

SCIENTIFIC REPORTS

OPEN

Near full length genome of a recombinant (E/D) cosavirus strain from a rural area in the central region of Brazil

Antonio Charlys da Costa¹, Adriana Luchs², Flavio Augusto de Pádua Milagres^{3,4,5,6}, Shirley Vasconcelos Komninakis^{7,8}, Danielle Elise Gill¹, Márcia Cristina Alves Brito Sayão Lobato^{4,6}, Rafael Brustulin^{4,5,6}, Rogério Togisaki das Chagas^{4,6}, Maria de Fátima Neves dos Santos Abrão^{4,6}, Cassia Vitória de Deus Alves Soares^{4,6}, Xutao Deng^{9,10}, Ester Cerdeira Sabino^{1,3}, Eric Delwart^{9,10} & Élcio Leal¹¹

In the present article we report the nearly full length genome of a Cosavirus strain (BRTO-83) isolated from a child with acute gastroenteritis, and who is an inhabitant of a rural area in the central region of Brazil. The sample was previously screened and negative for both: common enteric viruses (i.e. rotavirus and norovirus), bacteria, endoparasites and helminthes. Evolutionary analysis and phylogenetic inferences indicated that the Brazilian BRTO-83 Cosavirus strain was a recombinant virus highly related to the E/D recombinant NG385 strain (Genbank JN867757), which was isolated in Nigeria from an acute flaccid paralysis patient. This is the first report of a recombinant E/D Cosavirus strain detected in Brazil, and the second genome described worldwide. Further surveillance and molecular studies are required to fully understand the epidemiology, distribution and evolution of the Cosavirus.

The family *Picornaviridae* has undergone a significant expansion in recent years, due principally to the identification of previously unknown picornaviruses by next-generation sequencing (NGS) of clinical and environmental samples. The family is divided currently into 29 genera, and the natural hosts of these viruses are vertebrates, including mammals and birds. Enteroviruses, polioviruses, hepatoviruses and aphthoviruses are the most exhaustively characterized animal pathogens, and are associated with a wide spectrum of clinical manifestations, including undifferentiated febrile illness, respiratory illness, aseptic meningitis and acute flaccid paralysis (AFP)¹. Recently, metagenomics analysis made possible the identification of three new human picornaviruses: klassevirus, salivirus, and cosavirus²⁻⁴. The pathogenicity of Cosaviruses in humans remains unknown. However, Cosaviruses have been detected in human fecal specimens, and studies have suggested its association with gastroenteritis infection^{5,6}. The genus Cosavirus consists of five genotypes (Cosavirus A, Cosavirus B, Cosavirus D, Cosavirus E and Cosavirus F) and is most closely related to members of the Cardiovirus and Senecavirus genera^{2,7}.

We herein report the detection of a recombinant Cosavirus strain (BRTO-83) isolated from a fecal sample obtained from a 1-year-old-child patient presenting acute gastroenteritis in 2013, in the state of Tocantins, Central Brazil. Tocantins is characterized by dry grassland, scrub forests, vast rivers and soybean plantations. The modern capital, Palmas, was purposefully built in the geographic center of the state and is surrounded by forested

¹Institute of Tropical Medicine, University of São Paulo, São Paulo, Brazil. ²Enteric Disease Laboratory, Virology Center, Adolfo Lutz Institute, São Paulo, Brazil. ³LIM/46, Faculty of Medicine, University of São Paulo, São Paulo, Brazil. ⁴Secretary of Health of Tocantins, Tocantins, Brazil. ⁵Federal University of Tocantins, Tocantins, Brazil. ⁶Public Health Laboratory of Tocantins State (LACEN/TO), Tocantins, Brazil. ⁷Postgraduate Program in Health Science, Faculty of Medicine of ABC, Santo André, Brazil. ⁸Retrovirology Laboratory, Federal University of São Paulo, São Paulo, São Paulo, Brazil. ⁹Blood Systems Research Institute, San Francisco, USA. ¹⁰Department Laboratory Medicine, University of California San Francisco, San Francisco, USA. ¹¹Institute of Biological Sciences, Federal University of Pará, Pará, Brazil. Antonio Charlys da Costa, Adriana Luchs and Flavio Augusto de Pádua Milagres contributed equally to this work. Ester Cerdeira Sabino, Eric Delwart and Élcio Leal jointly supervised this work. Correspondence and requests for materials should be addressed to A.C.d.C. (email: charlysbr@yahoo.com.br) or É.L. (email: elcioleal@gmail.com)

hills. The patient was an inhabitant of Barra do Ouro, a small municipality (4,460 inhabitants) located 422 km from the state capital, Palmas.

