz, Beate Sodeik, Charles Wood and Yan Yuan for plasmids, cosmids, BACs and viruses. This manuscript is dedicated to C.A.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1116804/DC1
Materials and Methods
Figs. S1 to S10
Tables S1 to S10
References and Notes

30 June 2005; accepted 21 November 2005
Published online 8 December 2005; 10.1126/science.1116804
Include this information when citing this paper.

Fig. 1. Topology of the KSHV protein interaction network. (**a**) Protein interaction map of KSHV. KSHV proteins are indicated as nodes, protein interactions either as hatched (found only by Y2H) or solid (confirmed by CoIP) edges. Orthologous proteins interactors in KSHV and VZV are marked by circled nodes, orthologous interactions detected in both viruses by red edges. KSHV ORFs were assigned into five functional classes depicted in different colours based on GenBank annotations for the corresponding ORF or its

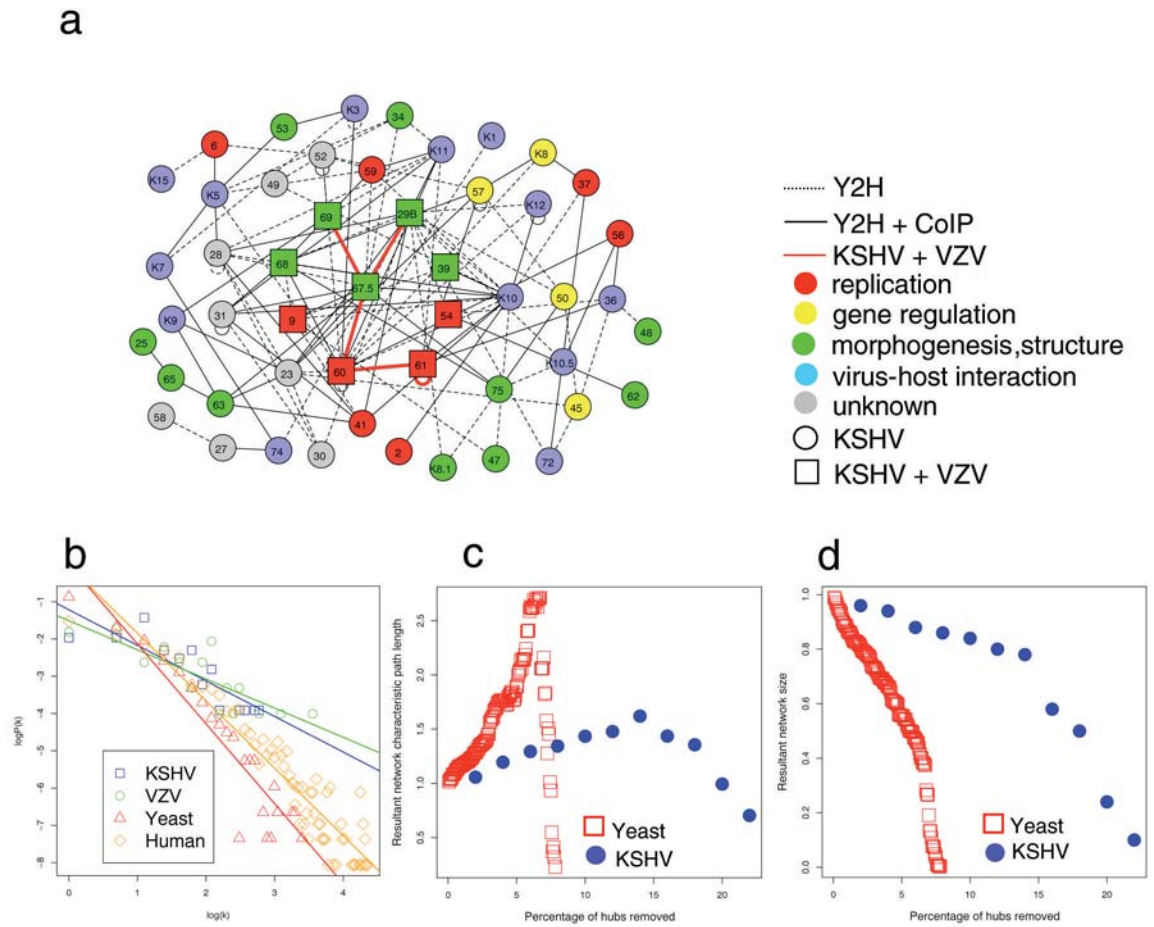
orthologs. (**b**) Comparison of the approximated power-law degree distribution of two herpesviral (KSHV and VZV) networks and two cellular (yeast and human) networks. The *yeast* dataset is derived from Uetz *et al*. (*9*), the *H. sapiens* (predicted) dataset from Lehner *et al*. (*12*). For each network, node degrees k and their relative frequency (i.e. probability) are plotted on a bilogarithmic scale and fitted by linear regression. (**c** and **d**) Simulations of deliberate attack on KSHV and yeast networks by removing their most highly connected nodes (in decreasing order). After each node is removed, the new network characteristic path length (average distance between any two nodes) and size (number of nodes) of the remaining single largest connected component (SLCC) are computed and plotted as a multiple or fraction of the original parameters. KSHV exhibits much higher attack tolerance, as the increase in path length and the decrease in network size are considerably smaller.

