Kalium 2.0, a comprehensive database of polypeptide ligands of potassium channels

Valentin M. Tabakmakher1,2, Nikolay A. Krylov1,3, Alexey I. Kuzmenkov1, Roman G. Efremov1,3,4 & Alexander A. Vassilevski1,4

Potassium channels are the most diverse group of ion channels in humans. They take vital parts in numerous physiological processes and their malfunction gives rise to a range of pathologies. In addition to small molecules, there is a wide selection of several hundred polypeptide ligands binding to potassium channels, the majority of which have been isolated from animal venoms. Until recently, only scorpion toxins received focused attention being systematically assembled in the manually curated Kalium database, but there is a diversity of well-characterized potassium channel ligands originating from other sources. To address this issue, here we present the updated and improved Kalium 2.0 that covers virtually all known polypeptide ligands of potassium channels and reviews all available pharmacological data. In addition to an expansion, we have introduced several new features to the database including posttranslational modification annotation, indication of ligand mode of action, BLAST search, and possibility of data export.

Background & Summary

Potassium (K⁺) channels are a superfamily of integral membrane proteins responsible for selective potassium ion permeation through cell membranes. Activity of K⁺ channels regulates cell excitability and controls the shape of the action potential. Being present in various cells they participate in processes as diverse as cognition, muscle contraction, and hormone secretion. K⁺ channels are composed of two or four major α subunits that form the pore and auxiliary β subunits. K⁺ channels of mammals are classified into four groups according to gene homology and structure of the α subunits: calcium- and sodium-activated (KCa and KNa), inwardly rectifying (Kir), two pore domain (K2P), and voltage-gated (KV) potassium channels.

A large number of various molecules can interact with K⁺ channels. Three major classes are often cited: metal ions, low-molecular-mass substances, and polypeptides. Despite structural differences most K⁺ channel ligands may either physically occlude the channel pore, or change channel properties through gating modification. Polypeptide ligands are of special interest to researchers due to high affinity (often active at nanomolar or even subnanomolar concentrations) and selectivity towards their targets. Most of these molecules are toxins from venomous animals but some are found in different sources. Polypeptide ligands play a key role in unravelling the functions of K⁺ channels and serve a pool of natural prototypes for drug discovery.

>95% of K⁺ channel polypeptide ligands have been identified in just five groups of organisms and scorpion toxins (KTx) provide >50% of this variability. They consist of ~20–75 amino acid residues and usually contain 2–4 disulfide bridges. Five structural folds are described for KTx: cysteine-stabilized α-helix/β-sheet (CSα/β), cysteine-stabilized helix-loop-helix (CSα/α) with two or three disulfide bonds, Kunitz-type, and inhibitor cystine knot (ICK) folds. KTx generally inhibit Kv and KCa channels through pore blockage. The most famous ligands of K⁺ channels from snakes are dendrotoxins that contain ~55–60 amino acid residues and form a Kunitz-type fold. Another important group is myotoxin-like polypeptides composed of ~40–50 amino acids, which assume a similar fold to human β-defensins and display versatile activities including Kv channel inhibition.

Spider toxins containing ~30–40 amino acid residues and forming the ICK motif inhibit mostly K⁺ channel
activation via interactions with the voltage sensor. The founding member of this group is hanatoxin and their peculiar ability is to partition into membranes and interact with the channels by lateral association within the membranes. Some weak pore blockers of Kv channels assuming the Kunitz-type fold have also been found in spider venom. The channel ligands from sea anemones are composed of ~35–65 amino acid residues and can be subdivided into three subgroups by structural features. Their spatial structures are presented by a combination of α and/or 310-helices, several β-strands, or the Kunitz-type fold. Sea anemone toxins often bear posttranslational modifications and inhibit Kv and KCa channels. Cone snails use a number of different structural classes of toxin to target Kv channels: κA-, κO-, κM-, κI-, κJ-, and κL-conotoxins. These polypeptides comprise ~20–30 amino acid residues and present diverse disulfide patterns and folds. Two toxins have a particularly unusual structure: conkunitzin-S1, a 60 residues-long polypeptide with the Kunitz-type fold, and contryphan-Vn of just nine amino acids. Conotoxins are also often subjected to posttranslational modifications and inhibit KV and KCa channels. Cone snails use a number of different structural classes of toxin to target Kv channels: κA-, κO-, κM-, κI-, κJ-, and κL-conotoxins. These polypeptides comprise ~20–30 amino acid residues and present diverse disulfide patterns and folds. Two toxins have a particularly unusual structure: conkunitzin-S1, a 60 residues-long polypeptide with the Kunitz-type fold, and contryphan-Vn of just nine amino acids. Conotoxins are also often subjected to posttranslational modifications. In addition, a comparatively small number of molecules affecting K+ channels has been found in some species of bees, worms, lizards, fungi, and scolopendra; moreover, human ß-defensin 4A displays activity against several Kv isoforms.