Results and Discussion

A survey conducted throughout in the northern state of Tocantins in Brazil was performed during the years 2010 through 2016 to screen human feces for the presence of enteric pathogens. Fecal specimens were screened for enteric viruses (i.e. rotavirus and norovirus), bacteria (i.e. *E. coli* and *Salmonella sp.*), endoparasites (i.e. *Giardia sp.*) and helminthes using conventional (i.e. culture techniques) and molecular methods (i.e. commercial enzyme immunoassays). NGS techniques were used to identify possible unscreened or undetected enteric viruses, and a recombinant Cosavirus strain was detected in sample BRTO-83.

The near full length genome (6278 nt) was sequenced and compared with previously described Cosavirus strains (Fig. 1 and Supplementary data). Recombination analysis indicated that the first 2680 bases of the BRTO-83 genome are related to the genotype E, while the remaining nucleotides are related to the genotype D (Fig. 2 and Supplementary data). This pattern of Cosavirus recombination has been previously reported only once, in 2007, from a Nigerian patient suffering from AFP (NG385 strain; JN867757). Additional phylogenetic analysis, performed using complete genomes and the partitions corresponding to the genome breakpoints, confirmed that the Brazilian BRTO-83 strain was closely related to the NG385 strain E/D from Nigeria. The recombination breakpoint detected in both isolates, BRTO-83 and NG385, coincided, suggesting that both sequences display identical mosaic patterns (Fig. 2 and Supplementary Material). The inferred trees, which were assembled using the partitions and composed by free recombination regions also suggest that the recombinant E/D strains detected in Brazil and Nigeria may share the same evolutionary ancestor (Fig. 1 and Supplementary data).

The child infected by the BRTO-83 strain was suffering from a severe episode of diarrhea, and the data strongly suggests that a Cosavirus was, in fact, the causative agent of the disease. A common feature of virome analysis is that the epidemiological background information on potential contact with animals, consumption of contaminated food and/or medical records of the patients is not generally available. The present study is no exception; there was no epidemiological connection between the patient and the virus. Cosaviruses are not pathogens recognized to be associated with gastroenteritis, and the patient may have been infected with other bacteria and parasites not recognized in the screening flow. In addition, the protocol applied focused on viral metagenomics and removed eukaryotic cells from the test samples, making impossible the use of the BRTO-83 sample library to screen for reads from bacteria, parasites or helminthes.

Human Cosavirus infection reports are still very limited worldwide, but the association of the viruses with gastroenteritis is increasing^{4,8,9}. In Brazil, there is only one study considering Cosaviruses role; nevertheless its association with gastroenteritis could not be established³. This is the first report of a recombinant E/D Cosavirus strain detected in Brazil, and the second genome described worldwide. Further surveillance and molecular studies are required to fully understand the epidemiology, distribution and evolution of the Cosavirus.