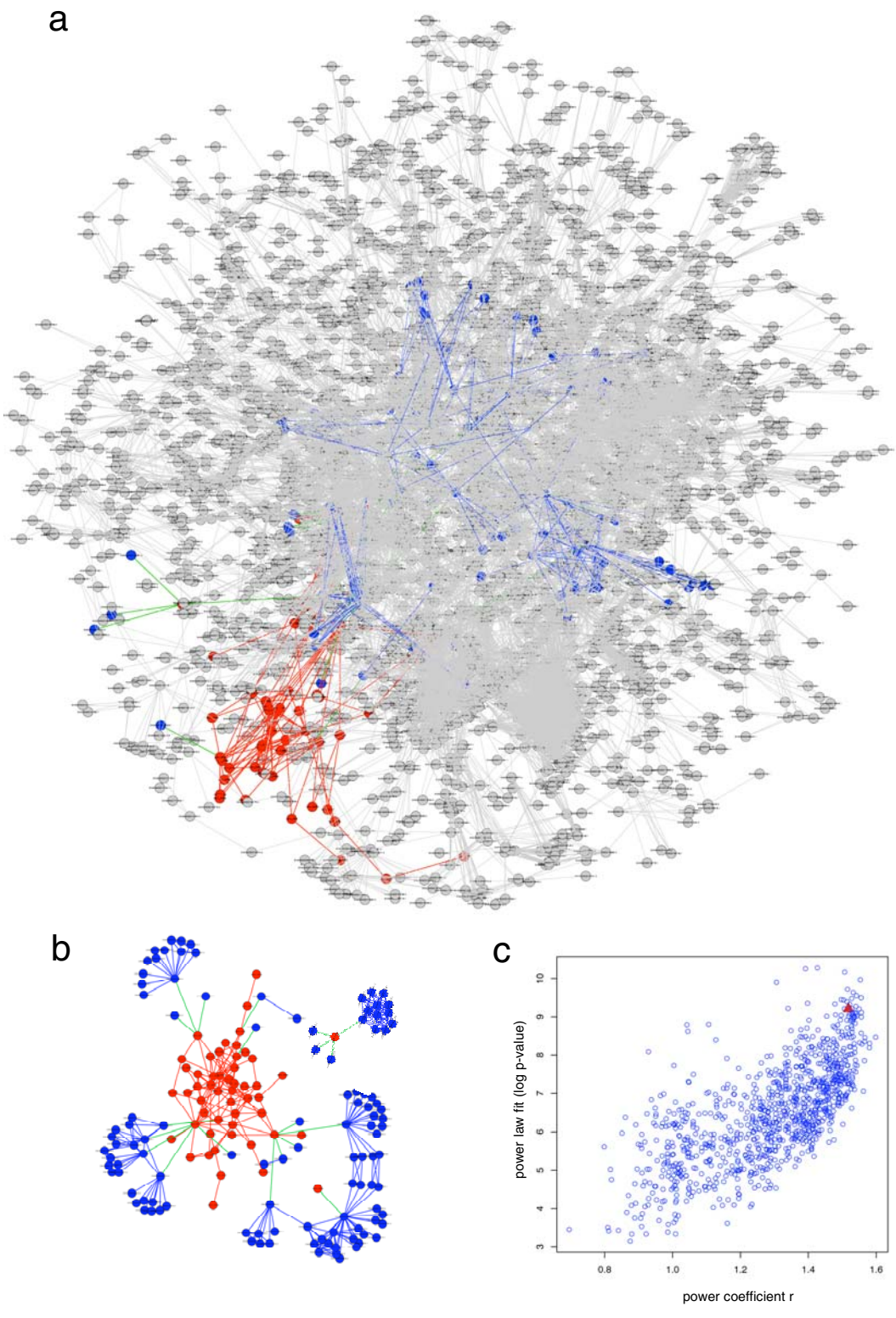
Fig. 2. Interplay between the KSHV and a predicted high-confidence human network. (**a**) Global view of the interplay between the KSHV and a predicted high-confidence human interaction network consisting of 10,636 edges among 3,169 nodes. Viral proteins are depicted as red nodes, cellular interacting proteins (level 1 and 2) as blue nodes and cellular proteins (level >2) as grey nodes. Interactions between viral proteins are depicted as red edges, between viral and cellular proteins as green edges and between cellular level 1 and 2 proteins as blue edges. (**b**) Local view of the combined KSHV and human interaction networks (level ≤ 2). (**c**) The combined viral-host network possesses a bigger and statistically more significant power coefficient and thus adopts scale-free features. The combined virus-host network was compared to 1,000 random networks which were generated by rewiring fixed virus interactors to swapped cellular proteins with the same degree as the actual target. The power coefficient and the power-law fit of the predicted (red triangle) and 1000 random (blue circles) KSHV-human networks are indicated on the x-axis and the y-axis, respectively. Among 1000 random networks, 65 possess a bigger power coefficient, 19 a more significant power-law fit, and only 8 a bigger power coefficient that is more significant (empirical p-value <0.1%).

