D244–D250 *Nucleic Acids Research, 2009, Vol. 37, Database issue* *Published online 30 October 2008*
doi:10.1093/nar/gkn834

Kinomer v. 1.0: a database of systematically classified eukaryotic protein kinases

David M. A. Martin, Diego Miranda-Saavedra and Geoffrey J. Barton*

College of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, Scotland, UK

Received August 15, 2008; Revised October 13, 2008; Accepted October 14, 2008

ABSTRACT

The regulation of protein function through reversible phosphorylation by protein kinases and phosphatases is a general mechanism controlling virtually every cellular activity. Eukaryotic protein kinases can be classified into distinct, well-characterized groups based on amino acid sequence similarity and function. We recently reported a highly sensitive and accurate hidden Markov model-based method for the automatic detection and classification of protein kinases into these specific groups. The Kinomer v. 1.0 database presented here contains annotated classifications for the protein kinase complements of 43 eukaryotic genomes. These span the taxonomic range and include fungi (16 species), plants (6), diatoms (1), amoebas (2), protists (1) and animals (17). The kinomes are stored in a relational database and are accessible through a web interface on the basis of species, kinase group or a combination of both. In addition, the Kinomer v. 1.0 HMM library is made available for users to perform classification on arbitrary sequences. The Kinomer v. 1.0 database is a continually updated resource where direct comparison of kinase sequences across kinase groups and across species can give insights into kinase function and evolution. Kinomer v. 1.0 is available at http://www.compbio.dundee.ac.uk/kinomer/.

INTRODUCTION

The regulation of protein function through reversible phosphorylation by protein kinases and phosphatases is a widespread cellular mechanism thought to control virtually every cellular activity (1), and abnormal levels of phosphorylation are known to be responsible for severe diseases (2).

Hanks and Hunter were the first to report that sequence similarity of kinase catalytic domains reflects protein kinase function and/or mode of regulation (3,4). Observation of distinct clades where function segregated with sequence similarity allowed Hanks and Hunter to divide the protein kinase superfamily into specific 'groups'. The currently accepted classification of the eukaryotic protein kinase superfamily considers eight 'conventional' protein kinase groups (ePKs) and four 'atypical' groups (aPKs) (5,6). Among the ePKs are the AGC group (including cyclic-nucleotide and calcium-phospholipid-dependent kinases, ribosomal S6-phosphorylating kinases, G protein-coupled kinases and all close relatives of these sets); the CAMKs (calmodulin-regulated kinases); the CK1 group (casein kinase 1, and close relatives); the CMGC group (including cyclin-dependent kinases, mitogen-activated protein kinases, glycogen synthase kinases and CDK-like kinases); the RGC group (receptor guanylate cyclase); the STEs (including many kinases functioning in MAP kinase cascades); the TKs (tyrosine kinases) and the TKLs (tyrosine kinase-like kinases). However, there is a significant proportion of kinases which, whilst exhibiting some degree of sequence similarity to the eight groups above, could not be classified easily into particular groups. These form a ninth group called 'Other'.

The aPKs are a small set of protein kinases that do not share clear sequence similarity with ePKs, but have been shown experimentally to have protein kinase activity. The *bona fide* aPKs (6) are the alpha-kinase group (exemplified by myosin heavy chain kinase of *Dictyostelium discoideum*), PIKK (phosphatidyl inositol 3′ kinase-related kinases), RIO and PHDK (pyruvate dehydrogenase kinases).

The sequencing of complete genomes for many eukaryotic species has allowed the determination and comparison of their complete kinase complements (kinomes). These include the kinomes of *Saccharomyces cerevisiae* (7), *Caenorhabditis elegans* (8), *Drosophila melanogaster* (9), *Mus musculus* (10), *Homo sapiens* (5), *Dictyostelium discoideum* (11), *Strongylocentrotus purpuratus* (12),

*To whom correspondence should be addressed. Tel: +44 1382 385860; Fax: +44 1382 385764; Email: geoff@compbio.dundee.ac.uk
Present address:
Diego Miranda-Saavedra, Cambridge Institute for Medical Research, Wellcome Trust/MRC Building, Addenbrooke's Hospital, Hills Road, Cambridge CB2 0XY, UK.

The authors wish it to be known that, in their opinion, the first two authors should be regarded as the joint First Authors.

Tetrahymena thermophila (13), and the plants *Arabidopsis thaliana* and *Oryza sativa* (14). Several parasite kinomes have been determined, including the malaria parasite *Plasmodium falciparum* (15), its comparison with *Plasmodium yoelii* (16) and those of the three Trypanosomatid species *Leishmania major*, *Trypanosoma brucei* and *Trypanosoma cruzi* (17). The kinomes of *H. sapiens*, *M. musculus*, *S. purpuratus*, *D. melanogaster*, *C. elegans*, *S. cerevisiae*, *D. discoideum* and *T. thermophila* are available through Kinbase (http://www.kinase.com/kinbase/). In particular, the observation that many important protein kinases of parasitic protozoa are significantly dissimilar from their eukaryotic counterparts has raised the prospects for therapeutics based on the selective inhibition of parasitic protein kinases (18–20).