Materials and Methods

The protocol used to perform deep-sequencing was a combination of several protocols normally applied to viral metagenomics and/or virus discovery, and has been partially described by da Costa *et al.*¹⁰. In summary, 50 mg of the human BRTO-83 fecal sample was diluted in 500 µl of Hanks' buffered salt solution (HBSS), added to a 2 ml impact-resistant tube containing lysing matrix C (MP Biomedicals, USA) and homogenized in a FastPrep-24 5G Homogenizer (MP biomedical, USA). The homogenized sample was centrifuged at 12,000 × g for 10 min, and approximately 300 µl of the supernatant was then percolated through a 0.45 µm filter (Merck Millipore, Billerica, MA, USA) in order to remove eukaryotic- and bacterial- cell-sized particles. Approximately 100 µl, roughly equivalent to one fourth of the volume of the tube of cold PEG-it Virus Precipitation Solution (System Biosciences, CA, USA), was added to the obtained filtrate, and the contents of the tube were gently mixed then incubated at 4 °C for 24 hours. After the incubation period, the mixture was centrifuged at 10000 × g for 30 minutes at 4 °C. Following centrifugation, the supernatant (~350 µl) was discarded. The pellet, rich in viral particles was treated with a combination of nuclease enzymes (TURBO DNase and RNase Cocktail Enzyme Mix - Thermo Fischer Scientific, CA, USA; Baseline-ZERO DNase - Epicentre, WI, USA; Benzonase - Darmstadt, Germany; and RQ1 RNase-Free DNase and RNase A Solution - Promega, WI, USA) in order to digest unprotected nucleic acids. The resulting mixture was subsequently incubated at 37 °C for 2 h.

After incubation, viral nucleic acids were extracted using a ZR & ZR-96 Viral DNA/RNA Kit (Zymo Research, CA, USA) according to the manufacturer's protocol. The cDNA synthesis was performed with AMV reverse transcription (Promega, WI, USA). A second strand of cDNA synthesis was performed using DNA Polymerase I Large (Klenow) Fragment (Promega, WI, USA). Subsequently, a Nextera XT Sample Preparation Kit (Illumina, CA, USA) was used to construct a DNA library, which was identified using dual barcodes. For size range, Pippin Prep (Sage Science, Inc.) was used to select a 300 bp insert (range 200–400 bp). The library was deep-sequenced using the Hi-Seq 2500 Sequencer (Illumina, CA, USA) with 126 bp ends. Bioinformatic analysis was performed according to the protocol previously described by Deng *et al.*¹¹. Contigs sharing a percent nucleotide identity of 95% or less were assembled from the obtained sequence reads by *de novo* assembly. The contigs included the rotavirus sequences, as well as others, such as enteric viruses, human, fungal and bacterial sequences. The resulting singlets and contigs were analyzed using BLASTx to search for similarity to viral proteins in GenBank's Virus RefSeq. The contigs were compared to the GenBank nonredundant nucleotide and protein database (BLASTn and BLASTx). After the identification of a Cosavirus, a reference template sequence was used for mapping the full-length genome with Geneious R9⁷.

BLASTn was initially used to identify viral sequences through their sequence similarity to annotated viral genomes in GenBank. Based on the best hits of the blastx searches, the following 14 polyprotein (6329 nt) sequences, listed by their Genbank numbers, were chosen to be used in the next analyses: JN867756, JN867757,

