PrionOme: A database of prions and other sequences relevant to prion phenomena

Djamel Harbi[1], Marimuthu Parthiban[5], Deena M.A. Gendoo[1], Sepehr Ehsani[4], Manish Kumar[2], Gerold Schmitt-Ulms[4], Ramanathan Sowdhamini[3] and Paul M. Harrison[1]*

[1] Dept. of Biology,
McGill University,
Stewart Biology Building, 1205 Dr. Penfield Ave.,
Montreal, QC, H3A 1B1, Canada.

[2] Dept. of Biophysics,
University of Delhi (South Campus),
Benito Juarez Road, New Delhi-110021, India.
Email: manish@south.du.ac.in

[3] National Centre for Biological Sciences (TIFR),
GKVK Campus,
Bellary Road, Bangalore, 560065, India.
Email: mini@ncbs.res.in

[4] Tanz Centre for Research in Neurodegenerative Diseases, and
Dept. of Laboratory Medicine and Pathobiology,
University of Toronto,
6 Queen's Park Crescent West,
Toronto, ON, M5S 3H2, Canada.
Email: g.schmittulms@utoronto.ca

[5] Dept. of Biosciences, Biochemistry,
Åbo Akademi University,
Tykistökatu 6A, Turku, Finland.

***corresponding author**
Tel: (514) 398 6420 **Fax:** (514) 398 5069 **Email:** paul.harrison@mcgill.ca

Abstract:

Prions are units of propagation of an altered state of a protein or proteins. Prions can propagate from cell to cell, and from organism to organism, through cooption of other protein copies. Prions contain no necessary nucleic acids, and are important both as both pathogenic agents, and as a potential force in epigenetic phenomena. The original prions were derived from a misfolded form of the mammalian Prion Protein PrP. Infection by these prions causes neurodegenerative diseases. Other prions cause non-Mendelian inheritance in budding yeast, and sometimes act as diseases of yeast. We have compiled a database of >2000 prion-related sequences, called the PrionOme. The database comprises seven PrionOme classification categories: prionogenic sequences (*i.e.*, sequences that can make prions), 'prionoids' (*i.e.*, phenomena that have some prion characteristics), orthologs, paralogs, pseudogenes, prion interactors, and prion-like molecules. Database entries list: supporting information for PrionOme classifications, prion-determinant areas (where relevant), and disordered and compositionally-biased regions. Also included are original references for the PrionOme classifications, transcripts and genomic coordinates, and structural data (including comparative models). We provide database usage examples for both vertebrate and fungal prion contexts. As development of this resource is on-going, we will be very happy to receive and act on any constructive comments from peer scientists in the areas of prion biology and protein misfolding, either by email or using the feedback form provided on the PrionOme website. We hope that this database will be a valuable experimental aid and reference resource. It is freely available at: http://libaio.biol.mcgill.ca/prion.

Background

Prions are alternative, propagating states of normal cellular proteins. Prions were originally defined as the causative agent of mammalian transmissible spongiform encephalopathies (TSEs), diseases which include scrapie in sheep and Creutzfeldt-Jakob disease (CJD) in humans **(1)**. CJD involves progressive dementia, and death within a year of diagnosis. TSEs can arise in inherited, sporadic or infectious forms. Infectious prions lack nucleic acids **(1)**, and rely on the presence of a host prion-protein gene for propagation **(2)**. Whereas the normal cellular form of the prion protein (PrP-C) is mostly alpha-helical **(3,4)**, the infectious form of the prion protein (PrP-Sc), is mostly beta-sheet **(5)**, which indicates a dramatic conformational change in the infectious protein.

The prion protein (PrP) is highly conserved across mammals, typically with >50% sequence identity relative to human PrP **(6,7)**. PrP maintains very high sequence conservation (>95%) in regions associated with disease, and also conserves a metal ion–binding repetitive region whose copy number is implicated in some human prion diseases **(9)**. Mammalian PrP paralogs have continued to be discovered and have been demonstrated to be of neurological relevance and functionally linked to PrP **(10)**. Doppel is a divergent PrP homolog (~25% sequence identity), that is mainly expressed in the testis, and can cause neurodegeneration when aberrantly expressed in the central nervous system (CNS) **(10)**. A second paralog of PrP, dubbed Shadoo, is more highly conserved across vertebrates, than either Doppel or PrP. Shadoo intriguingly shares a high degree of sequence identity with PrP in the short alanine-rich stretch that forms a transmembrane alpha-helix in some disease-associated PrP products **(11)**. Shadoo is expressed in the CNS, and both Shadoo and PrP-C can counteract Doppel neurotoxicity in a similar way **(12)**. PrP homologs have also been observed in fish **(13)**, and extensive genome-scale analyses in vertebrates have led to the discovery of additional gene and pseudogene family members for PrP **(14)**. Moreover, distant homology to ectodomains in the ZIP family of proteins linked to cytosolic divalent metal ion import, indicates that the PrP gene family may have originated from an ancestral ZIP sequence in an early common metazoan ancestor **(15)**.