We have recently exploited the sequence similarity of protein kinases in developing a multi-level Hidden Markov Model (HMM) library that is capable of classifying protein kinases into their correct functional group (6). The protein kinase HMM library was shown to be three times more sensitive than BLAST for identifying kinase catalytic domains. It was also shown to be more sensitive than a general Pfam model of the kinase catalytic domain, with the added advantage that the HMM library is capable of discriminating among protein kinase groups. The validated HMM library was applied to improve the group-level classification of the *S. cerevisiae* ePKs from 66.96% to 90.43% by classifying many of the yeast kinases previously assigned to the 'Other' group. In this article, we describe the extension of this analysis to the complete classification at the kinase group level of 43 curated eukaryotic kinomes and a web-based resource through which these annotations can be examined. In addition, we provide an interface to the HMM library, allowing for the classification of arbitrary sequences.

MATERIALS AND METHODS

Sequence data sources

The complete translated protein coding sequences were obtained for the fungi *Aspergillus fumigatus* (21), *Aspergillus nidulans* (22), *Aspergillus niger* (23), *Aspergillus oryzae* (24), *Candida glabrata* (25), *Cryptococcus neoformans* (26), *Debaryomyces hansenii* (25), *Kluyveromyces lactis* (25), *Magnaporthe grisea* (27), *Neurospora crassa* (28), *Phanerochaete chrysosporium* (29), *Ustilago maydis* (30) and *Yarrowia lipolytica* (25). Among the photosynthetic organisms we have included *A. thaliana* (31), the red alga *Cyanidioschyzon merolae* (32), the rice species *Oryza sativa ssp. Japonica* (33), the green algae *Ostreococcus lucimarinus* (34) and *Ostreococcus tauri* (35), and the poplar tree *Populus trichocarpa* (36). The metazoan genomes include the yellow fever mosquito *Aedes aegypti* (37), the malaria mosquito vector *Anopheles gambiae* (38), the silkworm *Bombyx mori* (39), the common dog *Canis familiaris* (40), the early chordate *Ciona intestinalis* (41), the chicken *Gallus gallus* (42), the Rhesus macaque *Macaca mulatta* (43), the marsupial *Monodelphis domestica* (Opossum) (44), the fishes medaka *Oryzias latipes* (45), *Takifugu rubripes* (46) and *Tetraodon nigroviridis* (47), the laboratory rat *Rattus norvegicus* (48) and the chimpanzee *Pan troglodytes* (49). Finally, we have also included the amoeba *Entamoeba histolytica* (50), the diatom *Thalassiosira pseudonana* (51) and the pathogenic protist *Trichomonas vaginalis* (52). The manually annotated kinomes of *Caenorhabditis elegans* (8), *Dictyostelium discoideum* (11), *Drosophila melanogaster*, *Homo sapiens* (5) and *M. musculus* (10) were downloaded from Kinbase (http://www.kinase.com/kinbase/) on 28 September 2008. The manually annotated kinomes of *Encephalitozoon cuniculi*, *Saccharomyces cerevisiae* and *Schyzosaccharomyces pombe* had previously been manually annotated and analysed in detail (53).

Kinase classification

The predicted peptide sequences for each of the genomes were searched individually against the Kinomer v. 1.0 multi-level HMM library (6) with the hmmpfam program of the HMMer package (54). Partial matches to the kinase catalytic domain were excluded through manual curation. Empirical cutoffs for association of kinase matches with each of the specific kinase groups were determined through analysis of the significance scores for the matches of the library HMMs to the well annotated kinases in Kinbase for the organisms *H. sapiens*, *C. elegans*, *D. melanogaster* and *S. cerevisiae* (6). The highest observed *E*-value for that group was taken as the cutoff for confident assignment. These are AGC ($2.7e^{-7}$), CAMK ($3.2e^{-14}$), CK1 ($3.2e^{-5}$), CMGC ($1.2e^{-7}$), RGC ($4.8e^{-5}$), STE ($1.4e^{-6}$), TK ($1.1e^{-9}$), TKL ($1.7e^{-12}$), Alpha ($8.5e^{-66}$), PDHK ($2.7e^{-10}$), PIKK ($8.4e^{-6}$) and RIO ($2.3e^{-3}$). Protein kinase catalytic domains that had *E*-values above this cutoff were automatically classified as belonging to the 'Other' group. Table 1 lists the protein kinase complements of the 43 eukaryotic genomes contained in Kinomer v.1.0, split by kinase group. All kinase matches were stored in a relational database, linking the sequence to the library matches and the subsequent assignments to a functional group.

User interface

The Kinomer v. 1.0 web server provides a comprehensive search interface for accessing the database. Sequences can be retrieved by kinase group, by species or by a combination of both. A summary table illustrates the quality of match of each sequence to the HMM library, as well as providing direct clickable links to the public databases (Figure 1). In addition, an option is available to allow data sets to be downloaded as FASTA format sequence files. The multiple sequence alignment analysis program Jalview (55) is integrated into the Kinomer v. 1.0 interface and allows visualization of the query results. Kinase sequences retrieved are grouped by type and aligned. Jalview allows colouring of the sequences by protein secondary structural properties or amino acid chemical character and on-the-fly calculation of Neighbour-Joining and average distance phylogenetic trees. The web-applet form of Jalview can launch the full Jalview application *via* the 'File->View in Full Application' option. This gives access to further tools for the generation of multiple sequence alignments by Muscle (56), MAFFT (57,58)

Table 1. The kinomes of the 43 genomes analysed split into the major kinase groups

