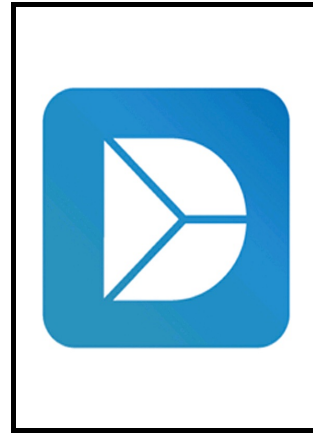


Author's Accepted Manuscript

A survey of sequences of KT-HAK-KUP transporters in green algae and basal land plants

Guillermo E. Santa María, Sonia Oliferuk, Jorge I. Moriconi



www.elsevier.com/locate/dib

PII: S2352-3409(18)30784-4S0176-1617(18)30102-0

DOI: <https://doi.org/10.1016/j.dib.2018.07.011>

Reference: DIB2828

To appear in: *Data in Brief*

Received date: 20 April 2018

Accepted date: 5 July 2018

Cite this article as: Guillermo E. Santa María, Sonia Oliferuk and Jorge I. Moriconi, A survey of sequences of KT-HAK-KUP transporters in green algae and basal land plants, *Data in Brief*, <https://doi.org/10.1016/j.dib.2018.07.011>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Data Article**Title:** A survey of sequences of KT-HAK-KUP transporters in green algae and basal land plants**Authors:** Guillermo E. Santa María^{a,*}, Sonia Oliferuk^a, Jorge I. Moriconi^a**Affiliations:** *Instituto Tecnológico Chascomús (INTECH), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) and Universidad Nacional de San Martín (UNSAM). Avda Intendente Marino km 8,2. Chascomús, 7130. Provincia de Buenos Aires. Argentina.****Corresponding author****E-mail address:** gsantama@intech.gov.ar**Abstract**

In this data article, information is provided on sequences of KT-HAK-KUP transporters from green algae and basal land plants. A data set is offered containing sequences corresponding to the chlorophyte algae *Chlamydomonas eustigma*, *Gonium pectorale* and *Coccomyxa subellipsoidea*, the charophyte algae *Coleochaete orbicularis* and *Klebsormidium flaccidum*, the bryophyte *Sphagnum fallax*, the marchantophyte *Marchantia polymorpha* and the gymnosperm *Pinus taeda*, which have been not formerly analyzed. In addition, an analysis of similarity scores among representatives of the clusters recognized in photosynthetic green organisms (namely, chlorophyte algae, charophyte algae, basal embryophytes and higher embryophytes) is performed as well as an analysis of membrane topology for them.

Key Words: algae, HAK, KUP, KT, transporter

Color should be not used in any of the figures in print

Specifications Table

Subject area	<i>Plant Biology</i>
More specific subject area	<i>Plant ion transport, transporter phylogeny</i>
Type of data	<i>Protein sequences retrieved from public databases, analysis of similarity and topological models.</i>
How data was acquired	<i>Retrieved from public databases and further bioinformatic analysis</i>
Data format	<i>Raw data in fasta (for sequence data set), analyzed data (Tables and</i>

	<i>Figures)</i>
Experimental factors	<i>No applicable</i>
Experimental features	<i>No applicable</i>
Data source location	<i>Chascomús, Buenos Aires, Argentina</i>
Data accessibility	The data are available within this article
Related article	<i>KT-HAK-KUP transporters in major terrestrial photosynthetic organisms: a twenty years tale</i> [1]

Value of the Data

- Data presented correspond to sequences of KT-HAK-KUP transporters retrieved from public databases which have been not formerly analyzed for the organisms *Coccomyxa subellipsoidea*, *Chlamydomonas eustigma*, *Gonium pectorale*, *Sphagnum fallax*, *Marchantia polymorpha*, *Pinus taeda*, *Coleochaete orbicularis* and *Klebsormidium flaccidum* and a new sequence for *Chlamydomonas reinhardtii*. The data set is complemented with sequences already available for *Chlamydomonas reinhardtii*, *Physcomitrella patens*, *Selaginella moellendorffii*, *Picea abies*, *Amborella trichopoda*, *Arabidopsis thaliana*, *Prunus persica*, *Oryza sativa* and *Zea mays*.
- These sequences can be used to build up phylogenetic trees and to perform complementary analyses based on them. In this data article, they have been used to analyze the similarity scores in pair comparisons among members of different KT-HAK-KUP clades as well as to predict their possible topology.

Data

Data introduced correspond to new KT-HAK-KUP sequences derived from public databases and the analysis performed for them. The description of the sources for all the sequences used for *Coccomyxa subellipsoidea*, *Chlamydomonas reinhardtii*, *Chlamydomonas eustigma*, *Gonium pectorale*, *Sphagnum fallax*, *Marchantia polymorpha*, *Pinus taeda*, *Coleochaete orbicularis* and *Klebsormidium flaccidum* is provided in Table 1, while the corresponding sequences are provided in the accompanying data set (Appendix A. Supplementary Material, Table). The data set also contains the KT-HAK-KUP sequences already posted by Nieves-Cordones et al. [2] for *Physcomitrella patens*, *Selaginella moellendorffii*, *Picea abies*, *Amborella trichopoda*, *Arabidopsis thaliana*, *Prunus persica*, *Oryza sativa* and *Zea mays* (Table 6), which have been used to construct a phylogenetic tree of KT-HAK-KUPs in green algae and land plants [1]. The data set includes only full length sequences that contain a region with homology to the first putative first transmembrane domain as well as the highly conserved GGT(A/L/I/P/S)F(L/A)A(S)L(V/I/M/A)YS(T/A) motif as determined following their alignment with MAFT. An alignment of chlorophyte and charophyte algae sequences together with those

