



Original article

KinetochoresDB: a comprehensive online resource for the kinetochore and its related proteins

Chen Li¹, Steve Androulakis², Ashley M. Buckle¹ and Jiangning Song^{1,3,4,*}

¹Infection and Immunity Program, Biomedicine Discovery Institute and the Department of Biochemistry and Molecular Biology, Faculty of Medicine, Monash University, Melbourne, Victoria, 3800, Australia, ²Monash Bioinformatics Platform, Monash University, Melbourne, Victoria, 3800, Australia, ³Monash Centre for Data Science, Faculty of Information Technology, Monash University, Victoria, 3800, Australia, and ⁴National Engineering Laboratory of Industrial Enzymes and Key Laboratory of Systems Microbial Biotechnology, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, 300308, China

*Corresponding author: Tel: +61 3 9902 9304; Fax: +61 3 9902 9500; E-mail: Jiangning.Song@monash.edu

Correspondence may also be addressed to Ashley M. Buckle. Tel: +61 3 9902 9313; Fax: +61 3 9902 9500; Email: Ashley.Buckle@monash.edu

Citation details: Li, C., Androulakis, S., Buckle, A.M., *et al.* KinetochoresDB: a comprehensive online resource for the kinetochore and its related proteins. *Database* (2016) Vol. 2016: article ID baw019; doi:10.1093/database/baw019

Received 5 June 2015; Revised 15 January 2016; Accepted 5 February 2016

Abstract

KinetochoresDB is an online resource for the kinetochore and its related proteins. It provides comprehensive annotations on 1554 related protein entries in terms of their amino acid sequence, protein domain context, protein 3D structure, predicted intrinsically disordered region, protein–protein interaction, post-translational modification site, functional domain and key metabolic/signaling pathways, integrating several public databases, computational annotations and experimental results. KinetochoresDB provides interactive and customizable search and data display functions that allow users to interrogate the database in an efficient and user-friendly manner. It uses PSI-BLAST searches to retrieve the homologs of all entries and generate multiple sequence alignments that contain important evolutionary information. This knowledgebase also provides annotations of single point mutations for entries with respect to their pathogenicity, which may be useful for generation of new hypotheses on their functions, as well as follow-up studies of human diseases.

Database URL: <http://lightning.med.monash.edu/kinetochoresDB2/>

Introduction

During cell mitosis and meiosis, the kinetochore plays a critical role of locating the attachments on chromosomes and pulling sister chromatids apart. Assembled on centromeric chromatin, kinetochore functions during the cell cycle (1–6). During the last few decades, numerous studies of the kinetochore and its related proteins have characterized its function, architecture and the repertoire of its related proteins using biochemistry, structural biology and cell biology techniques (4, 7–11). Both the stability of the kinetochore–microtubule interface and mutations occurring in the kinetochore and its related proteins are associated with a number of human diseases (12–15). Dynamics studies of the kinetochore have also shown that deregulation of the kinetochore–microtubule dynamics frequently results in chromosome instability, leading to the development of cancer (10, 11, 16). Other experimental studies reveal that mutations of the kinetochore and its related proteins are closely linked to human diseases. For example, the adenomatous polyposis coli protein, localized in both centrosome and kinetochore, contains ~15 disease-associated mutations that cause familial adenomatous polyposis (14, 15) and Medulloblastoma (12).

Despite its biological significance and our increasing awareness of its potential roles in human diseases, there is currently a paucity of publically available databases or resources that focus on providing comprehensive functional annotations of the kinetochore and its related proteins. The only available database is MiCroKiTS (17), an integrated online resource for kinetochore, midbody, telomere, centrosome and spindle proteins. However, important annotations of entries in MiCroKiTS are not available in terms of protein 3D structure, protein interaction partners, metabolic/signaling pathways etc., all of which are crucial aspects for follow-up functional studies of these proteins.

In an effort to address this knowledge gap, we created KinetochoreDB, which integrates several public databases, computational annotations and experimental results for currently 1554 related entries. KinetochoreDB contains several important features, the majority of which are not available in MiCroKiTS (Table 1):

- i. It provides annotations of protein 3D structure when structural information is available. For protein entries with available structural information, the corresponding Protein Data Bank (PDB) IDs and their related information are provided. In addition, information on predicted intrinsic disorder is provided, which is particularly important for providing structural insights into those entries in KinetochoreDB whose 3D structures have not been solved.