The first version of Kalium comprised only scorpion toxins, while its current expansion and update includes all known polypeptide ligands identified in living organisms. For all these compounds detailed activity data are provided collected from original manuscripts. Several major improvements have been introduced, such as the indication of toxin mode of action, BLAST search, and possibility to export data in .csv (comma-separated) or .txt (tab-delimited) format. Kalium is manually curated, and presents a comprehensive list of all known polypeptide K+ channel ligands available to users. Kalium is of primary utility to researchers investigating the structure and function of K+ channels, toxicologists addressing the variability and mode of action of natural toxins, pharmacologists and research and development managers involved in drug discovery targeting K+ channels, and biochemical community in general.

### Methods

**Data sources and curation.** Data for Kalium 2.0 were assembled from scorpion venom peptide entries already present in the first release of Kalium, which was updated and expanded by adding the available information on K+ channel ligands from other organisms. As a result, Kalium 2.0 contains twice as many entries as Kalium 1.0. The compiled data on all publically available sequences of polypeptide ligands of K+ channels were obtained from UniProt. Available PDB structures with links to the RCSB Protein Data Bank and location of disulfide bonds were also extracted from UniProt. The data set was then manually filtered and refined, including the following steps: removal of peptides with partial sequence, removal of entries supported by genomic or transcriptomic information only, and sorting by the source organism into six groups: snakes, scorpions, spiders, sea anemones, cone snails, and miscellaneous. Kalium 1.0 and 2.0 entries statistics is summarized in Table 1.

| Source organisms | Entries added | Kalium 1.0 | Kalium 2.0 | Current number of entries |
|------------------|---------------|------------|------------|--------------------------|
| Scorpions        | 174           | 19         | 193        |                          |
| Snakes           | —             | 29         | 29         |                          |
| Spiders          | —             | 50         | 50         |                          |
| Sea anemones     | —             | 35         | 35         |                          |
| Cone snails      | —             | 19         | 19         |                          |
| Miscellaneous    |               |            |            |                          |
| Nematodes        | —             | 2          | 2          |                          |
| Hymenopterans    | —             | 4          | 4          |                          |
| Lizards          | —             | 1          | 1          |                          |
| Human            | —             | 1          | 1          |                          |
| Fungi            | —             | 1          | 1          |                          |
| Centipedes       | —             | 5          | 5          |                          |
| Total            | 174           | 164        | 340        |                          |

Table 1. Kalium entries statistics.
Table 2. Comparison of ligand molecular masses measured experimentally and calculated in Kalium.

| UniProt ID | Experimental mass, Da | Kalium calculated mass, Da | Known modifications |
|------------|-----------------------|----------------------------|---------------------|
| Q9U3Z3     | 3569                  | 3571.74                    | Signal and propeptide cleavage, 4 disulfide bridges, 4 γ-carboxyglutamates, 1 γ-hydroxyproline, C-terminal amidation |
| P0CG45     | 2805.84               | 2807.25                    | 3 disulfide bridges, 7 γ-hydroxyprolines |
| Q86QT3     | 4730.8                | 4730.49                    | Signal peptide cleavage, 4 disulfide bridges |
| P84704     | 2872.5                | 2873.32                    | 2 disulfide bridges |
| P0C166     | 4082.8                | 4081.99                    | N-Terminal cyclization of glutamine, C-terminal amidation, 4 disulfide bridges |

Fig. 1 Data sources and curation. Schematic representation of the data stream and curation process in Kalium 2.0.

Implementation. Interface to the Kalium database is centered around the main table with data on K⁺ channel ligands, initially sorted according to source organism group, organism name and polypeptide family name or common name. The table supports searching, multi-column ordering and filtering, and multi-row selection. BLAST search and sequence alignment using the Clustal Omega program via UniProt web server is implemented, as well as data export for toxins selected by users; all these options are new in Kalium 2.0. Extended information including detailed activity data (the "Ligand card") is available for each entry as a special popup window.

