Archives of Virology (2023) 168:242
https://doi.org/10.1007/s00705-023-05879-3

ANNOTATED SEQUENCE RECORD

Complete genome sequence of a divergent strain of cardamom mosaic virus

M. Greeshma[1] · **A. I. Bhat**[1]

Received: 12 July 2023 / Accepted: 8 August 2023 / Published online: 5 September 2023

Abstract

Cardamom mosaic virus (CdMV; genus *Macluravirus*), which causes mosaic (*katte*) disease in cardamom, is a highly variable member of the family *Potyviridae*. So far, the complete genome sequence of one isolate from Karnataka (KS) has been reported. In the present study, we determined the complete genome sequence of a CdMV isolate from Kerala (KI) and the complete CP gene sequences of nine isolates of CdMV from Kerala, Karnataka, and Tamil Nadu, India. The complete genome of CdMV (KI) consists of 8255 nucleotides (nt) with two open reading frames (ORFs). The large ORF, potentially coding for a polyprotein of 2638 amino acids (aa), is further processed into nine mature proteins at eight cleavage sites. The second ORF, PIPO (pretty interesting *Potyviridae* ORF) starting with a $C(A)_6$ motif, encodes a small protein of 56 aa. The viral genome contains an additional 13 nt in the 5′ untranslated region (UTR) and 6 nt in the CP gene, as well as a deletion of 13 nt at the 3′ UTR in comparison to the KS isolate of CdMV. The complete viral genome and polyprotein share 76% and 85% sequence identity with the KS isolate of CdMV, indicating that the present isolate is highly divergent from the KS isolate. Sequencing and analysis of the CP sequences of 16 CdMV isolates from different regions revealed high heterogeneity among them, suggesting that they should be considered members of more than one species.

Keywords cardamom · virus · sequence · variability · identity · phylogeny

Cardamom mosaic virus (CdMV) (genus *Macluravirus*; family, *Potyviridae*), which is associated with mosaic or *katte* disease, is the most destructive virus infecting cardamom. CdMV is widely distributed in cardamom-cultivating regions of Guatemala and India, causing a gradual loss in yield and a complete decline of the plant in 3–5 years. The disease is characterized by the presence of intermittent light green stripes on the leaf lamina and pseudostem, stunting, and reduced vigour [1, 2]. CdMV is primarily transmitted through infected rhizomes, while the secondary spread in a plantation takes place through the banana aphid *Pentalonia caladii* in a non-persistent manner [3, 4]. The virus appears as flexuous filamentous particles of about 650 nm in length and 10–12 nm in diameter with a monopartite positive-sense single-stranded RNA. So far, the complete genome sequence of only one isolate of CdMV (KS) from Sakleshpur, Karnataka, India, has been reported [5]. The first ORF in the viral genome potentially codes for a large polyprotein, which is cleaved to yield nine functional proteins: HC-Pro (helper component protease), P3 (protein 3), 7K (7-kDa peptide), CI (cylindrical inclusion protein), 9K (9-kDa peptide), VPg (viral protein, genome-linked), NIa-Pro (nuclear inclusion a-protease), NIb (nuclear inclusion b), and CP (coat protein). The second putative small ORF called PIPO (pretty interesting *Potyviridae* ORF) located within the P3 cistron is produced by means of a polymerase slippage mechanism. The genome has Vpg linked to its 5′ end and is polyadenylated at its 3′ end [5–7].

Based on symptoms, partial CP sequences, and phylogenetic analysis, CdMV is considered one of the highly variable members of the family *Potyviridae.* Region-wise variability has been seen among isolates. The isolates from Karnataka (except Sirsi) have been found to be more closely related to each other than isolates from Kerala and Tamil Nadu [8, 9]. So far, the complete CP gene sequences of six CdMV isolates from different regions of Karnataka have

Communicated by Ralf Georg Dietzgen

✉ A. I. Bhat
Ishwarabhat.a@icar.gov.in; aib65@yahoo.co.in

[1] Division of Crop Protection, ICAR-Indian Institute of Spices Research, Marikunnu, Kozhikode, Kerala 673012, India

been reported (Supplementary Table S1), while either the complete genome sequence or the complete CP sequence of CdMV isolates from Kerala and Tamil Nadu is lacking. Hence, in the present study, the complete genome sequence of an isolate of CdMV from the Idukki region of Kerala (KI) and the CP gene sequences of four isolates from Kerala, four isolates from Karnataka, and one isolate from Tamil Nadu were determined to study the variability among CdMV isolates.