- ii. It provides comprehensive annotations of single point mutations and their pathogenic effects. These mutations are classified as either pathogenic or nonsense mutations in KinetochoreDB. For disease-associated pathogenic mutations, KinetochoreDB provides details of the disease caused by the mutation, allows users to search the entire database with the disease name of interest, and provides user-friendly options to browse the related kinetochore proteins that harbor such disease-associated mutations.
- iii. It provides metabolic/signaling pathway information for each entry by cross-referencing the KEGG database. Such information is important for understanding the functions of kinetochore proteins from a biochemical network perspective. In particular, the pathway information and the link to Kyoto Encyclopedia of Genes and Genomes (KEGG) will be provided if an entry has pathway information available in KEGG.
- iv. It provides multiple sequence alignments (MSAs) for all included entries, thereby allowing users to readily identify evolutionarily conserved regions within the family of a kinetochore protein. In addition, the visualization of MSAs implemented by Jalview is user-friendly and customizable.
- v. It provides convenient user enquiry and new entry submission options by enabling users to automatically upload their newly discovered sequences into the online database.

Materials and Methods for Database Construction

We define ‘kinetochore and its related proteins’ with respect to protein subcellular location and Gene Ontology (GO). The entries of KinetochoreDB originate from three major resources; QuickGo database (18), UniProt database (19) and MiCroKiTS, and the database was populated as follows. From MiCroKiTS, we obtained data entries that have been experimentally verified to be located in kinetochore. By searching GO terms from QuickGo with the keyword ‘kinetochore’, we obtained 64 GO terms related to kinetochore. For each GO term, we searched and filtered the reviewed entries from the UniProt database to ensure that all the downloaded entries contain the GO annotation. Applying this procedure resulted in 53 GO terms including 25 cellular component terms, 2 molecular function terms and 26 biological process terms (Supplementary Table S1). In addition, we queried ‘subcellular location’ with the keyword ‘kinetochore’ in UniProt and downloaded the entries with published experimental evidence from the search results. After the removal of redundant entries, the resulting dataset contained a total of 1554

Table 1. Comparison between KinetochoresDB and MiCroKiTS

Annotation category	KinetochoresDB	MiCroKiTS
Target	Kinetochores and its related proteins	Kinetochores, centrosome, midbody, telomere and spindle proteins
Biological model	Yes	No
Protein 3D structure	Yes, detailed structural information available; customizable display	No
Protein intrinsic disorder	Yes, predicted by VSL2B	No
Protein interaction partner	Yes, detailed information available	No
Metabolic/signaling pathway	Yes	No
Disease-associated mutations	Yes, incorporating both disease-associated and nonsense mutations	No
Evolutionary conservation	Yes, MSAs curated and visualized using Jalview	No MSA
User enquiry and submission	Yes	No

carefully reviewed entries. The detailed procedures of database construction and data collection are illustrated in [Figure 1](#) and a statistical summary can be found in [Figure 2](#) and [Table 2](#), respectively.

For each entry, KinetochoresDB integrates several public resources, including the UniProt database, NCBI Gene database (<http://www.ncbi.nlm.nih.gov/gene/>), 1000 Genome Project database (20) (for human proteins), BioModels database (21), Research Collaboratory for Structural Bioinformatics (RCSB) PDB (22), Online Mendelian Inheritance in Man (23), BioGRID 3.2 (24), Pfam database (25) and KEGG (26), in order to provide a comprehensive description with respect to basic protein information, protein structure, function, mutation and evolutionary conservation. An important feature of KinetochoresDB is the provision of 3D structure. To achieve this, we manually searched all the entries against the PDB database using their corresponding UniProt identifiers and protein names. For protein complex structures, we identified the PDB chain for each entry and annotated the entry with that chain. In addition, we also generated MSAs using all homologous sequences for each protein entry. Homologous sequences were retrieved by PSI-BLAST (27) search against the Swiss-Prot dataset obtained from UniProt. The alignments were generated using Clustal Omega (28). We also predicted natively disordered regions for all protein entries using one of the most widely used disorder predictors, namely VSL2B (29). A residue is annotated as disordered by VSL2B if its prediction score was >0.5 .

We used Jmol (<http://jmol.sourceforge.net/>) and pViz (30) for visualization of protein structures, and Jalview (31) for customizable editing and display of MSAs for each protein entry. In order to provide the annotations of the domain context including functional domains and sites for each protein entry in KinetochoresDB, another Javascript based plug-in, The Protein Feature View ([\[lic.github.io/proteinfeatureview/\]\(http://lic.github.io/proteinfeatureview/\)\), was used to display protein structural, functional domains and sites in a better and more interactive way. The display of protein domain context can be found in ‘Protein Domain Context’ section of the webpage. The information stored in KinetochoresDB resides in a MySQL relational database. A highly interactive web front-end to the data was implemented using the Javascript framework, JQuery. Apache Tomcat handles serving of data to users on the web, utilizing a set of Java Servlets and JavaServer Pages for data searching and viewing.](http://andreasp</p>
</div>
<div data-bbox=)