Kalium is an OpenUI 5 Model-View-Controller web application built upon a Django web framework and SQLite3 database engine. The web interface consists of single dynamically generated HTML5 page with JSON data being fetched from the server asynchronously via AJAX requests. Standard Django web admin interface is used for data access and curation. Modern HTML5-capable browsers (desktop and mobile variants) are supported.
Data Records

Original Kalium 1.0 was assembled as a database of K$^+$ channel toxins from scorpion venom$^3$. Due to database expansion following the addition of K$^+$ channel ligands from other organism sources, the structure of Kalium 2.0 was improved. A copy of Kalium database in CSV format can be accessed at Figshare.$^4$

The main window.

The main window of Kalium is presented by one large general table, in which all data about K$^+$ channel ligands from various sources are assembled (Fig. 2 and Table 3). “Home”, “About”, “Help”, “FAQ”, and “Contacts” located in the top right corner link to pages that contain information about developers and tips. Below those links come buttons “Clustal”, “BLAST”, and “Export as” (a drop-down list of export file format), and a search field. Buttons for source organism selection are located under the Kalium logo in the top left corner. Other control elements of the table are placed in the headers and function to filter information of interest as discussed below. Multiparameter filtering is now an available option in Kalium 2.0.

Ligand card.

For each polypeptide entry, detailed information is summarized in the “Ligand card” (Fig. 3) available by clicking on polypeptide name in the field “Name” of the general table. As it was implemented in the first Kalium release, all information presented in the general table is duplicated in the Ligand card in an expanded way.$^3$ All records of the renewed Ligand card are explained in Tables 3 and 4.

Export file format.

Downloadable text file containing data on Kalium entries is generated in the column-separated (default name is “export.csv”) or tabulation-separated (“export.txt”) format. For multiple selected entries, the file consists of truncated Ligand cards appended one by one. Each truncated Ligand card includes UniProt ID, sequence, list of PDB IDs (if available), molecular mass, and mode of action followed by a table of experimentally determined activity data (if available).

Technical Validation

Database generation process consisted of fetching, filtering and merging manually collected data from the literature and information from the UniProt.$^36$ UniProt data validation was not performed, since it is one of the most accurately curated biological resources. The records included in Kalium 2.0 are based on published material in peer-reviewed scientific journals; each specific data value is supported by the original references, so users can evaluate the validity and accuracy of the original source. The overall correctness of the database generation process was verified manually. Mass calculation for mature toxins containing 20 common amino acids and modified residues, was checked against the ExPASy server.$^{38,39}$
Usage Notes

Kalium 2.0 is freely available for users. Most of the original Kalium 1.0 features were upgraded and new features were implemented, we therefore describe all of them in detail below. Moreover, here we give an example of how Kalium 2.0 can be utilized by researchers with specific needs.

Organism selection buttons. A major new feature of Kalium 2.0 is buttons for organism group selection (Fig. 2). Clicking one or several buttons allows filtering data in the main table according to the source organism groups: snakes, scorpions, spiders, sea anemones, cone snails, and miscellaneous. The “Miscellaneous” group includes K⁺ channel ligands from fungi, worms, bees, wasps, centipedes, lizards, and humans.

Selecting and manipulating data: Clustal, BLAST, and Export. Check boxes on the left side of the general table permit selection of one or more entries; for all entries selection, users may click once on the column header. Multiple (two or more) entries selection allows performing Clustal alignment request. New features of Kalium 2.0 include an easy BLAST search for multiple sequences and data export for selected polypeptides in a text file.

To submit an alignment request, after entry selection, users need to click the “Clustal” button; the results of Clustal Omega pair/multiple sequence alignment will appear in a new browser tab. Similarly, to submit a BLAST search request, users are required to click the “BLAST” button; the results will appear in separate browser tab for

