

SCIENTIFIC REPORTS



OPEN

TryTransDB: A web-based resource for transport proteins in Trypanosomatidae

Krushna Sonar, Ritika Kabra & Shailza Singh

TryTransDB is a web-based resource that stores transport protein data which can be retrieved using a standalone BLAST tool. We have attempted to create an integrated database that can be a one-stop shop for the researchers working with transport proteins of Trypanosomatidae family. TryTransDB (Trypanosomatidae Transport Protein Database) is a web based comprehensive resource that can fire a BLAST search against most of the transport protein sequences (protein and nucleotide) from Trypanosomatidae family organisms. This web resource further allows to compute a phylogenetic tree by performing multiple sequence alignment (MSA) using CLUSTALW suite embedded in it. Also, cross-linking to other databases helps in gathering more information for a certain transport protein in a single website.

Trypanosomatidae Transport Protein Database (TryTransDB), is a web based comprehensive resource that harbors protein and nucleotide sequences belonging to various superfamilies of transport proteins found in the members of the Trypanosomatidae family. The species of this family are parasitic protozoans in nature and have a single flagellum along with kinetoplastid which forms as an important distinguishing feature. The parasites can be dixenous or monoxenous depending on the number of hosts needed to complete their life cycle. The type of hosts ranges from insects which are usually the vectors to hosts like plants (eg. *Phytomonas* genera) and animals (eg. *Trypanosoma* and *Leishmania* genera)¹. Organisms belonging to *Leishmania*² cause Leishmaniasis while *Trypanosoma cruzi* and *Trypanosoma brucei* cause Chagas³ and sleeping sickness⁴ respectively. These diseases have serious deleterious effects or sometimes even fatal leishmaniasis⁵.

The above-mentioned diseases can be localized in many parts of India, Africa and recently in South America⁶. Numerous drugs are available to protect against the infection but all of these have their respective pros and cons^{6,7}. Published research repertoire says that resistance has been observed against Trypanosomiasis⁸ and Leishmaniasis⁹. This resistance is usually attained by the virtue of transport proteins found on the cell membrane that is responsible for pumping out the drugs out of the cell^{9,10}. One way to inhibit the function of these transport protein is to have certain types of molecule, be designed that will hinder their function either by inhibition or reducing their efficacy.

Studies have shown that these transport proteins like Major Facilitator Superfamily (MFS) and ATP-binding Cassette (ABC) protein can be used as potential drug targets^{11,12}. MFS and ABC are Transport Proteins, that are responsible for the transportation of molecules into and outside of the cell. These proteins are located either on the cell membrane or buried inside it where they span the complete width of the membrane.

Six types of superfamilies of transport proteins that are found in Trypanosomatidae family. ATP binding Cassette transporters (ABC), protein is located on the extracellular region of membrane and transportation is done by the hydrolysis of an ATP molecule. Various models of mechanism have been proposed¹³. Major Facilitator Superfamily (MFS), is the most ubiquitous, present in many organisms. Originally assumed function was sugar uptake just to later find out there are more biomolecules that are transported. The “rocker-switch model” is the proposed mechanism of function¹⁴. Resistance-Nodulation-Division (RND), proteins belonging to this family actively transport. This protein is a part of a complex comprising two more proteins that together span the outer and the periplasmic membrane. The association of these three results in transport, usually the proteins in these family allow the bacteria to gain resistance to a drug¹⁵. Amino Acid-Polyamine-Organocation Family (APC), protein has 12 transmembranes and the mechanism is known as “Rocking-bundle”^{16,17}.

Multidrug/Oligosaccharidyl-lipid/Polysaccharide (MOP) superfamily: The structure of these proteins comprises of twelve transmembrane helices and they act as cation antiport. Voltage-gated Ion Channel (VIC), they form ion channels and are specific to ions like sodium, potassium, calcium and chloride ions. They are uniport type transporters and have six transmembrane spanning segments.