Database Utility

The ‘Search’ page (<http://lightning.med.monash.edu/kinetochoresDB2/Search.jsp>) ([Figure 3](#)) allows users to search the database in several different ways. These search options can be generally classified into two groups: search with ID or search with keywords. Examples are provided below to assist users to understand how to perform the search. When searching the database with IDs, KinetochoresDB provides two different kinds of IDs to facilitate the search: UniProt ID and KinetochoresDB ID. The latter is composed of ‘KD’ and five digits, e.g. KD00095. Considering that the database includes 1554 entries in total, each entry is accordingly numbered as KD00001–KD01554. In addition KinetochoresDB offers alternative search options with keywords. These include protein name, kinase name, post-translational modification (PTM) type, name of protein interaction partner and name of diseases caused by single point mutations ([Figure 3B](#)). After selecting the ‘Submit’ button, the corresponding search results will be shown on the webpage. For each entry, there are generally 10 sections of structural and functional categories, including general information, protein domain context, protein structure, disordered regions prediction,

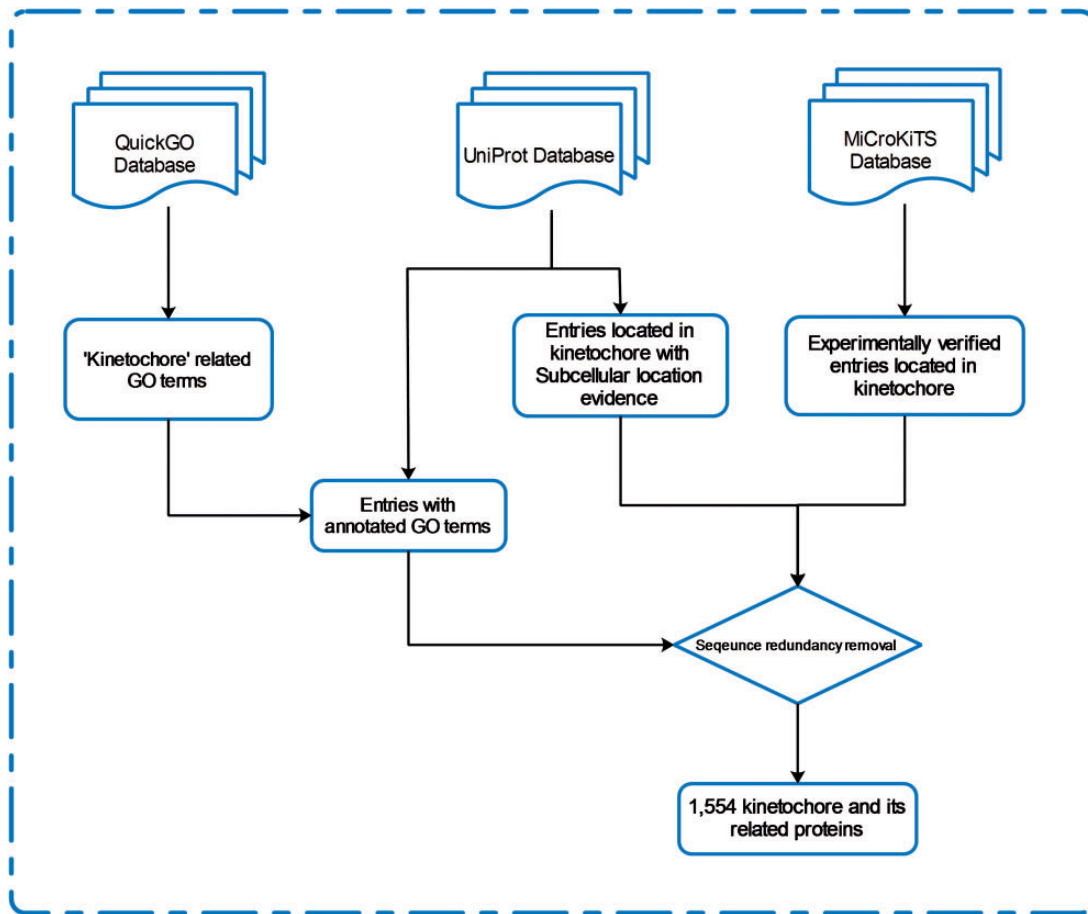


Figure 1. The schema of database construction and data collection processes.