| Table field       | Definition                                                                                           |
|-------------------|------------------------------------------------------------------------------------------------------|
| Organism          | The Latin name of the source organism.                                                               |
| Name              | The nomenclature name or conventionally used name of polypeptide.                                     |
| Synonyms          | Trivial name(s) of polypeptide.                                                                      |
| UniProt ID        | Unique UniProt ID of polypeptide.                                                                     |
| Sequence          | Amino acid sequences of mature polypeptides presented in the one-letter code. “–NH₂” indicates amidation of the C-terminal amino acid; “Z” is for the N-terminal pyroglutamic acid; “O” for 4-hydroxyproline; “E” for 4-carboxyglutamic acid; “K” for N6-formyllysine; “T, S” are for O-glycosylated threonine and serine; and “W” is for D-tryptophan. Cysteine residues are marked; different colors indicate the disulfide bond connectivity. |
| PDB               | Available PDB ID(s) of polypeptide.                                                                   |
| Mass              | Molecular mass of mature polypeptide calculated taking into account the post-translational modifications. Molecular masses for O-glycosylated polypeptides are marked with the “+” symbol. |
| Publication date  | The date when the polypeptide sequence was first published.                                          |
| Activity          | The list of all targets on which the polypeptide was ever tested.                                    |

Table 3. Description of Kalium 2.0 main window general table.
The type of data reported: dissociation constant (Kd) for ion channel activity.
The cell type used for channel expression: vertebrate nerve, muscle, and heterologous systems.
The origin organism of the ion channel that was used for measurements: the most common channels belong to subfamilies according to current nomenclature. As of February 2019, the filtering option is active for families of scorpion toxins only, since the nomenclature of just these molecules is the most conventional, clear and universally recognized (an updated Tytgat-Possani nomenclature [17,41]). “Name” enables selecting toxin family from a list and click the “Export as” button; the resulting file containing data from the selected entries will be generated and sent to the user’s browser.

Organism. The “Organism” header is the control element for filtering and sorting entries by species names listed according to current biological classification. One click on the column header opens a drop-down menu, where users can choose one or more species to filter the full data set. The Latin names in the table body are linked to the UniProt Taxonomy database ensuring valid classification.

Name. The “Name” header is the control element for filtering and sorting entries by polypeptide families and subfamilies according to current nomenclature. As of February 2019, the filtering option is active for families of scorpion toxins only, since the nomenclature of just these molecules is the most conventional, clear and universally recognized (an updated Tytgat-Possani nomenclature [17,41]). “Name” enables selecting toxin family from a drop-down menu. Ligand card opens when clicked on toxin name in the table body.

Synonyms. The “Synonyms” header is the control element for searching/filtering trivial names of polypeptides. Many scientists identify certain molecules using trivial names only; therefore their inclusion in Kalium 2.0 is a necessity.

UniProt ID. Click on UniProt ID switches to corresponding UniProt pages.

PDB. The “PDB” header is the control element for filtering entries by PDB ID (if available). Clicking this filter button will show entries with resolved spatial structure only. All PDB IDs are linked to corresponding Protein Data Bank [27] pages.

Mass. The “Mass” header is the control element for sorting entries according to molecular mass. One click on this button will sort entries by ascending order of masses, next click — by descending order.

Publication date. The “Publication date” header is the control element for sorting entries according to the date when the sequence was first published.

Activity. The “Activity” header is the control element for filtering and sorting entries by information about activities on different K+ channels. One click on the column header opens a drop-down menu, where users can select one or more channels. The header is used to sort entries according to specific targets. Ligand card can be opened for detailed information by clicking on a channel name.

Ligand card. For user convenience the information of the records “Organism”, “UniProt ID”, “PDB”, and “Ref. (PubMed/DOI)” is linked to corresponding web pages.

### Table 4: Description of Ligand card records. Those records that duplicate information of the main window general table are described in Table 3.

![Table](image)

For user convenience the information of the records “Organism”, “UniProt ID”, “PDB”, and “Ref. (PubMed/DOI)” is linked to corresponding web pages.
Kalium 2.0 application example. Kalium provides convenient tools to analyze the selectivity features of K^+ channel ligands. For instance, Kalium may help infer the molecular determinants underlying ligand specificity against particular channel isoforms. Investigators can identify all known polypeptides that were tested against chosen K^+ channel isoforms by selecting the appropriate channels in the “Activity” header. The most suitable entries may be selected and analyzed by Clustal or BLAST. As a result, assumptions may be made on potentially important residues and this information may be further used to produce artificial molecules with enhanced selectivity or affinity. To perform such analysis without using Kalium is difficult, because it is associated with deep literature search. This search has already been performed during data assembly and is central to manual data curation at Kalium.

Code Availability
Code is available upon request.