Apart from the original infectious prion phenomena just described, prions have been defined in other organisms. In the budding yeast *S. cerevisiae*, prions have been identified as cytoplasmic or nuclear elements inherited in a non-Mendelian fashion **(16-18)**. The first two yeast prions discovered were [PSI+] and [URE3] **(16,17,19)**. [PSI+] arises from propagation of a misfolded amyloid form of Sup35p, part of the translation termination complex. Formation of [PSI+] prions reduces the efficiency of translation termination and increases levels of nonsense-codon readthrough **(17,20,21)**. Such readthrough has been demonstrated to be a potential mechanism to uncover cryptic genetic variation **(22,23)**. [URE3], the prion form of the nitrogen catabolism protein Ure2p, functions to up-regulate poor nitrogen source usage, even when rich sources are

available **(16,18,24)**. Prions may also be considered as diseases of budding yeast, in certain contexts **(25,26)**. A defining characteristic of the known yeast prions is a region with an obvious bias for glutamine (Q) and/or asparagine (N) residues **(27-30)**. Mutation of these residues diminishes or abolishes prion formation **(29,30)**. Previously, we have shown that Q/N biases are maintained in fungi that are estimated to have diverged from each other ~1 billion years ago; furthermore, there is evidence for purifying selection, to varying degrees, on different prion domains or subdomains **(31)**. Several hundred prion-like domains (with pronounced Q/N bias) occur in the proteomes of diverse fungi and higher eukaryotes **(27,31,32)**.

The universe of prion phenomena continues to expand **(33-37).** For example, in budding yeast, a prion that is formed by the CYC8 protein is part of a transcriptional regulatory complex that regulates 7% of yeast genes, thus potentially regulating their expression *en masse* **(38)**. Although the vast majority of described prions arise *via* propagation of alternative amyloid states of proteins, other types of prions are possible, such as the [β] prion, which arises through propagation of the auto-activated state of yeast protease B **(39)**. Also, some amyloids, dubbed 'prionoids', may exhibit evidence of prion-like propagation in some contexts or experimental conditions **(40)**.

Here, we report the availability of a public-domain resource, the PrionOme database. This database enables tracking of the rapidly growing corpus of prion phenomena and prion-related sequence data. The PrionOme comprises curated information about prions, prionogenic sequences, ‘prionoids’, prion protein orthologs and paralogs, pseudogenes (*i.e.*, genes copies that appear to have lost their protein-coding ability), prion protein interactors and prion-like molecules.

Database Construction & Content

Database structure

The PrionOme database is a record-based database, similar in format to the DisProt database **(41)**. It is managed and updated using the SQL and PHP scripting languages. The individual database entries are for single sequences. A screen-shot example is depicted in Figure 1. This is the database entry for a duplicated transcribed pseudogene in the human genome that is homologous to Sprn, the gene encoding the Shadoo protein. Database entries are indexed by Prion Identifier (which is in the form PDxxxx, where xxxx is a four-digit number), and also by PrionOme Classification (which is explained below in the section describing the database entry format). The source data and its collation, is described below for each field of the entry format. Redundant identical sequence entries in the database (either for full-length or fragmentary sequences) were curated and removed.

Database Entry Format

There are nine sections for each database entry. The sections are as follows:

(i) General Information:

This first section contains the following data:

Prion ID: A unique identifier for each entry in the database. It is in the format PDxxxx, where xxxx is a four-digit number.

UniProt accession and SwissProt ID: This is the UniProt/SwissProt identification corresponding to protein sequences that can be found in UniProt/SwissProt. Some PrionOme entries (e.g., new genome annotations for Prion Protein homologs, or pseudogenes) do not have a UniProt/SwissProt identification.

Source Organism and Taxonomy: This is the full binomial name of the source organism. The equivalent English name is usually given in brackets.

Sequence length: This is in residues.

Protein Name: This is a full descriptive name of the protein.

Prion Name: This applies only to 'prionogenic' sequences, i.e., sequences that can make prions. The name of the prion phenomenon for which the sequence is prionogenic is indicated. Prion Names follow the standard convention for prions in fungal genetics **(42)**. The original pathogenic prion in mammals has been given the name [PrP TSE prion]. Entries labeled 'Alberti, et al (2009) data', are proteins from the analysis of Alberti, et al. **(33)**, which were indicated as likely prions via a combination of fluorescence microscopy, SDD-AGE analysis, Sup35C prion assay and in vitro assembly assay **(33)**.

In addition, the Prion Type is given. Prions are classified into four types based on the classification in the review article by Wickner, et al. **(43)**. These Types are as follows:

Type 1: The prion state of the protein is a pathogenic or toxic amyloid/amyloid-like form.

Type 2: The prion state of the protein is an inactive amyloid/amyloid-like form, but may cause partial loss of normal function of the protein. This may lead to epigenetic phenomena, such as uncovering hidden genetic variation, in the case of the [PSI+] prion.

Type 3: The prion state of the protein is an amyloid/amyloid-like form that is the major active functional form of the protein.

Type 4: The prion state is an active state of an enzyme.

Sequence: This is the amino-acid sequence in one-letter code.