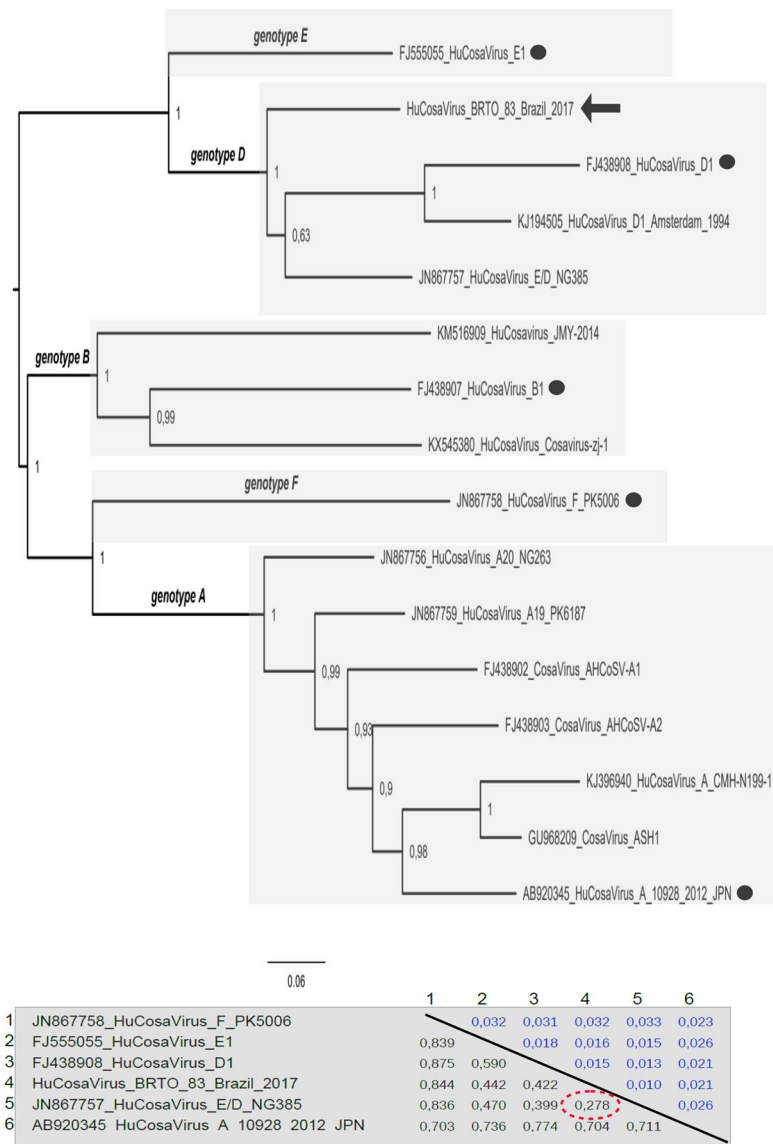


Figure 1. Phylogenetic tree constructed using near-full-length genomes of the human cosaviruses. The Maximum likelihood tree was constructed with genomes of human strains of cosaviruses (HuCosaVirus). The Brazilian strain BRTO-83, described in this study, is indicated by a filled arrow. The values on the tree indicate the statistical support of each node that was calculated using the Shimodaira-Hasegawa test. Genotype reference strains are indicated by filled dots on the genome tree. The panel in the figure is a matrix that shows the values of genetic distances between some isolates. Values below the diagonal are genetic distances and values above the diagonal are standard errors. These distances were estimated using Mega software version 7.0.

JN867758, JN867759, FJ555055, FJ438908, KJ194505, KM516909, AB920345, KJ396940, GU968209, FJ438903, FJ438902, FJ438907 and KX545380. These genomes were then aligned using Clustal \times software¹². Subsequently a phylogenetic tree was constructed by the Maximum Likelihood approach and branch support values were assessed by the Shimodaira-Hasegawa test. All trees were inferred using PhyML software¹³. The KHY model and gamma distribution were selected according to the likelihood ratio test (LRT) implemented in the jModeltest software¹⁴.

To determine the extent of recombination among sequences, we used software RDP v.4, which utilizes a collection of methods. An excellent and detailed explanation of each method implemented in the RDP software can be found in the user's manual (<http://darwin.uvigo.es/rdp/rdp.html>). Breakpoints were identified by means of software Simplot, version 2.5, which is available at <http://sray.med.som.jhmi.edu/RaySoft>. Recombination was checked by means of Bootscan, using the neighbor-joining approach and the F84 substitution model. A sliding window of 350 bp with 60-bp increments was used. Window trees were replicated 300 times to provide bootstrap support for permuted trees.