Protein kinase group	Number of predicted peptides	AGC	CAMK	CK1	CMGC	RGC	STE	TK	TKL	Other	Total ePKs	Alpha	PDHK	PIKK	RIO	Total aPKs
Fungi																
Ascomycete fungi																
Aspergillus fumigatus	9630	20	27	3	30	0	13	1	0	8	**102**	0	3	4	1	**8**
Aspergillus nidulans	10 701	19	23	2	27	0	12	0	0	17	**100**	0	3	4	1	**8**
Aspergillus niger	11 200	21	21	3	44	0	12	1	0	16	**118**	0	3	4	1	**8**
Aspergillus oryzae	12 074	18	23	3	32	0	13	1	0	8	**98**	0	3	4	1	**8**
Candida glabrata	5215	25	30	4	23	0	11	0	0	11	**104**	0	2	5	1	**8**
Debaryomyces hansenii	6319	19	18	3	23	0	13	0	0	15	**91**	0	3	3	1	**7**
Encephalitozoon cuniculi	1997	4	5	2	12	0	0	0	1	5	**29**	0	0	2	1	**3**
Kluyveromyces lactis	5327	22	22	3	23	0	12	0	0	8	**90**	0	3	4	1	**8**
Magnaporthe grisea	11 109	21	15	2	42	0	0	1	0	19	**100**	0	3	3	0	**6**
Neurospora crassa	9822	19	20	2	21	0	14	1	0	18	**95**	0	3	4	1	**8**
Saccharomyces cerevisiae	6717	20	36	4	25	0	14	0	0	18	**117**	0	2	5	2	**9**
Schyzosaccharomyces pombe	5021	20	28	5	26	0	13	0	0	17	**109**	0	1	5	2	**8**
Yarrowia lipolytica	6436	19	19	2	21	0	11	0	0	4	**76**	0	3	4	1	**8**
Basidiomycete fungi																
Cryptococcus neoformans	6578	19	19	4	25	0	13	0	1	9	**90**	0	3	5	2	**10**
Phanerochaete chrysosporium	10 048	33	23	5	25	0	16	1	3	10	**116**	0	3	4	1	**8**
Ustilago maydis	6522	17	19	2	18	0	16	0	2	10	**84**	0	3	2	1	**6**
Plants																
Streptophytes																
Arabidopsis thaliana	30 690	76	116	20	119	0	73	3	625	86	**1118**	0	1	4	3	**8**
Oryza sativa ssp. Japonica	66 710	72	131	31	147	0	74	5	1179	139	**1778**	0	4	8	2	**14**
Populus trichocarpa	58 036	56	107	19	96	0	76	3	1033	136	**1526**	0	1	7	2	**10**
Green algae																
Ostreococcus lucimarinus	7651	16	24	4	21	0	9	2	11	13	**100**	0	1	5	1	**7**
Ostreococcus tauri	7892	15	19	4	23	0	9	2	13	13	**98**	0	1	4	1	**6**
Red algae																
Cyanidioschyzon merolae	5014	10	9	2	16	0	7	0	9	9	**62**	0	1	3	1	**5**
Diatoms																
Thalassiosira pseudonana	11 390	33	39	3	24	0	8	0	4	26	**137**	0	2	4	2	**8**
Amoebozoa																
Dictyostelium discoideum	13 463	43	27	5	38	0	43	3	69	27	**255**	6	0	5	2	**13**
Entamoeba histolytica	9772	37	49	9	47	0	29	7	109	34	**321**	0	0	6	3	**9**
Excavates/Trichomonads																
Trichomonas vaginalis	59 681	154	321	64	131	1	39	1	90	86	**887**	0	0	42	2	**44**
Metazoans																
Arthropods/Nematodes																
Aedes aegypti	16 789	48	35	10	43	7	26	35	18	8	**230**	0	4	6	3	**13**
Anopheles gambiae	13 133	37	34	7	31	6	25	32	17	7	**196**	0	1	5	3	**9**
Bombyx mori	21 302	24	20	3	19	6	18	25	9	7	**131**	0	1	5	3	**9**
Caenorhabditis elegans	27 258	38	49	84	50	27	31	82	17	38	**416**	1	2	5	4	**12**
Drosophila melanogaster	20 815	41	41	10	38	6	21	33	22	11	**223**	0	1	5	3	**9**
Chordata/Fishes																
Ciona intestinalis	19 858	71	72	13	51	3	43	83	23	25	**384**	2	1	12	4	**19**
Oryzias latipes	25 107	116	146	16	98	10	79	135	56	25	**681**	2	5	5	4	**16**
Takifugu rubripes	21 974	92	111	13	100	12	62	113	54	25	**582**	1	6	6	4	**17**
Tetraodon nigroviridis	28 005	94	102	12	73	14	55	108	53	33	**544**	1	5	5	3	**14**
Chordata/Birds																
Gallus gallus	22 195	81	89	14	63	3	72	117	59	16	**514**	6	3	9	4	**22**
Chordata/Mammals																
Canis familiaris	25 559	99	116	22	98	9	78	124	61	14	**621**	7	5	6	4	**22**
Homo sapiens	46 704	82	95	12	68	5	61	91	48	16	**478**	6	5	6	3	**20**
Macaca mulatta	36 423	133	153	23	127	6	102	134	71	27	**776**	12	7	11	2	**32**
Monodelphis domestica	32 612	126	149	27	113	13	118	213	67	27	**853**	9	8	10	4	**31**
Mus musculus	39 667	79	118	11	67	7	60	91	49	16	**498**	6	5	6	3	**20**
Pan troglodytes	32 834	116	136	19	118	5	97	149	75	17	**732**	10	6	12	3	**31**
Rattus norvegicus	33 438	127	0	29	96	7	0	148	67	10	**484**	7	6	6	3	**22**