(ii) PrionOme Classification:

This section of the database entry contains the PrionOme classification, as well as a short description of the supporting information for this classification. The total numbers of entries with each PrionOme Classification is summarized in Table 1. The PrionOme Classification can be one or more of the following:

Prionogenic: A protein sequence that is known to form prions. These were collated by the database authors from the literature.

Ortholog: An ortholog of a sequence that is known to form prions. Orthologs are the mutually most similar sequences in different organisms. These were calculated as orthologs using the bi-directional best hits approach. Some orthologous proteins are from novel genome annotations, annotated in Harrison, et al. (2010) **(14)**.

Paralog: A paralog of a sequence that is known to form prions. Paralogs are duplications of protein sequences within the same genome. Some orthologous proteins are from novel genome annotations, annotated in Harrison, et al. (2010) **(14)**. ZIP metal-ion import proteins that contain a prion-like domain, that were detected through sensitive Hidden Markov Model analysis **(15)**, are classified as paralogs in the PrionOme database.

Prionoid: A sequence with incomplete evidence for prionogenicity, or which behaves in a prion-like way only in certain experimental contexts. The definition of 'prionoids' is discussed in the review Aguzzi and Rajendran **(40)**.

Prion-like: A protein sequence that contains a prion-like domain. Sequences are defined as prion-like if they have an obvious bias for glutamine and/or asparagine residues. Prion-like proteins in budding yeast were determined using the algorithms described in the ref. **(44)**, with a maximum binomial P-value of $1x10^{-10}$. Prion-like sequences defined for nematodes and other eukaryotes were taken from other sources (such as WormBase **(45)**, for nematode sequences).

Pseudogene: A copy of a prion-related gene that is formed via retrotransposition, or other processes of duplication, followed by coding-sequence disablement **(46)**. These annotations were taken from Harrison, et al. (2010) **(14),** and other literature.

Interactor: Proteins shown to interact with a prionogenic protein. Data in this first release of the PrionOme database are restricted to entries listed as binary interactors of prionogenic proteins, taken from the IntAct database **(47)**.

(iii) Prion Determinant:

This is the part of the protein sequence that forms the prion determinant, *i.e.,* that is required for prionogenic activity. In this section, the method of determination for the prion determinant (curated from the literature), and its start and end points in the sequence, are indicated.

(iv) Compositionally-Biased (CB) Regions:

These are regions of the protein sequence that have a bias towards a subset of amino-acid residue types. They are defined using the algorithm described in refs. **(27,44,48)**. In this section are indicated the following: the start and end points in the sequence; the binomial P-value for the CB region; the 'signature' of the compositionally-biased region. In the signature are listed the one-letter codes of amino-acid residue types that define the biased region, in decreasing order of importance **(44)**. That is, 'NY' indicates a region that is rich in N (asparagines) and Y (tyrosines). All compositional biases with binomial P-value $<=10^{-7}$ are listed.

(v) Known Disordered Regions:

These are taken from the DisProt database **(41)**. The ID from the DisProt database, and the start and end positions of the disordered region in the protein sequence are listed.

(vi) Predicted Disordered Regions:

These are predicted using the DISOPRED algorithm, and default parameter values **(49).** The start and end positions of the predicted disordered region in the protein sequence are listed.

(vii) References and Comments:

References were curated from the literature. Original references are given that describe the demonstration of a prionogenic determinant, or, if not appropriate, for the sequencing of a prionogenic sequence, or ortholog or paralog. Additional references are provided if the sequence has been updated, or if additional significant sequence annotations are derived. Original references for the determination of interactors are also given. Comments on the other sections of the database entry are recorded in this section. Specifically also, for any sequence database entry, are listed the PrionOme Database identifiers (PDxxxx) of other sequences in the database, for which that sequence is an interactor, paralog or ortholog. For example, if PD8888 is an interactor of the prionogenic sequence PD7777, then the phrase, 'Interacts with: PD7777' is listed in the Comment section of the entry for PD8888.

(viii) Genomic Information:

In this section, we list the transcripts IDs (taken from the EMBL database); and the coordinates of the gene of the protein sequence in a recent genome assembly. These genome coordinate annotations are taken from the work in

Harrison, et al. (2010) **(14)**, or otherwise transferred from the Ensembl database **(50)**. Listed are the name of the genome assembly, the chromosome, strand direction and start and end points of each exon.

(ix) Structure Information:

Coordinates of structures for the PrionOme database entries are listed here. These are either experimental structures, or comparative models. Experimental structures are taken from the PDB **(51)**. They are colour-coded according to the source of the data (Red for NMR spectroscopy; Green for X-ray crystal structure; Magenta for EM (Electron Microscopy); and Blue for a comparative model). Comparative models were made using the program MODELLER **(52)** for PrP-related proteins for which the sequence alignment cannot be made automatically. These proteins occur in fish species, and tend to have large, interspersed disordered loops.

User Interface

The user interface is designed so that investigators can select and download subsets of prion-related sequence and annotation data, that are of interest. A comprehensive 'Help' page is provided. The database can be accessed by:

(i) *Browsing:* A 'Browse' link on the home page enables users to browse the complete listing of PrionOme database entries and select any entries that they want. The selected entries can then be downloaded in complete database entry format, FASTA format (*i.e.*, just the protein sequence), or as a CLUSTALW multiple sequence alignment **(53)**, using the labeled buttons on the bottom of the display page.