References
1. Hille, B. Ion Channels of Excitable Membranes. (Sinauer Associates Inc Sunderland, 2001).
2. González, C. et al. K(+) channels: function-structural overview. Compr. Physiol. 2, 2087–2149 (2012).
3. Doyle, D. A. et al. The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. Science 280, 69–77 (1998).
4. Heinemann, S. H., Retting, J., Graack, H. R. & Pongs, O. Functional characterization of Kv channel beta-subunits from rat brain. J. Physiol. 493 (Pt 3), 625–633 (1996).
5. Kaczmarek, L. K. et al. International Union of Basic and Clinical Pharmacology. C. Nomenclature and properties of calcium-activated and sodium-activated potassium channels. Pharmacol. Rev. 69, 1–11 (2017).
6. Gutman, G. A. et al. International Union of Pharmacology. LII. Nomenclature and molecular relationships of voltage-gated potassium channels. Pharmacol. Rev. 57, 473–508 (2005).
7. Kubo, Y. et al. International Union of Pharmacology. LIX. Nomenclature and molecular relationships of inwardly rectifying potassium channels. Pharmacol. Rev. 57, 509–526 (2005).
8. Goldstein, S. A. N. et al. International Union of Pharmacology. LV. Nomenclature and molecular relationships of two-P potassium channels. Pharmacol. Rev. 57, 527–540 (2005).
9. Alexander, S. P. et al. The Concise Guide To Pharmacology 2017/18: Voltage-gated ion channels. Br. J. Pharmacol. 174, S160–S194 (2017).
10. Kuzmenkov, A. I., Grishin, E. V. & Vassilevski, A. A. Diversity of potassium channel ligands: focus on scorpion toxins. Biochem. (Mosc.) 80, 1764–1799 (2015).
11. Catterall, W. A. et al. Voltage-gated ion channels and gating modifier toxins. Toxicon 49, 124–141 (2007).
12. Norton, R. S. & Chandy, K. G. Venom-derived peptide inhibitors of voltage-gated potassium channels. Neuropharmacology 127, 124–138 (2017).
13. Chhabra, S. et al. Kv1.3 channel-blocking immunomodulatory peptides from parasitic worms: implications for autoimmune diseases. PASEB J. 28, 3952–3964 (2014).
14. Noble, M. et al. The toxin helothermine affects potassium currents in newborn rat cerebellar granule cells. J. Membr. Biol. 139, 49–55 (1994).
15. Yang, W. et al. Endogenous animal toxin-like human β-defensin 2 inhibits own potassium channels through interaction with channel extracellular pore region. Cell. Mol. Life Sci. 72, 845–853 (2015).
16. Norton, R. S. Enhancing the therapeutic potential of peptide toxins. Expert Opin. Drug Discov. 12, 611–623 (2017).
17. Trygat, J. et al. A unified nomenclature for short-chain peptides isolated from scorpion venoms: alpha-KTx molecular subfamilies. Trends Pharmacol. Sci. 20, 444–447 (1999).
18. Mouhat, S., Andreotti, N., Jouirou, B. & Sabatier, J.-M. Animal toxins acting on voltage-gated potassium channels. Curr. Pharm. Des 14, 2503–2518 (2008).
19. Harvey, A. L. & Anderson, A. J. Dendrotoxins: snake toxins that block potassium channels and facilitate neurotransmitter release. Pharmacol. Ther. 31, 33–55 (1985).
20. Berndt, K. D., Güntert, P. & Wuthrich, K. Nuclear magnetic resonance solution structure of dendrotoxin K from the venom of Dendroaspis polylepis polylepis. J. Mol. Biol. 234, 735–750 (1993).
21. Peigneur, S. et al. Crotamine pharmacology revisited: novel insights based on the inhibition of KV channels. Mol. Pharmacol. 82, 90–96 (2012).
22. Wang, J. M. et al. Molecular surface of tarantula toxins interacting with voltage sensors in K(v) channels. J. Gen. Physiol. 123, 455–467 (2004).
23. Swartz, K. J. & MacKinnon, R. An inhibitor of the Kv2.1 potassium channel isolated from the venom of a Chilean tarantula. Neuron 15, 941–949 (1995).
24. Swartz, K. J. & MacKinnon, R. Hanatoxin modifies the gating of a voltage-dependent K+ channel through multiple binding sites. Neuron 18, 665–673 (1997).
25. Liang, S. An overview of peptide toxins from the venom of the Chinese bird spider Selenocosmia huwena Wang [=Ornithoctonus huwena (Wang)]. Toxicon 43, 575–585 (2004).
26. Honma, T. & Shiomi, K. Peptide toxins in sea anemones: structural and functional aspects. Mar. Biotechnol. (NY) 8, 1–10 (2005).
27. Terlau, H. & Olivera, B. M. Conus venom: a rich source of novel ion channel-targeted peptides. Physiol. Rev. 84, 41–68 (2004).
28. Prashanth, J. R., Dutertre, S. & Lewis, R. J. Pharmacology of predatory and defensive venom peptides in cone snails. Mol. Biosyst. 13, 2453–2465 (2017).
29. Livett, B. G., Gayler, K. R. & Khalil, Z. Drugs from the sea: conopeptides as potential therapeutics. Curr. Med. Chem. 11, 1715–1723 (2004).
30. Bayrhuber, M. et al. Conkunitzitin-21 is the first member of a new Kunzite-type neurotoxin family. Structural and functional characterization. J. Biol. Chem. 280, 23766–23770 (2005).
31. Massilia, G. R., Schininà, M. E., Ascenzi, P. & Poliatti, C. Fortunypan-Vn: a novel peptide from the venom of the Mediterranean snail Conus ventricosus. Biochem. Biophys. Res. Commun. 288, 908–913 (2001).
32. Lazdunski, M. Apamin, a neurotoxin specific for one class of Ca2+-dependent K+ channels. Cell Calcium 4, 421–428 (1983).
33. Xiang, F. et al. Plectasin, first animal toxin-like fungal defensin blocking potassium channels through recognizing channel pore region. Toxins (Basel) 7, 34–42 (2015).
34. Yang, S. et al. Chemical punch packed in venoms makes centipedes excellent predators. Mol. Cell. Proteomics 11, 640–650 (2012).
35. Kuzmenkov, A. I., Krylov, N. A., Chugunov, A. O., Grishin, E. V. & Vassilevski, A. A. Kalium: a database of potassium channel toxins from scorpion venom. Database (Oxford) 2016, baw058 (2016).
36. The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45, D158–D169 (2017).
37. Berman, H. M. et al. The Protein Data Bank, Nucleic Acids Res. 28, 235–242 (2000).