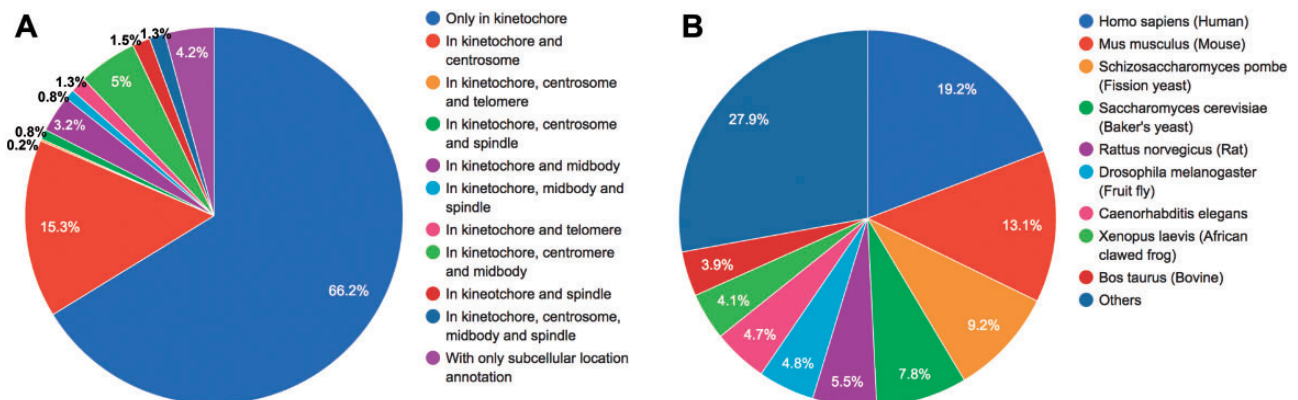


Figure 2. Statistical summary of locations and species of KinetochoreDB entries. **(A)** Distribution of protein locations according to MiCroKiTS and protein subcellular location annotations from UniProt (entries with GO annotations only were excluded). **(B)** Distribution of species of all KinetochoreDB entries.

interaction partner, PTMs, Pfam domain, protein mutation, metabolic/signaling pathway and protein alignment with homologs. To provide an illustration of the annotations for each entry in KinetochoreDB, we use 'UniProt ID = O14965' (KinetochoreDB ID = 'KD01531') as an example query. The resulting page with ten sections is shown in Figure 4.

For protein overview and domain context, KinetochoreDB uses The Protein Feature View plug-in to facilitate the general description of protein entries including functional sites and domains (Supplementary Figure S1A), allowing a more detailed inspection of the constituent domains of the protein. For protein structure overview, KinetochoreDB provides two different ways to inspect the

3D structures. A single structure for the current protein entry can be examined by clicking the ‘View’ button to launch Jmol, a Java applet for displaying 3D structures (Supplementary Figure S1C). Multiple structures can also be viewed together as an ensemble using pViz (Supplementary Figure S1B).

With respect to protein–protein interaction, the interaction partner is highlighted if this protein is also an entry of KinetochoreDB. As a result of our search strategy (see ‘Materials and Methods for Database Construction’ for details), certain proteins in MSAs might not be included in the current KinetochoreDB. To facilitate the comparison between entries in KinetochoreDB and their homologs, we

archived the homologs by extracting protein UniProt IDs. Detailed information for these files is available in the ‘Protein alignment’ section of the webpage.

KinetochoreDB will be updated on a regular basis (typically every 2–3 months) to include newly available entries from various resources, in order to allow an up-to-date archive of recent results of the kinetochore and its related proteins. All the updates will be carefully reviewed prior to their release. In addition, we also allow users to submit their research findings related to kinetochore to our KinetochoreDB. New submission should include protein/gene detailed information (such as sequence, molecular weight and protein function description), protein structural information, protein functional annotation, single point mutation and its pathogenicity and metabolic/signaling pathway (Supplementary Figure S2). After careful review and validation, new data will be included in KinetochoreDB and made publically available.

Table 2. Statistical summary of the information contained in KinetochoreDB

Number of entries	1554
Number of protein structures	1232
Number of protein interactions	49 931
Number of biological models	55
Number of mutations (including residue substitution, insertion and deletion)	1424
Number of KEGG pathways	202
Number of Pfam domains	1173
Number of PTMs sites	4027

Discussion

Certain protein entries in KinetochoreDB harbor mutations. We thus provide a brief statistical analysis of these mutations. A total of 1424 mutations were found to occur in the 206 entries in KinetochoreDB. Among these, 689

A

ID Search

Search with Uniprot ID or database ID.

UniprotID

B

Keyword Search

Use different kinds of keywords to search database

Protein Name

Kinase

Post-translational Modification Type

Interaction Partner Name

Disease caused by mutation

Species

Figure 3. The search interface and options provided by KinetochoreDB. (A) Protein ID search with either UniProt ID or KinetochoreDB ID. (B) Keyword search with the protein name, kinase, PTM type, interaction partner name, disease or species.



Downloaded from https://academic.oup.com/database/article/doi/10.1093/database/baw019/2630212 by guest on 19 May 2024

Figure 4. A webpage showing protein detailed information for search results in KinetochoresDB using the UniProt ID O14965 as the query. The results are classified and organized in ten sections including protein information, protein domain context, protein structure, disordered region prediction, PTMs, protein interaction, protein mutation, Pfam domain, metabolic/signaling pathway and MSA.