(ii) *Searching:* Users can 'Search' either by keyword, by SQL query, or by BLAST interface. For keyword search, the database or field in which to search must be selected from a pull-down menu. For searching the sequences in the database, users can specify a regular expression of the sort used by the ELM database of linear motifs **(54)**. A screenshot of a keyword search result for the phrase *Anolis* (a lizard) in the ‘Organism' field is shown in Figure 2. Database entries can be downloaded in one of three formats (complete database entry format, FASTA format or CLUSTALW multiple sequence alignment), as detailed above for 'Browsing'. For searching the database using the BLAST query box, the user can either paste in a single sequence in FASTA format, or upload one from a file, and then specify the parameters of the BLAST search.

In addition, on the home page, there are links to convenient lists of database entries, grouped by: *(i)* Prion Name, *(ii)* PrionOme Classification, *(iii)* Organism, and *(iv)* Supporting Information for PrionOme Classification. In some cases, the database entries belong to more than one PrionOme Classification. For example, there are eleven database entries that are classified as both 'Paralog' and 'Interactor'. There are separate links for these database entries that

have multiple PrionOme Classifications. A 'PrionOme News' window is also provided, in which updates and changes to the PrionOme database will be documented.

Examples of Database Usage

We anticipate that the database will be useful as a reference resource for prion biologists, since it enables access to curated, up-to-date listings of a variety of prion-related sequences. In addition, the database will be useful to investigators as a bioinformatic aid to experimental design. Database entries contain cross-references to other database entries (that are interactors or homologs), and to relevant entries in other databases (*e.g.*, DisProt **(41)**).

Database entry browsing can be informative for experimental design. For example, examination of the interactors of Ure2p, the protein that forms the [URE3] prion, produces some interesting results. The entry accession for Ure2p, the protein that forms the [URE3] prion, is PD0040. By keyword-searching in the 'Comment' field for 'Interacts with: PD0040', we find a list of 70 proteins that interact with Ure2p. Browsing through these, we find several that are also candidate prions according to the survey of Alberti, *et al.* **(33)**, *e.g.*, the PrionOme database entry PD0734. This is the G-protein-coupled receptor GPR1, which was shown, *via* the dihydrofolate reductase reconstruction technique, to interact with Ure2p **(55)**. The database entry indicates that GPR1 contains a long (>60 amino acid residues) region that is compositionally-biased for asparagine (N), similar to the prion determinant in Ure2p. This protein is thus a candidate for prion 'cross-seeding', a potential mechanism through which prions can act as inducers of each other **(56)**.

In Tables 2 and 3, we have presented some data analysis of prion-related sequences resulting from queries to the PrionOme database, made using the SQL query interface. Firstly, for Table 2, we queried the PrionOme database for mammalian sequences of the Prion Protein (PrP) family, that have the sequence motif 'YYR'. This motif is an epitope for an antibody that is specific for the disease-related prion form of the PrP protein, PrP^{Sc} **(57)**. We find that all of the PrP sequences that have so far been shown to be prionogenic, retain this YYR sequence motif, as well as the substantial majority of other sequences orthologous to PrP (Table 2). However, none of the paralogous sequences contain this tripeptide (Table 2).

Secondly, as presented in Table 3, we investigated the composition of prion-related domains in *Saccharomyces cerevisiae* (budding yeast). Known prionogenic sequences in *S. cerevisiae* tend to have a pronounced bias for asparagine (N) and/or glutamine (Q) residues **(31,44)**. A microcrystal structure of a small peptide fragment of the Sup35p prion determinant domain indicated that prion domains can be stabilized by hydrogen bonds between Q and N sidechains **(58)**. Also, sidechains that contain six-membered rings, such as phenylalanine (F) and tyrosine (Y) may be able to interact in an arrangement termed π stacking **(59)**. We checked how many of the prionogenic and prion-like sequences from

budding yeast in the PrionOme database also have a compositional bias for F or Y residues (Table 3). A significant fraction of the domains classed as prionogenic contain an F/Y compositional bias (5/28, 18%). Also, a smaller fraction of prion-like domains (as defined in *Methods*), contain an F/Y compositional bias in their prion-like regions (21/306, 7%). Thus, these domains could use π stacking for amyloid stabilization, in addition to Q/N sidechain hydrogen-bonding interactions.

Relationship to other databases

There are no other published, currently maintained databases for prions and for the wide array of prion-related sequences that are encompassed in PrionOme. Two related published databases do however exist. The Prion Disease Database (PDDB) is a repository of time-course mRNA measurements for genes that may be related to prion disease **(60)**. Also, AmyPDB, the amyloid precursor database, contains some information about a small subset of prionogenic sequences that have been demonstrated to propagate through amyloid formation **(61)**. In the future, we will cross-reference our database with these two databases, where possible.