Scientific Data | (2019) 6:73 | https://doi.org/10.1038/s41597-019-0074-x
38. Wilkins, M. R. et al. High-throughput mass spectrometric discovery of protein post-translational modifications. J. Mol. Biol. 289, 645–657 (1999).
39. Bairoch, A. et al. In The Proteomics Protocols Handbook (ed. Walker, J.) Protein identification and analysis tools on the ExPASy server (Humana Press, 2009).
40. Vassilevski, A., Tabakmakher, V. M., Krylov, N. A., Kuzmenkov, A. I. & Efremov, R. G. Kalium database. figshare https://doi.org/10.6084/m9.figshare.6913259.v3 (2019).
41. Santibáñez-López, C. E. & Possani, L. D. Overview of the Knottin toxin-like peptides in scorpion venoms: Insights on their classification and evolution. Toxicon 107, 317–326 (2015).
42. Kuzmenkov, A. I. et al. C-Terminal residues in small potassium channel blockers OdK1 and OSK3 from scorpion venom fine-tune the selectivity. Biochim. Biophys. Acta - Proteins Proteomics 1865, 465–472 (2017).
43. Kuzmenkov, A. I. et al. KV1.2 channel-specific blocker from Mesobuthus eupeus scorpion venom: Structural basis of selectivity. Neuropharmacology 143, 228–238 (2018).

Acknowledgements
We thank Anton O. Chugunov (Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia) for discussions concerning the database development. This work was funded by the Molecular and Cell Biology Program of the Russian Academy of Sciences, the Basic Research Program at the National Research University Higher School of Economics and by the Russian Academic Excellence Project ‘5–100’. Literature searches and data collection were performed by V.M.T. supported by the Russian Science Foundation (grant no. 17-74-10172).

Author Contributions
Working group organisers: R.G.E. and A.A.V. A.A.V. provided the idea of the database and upgrades. V.M.T. performed literature searches and data collection. A.I.K. provided the design of the database. N.A.K. performed programing and implementation of the database in the web interface. V.M.T. and A.I.K. performed testing the database efficiency. V.M.T., N.A.K., A.I.K. and A.A.V. wrote the manuscript and all authors contributed to its editing.

Additional Information
Competing Interests: The authors declare no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

The Creative Commons Public Domain Dedication waiver http://creativecommons.org/publicdomain/zero/1.0/ applies to the metadata files associated with this article.

© The Author(s) 2019