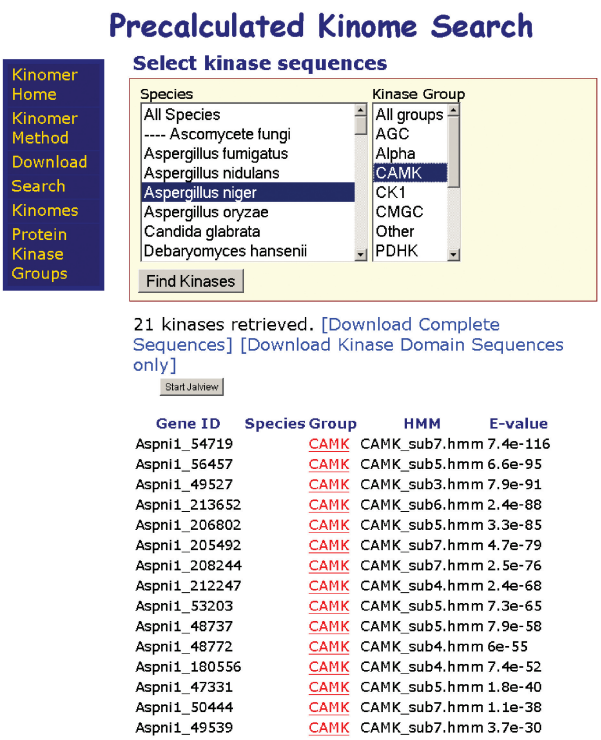

Figure 1. The precalculated kinomes may be downloaded from the Kinomer v. 1.0 website and select by species, kinase group or a combination of both.

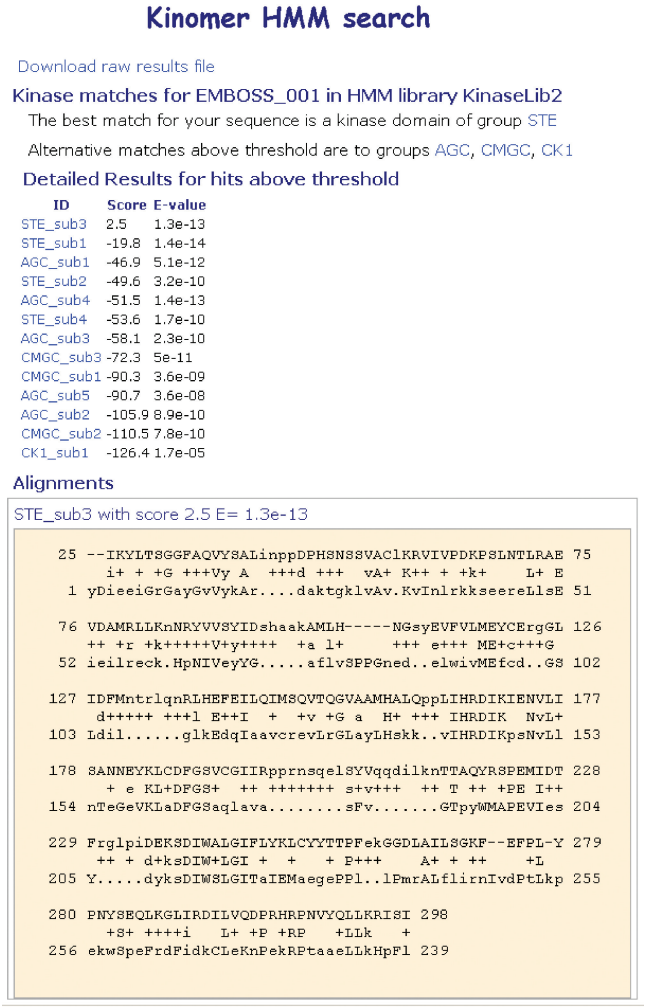

Figure 2. Results of searching a peptide sequence for kinase catalytic domains using the Kinomer v. 1.0 HMM library. A list of hits is displayed at the top followed by the alignment of the peptide sequence to the individual sub-group HMMs that constitute the HMM library.

or ClustalW (59) and secondary structure prediction by JNet (60,61).

In addition, a separate web interface allows users to classify arbitrary sequences with the HMM library. This web based tool allows a user to upload a sequence in any of the many sequence formats supported by EMBOSS (62), including the popular FASTA, GCG, PIR and SwissProt (62) formats. This sequence is subjected to basic quality assurance checks before the hmmpfam search job is queued for execution on a multi-node Linux cluster. The user is then provided with a job ID, and the interface is asynchronous, returning a status page to the user which is updated automatically. The user can bookmark the results page and return at a later time. In addition, an optional field allows the user to associate arbitrary comments with their job, a useful feature to allow otherwise similar jobs to be distinguished. There are no additional parameters that are user-selectable. This allows for a clean and straightforward interface form.

The results are displayed as a formatted HTML page (Figure 2) with the group classification clearly indicated. This shows to which protein kinase group Kinomer v. 1.0 has assigned the sequence. In addition, alternative assignments are given and a summary of all potential significant matches shown. Kinomer v. 1.0 will typically show matches to many kinase group HMMs spanning several kinase groups. All the top-scoring HMMs for one particular group will be the most significant matches, followed by closely related groups. The detailed alignment for each HMM match is linked further down the screen. As some users may wish for more details, the Kinomer v. 1.0 results page also provides a link to the raw HMMer output.