Conclusions and future developments

We have constructed PrionOme, a comprehensive database resource for prions and other prion-related molecules. We presented some examples of database utility for prion researchers. As PrionOme develops further, it will become increasingly useful for investigators as a reference database, and also as an aid for further experimental inquiry. Future developments will include the addition of mutation and polymorphism data for all entries in the database, some of which are not deposited in existing standard databases for protein sequence and genetic variation. We are currently performing quality control on the polymorphism data, and it will be added to the second version of the database in the near future. Also, the literature will be curated for new or overlooked reports of interactors of prionogenic proteins. A graphical display of features in the database sequences will also be added to each database entry. The database will be updated on a weekly basis to incorporate user-submitted data, to add new/overlooked data, and to cover new prions and prion-related phenomena, as they are discovered. More detailed comments, and more comprehensive lists of literature references, will also be added to the database entries, on a regular basis.

As development of this resource is on-going, we will be very happy to receive and act on any constructive comments from peer scientists in the areas of prion biology and protein misfolding, either by email or via the ‘FeedBack’ page on the PrionOme website.

Acknowledgements

This work was funded to by a grant from the PrioNet Canada Network of Centres of Excellence to PH and RS. RS was also funded by the National Centre for Biological Sciences, and was a Senior Fellow of the Wellcome Trust (UK).

References

1. **Prusiner, S.B. (1998) Prions. *Proc Natl Acad Sci U S A*, 95, 13363-13383.**
2. **Prusiner, S.B., Groth, D., Serban, A., Koehler, R., Foster, D., Torchia, M., Burton, D., Yang, S.L. and DeArmond, S.J. (1993) Ablation of the prion protein (PrP) gene in mice prevents scrapie and facilitates production of anti-PrP antibodies. *Proc Natl Acad Sci U S A*, 90, 10608-10612.**
3. **Donne, D.G., Viles, J.H., Groth, D., Mehlhorn, I., James, T.L., Cohen, F.E., Prusiner, S.B., Wright, P.E. and Dyson, H.J. (1997) Structure of the recombinant full-length hamster prion protein PrP(29-231): the N terminus is highly flexible. *Proc Natl Acad Sci U S A*, 94, 13452-13457.**
4. **Riek, R., Hornemann, S., Wider, G., Billeter, M., Glockshuber, R. and Wuthrich, K. (1996) NMR structure of the mouse prion protein domain PrP(121-321). *Nature*, 382, 180-182.**
5. **Pan, K., Baldwin, M., Nguyen, J., Gasset, M., Serban, A., Groth, D., Mehlhorn, I., Huang, Z., Fletterick, R.J., Cohen, F.E. *et al.* (1993) Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins. *Proc Natl Acad Sci U S A.* , 90, 10962-10966**
6. **Schatzl, H.M., Da Costa, M., Taylor, L., Cohen, F.E. and Prusiner, S.B. (1995) Prion protein gene variation among primates. *J Mol Biol*, 245, 362-374.**
7. **Wopfner, F., Weidenhofer, G., Schneider, R., von Brunn, A., Gilch, S., Schwarz, T.F., Werner, T. and Schatzl, H.M. (1999) Analysis of 27 mammalian and 9 avian PrPs reveals high conservation of flexible regions of the prion protein. *J Mol Biol*, 289, 1163-1178.**

9. **Goldfarb, L.G., Brown, P., McCombie, W.R., Goldgaber, D., Swergold, G.D., Wills, P.R., Cervenakova, L., Baron, H., Gibbs, C.J., Jr. and Gajdusek, D.C. (1991) Transmissible familial Creutzfeldt-Jakob disease associated with five, seven, and eight extra octapeptide coding repeats in the PRNP gene. *Proc Natl Acad Sci U S A*, 88, 10926-10930.**
10. **Moore, R., Lee, I., Silverman, G., Harrison, P., Strome, R., Heinrich, C., Karunaratne, A., Pasternak, S., Chishti, M., Liang, Y. *et al.* (1997) Ataxia in prion protein (PrP)-deficient mice is associated with upregulation of the novel PrP-like protein doppel. *J. Mol. Biol.* , 292, 797-817.**
11. **Hegde, R.S., Mastrianni, J.A., Scott, M.R., DeFea, K.A., Tremblay, P., Torchia, M., DeArmond, S.J., Prusiner, S.B. and Lingappa, V.R. (1998) A transmembrane form of the prion protein in neurodegenerative disease. *Science (New York, N.Y.*, 279, 827-834.**
12. **Watts, J., Drisaldi, B., Ng, V., Yang, J., Strome, B., Horne, P., Sy, M., Yoong, L., Young, R., Mastrangelo, P. *et al.* (2007) The CNS glycoprotein Shadoo has PrP(C)-like protective properties and displays reduced levels in prion infections. *EMBO J.* , e-pub. .**