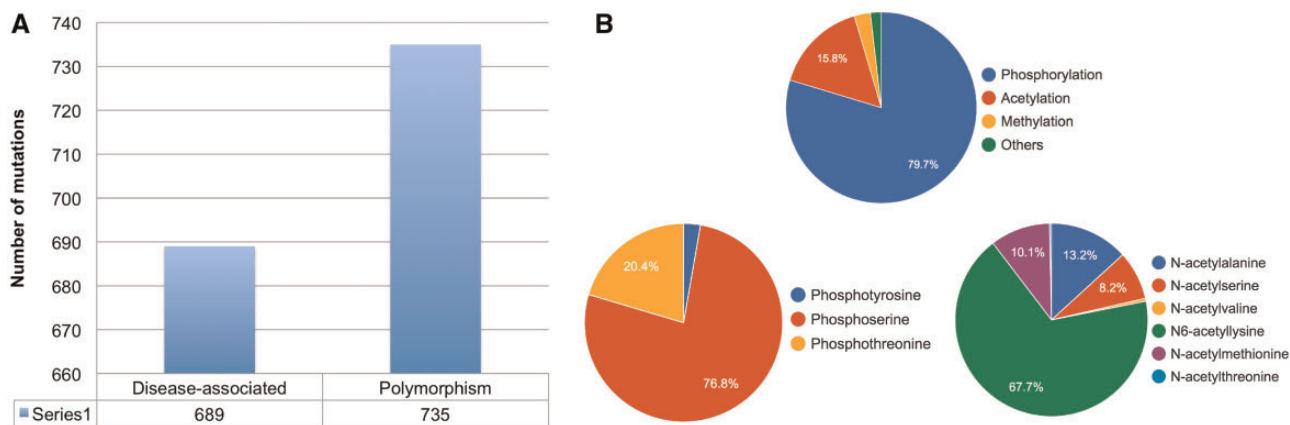


Figure 5. Statistical analysis of single point mutations and protein PTM types in KinetochoreDB. **(A)** Distribution of disease-associated mutations and polymorphisms. **(B)** Distribution of different major types of protein PTM, e.g. phosphorylation, acetylation, methylation and others. The distribution of sub-types of phosphorylation and acetylation is also shown.

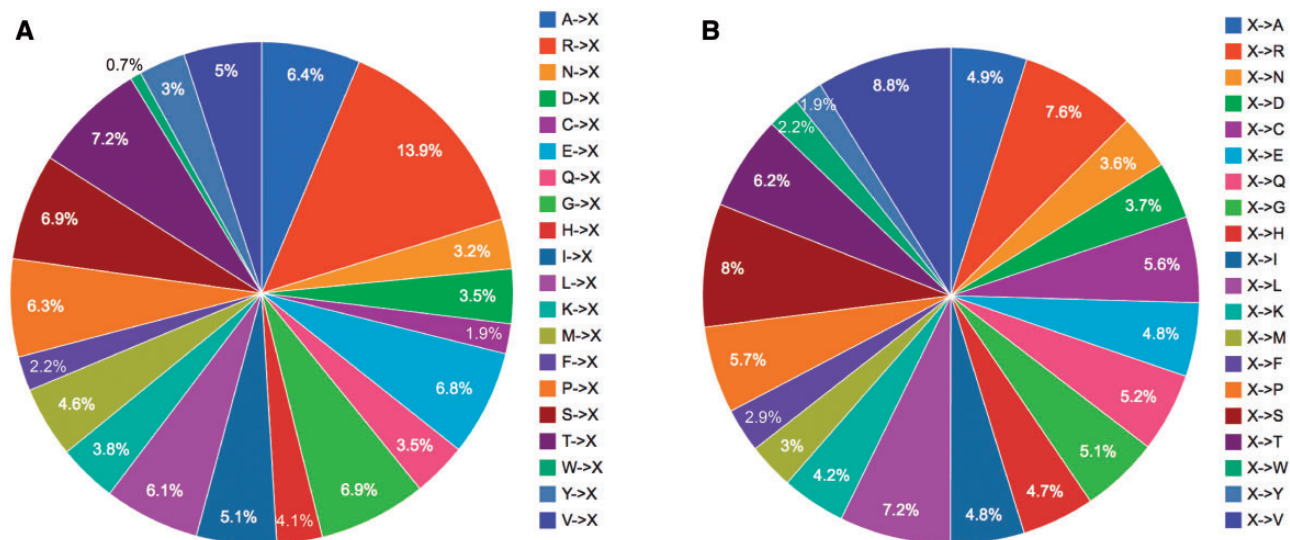


Figure 6. Statistical analysis of single point mutations in KinetochoreDB according to two mutation patterns **(A)** (A..V)->X and **(B)** X->(A..V), where X denote any type of amino acid).