CdMV-infected leaves showing typical symptoms were collected from the Thiruthali variety of cardamom from Santhanpara, Idukki, Kerala, and used for complete genome sequencing. Similarly, isolates from Nedumkandam, Kattappana, and Pampadumpara in Idukki District and Meppadi in Wayanad District of Kerala, Sakleshpur (two isolates) in Hassan District, Appangala in Kodagu District, and Sirsi in Uttara Kannada District of Karnataka, and Thadiyankudisai in Dindigul District of Tamil Nadu was used for sequencing of the complete CP gene (Supplementary Table S1). Total RNA was extracted from cardamom leaf samples using an RNeasy Plant Mini Kit (QIAGEN, Germany) as per the instructions of the manufacturer. The complete genome of CdMV (KI) (except the 5′ and 3′ ends) was amplified by RT-PCR using eight primer pairs producing overlapping sequences at the ends, designed based on a multiple alignment of the complete genome sequence of the CdMV (KS) isolate (MF622947) and those of other macluraviruses available in the GenBank database (Fig. 1a; Supplementary Table S2) [9]. The primer pair AIB 773 – 44 was used for amplification of the complete CP gene [10]. The 5′ and 3′ ends of the viral genome were amplified using a 5′/3′ RACE Kit, 2nd Generation (Roche Diagnostics GmbH, Germany) as per the instructions of the manufacturer, using a CdMV-specific primer (Fig. 1a, Supplementary Table S2). The resulting amplified products were cloned into the pTZ57R/T vector and sequenced using the facility at Eurofins Pvt Ltd, Bangalore, India.