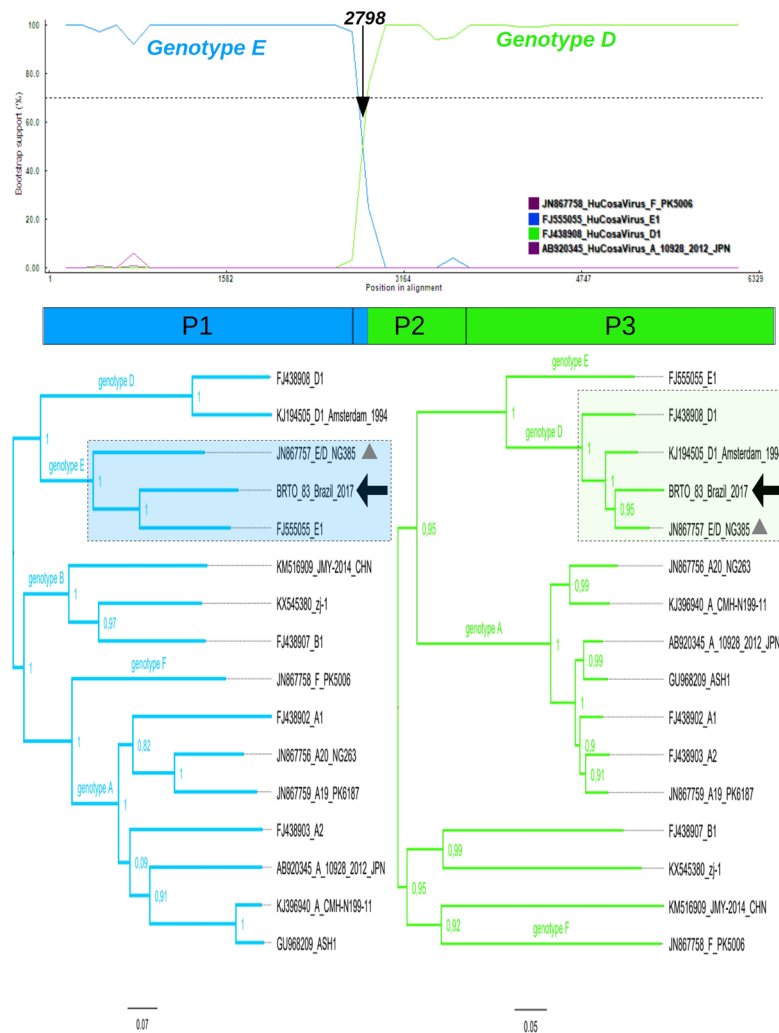


Figure 2. Recombination pattern of the Brazilian isolate BRTO-83. Colored lines represent the probability (given in bootstrap value) that genomic regions belong to a certain parental genotype. The x-axis represents the sequence length in base pairs (bp). The y-axis represents the statistical support (bootstrap) based on 500 replicates. Each plotted line refers to a certain genotype (see the code color on the small panel). A vertical arrow indicates the recombination break point on the genome region. The plot indicates a single breakpoint in the polyprotein region of the isolate BRTO-83 at the position 2798. This breakpoint corresponds to the initial part of the Cosavirus protein P1 (see diagram below the recombination plot). Strains used as parental references are also shown in the recombination plot. Trees were constructed using the partitioned genome. The first partition tree (nucleotides; 1–2798) demonstrates the isolate BRTO-83 clustering with genotype E (left tree), and the second tree (nucleotides 2799–6269) demonstrates the isolate BRTO-83 within the clade formed by genotype D strains (right tree). Genotype reference strains are indicated by filled dots on the genome tree. The recombinant strain NG385 is indicated by gray triangles in the trees. Analyses were performed using a neighbor-joining method and a Kimura 2 parameters model in windows of 250 bp sliding along sequences in increments of 40 bp. For parental references, non-recombinant sequences were used. The bootscan plot was obtained by analysis performed using RDP v.4 software.

Accession number. The BRTO-83 strain has been deposited in GenBank under accession number MG979713.

Ethical approval. Previous Ethics Committee approval was granted by the Adolfo Lutz Institute, São Paulo, Brazil (CEP 965.723; CTC 45G-2014), the Faculdade de Medicina da Universidade de São Paulo (CAAE: 53153916.7.0000.0065) and the Centro Universitário Luterano de Palmas - ULBRA (CAAE 53153916.7.3007.5516). This was an anonymous unlinked study and informed consent was not required according to resolution 466/12 concerning research involving humans (Conselho Nacional de Saúde/Ministério da Saúde, Brasília, 2012).