National Centre for Cell Science, NCCS Complex, Ganeshkhind, SP Pune University Campus, Pune, 411007, India. Correspondence and requests for materials should be addressed to S.S. (email: singhs@nccs.res.in)

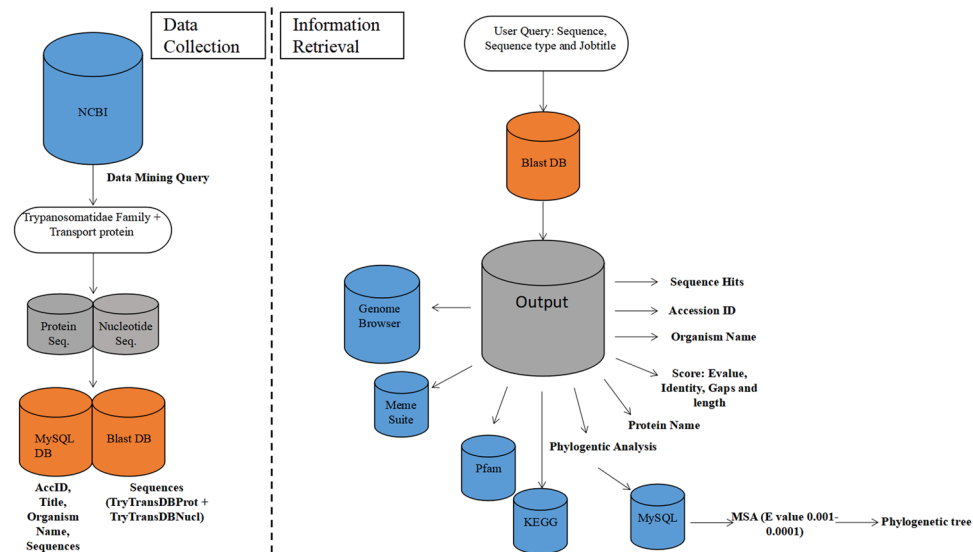


Figure 1. TryTransDB architecture.

Owing to the importance of these transport protein not only as drug target but also as a medium for the cell to be in contact with the environment, TryTransDB is made to cater the needs of researchers who are working in the respective fields. Since TryTransDB comprises of sequences only belonging to the Trypanosomatidae family, it becomes easy to filter out sequences of user's interest using the standalone BLAST incorporated in it. CLUSTALW¹⁸, using its default values, helps in calculating the phylogeny between the BLAST hits. Crosslinking is provided in order to get a wider spectrum of information from databases like KEGG¹⁹, NCBI Pfam²⁰, EMBL-EBI Pfam²⁰ and NCBI Genome Browser²¹ for the respective hits.

Information regarding the organism family with its medicinal importance along with the lifecycle of Trypanosomiasis, Leishmaniasis and the transport protein family is also available which gives a small gist of what the database is all about. Post BLAST run, a user can also download sequences from the BLAST output using the "DOWNLOAD" option which can be further used for more resource intensive phylogeny algorithms or for MEME suite which is cross-linked in the navigation bar. Users with novel sequences can even submit their sequences using the SeqSub tool in TryTransDB.

Materials and Methods

The front end of TryTransDB is developed using HTML. Apache was used to host the website and back end handling was done using PHP5 and PERL. Two databases were made, one for BLAST standalone software while one for phylogeny and downloading of sequences was made using MySQL.

Data collection and compilation. Trypanosomatidae family members were selected from the taxonomy database of NCBI. Transport proteins; ATP-binding Cassette (ABC) Superfamily, Major Facilitator Superfamily (MFS), Resistance-Nodulation-Cell Division (RND) Superfamily, Amino Acid - Polyamine - Organocation (APC) Family, Multidrug/Oligosaccharidyl - lipid/Polysaccharide (MOP) Flippase Superfamily, Voltage-gated Ion Channel (VIC) Superfamily were selected. An appropriate query was fired onto NCBI (<https://www.ncbi.nlm.nih.gov/>) using an in house script to download the protein and nucleotide. Around 754 protein sequences and 254 nucleotide sequences were retrieved and stored in TryTransDB as shown in Fig. 1.

Data Availability. The data present in the database can be accessed on the website <http://43.227.133.106/homePage.php>.

Result

The home page gives an introduction to TryTransDB, wherein, on the right side is the navigation bar.