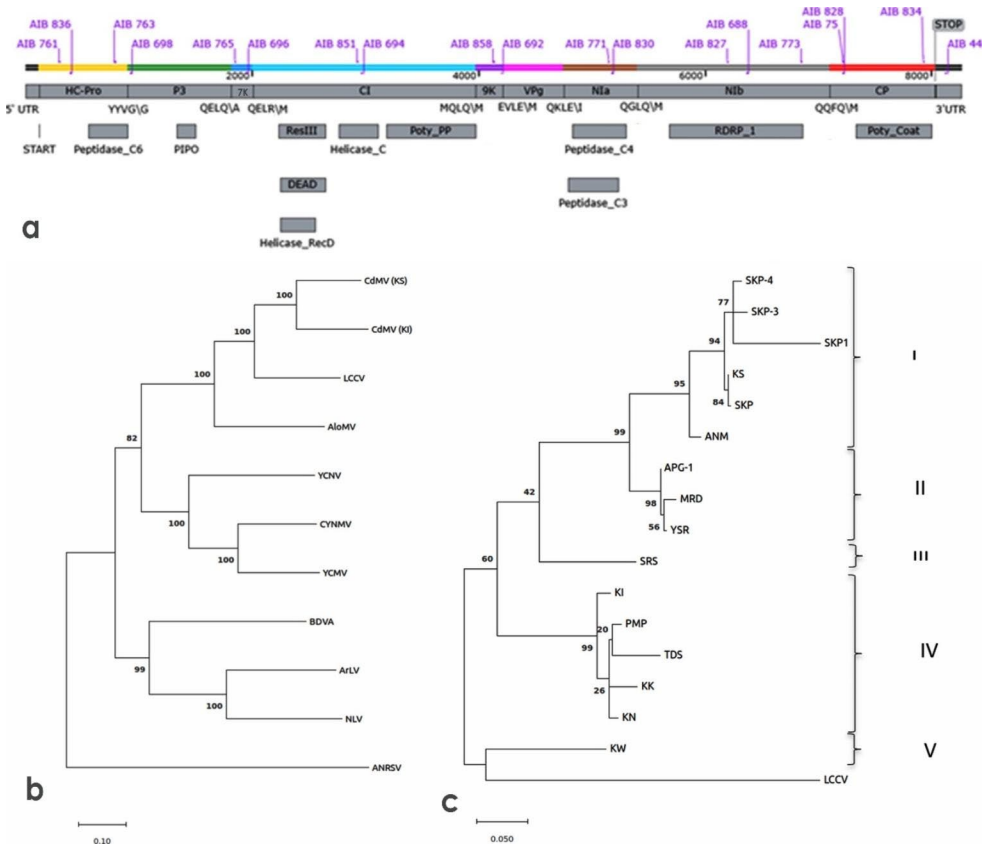

Fig. 1 (**a**) Genome organization of the KI isolate of cardamom mosaic virus. The positions of the mature proteins within the polyprotein are indicated by boxes, and the corresponding predicted protein motifs and cleavage sites are indicated below the boxes. The positions of primers used for the amplification are shown, and details about these primers are provided in Supplementary Table S2. (**b**) Phylogenetic tree based on the complete genome sequence of cardamom mosaic virus (KI) and those of other macluraviruses, constructed in MEGA X by the maximum-likelihood method with the Kimura 2-parameter model and 1000 bootstrap replicates. ANRSV was used as an outgroup. (**c**) Phylogenetic tree based on the complete coat protein amino acid sequences of 16 CdMV isolates, constructed in MEGA X using the maximum-likelihood method with Jones-Taylor-Thornton (JTT) model and 1000 bootstrap replicates. LCCV was used as an outgroup. Details about the isolates and their accession numbers are provided in Supplementary Table S1

CdMV Isolate	MRD	APG-1	YSR	SKP-1	ANM	SKP-3	SKP-4	SKP	KS	KW	SRS	TDS	KI	KK	KN	PMP
MRD		**97**	**98**	79	83	81	82	81	81	70	73	74	74	75	73	74
APG-1	**98**		**99**	79	83	81	82	81	81	70	73	75	74	74	73	73
YSR	**98**	**99**		79	83	81	82	81	81	71	73	75	74	74	73	74
SKP1	80	81	81		**93**	**96**	**95**	**95**	**95**	69	72	71	71	72	71	71
ANM	90	91	90	**87**		**94**	**95**	**96**	**96**	71	73	72	72	73	73	73
SKP-3	88	89	88	**91**	**93**		**97**	**97**	**97**	70	74	73	73	74	73	73
SKP-4	88	89	88	**91**	**94**	**98**		**97**	**97**	72	74	73	73	74	73	73
SKP	87	88	87	**91**	**95**	**97**	**98**		**98**	71	73	73	72	73	72	72
KS	87	88	87	**91**	**95**	**97**	**98**	**100**		71	74	73	73	74	73	73
KW	79	79	79	70	78	77	78	77	77		73	73	73	73	73	73
SRS	82	83	83	74	81	82	82	81	82	77		74	75	75	75	75
TDS	80	81	80	72	78	79	79	79	79	78	80		**89**	**89**	**90**	**89**
KI	79	81	80	72	78	79	79	79	79	78	81	**93**		**95**	**95**	**95**
KK	79	80	80	71	78	79	79	79	79	79	81	**92**	**95**		**95**	**96**
KN	79	80	79	71	78	79	79	79	79	78	80	**94**	**96**	**96**		**96**
PMP	79	80	80	71	78	79	79	79	79	79	80	**94**	**96**	**96**	**98**	

Table 1 Percent identity in the coat protein nucleotide (above the diagonal line) and amino acid (below the diagonal line) sequences of cardamom mosaic virus (KI) and other isolates of CdMV. Details about the isolates and their GenBank accession numbers are provided in Supplementary Table S1.