DISCUSSION

The 43 species considered here span a number of phylogenetic lineages, genome sizes and display a range of adaptations to their environment. The genome-wide kinase group assignments are consistent with our previously published results (6) in that seven protein kinase groups (AGC, CAMK, CK1, CMGC, STE, PIKK and RIO) are present in all species surveyed (Table 1) and some kinases in these groups are likely to be essential. Kinases of the groups RGC, TK, TKL, Alpha and PDHK are late innovations in specific phyla or have been lost secondarily in specific lines of descent. The presence of a discrete number of putative TKs in photosynthetic organisms and the pathogen *Entamoeba histolytica* suggests that TKs are also likely to have had an ancient origin. This observation

has recently been strengthened by the finding of animal-like signalling molecules in the green alga *Chlamydomonas reinhardtii* (63). These include scavenger receptor cysteine rich (SRCR) and C-type lectin domain (CTLD) proteins, both of which play key roles in the innate immune system of metazoa. The identification of SH2 domain proteins in photosynthetic organisms (63,64) suggests that phosphotyrosine-SH2 domain signalling also has an ancient origin and that important cell signalling and adhesion domains evolved before the divergence of the animal lineage.

The observation that many species outside the Opisthokont group lack important kinase groups, as is the case of TKs in Apicomplexa (Miranda-Saavedra, D. *et al.*, manuscript submitted for publication), and which have many lineage-specific groups of kinases, suggests that the group level is the most specific level for the automatic classification of kinomes based on models constructed from sequences outside the taxonomic clade under investigation. With the availability of a number of Deuterostome, Protostome and pre-bilaterian genome sequences, having all kinases belonging to a particular kinase group enables novel analyses to be performed. For example, it is now possible to trace the evolution of receptor tyrosine kinase families and that of their ligands. Since receptor tyrosine kinases are multi-domain proteins, diverging rates of evolution of the various domains, and their incorporation in the receptor molecule in select phylogenetic lineages, is informative of distinct selection pressures and can be informative of newly acquired functions through the acquisition of new ligand-binding domains. This is the case with the Trk family of receptor tyrosine kinases, which encode the neurotrophin receptors [nerve growth factor (NGF), brain-derived neurotrophic factor (BDNF), neurotrophin-3 (NT-3) and neurotrophin-4 (NT-4)]. The neurotrophin receptors are an ancient family whose function has been lost in multiple lineages and the roles of the receptors have been modified over time (65).

Kinomer v. 1.0 also includes the manually annotated kinomes of the model fungi *S. cerevisiae* and *S. pombe*, and that of the unicellular fungi-like parasite *Encephalitozoon cuniculi* (53). We have recently shown that the two model fungi share ~85% of their kinomes (53), a degree of similarity much higher than that previously reported. The kinomes of budding and fission yeasts are therefore a useful dataset for annotating the kinomes of other fungi, among which we have included species of importance in basic and medical research, and in biotechnology. The manually annotated kinomes of *C. elegans*, *D. discoideum*, *D. melanogaster*, *H. sapiens* and *M. musculus*, as provided in Kinbase (http://www.kinase.com/kinbase/), have also been included in the Kinomer v. 1.0 database. These will facilitate the manual annotation of other kinomes included in the database and which belong to the same taxonomic clade. The classification of a number of kinases in the kinomes of *C. elegans*, *D. discoideum*, *D. melanogaster*, *H. sapiens* and *M. musculus* could be improved as suggested by the Kinomer v. 1.0 HMM group scores. However, careful manual annotation of the kinomes of other species in the same taxonomic clades will be performed in the future to make a more informed decision about the re-classification of such kinases.

To our knowledge, Kinomer v. 1.0 is unique in being based on a high-accuracy validated kinase-group classification method (6). Other databases of protein kinases exist, but none of these offer the combination of breadth and accuracy of kinase classification that is present in Kinomer v. 1.0. These include KinMutBase (66), a database of clinically validated mutations in human kinases that lead to disease, and RTK.db (67), a database of receptor tyrosine kinases. The Protein Kinase Resource (68) collates data from several databases and includes a subset of protein kinase 3D structures to produce high-quality multiple structure-based alignments. Kinbase (http://www.kinase.com/kinbase/) contains manually curated kinomes classified according to the Hanks and Hunter classification of protein kinases (4). Although of high quality, Kinbase only contains kinomes for nine species. Finally, KinG (69) includes protein kinases identified in completed genomes that have been classified by a variety of metazoan kinome-based sequence search methods, but do not provide the confidence in kinase classification that is seen in Kinomer v. 1.0. Different eukaryotic lineages possess lineage-specific kinase groups and families that are just beginning to be characterized and which constitute as much as 50% of their kinomes (17). The applicability of the KinG approach to non-metazoan kinases needs further testing. A similar limitation is encountered by the PANTHER (70) database. Although not specific to protein kinases, PANTHER provides an extensive and detailed HMM library for kinase families and sub-families. These family and sub-family HMM libraries are trained on metazoan sequences and thus preclude their use to annotate non-metazoan sequences confidently into kinase families and sub-families which may not exist in non-metazoan species. Kinomer v. 1.0 annotates to the group level only and in our view annotating to the family/sub-family level requires manual curation.