13. **Rivera-Milla, E., Stuermer, C. and Malaga-Trillo, E. (2003) An evolutionary basis for scrapie disease: identification of a fish prion mRNA. *Trends Genet.* , 19, 72-75.**
14. **Harrison, P.M., Khachane, A. and Kumar, M. (2010) Genomic assessment of the evolution of the prion protein gene family in vertebrates. *Genomics*, 95, 268-277.**
15. **Schmitt-Ulms, G., Ehsani, S., Watts, J.C., Westaway, D. and Wille, H. (2009) Evolutionary descent of prion genes from the ZIP family of metal ion transporters. *PloS one*, 4, e7208.**
16. **Lacroute, F. (1971) Non-Mendelian mutation allowing ureidosuccinic acid uptake in yeast. *J. Bacteriol.* , 106, 519-522.**
17. **Cox, B. (1965) [PSI], a cytoplasmic suppressor of super-suppression in yeast. *Heredity*, 20, 505-521.**
18. **Wickner, R. (1994) [URE3] as an altered URE2 protein: evidence for a prion analog in Saccharomyces cerevisiae. *Science (New York, N.Y*, 264, 528-530.**
19. **Sondheimer, N. and Lindquist, S. (2000) Rnq1: An epigenetic modifier of protein function in yeast. *Mol. Cell*, 5, 163-172.**
20. **Shorter, J. and Lindquist, S. (2005) Prions as adaptive conduits of memory and inheritance. *Nat Rev Genets*, 6, 435-450.**
21. **Harrison, P. and Gerstein, M. (2002) Studying genomes through the aeons: protein families, pseudogenes and proteome evolution. *J. Mol. Biol.* , 318, 1155-1174.**
22. **True, H., Berlin, I. and Lindquist, S. (2004) Epigenetic regulation of translation reveals hidden genetic variation to produce comlex traits. *Nature*, 431, 184-187.**
23. **True, H. and Lindquist, S. (2000) A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature*, 407, 477-483.**
24. **Ross, E., Baxa, U. and Wickner, R. (2004) Scrambled prion domains form prions and amyloid. *Mol. Cell. Biol.*, 24, 7206-7213.**
25. **Nakayashiki, T., Kurtzman, C.P., Edskes, H.K. and Wickner, R.B. (2005) Yeast prions [URE3] and [PSI+] are diseases. *Proc Natl Acad Sci U S A*, 102, 10575-10580.**
26. **McGlinchey, R.P., Kryndushkin, D. and Wickner, R.B. (2011) Suicidal [PSI+] is a lethal yeast prion. *Proc Natl Acad Sci U S A*, 108, 5337-5341.**
27. **Harrison, P. and Gerstein, M. (2003) A method to assess compositional bias in biological sequences and its application to prion-like glutamine/asparagine -rich domains in eukaryotic proteomes. *Genome Biol.*, 4, R40 [E-pub].**
28. **Santoso, A., Chien, P., Osherovich, L. and Weissman, J. (2000) Molecular basis of a yeast prion species barrier. *Cell*, 100, 277-288.**
29. **DePace, A., Santoso, A., Hillner, P. and Weissman, J. (1998) A critical role for amino-terminal glutamine/asparagine repeats in the formation and propagation of a yeast prion. *Cell*, 93, 1241-1252.**
30. **Maddelein, M. and Wickner, R. (1999) Two Prion-Inducing Regions of Ure2p Are Nonoverlapping. *NMol. Cell. Biol.*, 19, 4516-4524.**

31. **Harrison, L.B., Yu, Z., Stajich, J.E., Dietrich, F.S. and Harrison, P.M. (2007) Evolution of budding yeast prion-determinant sequences across diverse fungi. *J Mol Biol*, 368, 273-282.**

32. **Michelitsch, M.D. and Weissman, J.S. (2000) A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proc Natl Acad Sci U S A*, 97, 11910-11915.**

33. **Alberti, S., Halfmann, R., King, O., Kapila, A. and Lindquist, S. (2009) A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell*, 137, 146-158.**

34. **Rogoza, T., Goginashvili, A., Rodionova, S., Ivanov, M., Viktorovskaya, O., Rubel, A., Volkov, K. and Mironova, L. Non-Mendelian determinant [ISP+] in yeast is a nuclear-residing prion form of the global transcriptional regulator Sfp1. *Proc Natl Acad Sci U S A*, 107, 10573-10577.**

35. **Nemecek, J., Nakayashiki, T. and Wickner, R.B. (2009) A prion of yeast metacaspase homolog (Mca1p) detected by a genetic screen. *Proc Natl Acad Sci U S A*, 106, 1892-1896.**

36. **Du, Z., Park, K.W., Yu, H., Fan, Q. and Li, L. (2008) Newly identified prion linked to the chromatin-remodeling factor Swi1 in Saccharomyces cerevisiae. *Nature genetics*, 40, 460-465.**

37. **Brown, J.C. and Lindquist, S. (2009) A heritable switch in carbon source utilization driven by an unusual yeast prion. *Genes Dev*, 23, 2320-2332.**

38. **Patel, B.K., Gavin-Smyth, J. and Liebman, S.W. (2009) The yeast global transcriptional co-repressor protein Cyc8 can propagate as a prion. *Nat Cell Biol*, 11, 344-349.**

39. **Roberts, B.T. and Wickner, R.B. (2003) Heritable activity: a prion that propagates by covalent autoactivation. *Genes Dev*, 17, 2083-2087.**