Usage of TryTransDB. To visualize the working of TryTransDB, let's consider a case where the user wants to know if there are any organism and transport protein available for the MFS transport protein, user has sequenced. The input is protein sequence as shown in Fig. 2 and is taken from the existing entry from NCBI (CBZ12003.1).

BLAST output and KEGG. After entering the required fields in the "BLAST" page. The standalone BLAST using its local search algorithm checks sequences for similarity and identity against the query sequences and displays information as represented in Fig. 3.

BLAST output. Information like the hits along with Accession ID (linked to NCBI), Organism name, Protein name, Length, E value, Target Identity(%), Gaps and Pfam, Genome Browser (linked to NCBI) and KEGG are mentioned. The first hit confirms the BLAST has correctly identified the organism.

TryTransDB

Please enter the required information

Enter the sequence

```

->CBZ12003.1 Sequence
MPLDSEKPHRRRAAVDAAGQPSASVVKPKTSPNTVLAFALEEEENHIDDAALVAAMODENADPLLH
PAQTRALLHIEQREENRLRGDPWRFAVAFICLLLSISAMQWIAFSPVQEVRTYEHMNATQVN
FLATTYVIAVYIVLVELSCKLYEVTGKIGVLIAAABANALGALLKITALYAWPNAALLYLAQVSN
SETEVLTIATPPLISNRWFEKEKRWANTVMSSALNFGCGIIV
    
```

Select the BLAST type

Protein
 Nucleotide

Please enter job title

Leishmania major

Please enter username

test |

All rights are reserved © 2017 to Computational and Systems Biology Lab, National Center for Cell Science, SPPU Campus, Ganeshkhind, Pune 411007.

-
-
-
-
-
-
-
-
-

Figure 2. BLAST page. Details sequence, BLAST type, job title and username are to be filled in the given space.

TryTransDB

Leishmaniamajor

| Accession ID | Organism Name | Protein Name | Lenght | E value | Target Identity(%) | Gaps | Pfam | Genome Browser |
|--------------------------------|--|--------------|--------|--------------|--------------------|------|----------|----------------|
| CBZ12003.1 | Leishmania major strain Friedlin | MFS | 683 | 9.25738e-169 | 34.4 | 0 | P | G |
| XP_003721717.1 | Leishmania major strain Friedlin | MFS | 683 | 9.25738e-169 | 34.4 | 0 | P | G |
| CAM60014.1 | Leishmania infantum JPCM5 | MFS | 680 | 1.99072e-155 | 33.8 | 0 | P | G |
| XP_003858004.1 | Leishmania donovani | MFS | 680 | 1.99072e-155 | 33.8 | 0 | P | G |
| CBZ31280.1 | Leishmania donovani | MFS | 680 | 1.99072e-155 | 33.8 | 0 | P | G |
| XP_001462793.1 | Leishmania infantum JPCM5 | MFS | 680 | 1.99072e-155 | 33.8 | 0 | P | G |
| CAM36773.1 | Leishmania braziliensis MHOM/BR/75/M2904 | MFS | 687 | 1.43827e-145 | 30.1 | 1 | P | G |
| XP_001561627.1 | Leishmania braziliensis MHOM/BR | MFS | 687 | 1.43827e-145 | 30.1 | 1 | P | G |

All rights are reserved © 2017 to Computational and Systems Biology Lab, National Center for Cell Science, SPPU Campus, Ganeshkhind, Pune 411007.

-
-
-
-
-
-
-
-

Figure 3. Output page.

TryTransDB

[Leishmania major strain Friedlin](#)

[Leishmania major strain Friedlin](#)

[Leishmania infantum JPCM5](#)

[Leishmania donovani](#)

[Leishmania donovani](#)

[Leishmania infantum JPCM5](#)

[Leishmania braziliensis MHOM/BR/75/M2904](#)

[Leishmania braziliensis MHOM/BR/75/M2904](#)

[Trypanosoma cruzi Dm28c](#)

[Trypanosoma cruzi](#)

[Trypanosoma cruzi strain CL Brener](#)

[Trypanosoma cruzi](#)

[Trypanosoma cruzi](#)

[Trypanosoma cruzi strain CL Brener](#)