In summary, Kinomer v. 1.0 is an easy-to-use interface to a novel database of both manually and automatically annotated kinomes. The availability of 43 eukaryotic kinomes in a relational database allows the easy querying of protein kinases by species and/or protein kinase group. In addition, the Kinomer v. 1.0 website includes a web server interface to the previously validated HMM library for the classification of peptide sequences into protein kinase groups. In the future, Kinomer v. 1.0 will be enhanced with the addition of a number of manually annotated kinomes of fungal, metazoan and photosynthetic organisms (Miranda-Saavedra, D., *et al.*, manuscript in preparation). These will include the kinomes of pathogenic fungi of the *Rhizopus* and *Fusarium* geni, and the kinomes of several unicellular and multicellular photosynthetic organisms including diatoms, red, brown and green algae, and vascular plants. Thus, Kinomer v. 1.0 is a useful and developing repository of expert and automatically annotated kinomes.

ACKNOWLEDGEMENTS

We thank Drs Tom Walsh and Jonathan Monk for assistance with computing.

FUNDING

D.M.S. was a Wellcome Trust Prize Student at the University of Dundee. Funding for open access charge: Wellcome Trust.

Conflict of interest statement. None declared.

REFERENCES

1. Cohen,P. (2000) The regulation of protein function by multisite phosphorylation—a 25 year update. *Trends Biochem Sci.*, **25**, 596–601.
2. Cohen,P. (2001) The role of protein phosphorylation in human health and disease. The Sir Hans Krebs Medal Lecture. *Eur. J. Biochem.*, **268**, 5001–5010.
3. Hanks,S.K., Quinn,A.M. and Hunter,T. (1988) The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science*, **241**, 42–52.
4. Hanks,S.K. and Hunter,T. (1995) Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J.*, **9**, 576–596.
5. Manning,G., Whyte,D.B., Martinez,R., Hunter,T. and Sudarsanam,S. (2002) The protein kinase complement of the human genome. *Science*, **298**, 1912–1934.
6. Miranda-Saavedra,D. and Barton,G.J. (2007) Classification and functional annotation of eukaryotic protein kinases. *Proteins*, **68**, 893–914.
7. Hunter,T. and Plowman,G.D. (1997) The protein kinases of budding yeast: six score and more. *Trends Biochem Sci.*, **22**, 18–22.
8. Plowman,G.D., Sudarsanam,S., Bingham,J., Whyte,D. and Hunter,T. (1999) The protein kinases of *Caenorhabditis elegans*: a model for signal transduction in multicellular organisms. *Proc. Natl Acad. Sci. USA*, **96**, 13603–13610.
9. Morrison,D.K., Murakami,M.S. and Cleghon,V. (2000) Protein kinases and phosphatases in the *Drosophila* genome. *J. Cell Biol.*, **150**, F57–F62.
10. Caenepeel,S., Charydczak,G., Sudarsanam,S., Hunter,T. and Manning,G. (2004) The mouse kinome: discovery and comparative genomics of all mouse protein kinases. *Proc. Natl Acad. Sci. USA*, **101**, 11707–11712.
11. Goldberg,J.M., Manning,G., Liu,A., Fey,P., Pilcher,K.E., Xu,Y. and Smith,J.L. (2006) The *dictyostelium* kinome—analysis of the protein kinases from a simple model organism. *PLoS Genet.*, **2**, e38.
12. Bradham,C.A., Foltz,K.R., Beane,W.S., Arnone,M.I., Rizzo,F., Coffman,J.A., Mushegian,A., Goel,M., Morales,J., Geneviere,A.M. *et al.* (2006) The sea urchin kinome: a first look. *Dev. Biol.*, **300**, 180–193.
13. Eisen,J.A., Coyne,R.S., Wu,M., Wu,D., Thiagarajan,M., Wortman,J.R., Badger,J.H., Ren,Q., Amedeo,P., Jones,K.M. *et al.* (2006) Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol.*, **4**, e286.
14. Krupa,A., Anamika and Srinivasan,N. (2006) Genome-wide comparative analyses of domain organisation of repertoires of protein kinases of *Arabidopsis thaliana* and *Oryza sativa*. *Gene*, **380**, 1–13.
15. Ward,P., Equinet,L., Packer,J. and Doerig,C. (2004) Protein kinases of the human malaria parasite *Plasmodium falciparum*: the kinome of a divergent eukaryote. *BMC Genomics*, **5**, 79.
16. Anamika,K. and Srinivasan,N. (2007) Comparative kinomics of *Plasmodium* organisms: unity in diversity. *Protein Pept. Lett.*, **14**, 509–517.
17. Parsons,M., Worthey,E.A., Ward,P.N. and Mottram,J.C. (2005) Comparative analysis of the kinomes of three pathogenic trypanosomatids: *Leishmania major*, *Trypanosoma brucei* and *Trypanosoma cruzi*. *BMC Genomics*, **6**, 127.