The sequences that were obtained were compared with those in the GenBank database using BLASTn to confirm the identity of CdMV. The selected sequences were assembled manually based on the consensus sequences of the overlapping regions. Open reading frames (ORFs) in the assembled sequence were identified using the NCBI ORF Finder program (https://www.ncbi.nlm.nih.gov/orffinder/). Translation of the predicted ORF into an amino acid sequence and molecular weight estimation were done using the Expasy Translate and Compute pI/Mw Tool (SIB) (https://web.expasy.org/translate/). Polyprotein sequence identity with CdMV (KS) was confirmed using a BLASTp search. The predicted polyprotein sequence was subjected to protein motif prediction against the Pfam database using the motif search tool (https://www.genome.jp/tools/motif/). The complete genome and polyprotein sequences of CdMV (KS) and other macluraviruses and the CP sequences of other isolates of CdMV used for analysis were obtained from GenBank (Supplementary Table S1). Cleavage sites within the polyprotein were predicted based on comparison with the cleavage sites of CdMV (KS) and other macluraviruses. Annotations of individual predicted cleavage products were confirmed using predicted motifs and BLASTp analysis. A multiple alignment of the complete genome and polyprotein sequences and individual protein sequenced was made, and percent identity calculations for different CdMV isolates and members of the genus *Macluravirus* were done using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). A phylogenetic tree was constructed in MEGA X by the maximum-likelihood method with 1000 bootstrap replicates, using the Kimura 2-parameter model for the complete nucleotide sequence and the Jones-Taylor-Thornton (JTT) model for the CP amino acid sequence. The possibility of recombination among the different isolates of CdMV was investigated using Recombination Detection Program 4 (RDP4) [11].

The complete genome of the KI isolate of CdMV was amplified, sequenced using nine pairs of primers (Fig. 1a, Supplementary Fig. S1), and deposited in the NCBI database with the accession number OR076060. The CdMV

(KI) genome consists of 8255 nucleotides (nt) (GC content, 41%) with two ORFs and has 117 nt in 5′ UTR and 221 nt in the 3′ UTR (Fig. 1a). The first ORF (nt 118–8034) potentially codes for a single large polyprotein of 2638 amino acids (aa) with a molecular weight of 299 kDa. The protein motifs predicted within the polyprotein are Peptidase_C6, ResIII, DEAD, Helicase_RecD, Helicase C, Poty_PP, Peptidase_C3, Peptidase_C4, RdRP_1, and Poty_coat (Fig. 1a, Supplementary Table S3). The eight cleavage sites (YYVG\G, QELQ\A, QELR\M, MQLQ\M, EVLE\M, QKLE\I, QGLQ\M, QQFQ\M) predicted based on comparison with the cleavage sites of CdMV (KS) and other macluraviruses yielded nine cleavage products, namely, HC-Pro (nt 118–903), P3 (nt 904–1815), 7K (nt 1816–2007), CI (nt 2008–3975), 9K (nt 3976–4215), VPg (nt 4216–4752), NIa-Pro (nt 4753–5403), NIb (nt 5404–7101), and CP (nt 7102–8031) (Fig. 1a, Supplementary Tables S4, S5). Compared to the KS isolate, the KI isolate contains an additional 13 nt in the 5′ UTR and 6 nt (2 aa) in the CP region and a deletion of 13 nt in 3′ UTR [5]. A comparison of the cleavage sites of CdMV (KI) with those of CdMV (KS) and other macluraviruses revealed that R in QELR/M and K in QKLE/I are different from the corresponding amino acids in CdMV (KS) and that the amino acid R in QELR/M is unique to CdMV (KI) (Supplementary Table S5). The presence of a second ORF, called PIPO (nt 1332–1502), located within the P3 cistron, coding a small protein with 56 aa (6.9 kDa), was also predicted based on a comparison with CdMV (KS) and other macluraviruses (Fig. 1a, Supplementary Table S4) [5]. Similar to CdMV (KS) and AloMV, PIPO starts with a $C(A)_6$ motif instead of the $G(A)_6$ motif present in most members of the family *Potyviridae* [5, 6, 12]. Among the different proteins of CdMV (KI), 7K shared the most similarity with other macluraviruses, while PIPO shared the least similarity (Supplementary Table S6). Similar results were also observed for CdMV (KS), alpinia oxypylla mosaic virus (AloMV), and large cardamom chirke virus (LCCV) [5, 12, 13].

Potyvirid-specific motifs and regions predicted within the CdMV (KI) genome include GSGKSXXXP (aa 717–725) and DEXH (aa 810–813) motifs in CI (helicase motif), H-X31-D-X61-T-X4-C-X15-H (aa 1587–1702) in NIa-Pro (protease motif), GDD (aa 2155–2157) of RNA-dependent RNA polymerase (RdRp) in NIb (for replicase activity), and conserved tyrosine (aa 1430) in VPg (for attachment of VPg to the genome). Similar to the other macluraviruses, CdMV (KI) lacks the P1 coding region, the N-terminus of HC-Pro, and aphid transmission motifs such as PTK and KITC in HC-Pro and DAG in CP. Macluravirus-specific motifs such as NGTS (aa 2490–2493), WCANNGTSSE (aa 2486–2495), AFDF (aa 2567–2570) and S2493-R2529-D2569 (a viral RNA-binding motif specific to ssRNA filamentous viruses) were also predicted within the CP region of the CdMV (KI) genome [5, 7, 14].