(48.4%) mutations were found to cause diseases, while 735 (51.6%) were nonsense mutations (Figure 5A). PTMs, on the other hand, extend the chemical repertoire of amino acids by attaching new chemical groups and small molecules to the side chains of amino acids. Using the available PTM annotations in KinetochoreDB, we analysed the distribution of different types of PTMs for all the entries. We note that the kinetochore and its related proteins possess many PTM sites, the top three of which are phosphorylation, acetylation and methylation, respectively (Figure 5B). The distribution of different subtypes of acetylation and phosphorylation is also shown in Figure 5B. We further plotted the distributions of different types of mutations in Figure 6. We can see that there is no apparent difference between the two types of mutation patterns (Figure 6).

In addition, with the comprehensive dataset from KinetochoreDB, we conducted a statistical analysis of the number of proteins involved in different GO terms including cellular component, molecular function and biological process. The results are shown in Figure 7. For cellular component, the 10 top ranked GO terms are condensed chromosome kinetochore (GO:0000777), kinetochore (GO:0000776), condensed nuclear chromosome kinetochore (GO:0000778), cytoplasmic dynein complex (GO:0005868), condensed nuclear chromosome, centromeric region (GO:0000780), Ndc80 complex (GO:0031262), DASH complex (GO:0042729), Chromosome passenger complex (GO:0032133), Condensed chromosome outer kinetochore (GO:0000940) and kinetochore microtubule (GO:00005828). For molecular function, the top ranked GO terms are microtubule motor activity

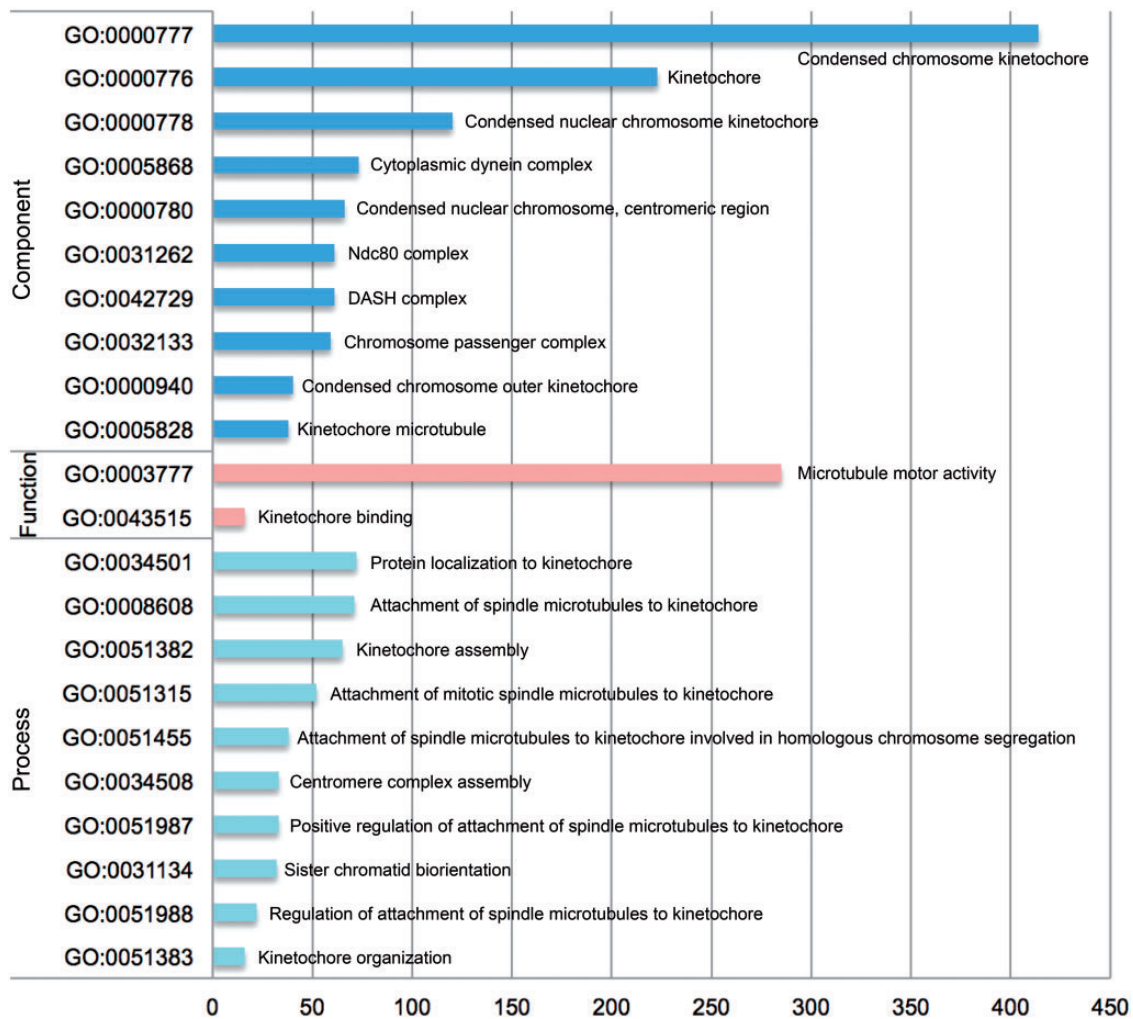


Figure 7. Statistical analysis of the number of entries from KinetochoreDB involved in different GO terms, which were grouped according to three categories: cellular component, molecular function and cellular process.