[Trypanosoma cruzi marinkellei](#)

[Trypanosoma cruzi](#)

[Trypanosoma cruzi Dm28c](#)

[Trypanosoma brucei brucei TREU927](#)

[Trypanosoma brucei brucei TREU927](#)

[Trypanosoma brucei gambiense DAL972](#)

[Trypanosoma brucei gambiense DAL972](#)

Home

Trypanosomatidae

Transport Protein

BLAST

SeqSub

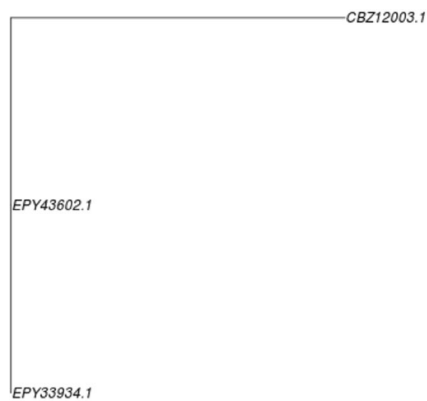
Genome Browser

PFam

MEME

Contact Us

Figure 4. KEGG output. Hyperlinks to the KEGG database are displayed.



The hits selected are based on the evalue ranging between 0.001 - 0.0001

The default parameters used for alignment and the tree generation are as follows:

Alignment:

Matrix = BLOSUM

Tree:

1. Seed = 1000
2. Kimura = ON
3. Clustering method = NJ
4. Output Tree = NJ

Figure 5. Phylogeny output. Newick form phylogenetic tree generated using CLUSTALW.

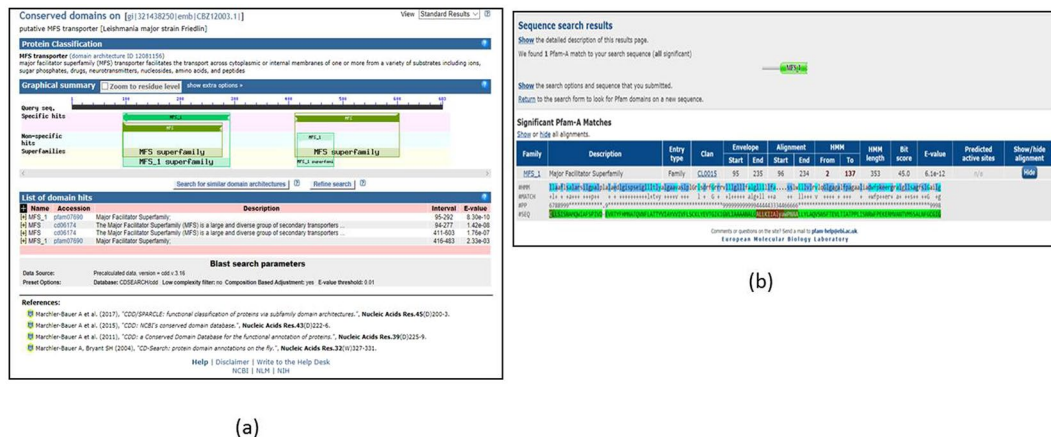


Figure 6. (a) Pfam output from NCBI. (b) Pfam output from EMBL database.