18. Doerig,C., Billker,O., Pratt,D. and Endicott,J. (2005) Protein kinases as targets for antimalarial intervention: kinomics, structure-based design, transmission-blockade, and targeting host cell enzymes. *Biochim. Biophys. Acta*, **1754**, 132–150.
19. Doerig,C. and Meijer,L. (2007) Antimalarial drug discovery: targeting protein kinases. *Expert Opin. Ther. Targets*, **11**, 279–290.
20. Naula,C., Parsons,M. and Mottram,J.C. (2005) Protein kinases as drug targets in trypanosomes and *Leishmania*. *Biochim. Biophys. Acta*, **1754**, 151–159.
21. Nierman,W.C., Pain,A., Anderson,M.J., Wortman,J.R., Kim,H.S., Arroyo,J., Berriman,M., Abe,K., Archer,D.B., Bermejo,C. *et al.* (2005) Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature*, **438**, 1151–1156.
22. Galagan,J.E., Calvo,S.E., Cuomo,C., Ma,L.J., Wortman,J.R., Batzoglou,S., Lee,S.I., Basturkmen,M., Spevak,C.C., Clutterbuck,J. *et al.* (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature*, **438**, 1105–1115.
23. Pel,H.J., de Winde,J.H., Archer,D.B., Dyer,P.S., Hofmann,G., Schaap,P.J., Turner,G., de Vries,R.P., Albang,R., Albermann,K. *et al.* (2007) Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. *Nat. Biotechnol.*, **25**, 221–231.
24. Machida,M., Asai,K., Sano,M., Tanaka,T., Kumagai,T., Terai,G., Kusumoto,K., Arima,T., Akita,O., Kashiwagi,Y. *et al.* (2005) Genome sequencing and analysis of *Aspergillus oryzae*. *Nature*, **438**, 1157–1161.
25. Dujon,B., Sherman,D., Fischer,G., Durrens,P., Casaregola,S., Lafontaine,I., De Montigny,J., Marck,C., Neuveglise,C., Talla,E. *et al.* (2004) Genome evolution in yeasts. *Nature*, **430**, 35–44.
26. Loftus,B.J., Fung,E., Roncaglia,P., Rowley,D., Amedeo,P., Bruno,D., Vamathevan,J., Miranda,M., Anderson,I.J., Fraser,J.A. *et al.* (2005) The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science*, **307**, 1321–1324.
27. Dean,R.A., Talbot,N.J., Ebbole,D.J., Farman,M.L., Mitchell,T.K., Orbach,M.J., Thon,M., Kulkarni,R., Xu,J.R., Pan,H. *et al.* (2005) The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature*, **434**, 980–986.
28. Galagan,J.E., Calvo,S.E., Borkovich,K.A., Selker,E.U., Read,N.D., Jaffe,D., FitzHugh,W., Ma,L.J., Smirnov,S., Purcell,S. *et al.* (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature*, **422**, 859–868.
29. Martinez,D., Larrondo,L.F., Putnam,N., Gelpke,M.D., Huang,K., Chapman,J., Helfenbein,K.G., Ramaiya,P., Detter,J.C., Larimer,F. *et al.* (2004) Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nat. Biotechnol.*, **22**, 695–700.
30. Kamper,J., Kahmann,R., Bolker,M., Ma,L.J., Brefort,T., Saville,B.J., Banuett,F., Kronstad,J.W., Gold,S.E., Muller,O. *et al.* (2006) Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature*, **444**, 97–101.
31. Arabidopsis.Genome.Initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.
32. Matsuzaki,M., Misumi,O., Shin,I.T., Maruyama,S., Takahara,M., Miyagishima,S.Y., Mori,T., Nishida,K., Yagisawa,F., Yoshida,Y. *et al.* (2004) Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature*, **428**, 653–657.
33. Goff,S.A., Ricke,D., Lan,T.H., Presting,G., Wang,R., Dunn,M., Glazebrook,J., Sessions,A., Oeller,P., Varma,H. *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa L. ssp. japonica*). *Science*, **296**, 92–100.
34. Palenik,B., Grimwood,J., Aerts,A., Rouze,P., Salamov,A., Putnam,N., Dupont,C., Jorgensen,R., Derelle,E., Rombauts,S. *et al.* (2007) The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc. Natl Acad. Sci. USA*, **104**, 7705–7710.
35. Derelle,E., Ferraz,C., Rombauts,S., Rouze,P., Worden,A.Z., Robbens,S., Partensky,F., Degroeve,S., Echeynie,S., Cooke,R. *et al.* (2006) Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proc. Natl. Acad. Sci. USA*, **103**, 11647–11652.
36. Tuskan,G.A., Difazio,S., Jansson,S., Bohlmann,J., Grigoriev,I., Hellsten,U., Putnam,N., Ralph,S., Rombauts,S., Salamov,A. *et al.* (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, **313**, 1596–1604.
37. Nene,V., Wortman,J.R., Lawson,D., Haas,B., Kodira,C., Tu,Z.J., Loftus,B., Xi,Z., Megy,K., Grabherr,M. *et al.* (2007) Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science*, **316**, 1718–1723.