(GO:0003777) and kinetochore binding (GO:0043515). For biological process, the 10 top ranked GO terms are protein localization to kinetochore (GO:0034501), attachment of spindle microtubules to kinetochore (GO:0008608), kinetochore assembly (GO:0051382), attachment of mitotic spindle microtubules to kinetochore (GO:0051315), attachment of spindle microtubules to kinetochore involved in homologous chromosome segregation (GO:0051455), centromere complex assembly (GO:0034508), positive regulation of attachment of spindle microtubules to kinetochore (GO:0051987), sister chromatid biorientation (GO:0031134), regulation of attachment of spindle microtubules to kinetochore (GO:0051988) and kinetochore organization (GO:0051383). More specifically, 414, 285 and 72 proteins contain the annotation of condensed chromosome kinetochore (GO:0000777), microtubule motor activity (GO:0003777) and protein localization to kinetochore (GO:0034501).

It should be noted that the statistical results regarding the PTM sites, mutations and GO terms are merely observations based on the entries in KinetochoreDB, rather than being interpreted as a biologically significant finding, as the entries included in KinetochoreDB only represent a subset of the entire proteome and are from different species.

Conclusions

The kinetochore and its related proteins play extremely important roles during cell division and mitosis. In the past few decades, research on this topic has attracted a great deal of interest, not only because they are important for the cell cycle, mitosis and meiosis (1–6), but also because they harbor mutations that can cause human diseases (12–15). In this context, databases such as KinetochoreDB that provide comprehensive annotations on the repertoire of kinetochore-related proteins will greatly facilitate in-depth

functional investigation of these proteins and their relationships with human diseases. Through effective data integration from multiple public resources, KinetochoreDB has collected large amounts of information for related protein entries with respect to their amino acid sequence, protein 3D structure, biological function and evolutionary conservation. By providing comprehensive functional annotations of all available kinetochore-related proteins, we believe that this online resource will be used as a powerful tool to bridge functional characterization and disease-associated mutation studies of this important class of proteins.

In the future, we will endeavor to improve and update the annotations and analysis of data entries in KinetochoreDB by the following means: (i) We will keep the database updated and provide up-to-date information to synchronize with the research progress in the kinetochore and its related proteins; (ii) We will integrate genomic information into our database and source these data from publicly available information or bioinformatics programs. These include coding sequence, transcription factor binding site, enhancer, promoter and other upstream or downstream regulatory information; (iii) We will combine other state-of-the-art predictors to annotate the natively disordered regions of all entries in the database, while highlighting the consensus prediction. Meanwhile, we will also collect experimentally verified disordered regions from DisProt (32), the most comprehensive resource dedicated to annotating disordered region of proteins; (iv) we will encourage experimental biologists to contribute to the development of KinetochoreDB by submitting their recent findings by making available newly added entries in the database after careful review.

In addition we will continue to improve and update the annotations and analysis of all entries in KinetochoreDB by implementing secondary analysis functions of the database and by integrating high-throughput experimental data. In particular, we will explore gene expression microarray data, transcriptomics and proteomics and detailed functional pathway data, so as to provide a comprehensive useful resource for the wider research community.

Supplementary data

Supplementary data are available at Database Online.

Acknowledgements

We would like to thank Dr Andreas Prlić for his kind help with the ‘The Protein Feature View’ plug-in to display protein domain context in KinetochoreDB.

Funding

This work was supported by grants from the National Natural Science Foundation of China (61202167, 61303169), the Hundred

Talents Program of the Chinese Academy of Sciences (CAS), Knowledge Innovative Program of CAS (KSCX2-EW-G-8) of CAS, and the National Health and Medical Research Council of Australia (NHMRC). JS is a Recipient of the Hundred Talents Program of CAS. AMB is an NHMRC Senior Research Fellow (1022688). Funding for open access charge: National Health and Medical Research Council (490989).

Conflict of interest. None declared.