The ICTV species demarcation criteria for the members of the family *Potyviridae* are < 76% identity at the nt level and < 82% at the aa level for the whole genome and < 76% identity at the nt level and < 80% at the aa level in the CP [7]. In the present study, the whole-genome and polyprotein sequences of CdMV (KI) shared 76% and 85% identity, respectively, with CdMV (KS) and 54–70% and 49–75% identity with other macluraviruses, indicating that CdMV (KI) is a strain of CdMV that is highly divergent from CdMV (KS) (Supplementary Table S6). CdMV (KI) and CdMV (KS) formed a distinct subclade within the *Macluravirus* group in the phylogenetic tree constructed based on the complete nt sequence. Sequence comparisons and phylogenetic analysis also showed that, among the macluraviruses, CdMV is closely related to large cardamom chirke virus (Fig. 1b). CdMV (KI) produced intermittent light green stripes on the leaf lamina, in contrast to the continuous dark green stripes along the veins produced by CdMV (KS) [5]. Comparisons of the CP genes suggests that CdMV (KI) may belong to a separate species, as it shares only 73% nt and 79% aa sequence identity with CdMV (KS) and 50–66% nt and 39–64% aa sequence identity with other macluraviruses (Supplementary Table S6). In order to examine the heterogeneity among different isolates of CdMV, we determined the complete CP sequence of nine CdMV isolates (four each from Kerala and Karnataka and one from Tamil Nadu) and deposited them in the GenBank database under the accession numbers OR076051 to OR076059 (Supplementary Table S1). A multiple alignment of the CP sequences of these nine isolates together with those of CdMV (KI) and six isolates from Karnataka (SKP1, SKP-3, MRD, YSR, ANM, and KS) available in the GenBank database (Supplementary Table S1) revealed the occurrence of a 2-aa deletion in all of the Karnataka isolates except SRS at aa position 72–73 and a 4-aa deletion in SRS at aa position 3 to 6 (Supplementary Fig. S2). Analysis using RDP4 indicated that among the different CdMV isolates, the Anemahal (ANM) isolate is a recombinant of the SKP isolate (major parent) and the Margodu (MRD) isolate (minor parent), and these results were supported by all seven detection methods used.

Comparison of the CP sequences of all 16 CdMV isolates showed that CdMV (KI) is closely related to isolates from Idukki and Dindigul (89–96% identity at the nt level and 92–98% at the aa level), while only two isolates from Kodagu and the SRS isolate from Uttara Kannada showed > 80% identity at the aa level. The remaining isolates showed < 75% nt and < 80% aa sequence identity, indicating a high degree of heterogeneity among the isolates (Table 1). Sequence comparisons and phylogenetic analysis using complete CP sequences grouped the CdMV isolates

into five clusters corresponding to the geographical origin of the isolates (Fig. 1c, Supplementary Fig. S3, Table 1). A high level of sequence similarity was observed among the isolates within each cluster (>89% and >87% nt and aa sequence identity), indicating the homogeneity of isolates within a region. Notably, the isolates did not group according to the symptoms they caused, as variability in the symptoms was observed among CdMV isolates both within and between clusters. The only isolate from cluster 5 that was sequenced (KW) was the most divergent, sharing only 69 to 73% nt and 70 to 79% aa sequence identity with the remaining isolates of CdMV, indicating that it may belong to a new species. The study revealed a high degree of sequence variability among CdMV isolates collected from different regions of India. The CdMV isolates within a region were more closely related to each other than to isolates from other regions, irrespective of the symptoms with which they were associated, indicating independent evolution of the virus within a region. Complete genome sequences of more isolates of CdMV and their biological characterization are needed to resolve the taxonomy and classify different CdMV isolates.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00705-023-05879-3.