38. Holt,R.A., Subramanian,G.M., Halpern,A., Sutton,G.G., Charlab,R., Nusskern,D.R., Wincker,P., Clark,A.G., Ribeiro,J.M., Wides,R. *et al.* (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science*, **298**, 129–149.
39. Xia,Q., Zhou,Z., Lu,C., Cheng,D., Dai,F., Li,B., Zhao,P., Zha,X., Cheng,T., Chai,C. *et al.* (2004) A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science*, **306**, 1937–1940.
40. Lindblad-Toh,K., Wade,C.M., Mikkelsen,T.S., Karlsson,E.K., Jaffe,D.B., Kamal,M., Clamp,M., Chang,J.L., Kulbokas,E.J. III, Zody,M.C. *et al.* (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, **438**, 803–819.
41. Dehal,P., Satou,Y., Campbell,R.K., Chapman,J., Degnan,B., De Tomaso,A., Davidson,B., Di Gregorio,A., Gelpke,M., Goodstein,D.M. *et al.* (2002) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science*, **298**, 2157–2167.
42. International.Chicken.Genome.Sequencing.Consortium. (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, **432**, 695–716.
43. Gibbs,R.A., Rogers,J., Katze,M.G., Bumgarner,R., Weinstock,G.M., Mardis,E.R., Remington,K.A., Strausberg,R.L., Venter,J.C., Wilson,R.K. *et al.* (2007) Evolutionary and biomedical insights from the rhesus macaque genome. *Science*, **316**, 222–234.
44. Mikkelsen,T.S., Wakefield,M.J., Aken,B., Amemiya,C.T., Chang,J.L., Duke,S., Garber,M., Gentles,A.J., Goodstadt,L., Heger,A. *et al.* (2007) Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature*, **447**, 167–177.
45. Kasahara,M., Naruse,K., Sasaki,S., Nakatani,Y., Qu,W., Ahsan,B., Yamada,T., Nagayasu,Y., Doi,K., Kasai,Y. *et al.* (2007) The medaka draft genome and insights into vertebrate genome evolution. *Nature*, **447**, 714–719.
46. Aparicio,S., Chapman,J., Stupka,E., Putnam,N., Chia,J.M., Dehal,P., Christoffels,A., Rash,S., Hoon,S., Smit,A. *et al.* (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science*, **297**, 1301–1310.
47. Jaillon,O., Aury,J.M., Brunet,F., Petit,J.L., Stange-Thomann,N., Mauceli,E., Bouneau,L., Fischer,C., Ozouf-Costaz,C., Bernot,A. *et al.* (2004) Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature*, **431**, 946–957.
48. Gibbs,R.A., Weinstock,G.M., Metzker,M.L., Muzny,D.M., Sodergren,E.J., Scherer,S., Scott,G., Steffen,D., Worley,K.C., Burch,P.E. *et al.* (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, **428**, 493–521.
49. Chimpanzee.Sequencing.and.Analysis.Consortium. (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, **437**, 69–87.
50. Loftus,B., Anderson,I., Davies,R., Alsmark,U.C., Samuelson,J., Amedeo,P., Roncaglia,P., Berriman,M., Hirt,R.P., Mann,B.J. *et al.* (2005) The genome of the protist parasite *Entamoeba histolytica*. *Nature*, **433**, 865–868.
51. Armbrust,E.V., Berges,J.A., Bowler,C., Green,B.R., Martinez,D., Putnam,N.H., Zhou,S., Allen,A.E., Apt,K.E., Bechner,M. *et al.* (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science*, **306**, 79–86.
52. Carlton,J.M., Hirt,R.P., Silva,J.C., Delcher,A.L., Schatz,M., Zhao,Q., Wortman,J.R., Bidwell,S.L., Alsmark,U.C., Besteiro,S. *et al.* (2007) Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science*, **315**, 207–212.
53. Miranda-Saavedra,D., Stark,M.J., Packer,J.C., Vivares,C.P., Doerig,C. and Barton,G.J. (2007) The complement of protein kinases of the microsporidium *Encephalitozoon cuniculi* in relation to those of *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *BMC Genomics*, **8**, 309.
54. Eddy,S.R. (1998) Profile hidden Markov models. *Bioinformatics*, **14**, 755–763.
55. Clamp,M., Cuff,J., Searle,S.M. and Barton,G.J. (2004) The Jalview Java alignment editor. *Bioinformatics*, **20**, 426–427.
56. Edgar,R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5**, 113.
57. Katoh,K., Kuma,K., Toh,H. and Miyata,T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.*, **33**, 511–518.
58. Katoh,K., Kuma,K., Miyata,T. and Toh,H. (2005) Improvement in the accuracy of multiple sequence alignment program MAFFT. *Genome Inform.*, **16**, 22–33.
59. 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.
60. Cole,C., Barber,J.D. and Barton,G.J. (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.*, **36**, W197–W201.
61. Cuff,J.A. and Barton,G.J. (2000) Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins*, **40**, 502–511.
62. Rice,P., Longden,I. and Bleasby,A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.
63. Wheeler,G.L., Miranda-Saavedra,D. and Barton,G.J. (2008) Genome analysis of the unicellular green alga *Chlamydomonas reinhardtii* Indicates an ancient evolutionary origin for key pattern recognition and cell-signaling protein families. *Genetics*, **179**, 193–197.
64. Williams,J.G. and Zvelebil,M. (2004) SH2 domains in plants imply new signalling scenarios. *Trends Plant Sci.*, **9**, 161–163.
65. Lanave,C., Colangelo,A.M., Saccone,C. and Alberghina,L. (2007) Molecular evolution of the neurotrophin family members and their Trk receptors. *Gene*, **394**, 1–12.
66. Ortutay,C., Valiaho,J., Stenberg,K. and Vihinen,M. (2005) KinMutBase: a registry of disease-causing mutations in protein kinase domains. *Hum. Mutat.*, **25**, 435–442.
67. Grassot,J., Mouchiroud,G. and Perriere,G. (2003) RTKdb: database of Receptor Tyrosine Kinase. *Nucleic Acids Res.*, **31**, 353–358.
68. Niedner,R.H., Buzko,O.V., Haste,N.M., Taylor,A., Gribskov,M. and Taylor,S.S. (2006) Protein kinase resource: an integrated environment for phosphorylation research. *Proteins*, **63**, 78–86.
69. Krupa,A., Abhinandan,K.R. and Srinivasan,N. (2004) KinG: a database of protein kinases in genomes. *Nucleic Acids Res.*, **32**, D153–D155.
70. Mi,H., Lazareva-Ulitsky,B., Loo,R., Kejariwal,A., Vandergriff,J., Rabkin,S., Guo,N., Muruganujan,A., Doremieux,O., Campbell,M.J. *et al.* (2005) The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.*, **33**, D284–D288.