References

1. Brinkley,B.R., Tousson,A., and Valdivia,M.M. (1985) The kinetochore of mammalian chromosomes: structure and function in normal mitosis and aneuploidy. *Basic Life Sci.*, 36, 243–267.
2. Chan,G.K., Liu,S.T., and Yen,T.J. (2005) Kinetochore structure and function. *Trends Cell Biol.*, 15, 589–598.
3. McAinsh,A.D., Tytell,J.D., and Sorger,P.K. (2003) Structure, function, and regulation of budding yeast kinetochores. *Annu Rev. Cell Dev. Biol.*, 19, 519–539.
4. Westermann,S., Drubin,D.G., and Barnes,G. (2007) Structures and functions of yeast kinetochore complexes. *Annu. Rev. Biochem.*, 76, 563–591.
5. Stankovic,A. and Jansen,L.E. (2013) Reductionism at the vertebrate kinetochore. *J Cell Biol.*, 200, 7–8.
6. Rago,F. and Cheeseman,I.M. (2013) Review series: The functions and consequences of force at kinetochores. *J. Cell Biol.*, 200, 557–565.
7. Yang,Y., Wu,F., Ward,T. *et al.* (2008) Phosphorylation of HsMis13 by Aurora B kinase is essential for assembly of functional kinetochore. *J. Biol. Chem.*, 283, 26726–26736.
8. Wan,X., O’Quinn,R.P., Pierce,H.L. *et al.* (2009) Protein architecture of the human kinetochore microtubule attachment site. *Cell*, 137, 672–684.
9. Sakuno,T., Tada,K., and Watanabe,Y. (2009) Kinetochore geometry defined by cohesion within the centromere. *Nature*, 458, 852–858.
10. Tanaka,T.U. and Desai,A. (2008) Kinetochore-microtubule interactions: the means to the end. *Curr. Opin. Cell Biol.*, 20, 53–63.
11. Bakhom,S.F., Thompson,S.L., Manning,A.L. *et al.* (2009) Genome stability is ensured by temporal control of kinetochore-microtubule dynamics. *Nat. Cell Biol.*, 11, 27–35.
12. Huang,H., Mahler-Araujo,B.M., Sankila,A. *et al.* (2000) APC mutations in sporadic medulloblastomas. *Am. J. Pathol.*, 156, 433–437.
13. Miyaki,M., Nishio,J., Konishi,M. *et al.* (1997) Drastic genetic instability of tumors and normal tissues in Turcot syndrome. *Oncogene*, 15, 2877–2881.
14. Stella,A., Montera,M., Resta,N. *et al.* (1994) Four novel mutations of the APC (adenomatous polyposis coli) gene in FAP patients. *Hum. Mol. Genet.*, 3, 1687–1688.
15. van der Luijt,R.B., Khan,P.M., Vasen,H.F. *et al.* (1997) Molecular analysis of the APC gene in 105 Dutch kindreds with familial adenomatous polyposis: 67 germline mutations identified by DGGE, PTT, and southern analysis. *Hum. Mutat.*, 9, 7–16.
16. Kops,G.J., Weaver,B.A., and Cleveland,D.W. (2005) On the road to cancer: aneuploidy and the mitotic checkpoint. *Nat. Rev. Cancer*, 5, 773–785.
17. Huang,Z., Ma,L., Wang,Y. *et al.* (2015) MiCroKiTS 4.0: a database of midbody, centrosome, kinetochore, telomere and spindle. *Nucleic Acids Res.*, 43, D328–D334.

18. Huntley,R.P., Sawford,T., Mutowo-Meullenet,P. *et al.* (2015) The GOA database: gene ontology annotation updates for 2015. *Nucleic Acids Res.*, 43, D1057–D1063.
19. Consortium, T.U. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, 43, D204–D212.
20. Abecasis,G.R., Auton,A., Brooks,L.D. *et al.* (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491, 56–65.
21. Juty,N., Ali,R., Glont,M. *et al.* (2015) BioModels: Content, Features, Functionality, and Use. *CPT Pharmacometrics Syst. Pharmacol.*, 4, e3.
22. Rose,P.W., Beran,B., Bi,C. *et al.* (2011) The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res.*, 39, D392–D401.
23. Hamosh,A., Scott,A.F., Amberger,J.S. *et al.* (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, 33, D514–D517.,
24. Chatr-Aryamontri,A., Breitkreutz,B.J., Heinicke,S. *et al.* (2013) The BioGRID interaction database: 2013 update. *Nucleic Acids Res.*, 41, D816–D823.
25. Finn,R.D., Bateman,A., Clements,J. *et al.* (2014) Pfam: the protein families database. *Nucleic Acids Res.*, 42, D222–D230.
26. Kanehisa,M., Goto,S., Sato,Y. *et al.* (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.*, 42, D199–D205.
27. Altschul,S.F., Madden,T.L., Schaffer,A.A. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25, 3389–3402.
28. Sievers,F., Wilm,A., Dineen,D. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, 7, 539.
29. Peng,K., Radivojac,P., Vucetic,S. *et al.* (2006) Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics*, 7, 208.
30. Mukhyala,K. and Masselot,A. (2014) Visualization of protein sequence features using JavaScript and SVG with pViz.js. *Bioinformatics*, 30, 3408–3409.
31. Waterhouse,A.M., Procter,J.B., Martin,D.M. *et al.* (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25, 1189–1191.
32. Vucetic,S., Obradovic,Z., Vacic,V. *et al.* (2005) DisProt: a database of protein disorder. *Bioinformatics*, 21, 137–140.