corresponding to *Arabidopsis thaliana* is provided (Appendix A. Supplementary Material, Figure 1). Percent similarity scores among representatives of KT-HAK-KUPs corresponding to the 12 clades identified in green photosynthetic organisms were estimated (Table 2). In the same way percent similarity scores among the members identified in each green algal clade are shown (Table 3). The capacity of four topology prediction services to generate satisfactory topological models for the bacterial KUP transporter, for which the topology has been experimentally determined [3], is summarized in Appendix A. Supplementary Material, Figure 2. It was found that all the services (TOPCONS, TOPCONS-single, THMH 2.0 and TMPred) provided an accurate agreement with the topology of KUP. Therefore, these four services were used to analyze the number and orientation of transmembrane domains in selected members of the above mentioned clades of green photosynthetic organisms. The predicted molecular weights of these representative KT-HAK-KUP transporters are given along with a prediction of the number of transmembrane domains derived from the use of the four topology prediction services (Table 4). In Table 5 a more detailed analysis of the residues involved in each transmembrane domain is provided for canonical members of the KT-HAK-KUP family of transporters, namely HvHAK1, AtHAK5, AtKUP4(TRH1) and AtKUP7.

Experimental Design, Materials, and Methods description

Sequences corresponding to the chlorophytes *Chlamydomonas reinhardtii* and *Coccomyxa subellipsoidea*, identified by performing Blast, were retrieved from Phytozome 12. It was also the source for sequences corresponding to the bryophyte *Sphagnum fallax* and to the marchantophyte *Marchantia polymorpha*. Sequences from the gymnosperm *Pinus taeda* were retrieved from Congenie.org. In turn, sequences from the charophyte algae *Coleochaete orbicularis* and *Klebsormidium flaccidum* were retrieved by performing Blast on the transcriptome shotgun assembly database. Sequences from *Chlamydomonas eustigma* and *Gonium pectorale* were retrieved from NCBI. The remaining sequences corresponding to the moss *Physcomitrella patens*, the lycopodiophyte *Selaginella moellendorffii*, the gymnosperm *Picea abies*, the basal angiosperm *Amborella trichopoda*, the dicots *Arabidopsis thaliana* and *Prunus persica* and the monocots *Oryza sativa* and *Zea mays* were obtained from Nieves-Cordones et al. [2]. Sequences for *Chlamydomonas reinhardtii* were denoted as proposed by He et al. [4], being an additional sequence retrieved from Phytozome 12. Following retrieval of sequences a phylogenetic tree was built up, being 6 clades recognized in embryophytes and 6 corresponding to chlorophyte and charophyte algae (Santa-María et al., submitted). A subset containing only green algae sequences and those of *Arabidopsis thaliana* were aligned. The multiple alignment was generated with the MAFFT program (version 7) at <https://mafft.cbrc.jp/alignment/server/index.html>. Pair comparisons of similarity among representatives of the 12 clades, and subgroups of clades II and XII, were performed at https://www.ebi.ac.uk/Tools/psa/emboss_needle/. In order to advance on the structure of

these putative transporters four services for the prediction of transmembrane domains were used to predict the transmembrane domains of KUP (TOPCONS, TOPCONS-single, THMH 2.0 and TMPred). These prediction services were next used to analyze the possible topology of potential representatives of the above mentioned clades of green photosynthetic organisms.

Acknowledgments

The authors express gratitude to the public repositories mentioned for share the information of the sequences used in this work. This work was funded by CONICET and the PICT 2014-1887 by the Agencia Nacional de Promoción Científica y Tecnológica (ANPCYT) to G.E. Santa María. S. Oliferuk acknowledges to the Consejo Nacional de Investigación Científica y Técnica, CONICET, for a post-doctoral fellowship.

Transparency document. Supplementary Material

Supplementary data associated with this article can be found in the online version at :

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at :

References

- [1] G.E. Santa-María, S. Oliferuk, J.I. Moriconi, KT-HAK-KUP transporters in major terrestrial photosynthetic organisms: a twenty years tale, *Journal of Plant Physiology* (accepted).
- [2] M. Nieves-Cordones, R. Ródenas, A. Chavanieu, R.M. Rivero, V. Martínez, I. Gaillard, F. Rubio, Uneven HAK/KUP/KT protein diversity among angiosperms: species distribution and perspectives, *Front. Plant Sci.* 7 (2016) 127.
- [3] Y. Sato, K. Nanatani, S. Hamamoto, M. Shimizu, M. Takahashi, M. Tabuchi-Kobayashi, A. Mizutani, J.I. Schroeder, S. Souma, N. Uozumi, Defining membrane spanning domains and crucial membrane-localized acidic amino acid residues for K⁺ transport of a Kup/HAK/KT-type *Escherichia coli* potassium transporter, *J. Biochem.* 155 (2014) 315–323.
- [4] C. He, K. Cui, A. Duan, Y. Zeng, J. Zhang, Genome-wide and molecular evolution analysis of the Poplar KT/HAK/KUP potassium transporter gene family, *Ecol. Evol.* 2 (2012) 1996–2004.