Acknowledgments The authors are thankful to the Head of the Division of Crop Protection and the Director of the ICAR-Indian Institute of Spices Research, Kozhikode, Kerala, India, for facilities. The first author is thankful to the University Grants Commission (UGC) for a Junior Research Fellowship.

Authors' contributions A.I. Bhat conceptualized, designed experiments, and finalized the manuscript. M. Greeshma performed all of the experiments and wrote the manuscript. Both authors have read and approved the final manuscript.

Funding Funding was received from the Indian Council of Agricultural Research, New Delhi.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval The present research did not involve any experimentation on humans or animals.

References

1. Jacob T, Usha R (2001) 3′-Terminal sequence analysis of the RNA genome of the Indian isolate of Cardamom mosaic virus: a new member of genus *Macluravirus* of *Potyviridae*. Virus Genes 23(1):81–88. https://doi.org/10.1023/a:1011191614839
2. Bhat AI, Biju CN, Srinivasan V, Ankegowda SJ, Krishnamurthy KS (2018) Current status of viral diseases affecting black pepper and cardamom. J Spices Aromat Crops 27(1):1–16. https://doi.10.25081/josac.2018.v27.i1.1009
3. Siddappaji C, Reddy KNRN (1972) A note on the occurrence of the aphid, *Pentalonia nigronervosa* form *caladii* Vander Goot,(Aphididae-Homoptera) on cardamon (*Elettaria cardamomum* Salisb). Mysore J Agr Sci 6(2):194–195
4. Venugopal MN (1995) Viral diseases of cardamom (*Elettaria cardamomum* Maton) and their management. J Spices Aromat Crops 4:32–39
5. Elangovan S, Srikakulam N, Pandi G, Jacob T, Ramakrishnan U, Madhugiri Narayanarao V, Tennyson J (2019) The first complete genomic sequence of cardamom mosaic virus, a member of the genus *Macluravirus* (family *Potyviridae*). Arch Virol 164(6):1723–1726. https://doi.org/10.1007/s00705-019-04203-2
6. Chung BYW, Miller WA, Atkins JF, Firth AE (2008) An overlapping essential gene in the *Potyviridae*. PNAS 105:5897–5902. https://doi.org/10.1073/pnas.0800468105
7. Inoue-Nagata AK, Jordan R, Kreuze J, Li F, Mäkinen K, Wylie SJ (2022) ICTV virus taxonomy profile: *Potyviridae* 2022. J Gen Virol 103:001738. https://doi.org/10.1099/jgv.0.001738
8. Jacob T, Jebasingh T, Venugopal MN, Usha R (2003) High genetic diversity in the coat protein and 3′ untranslated regions among geographical isolates of *Cardamom mosaic virus* from south India. J Biosci 28:589–595. https://doi.org/10.1007/BF02703334
9. Siljo A, Bhat AI, Biju CN (2013) Analysis of variability in *Cardamom mosaic virus* isolates occurring in India using symptomatology and coat protein gene sequence. J Spices Aromat Crops 22(1):70–75
10. Biju CN, Siljo A, Bhat AI (2010) Survey and RT-PCR based detection of *Cardamom mosaic virus* affecting small cardamom in India. Indian J Virol 21(2):148–150. https://doi.org/10.1007/s13337-011-0026-3
11. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol 1(1). https://doi.org/10.1093/ve/vev003
12. Hu W, Li Z, Wang X, Liu W, Huang C, Miao W, Cui H (2018) Complete genomic sequence of a novel macluravirus, alpinia oxyphylla mosaic virus (AloMV), identified in *Alpinia oxyphylla*. Arch Virol 163:2579–2582. https://doi.org/10.1007/s00705-018-3879-6
13. Sidharthan VK, Sharma SK, Baranwal VK (2021) The first near-complete genome sequence of large cardamom chirke virus mined from the transcriptome dataset of large cardamom. Plant Gene 28:100324. https://doi.org/10.1016/j.plgene.2021.100324
14. Zamora M, Méndez-López E, Agirrezabala X, Cuesta R, Lavín JL, Sánchez-Pina MA, Aranda MA, Valle M (2017) Potyvirus virion structure shows conserved protein fold and RNA binding site in ssRNA viruses. Sci Adv 3(9):eaao2182. https://doi.org/10.1126/sciadv.aao2182