40. **Aguzzi, A. and Rajendran, L. (2009) The transcellular spread of cytosolic amyloids, prions, and prionoids. *Neuron*, 64, 783-790.**

41. **Sickmeier, M., Hamilton, J.A., LeGall, T., Vacic, V., Cortese, M.S., Tantos, A., Szabo, B., Tompa, P., Chen, J., Uversky, V.N. *et al.* (2007) DisProt: the Database of Disordered Proteins. *Nucleic Acids Res*, 35, D786-793.**

42. **Wickner, R.B., Edskes, H.K., Ross, E.D., Pierce, M.M., Baxa, U., Brachmann, A. and Shewmaker, F. (2004) Prion genetics: new rules for a new kind of gene. *Annual review of genetics*, 38, 681-707.**

43. **Wickner, R.B., Edskes, H.K., Ross, E.D., Pierce, M.M., Shewmaker, F., Baxa, U. and Brachmann, A. (2004) Prions of yeast are genes made of protein: amyloids and enzymes. *Cold Spring Harb Symp Quant Biol*, 69, 489-496.**

44. **Harrison, P.M. (2006) Exhaustive assignment of compositional bias reveals universally prevalent biased regions: analysis of functional associations in human and Drosophila. *BMC Bioinformatics*, 7, 441.**

45. **Rogers, A., Antoshechkin, I., Bieri, T., Blasiar, D., Bastiani, C., Canaran, P., Chan, J., Chen, W.J., Davis, P., Fernandes, J. *et al.* (2008) WormBase 2007. *Nucleic Acids Res*, 36, D612-617.**

46. **Harrison, P., Zheng, D., Zhang, Z., Carriero, N. and Gerstein, M. (2005) Transcribed processed pseudogenes in the human genome: an intermediate**

form of expressed retrosequence lacking protein-coding ability. *Nucleic Acids Res.*, 33, 2374-2383.

47. Hermjakob, H., Montecchi-Palazzi, L., Lewington, C., Mudali, S., Kerrien, S., Orchard, S., Vingron, M., Roechert, B., Roepstorff, P., Valencia, A. *et al.* (2004) IntAct: an open source molecular interaction database. *Nucleic Acids Res*, 32, D452-455.

48. Harbi, D., Kumar, M. and Harrison, P.M. (2011) LPS-annotate: complete annotation of compositionally biased regions in the protein knowledgebase. *Database (Oxford)*, 2011, baq031.

49. Ward, J.J., McGuffin, L.J., Bryson, K., Buxton, B.F. and Jones, D.T. (2004) The DISOPRED server for the prediction of protein disorder. *Bioinformatics*, 20, 2138-2139.

50. Hubbard, T.J., Aken, B.L., Ayling, S., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Clarke, L. *et al.* (2009) Ensembl 2009. *Nucleic Acids Res*, 37, D690-697.

51. Berman, H., Henrick, K., Nakamura, H. and Markley, J.L. (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res*, 35, D301-303.

52. Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*, 234, 779-815.

53. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, 23, 2947-2948.

54. Gould, C.M., Diella, F., Via, A., Puntervoll, P., Gemund, C., Chabanis-Davidson, S., Michael, S., Sayadi, A., Bryne, J.C., Chica, C. *et al.* ELM: the status of the 2010 eukaryotic linear motif resource. *Nucleic Acids Res*, 38, D167-180.

55. Tarassov, K., Messier, V., Landry, C.R., Radinovic, S., Serna Molina, M.M., Shames, I., Malitskaya, Y., Vogel, J., Bussey, H. and Michnick, S.W. (2008) An in vivo map of the yeast protein interactome. *Science (New York, N.Y*, 320, 1465-1470.

56. Derkatch, I., Bradley, M., Zhou, P., Chernoff, Y. and Liebman, S. (2001) Prions affect the appearance of other prions: The story of [PIN+]. *Cell*, 106, 171-182.

57. Paramithiotis, E., Pinard, M., Lawton, T., LaBoissiere, S., Leathers, V.L., Zou, W.Q., Estey, L.A., Lamontagne, J., Lehto, M.T., Kondejewski, L.H. *et al.* (2003) A prion protein epitope selective for the pathologically misfolded conformation. *Nature medicine*, 9, 893-899.

58. Nelson, R., Sawaya, M., Balbirnie, M., Madsen, A., Riekel, C., Grothe, R. and Eisenberg, D. (2005) Structure of the cross-beta spine of amyloid-like fibrils. *Nature*, 435, 773-778.

59. Gazit, E. (2002) A possible role for pi-stacking in the self-assembly of amyloid fibrils. *FASEB J.*, 16, 77-83.

60. Gehlenborg, N., Hwang, D., Lee, I.Y., Yoo, H., Baxter, D., Petritis, B., Pitstick, R., Marzolf, B., Dearmond, S.J., Carlson, G.A. *et al.* (2009) The

Prion Disease Database: a comprehensive transcriptome resource for systems biology research in prion diseases. *Database (Oxford)*, 2009, bap011.