Table 1. Comparison of network parameters of cellular and viral protein interaction networks. The table indicates key parameters of eight viral and cellular networks (all analyses were done using the SLCC of the respective network). Among 123 nonredundant KSHV protein interactions, 8 are self-interactions and the SLCC consists of 115 edges; among 173 nonredundant VZV protein interactions, 13 are self-interactions and one edge is isolated, leaving the SLCC with 159 edges. Note that only one of the 123 interactions in KSHV and 19 of the 173 in VZV were detected bidirectionally and that one interaction in KSHV and 13 in VZV were redundantly detected by distinct fragments of the same proteins. Interactions only detected in one direction are a common phenomenon in two-hybrid assays and most likely due to steric hindrance of either bait or prey fusion proteins. The dataset for Vaccinia virus is derived from McCraith *et al*. (*13*), for *S. cerevisiae I* from Uetz *et al*. (*9*), for *S. cerevisiae II* from the DIP database (October 2004 release), for *H. sapiens I* from Rual *et al*. (*8*), for *H. sapiens II* from Stelzl *et al*. (*7*) and for *H. sapiens* (predicted) from Lehner *et al*. (*12*).The table includes the number of nodes and edges in the SLCC; the average node degree (i.e. number of neighbors); the power coefficient γ and its p-value (the slope and its significance under linear regression) as fitted by a power-law degree distribution ("scale-free" property); the characteristic path length and diameter, as well as the clustering coefficient and its fold enrichment over comparable random networks ("small-world" property) (*14*). For each real network, a corresponding ER (Erdös-Rényi) randomization has the same number of nodes and edges, while an ES randomization, generated through an edge-swapping algorithm, also has the same degree distribution. The fold enrichments shown are over the theoretical clustering coefficient under the ER model and the median clustering coefficient of 1000 ES randomizations, respectively (see supplemental data). Note that the network parameters of the two yeast networks are surprisingly stable, although a large number of interactions have been included additionally into the DIB database in comparison to the initial dataset generated by Uetz and colleagues (*9*). The recently reported *H. sapiens* networks have rather low levels of local clustering, which was discussed as being caused by their incompleteness (*8*).

	KSHV	**VZV**	**Vaccinia virus**	***S. cerevisiae* I**	***S. cerevisiae* II**	***H. sapiens* I**	***H. sapiens* II**	***H. sapiens* (predicted)**
nodes	50	55	7	1,548	2,397	1,307	1,598	3,169
edges	123	173	6	2,358	6,101	2,483	3,072	10,636
average degree	4.60	5.78	1.71	3.05	5.09	3.80	3.84	6.71
power coefficient	0.95	0.78	n.a.	2.14	2.01	1.54	1.66	1.81
p-value	1.2E-4	1.1E-4	n.a.	3.6E-11	7.7E-23	1.2E-20	5.8E-25	1.4E-30
characteristic path length	2.84	2.34	n.a.	7.28	5.10	4.36	4.85	6.40
diameter	7	5	n.a.	23	13	12	13	20
clustering coefficient	0.146	0.393	n.a.	0.213	0.296	0.060	0.012	0.186
enrichment over ER	1.55	3.67	n.a.	108.1	139.5	20.6	4.8	87.9
enrichment over ES	0.76	1.01	n.a.	29.2	29.5	0.92	0.28	11.7

n.a. not applicable due to the low number of edges


a
Y2H
Y2H + CoIP
KSHV + VZV
replication
gene regulation
morphogenesis,structure
virus-host interaction
unknown
KSHV
KSHV + VZV
b
KSHV
VZV
Yeast
Human
log(k)
c
Resultant network characteristic path length
Yeast
KSHV
Percentage of hubs removed
d
Resultant network size
Yeast
KSHV
Percentage of hubs removed


a
b
c
power law fit (log p-value)
power coefficient r
3
4
5
6
7
8
9
10
0.8
1.0
1.2
1.4
1.6