References

1. Tapparel, C., Siegrist, F., Petty, T. J. & Kaiser, L. Picornavirus and enterovirus diversity with associated human diseases. *Infect Genet Evol* **14**, 282–293, <https://doi.org/10.1016/j.meegid.2012.10.016> (2013).
2. Greninger, A. L. *et al.* The complete genome of klassevirus - a novel picornavirus in pediatric stool. *Virology* **6**, 82, <https://doi.org/10.1186/1743-422X-6-82> (2009).
3. Stocker, A. *et al.* Cosavirus infection in persons with and without gastroenteritis, Brazil. *Emerg Infect Dis* **18**, 656–659, <https://doi.org/10.3201/eid1804.111415> (2012).
4. Yu, J. M., Ao, Y. Y., Li, L. L. & Duan, Z. J. Identification of a novel cosavirus species in faeces of children and its relationship with acute gastroenteritis in China. *Clin Microbiol Infect* **23**, 550–554, <https://doi.org/10.1016/j.cmi.2017.02.018> (2017).
5. Holtz, L. R., Finkbeiner, S. R., Kirkwood, C. D. & Wang, D. Identification of a novel picornavirus related to cosaviruses in a child with acute diarrhea. *Virology* **5**, 159, <https://doi.org/10.1186/1743-422X-5-159> (2008).
6. Kapoor, A. *et al.* A highly prevalent and genetically diversified Picornaviridae genus in South Asian children. *Proc Natl Acad Sci USA* **105**, 20482–20487, <https://doi.org/10.1073/pnas.0807979105> (2008).
7. Kapusinszky, B., Phan, T. G., Kapoor, A. & Delwart, E. Genetic diversity of the genus Cosavirus in the family Picornaviridae: a new species, recombination, and 26 new genotypes. *Plos One* **7**, e36685, <https://doi.org/10.1371/journal.pone.0036685> (2012).
8. Menage, L., Yodmeeklin, A., Khamrin, P., Kumthip, K. & Maneekarn, N. Prevalence of human cosavirus and scaffold virus with an emergence of scaffold virus genotype 6 in patients hospitalized with acute gastroenteritis in Chiang Mai, Thailand, 2014–2016. *Infect Genet Evol* **53**, 1–6, <https://doi.org/10.1016/j.meegid.2017.05.005> (2017).
9. Okitsu, S. *et al.* Detection and molecular characterization of human cosavirus in a pediatric patient with acute gastroenteritis, Japan. *Infect Genet Evol* **28**, 125–129, <https://doi.org/10.1016/j.meegid.2014.09.019> (2014).
10. Charlys da Costa, A. *et al.* Spread of Chikungunya Virus East/Central/South African Genotype in Northeast Brazil. *Emerg Infect Dis* **23**, 1742–1744, <https://doi.org/10.3201/eid2310.170307> (2017).
11. Deng, X. *et al.* An ensemble strategy that significantly improves de novo assembly of microbial genomes from metagenomic next-generation sequencing data. *Nucleic Acids Res* **43**, e46, <https://doi.org/10.1093/nar/gkv002> (2015).
12. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948, <https://doi.org/10.1093/bioinformatics/btm404> (2007).
13. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307–321, <https://doi.org/10.1093/sysbio/syq010> (2010).
14. Posada, D. jModelTest: phylogenetic model averaging. *Mol Biol Evol* **25**, 1253–1256, <https://doi.org/10.1093/molbev/msn083> (2008).

Acknowledgements

We thank Luciano Monteiro da Silva, Maria da Aparecida R. Teles, Renan Alves, Vanessa Silva and Renata Palma for support. We also thank the Illumina Inc, Sage Science Inc, MP Biomedicals Inc, Zymo Research Corporation, Promega Biotecnologia do Brasil, for the donation of reagents for this project. This study was supported by FAPESP #2016/01735-2 and CNPq #400354/2016-0. Antônio Charlys da Costa is funded by FAPESP #2017/00021-9 and Adriana Luchs is funded by FAPESP #2015/12944-9.

Author Contributions

Conceptualization: A.C.C., A.L., F.A.P.M., S.V.K., D.E.G., M.C.A.B.S.L., R.B., R.T.C., M.F.N.S.A., C.V.D.A.S., X.D., E.C.S. and E.D.; Sequence processing: A.C.C. and A.L.; Phylogenetic analysis: E.L.; Data Investigation: A.C.C., A.L., F.A.P.M., S.V.K., D.E.G., M.C.A.B.S.L., R.B., R.T.C., M.F.N.S.A., C.V.D.A.S. and E.C.S.; Supervision: E.C.S. and A.C.C.; Writing original draft: A.C.C., A.L., D.E.G. and E.L.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-30214-1>.

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