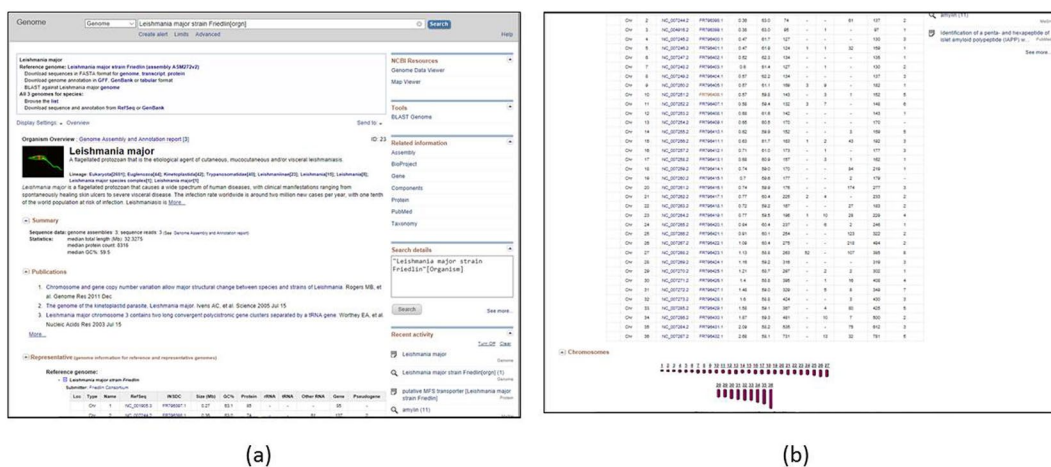


Figure 7. (a),(b) Genome output from NCBI for the selected entry.

KEGG links can be checked by clicking on the “KEGG” hyperlink (Fig. 4). A new tab is opened which comprises of only those entries for organisms that are present in KEGG’s database. So, it is likely that some organisms and their strains that have come as a hit will not appear in the KEGG list.

Phylogeny. A phylogenetic tree can also be generated by clicking on the “Phylogeny” link. TryTransDB computes the phylogenetic by invoking ClustalW and runs by using some default values. The database comprises of protein sequences whose E value is very close to each other and hence for the ease of crowding an E value cutoff (0.001–0.0001) has been applied for protein phylogenetic tree generation. This is not applied while calculating nucleotide phylogenetic tree generation since the number of sequences submitted are very less. Trees generated are represented in Newick format with the default conditions used and mentioned below (Fig. 5).

Pfam. From TryTransDB, Protein family, information can be accessed for the sequence in the output in this database. Pfam information is taken from NCBI and EMBL. In the output page, Pfam NCBI data can be accessed by clicking on “P” symbol (Fig. 6a) present in every row. To access the Pfam information from the EMBL database, there is Pfam button in the right-hand panel. This will redirect the user to the EMBL-EBI Pfam webpage. Of many types of input the database can take, one is sequence which can be downloaded using the “Download” in the top panel of the output page (Fig. 6b).

Genome Browser. Information like the location of the protein gene in the chromosome and additional details of the chromosome can be accessed through “Genome Browser”. By clicking on “G” symbol in the output page on every row of the result. This will redirect to the genome information of the respective entry in the NCBI (Figs. 7a,b).

Sequence submission. If the user wishes to submit a novel sequence, it can be done via the “SeqSub tool”. Sequence and the job title should be entered in the respective fields and a BLAST search is fired. If the sequence already exists in the database, it will prompt about the same, whereas, if the sequence is not present, the sequence will be manually curated and later added to the database, in a period of three weeks.

Discussion

Transport proteins are the potential drug target site since their function can be hindered by the use of some small molecule which leads to no exchange of molecules or important environmental signals or the administered drug like miltefosine in terms of leishmaniasis can function and not thrown out. Using this sequence repertoire, one can start with post-sequencing processes like structure prediction and molecular dynamics simulation techniques to understand the functioning and mechanism of these proteins in great details along with additional information available from cross-linked databases. The sites content would be kept updating for greater contributions towards the significant functionality of the web resource.

