

## SUPPLEMENTARY INFORMATION

### Transcriptional analysis of CRISPR I-B arrays of *Leptospira interrogans* serovar Lai and its processing by Cas6

Aman Prakash and Manish Kumar#

Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati,

Guwahati-781039, Assam, India

#corresponding author: Manish Kumar

Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati,

Guwahati-781039, Assam, India

Email: mkumar1@iitg.ac.in

Phone: +91-361-258-2230

Fax: +91-361-258-2249

## METHODS

### Bioinformatics analysis

Information of predicted CRISPRs (**Table S1**) in the I-B locus of *L. interrogans* sv. Copenhageni and sv. Lai was retrieved via a taxonomic-based search in the CRISPRCasdb database, a repository built with CRISPRCasFinder [1]. Nucleotide sequences of interCRISPR regions at the hypervariable region of sv. Lai's genome was extracted from the NCBI nucleotide database.

Multiple sequence alignments of interCRISPR regions were performed using the MUSCLE program [2], and graphic images of aligned sequences were obtained using the ESPript program (version 3.0) [3]. Sequences of repeat-spacer segments retrieved from the redefined CRISPR arrays of sv. Lai (LA\_Cr<sup>6-12</sup>) are tabulated in **Table S2**.

### **RT-PCR**

PCR was performed using a cDNA or negative control (RNA) template prepared in reverse transcription (random hexamers) with or without reverse transcriptase, respectively. Primer pairs selected to use in PCR were specific to spacer regions, as mentioned in **Table S3**. PCR products were resolved on 2% agarose gel.

**Table S1. CRISPRs predicted by the CRISPRCasdb in the I-B locus of *L. interrogans* serovars Copenhageni (strain Fiocruz L1-130) and Lai (strain 56601)**

| <i>L. interrogans</i><br>serovar<br>(strain) | CRISPR (I-B)<br>Id                      | Coordinate<br>(start-end) | Spacer | Repeat consensus                             | CRISPR<br>(I-B)<br>Direction | Evidence<br>level | <i>cas</i> (I-B)<br>ORFs<br>Direction |
|----------------------------------------------|-----------------------------------------|---------------------------|--------|----------------------------------------------|------------------------------|-------------------|---------------------------------------|
| Copenhageni<br>(Fiocruz L1-130)              | AE016823.1_2<br>(LIC_Cr <sup>2</sup> )  | 1,133,848-<br>1,134,101   | 3      | GTGCTCAACGCCTA<br>ACGGCATCAAAGT<br>TATATTCAG | ND                           | 4                 | –                                     |
| Lai (56601)                                  | AE010300.2_6<br>(LA_Cr <sup>6</sup> )   | 3,163,254-<br>3,163,495   | 3      | CTGAATATAACTTT<br>GATGCCGTTAGGC<br>G         | –                            | 4                 | +                                     |
|                                              | AE010300.2_7<br>(LA_Cr <sup>7</sup> )   | 3,163,731-<br>3,163,973   | 3      | CTGAATATAACTTT<br>GATGCCGTTAGGC<br>G         | –                            | 4                 |                                       |
|                                              | AE010300.2_8<br>(LA_Cr <sup>8</sup> )   | 3,164,138-<br>3,164,239   | 1      | CTGAATATAACTTT<br>GATGCCGTTAGGC<br>G         | –                            | 4                 |                                       |
|                                              | AE010300.2_9<br>(LA_Cr <sup>9</sup> )   | 3,164,476-<br>3,164,719   | 3      | CTGAATATAACTTT<br>GATGCCGTTAGGC<br>G         | –                            | 4                 |                                       |
|                                              | AE010300.2_10<br>(LA_Cr <sup>10</sup> ) | 3,164,839-<br>3,165,154   | 4      | CTGAATATAACTTT<br>GATGCCGTTAGGC<br>G         | –                            | 3                 |                                       |
|                                              | AE010300.2_11<br>(LA_Cr <sup>11</sup> ) | 3,165,387-<br>3,165,487   | 1      | CTGAATATAACTTT<br>GATGCCGTTAGGC<br>G         | –                            | 4                 |                                       |
|                                              | AE010300.2_12<br>(LA_Cr <sup>12</sup> ) | 3,165,651-<br>3,165,821   | 2      | CTGAATATAACTTT<br>GATGCCGTTAGGC<br>G         | –                            | 4                 |                                       |