61. Pawlicki, S., Le Bechec, A. and Delamarche, C. (2008) AMYPdb: a database dedicated to amyloid precursor proteins. *BMC Bioinformatics*, 9, 273.

Figure Legends

Figure 1: A screenshot of an entry in the PrionOme database, for the Shadoo transcribed pseudogene in the human genome.

Figure 2: An example of the results of a keyword search, for the term 'Anolis' in the 'Organism' field. Five entries for *Anolis caroliensis* (a lizard), appear listed.

Tables

Table 1: Summary of database content

PrionOme Classification	Number of Database Entries
Interactor	460
Ortholog	949
Paralog	205
Prion-like	417
Prionogenic	56
Prionoid	6
Pseudogene	13
TOTAL*	2019

** This is not an arithmetic sum of the individual PrionOme Classification categories, since some database entries have multiple PrionOme Classifications.*

Table 2: Counts of mammalian sequences in the PrionOme database with the sequence motif tyrosine-tyrosine-arginine (YYR)

PrionOme Classification	# of sequences with the YYR sequence motif	# of sequences without the YYR sequence motif
Prionogenic PrP sequences (Major Prion Protein)	20	0
Orthologs of Major Prion Protein	413†	9
Paralogs of Major Prion Protein	0	120

† As an example, the SQL query to obtain this data is: *SELECT prion_id , name FROM main WHERE sequence REGEXP 'YYR' AND prion_type='Ortholog' AND taxonomy LIKE '%mammal%' .*

Table 3: A summary of prionogenic and prion-like sequences in the PrionOme database from *Saccharomyces cerevisiae,* that have Q/N (glutamine/asparagine) and F/Y (phenylalanine/tyrosine) compositional biases

PrionOme Classification	# of sequences that have Q/N bias, but no F/Y bias	# of sequences that have Q/N bias, and F/Y bias*
Prionogenic**	23	5 ***
Prion-Like****	285	21

* F/Y compositional biases with binomial P-values <=10^{-7} are considered (see Methods for details).

** This data includes the Alberti, et al. (2009) data from screens for candidate prionogenic sequences.

*** As an example, the SQL query to obtain this data is: *SELECT prion_id , name FROM main WHERE prion_type LIKE'%Prionogenic%' AND bias REGEXP 'NY|QY|NF|QF|YN|FN|YQ|FQ' AND organism LIKE '%cerevisiae%'*

**** Some polymorphic sequences have been removed from these totals.

General Information	
Prion Id	PD1167
UniProt	
SwissProt	
Source Organism	Homo sapiens (Human)
Taxonomy	Eukaryota › Metazoa › Chordata › Craniata › Vertebrata › Euteleostomi › Mammalia › Eutheria › Euarchontoglires › Primates › Haplorrhini › Catarrhini › Hominidae › Homo
Sequence Length	124
Protein Name	Shadoo transcribed pseudogene (SprnP1)
Prion Name	N/A
Sequence	MNWAPANGWALLQEAAFLC#CGRGGLRGGSRGASRVCVRPALCYGAPGSSLRVAAAGAAG SGWRRAAGPRERGLEDEEDWVPGGNRTGS#GIYSYCTWTSGAGPTGGLCLCVGLGGALGA LGLLRP

PrionOme Classification	
PrionOme Classification	**Supporting Information for PrionOme Classification**
Pseudogene	sequence homology

Prion Determinant		
Method of Determination	**Start**	**End**

Compositionally Biased Regions			
Start	**End**	**P-value**	**Signature**

Known Disordered Regions		
DisProt	**Start**	**End**

Predicted Disordered Regions	
Start	**End**

References and Comments	
Reference	1. Harrison PM, Khachane A, Kumar M."**Genomic assessment of the evolution of the prion protein gene family in vertebrates.**" Genomics. 2010 May;95(5):268-77. Epub 2010 Mar 3. PubMed: 20206252 2. Premzl M, Gready JE, Jermiin LS, Simonic T, Marshall Graves JA."**Evolution of vertebrate genes related to prion and Shadoo proteins--clues from comparative genomic analysis.**"Mol Biol Evol. 2004 Dec;21(12):2210-31. Epub 2004 Sep 1. PubMed: 15342797
Comment	This is a transcribed pseudogene that overlaps the SYCE1 gene in human.

Genomic Information					
Transcript	**Assembly**	**Chromosome**	**Strand**	**Start**	**End**
FLJ44653	Homo_sapiens_NCBI36.53	10	-	135231597	135231960

Structure Information

5 results

Prion Id	Protein Name	Organism	Check ALL
PD1125	Prion Protein 1 (PrP1)	Anolis carolinensis (Green anole) (American chameleon)	☐
PD1126	Prion Protein 2 (PrP2) pseudogene	Anolis carolinensis (Green anole) (American chameleon)	☐
PD1127	Prion Protein 3 (PrP3)	Anolis carolinensis (Green anole) (American chameleon)	☐
PD1128	Doppel	Anolis carolinensis (Green anole) (American chameleon)	☐
PD1129	Shadoo (Sprn)	Anolis carolinensis (Green anole) (American chameleon)	☐

Multiple Alignment

Fasta

Complete Database Entry Format