References

- Kaufer, A., Ellis, J., Stark, D. & Barratt, J. The evolution of trypanosomatid taxonomy. *Parasites & Vectors*. **10**, 287, <https://doi.org/10.1186/s13071-017-2204-7> (2017).
- World Health Organization. Weekly epidemiological record. <http://www.who.int/wer>. No 22, 91, 285–296 (2016).
- Prata, A. The Lancet Infectious Diseases, **1**, 2, 92–100.
- Malvy, D. *et al.* Sleeping sickness *Clinical Microbiology and Infection*, **17**, 7, 986–995, <https://doi.org/10.1111/j.1469-0691.2011.03536.x>
- Shyam Sundar & Rai, M. Laboratory Diagnosis of Visceral Leishmaniasis. *Clin Vaccine Immunol* **9**(5), 951–958, <https://doi.org/10.1128/CDLI.9.5.951-958.2002> (2002).
- Barrett, M. P. & Croft, S. L. Management of trypanosomiasis and leishmaniasis. *British Medical Bulletin*. **104**(1), 175–196, <https://doi.org/10.1093/bmb/lds031> (2012).
- Bernardes, L., Zani, C. & Ivone, C. Trypanosomatidae Diseases: From the Current Therapy to the Efficacious Role of Trypanothione Reductase in Drug Discovery. *Current medicinal chemistry*. <https://doi.org/10.2174/0929867311320210005> (2013).
- Gehrig, S. & Gehrig, S. Development of drug resistance in *Trypanosoma brucei rhodesiense* and *Trypanosoma brucei gambiense*. Treatment of human African trypanosomiasis with natural products (Review). *International Journal of Molecular Medicine* **22**, 411–419, <https://doi.org/10.3892/ijmm.00000037> (2008).
- Ouellette, M. *et al.* ABC transporters in Leishmania and their role in drug resistance. *Drug Resistance Updates* **1**(1), 43–48, [https://doi.org/10.1016/S1368-7646\(98\)80213-6](https://doi.org/10.1016/S1368-7646(98)80213-6) (1998).
- Munday, J. C., Settimo, L. & de Koning, H. P. Transport proteins determine drug sensitivity and resistance in a protozoan parasite, *Trypanosoma brucei*. *Frontiers in Pharmacology*. **6**, 32, <https://doi.org/10.3389/fphar.2015.00032> (2015).
- Ravna A. W., Sager G., Dahl S. G., Sylte I. Membrane Transporters: Structure, Function and Targets for Drug Design. In: Napier S., Bingham M. (eds) *Transporters as Targets for Drugs*. *Topics in Medicinal Chemistry*, volume 4. Springer, Berlin, Heidelberg (2008).
- Landfear, S. M. Transporters for Drug Delivery and as Drug Targets in Parasitic Protozoa. *Clinical Pharmacology & Therapeutics* <https://doi.org/10.1038/clpt.2009.116>.
- Wilkens, S. Structure and mechanism of ABC transporters. *F1000Prime Reports*. **7**, 14, <https://doi.org/10.12703/P7-14> (2015).
- Esen, M. Q., Christian, L., Fatma, G. & Pär, N. Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nature Reviews Molecular Cell Biology* <https://doi.org/10.1038/nrm.2015.25>.
- Nikaido, H. Structure And Mechanism Of Rnd-Type Multidrug Efflux Pumps. *Advances in enzymology and related areas of molecular biology*. **77**, 1–60 (2011).
- Jack, D. L., Paulsen, I. T. & Saier, M. H. The amino acid/polyamine/organocation (APC) superfamily of transporters specific for amino acids, polyamines and organocations *Microbiology* **146**, 1797–1814, <https://doi.org/10.1099/00221287-146-8-1797>.
- Rudnick, G. The Cytoplasmic Permeation Pathway of Neurotransmitter Transporters. *Biochemistry*. **50**(35), 7462–7475, <https://doi.org/10.1021/bi200926b> (2011).
- Chenna, R. *et al.* Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research*. **31**(13), 3497–3500 (2003).
- Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*. **28**(1), 27–30 (2000).
- Finn R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Research*. **42**(Database issue): D222–D230. <https://doi.org/10.1093/nar/gkt1223> (2014).
- Karolchik, D., Hinrichs, A. S. & Kent, W. J. The UCSC Genome Browser. Current protocols in bioinformatics/editorial board, Andreas D Baxevanis. [et al.]. CHAPTER: Unit1.4. <https://doi.org/10.1002/0471250953.bi0104s28> (2009).

Acknowledgements

Krushna Sonar and Ritika Kabra acknowledges their fellowships from Department of Biotechnology, Government of India. Authors would also like to thank the Director, National Centre for Cell Science (NCCS) for supporting the Bioinformatics and High Performance Computing Facility (BHPCF) at NCCS, Pune, India.

Author Contributions

Conceived and designed the experiments: S.S. Performed the experiments: K.S., R.K. Analyzed the data: K.S., R.K. and S.S. Contributed to the writing of the manuscript: K.S., R.K. and S.S.

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 Author(s) 2018