ND: not defined

**Table S2. Repeats and spacers of redefined CRISPR I-B arrays at the hypervariable region of serovar Lai**

| CRISPRs             | Repeats (5'-3')                                                                                                                                                                                                         | Spacers (5'-3')                                                                                                                                                              |
|---------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| LA_Cr <sup>6</sup>  | R1: CTGAATATAACTTTGATGCCGTTAGGCGTTGAGCAC<br>R2: CTGAATATAACTTTGATGCCGTTAGGCGTTGAGCAC<br>R3: CTGAATATAACTTTGATGCCGTTAGGCGTTGAGCAC<br>R4: CTGAATATAACTTTGATGCCGTTAGGCGGTTAGAA                                             | S1: CCGTTCTGATTTTTTCTTTTCCTTCCTTTTGTAT<br>S2: ACCCCACGATACTACCTGTCAGACCGTGCCCGGAT<br>S3: GCACTCCTCGAACTGGTAAAACTACCGATGCTCGC                                                 |
| LA_Cr <sup>7</sup>  | R1: CTTACAAAAATCGGGTGCCGTTAGGCGTTGAGCAC<br>R2: CTGAATATAACTTTGATGCCGTTAGGCGTTGAGCAC<br>R3: CTGAATATAACTTTGATGCCGTTAGGCGTTGAGCAC<br>R4: CTGAATATAACTTTGATGCCGTTAGGCGTTGAGCAC<br>R5: CTGAATATAACTTTGATGCCGTTAGGCGATTAGAT  | S1: TTGATTGGTGCAGTTGTGCTTGTTTTGATTGGTGT<br>S2: CCGTTCTGATTTTTTCTTTTCCTTCCTTTTGTAT<br>S3: CAGTAGATTTTGGATACACAAACCGTTTGTGTTTC<br>S4: GAATACAACCTCTTCAAAAAAGAGGAACCGGCGTA      |
| LA_Cr <sup>8</sup>  | R1: CTTACAAAAATCGGGATGCCGTTAGGCGTTGAGCAC<br>R2: CTGAATATAACTTTGATGCCGTTAGGCGGTTAGAA                                                                                                                                     | S1: AAGGAAACAGGTCTCTTAAGTTTAGCACCTGCTGCAG                                                                                                                                    |
| LA_Cr <sup>9</sup>  | R1: CTTACAAAAATCGGGATGTCGGTAGGCGTTGAGTAC<br>R2: CTGAATATAACTTTGATGCCGTTAGGTGTTGAGCAC<br>R3: CTGAATATAACTTTGATGCCGTTAGGCGTTGAGCAC<br>R4: CTGAATATAACTTTGATGCCGTTAGGCGTTGAGCAC<br>R5: CTGAATATAACTTTGATGCCGTTAGGCGGTTAGAA | S1: TGAGTATGCAAATGAGCTTCGGCTTCGAATTCCT<br>S2: AAAGACAAATCGGTTTCATTGATTTTTTCGGATCT<br>S3: GAGTTAAATTCTGCCCACTCCATGGCCTAAATCAG<br>S4: AGGGGCTATAAAATTGAGGTATCCTCAAAACGTAAA     |
| LA_Cr <sup>10</sup> | R1: CTTACAAAAATCGGGATGCCGTTAGGCGTTGAGCAC<br>R2: CTGAATATAACTTTGATGCCGTTAGGCGTTGAGCAC<br>R3: CTGAATATAACTTTGATGCCGTTAGGCGTTGAGCAC<br>R4: CTGAATATAACTTTGATGCCGTTAGGCGTTGAGCAC<br>R5: CTGAATATAACTTTGATGCCGTTAGGCGATTAGAC | S1: GAATAACTCGTTCGAAAGCGTTCTGCGGATCTT<br>S2: TCGTAAAGATCGTCTGCGTGTTTCGTCGTGATACGTGT<br>S3: AGCATAGCGGACGTGTTCTTTGTTTCGTTTTATCGAA<br>S4: AGCGTGCAACAAAACACGGAATACAATTCTCTTTGC |
| LA_Cr <sup>11</sup> | R1: CTTACAAAAATCGGGTGCCGTTAGGCGTTGAGCAC<br>R2: CTGAATATAACTTTGATGCCATTAGGCGTTGAGCAC<br>R3: CTGAATATAACTTTGATGCCGTTAGGCGATTAGAC                                                                                          | S1: CACAACCGTGACAAATATTTGAAATCGTTCTGACT<br>S2: GTGCCTTGAGAGACCTTAAGTTTATCCGCGACATTG                                                                                          |
| LA_Cr <sup>12</sup> | R1: CTTACAAAAATCGGGATGCCGTTAGGCGTTGAGCAC<br>R2: CTGAATATAACTTTGATGCCGTTAGGCGTTGAGCAC<br>R3: CTGAATATAACTTTGATGCCGTTAGGCGATTAGAC                                                                                         | S1: ACCCGGTTTGCATTTACCGAAGTCCAAATCAATTTC<br>S2: TAAGAGTAAGTGAGGCTAGATTTAGAGGGATTGT                                                                                           |

**Table S3. Primer pairs used in the RT-PCR experiment**

| <b>Primer pair</b>                                                                                                                 | <b>Sequence (5'-3')</b>                          | <b>Region amplified</b>                      |
|------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------|----------------------------------------------|
| LA_Cr <sup>6</sup> S1 forward ( <sup>6</sup> S1 <sub>f</sub> )<br>LA_Cr <sup>7</sup> S4 reverse ( <sup>7</sup> S4 <sub>r</sub> )   | CCGTTCTGATTTTTCTTTTCCT<br>TACGCCGGTTCCTCTTTTTTG  | LA_Cr <sup>6</sup> S1 to Cr <sup>7</sup> S4  |
| LA_Cr <sup>7</sup> S2 forward ( <sup>7</sup> S2 <sub>f</sub> )<br>LA_Cr <sup>9</sup> S4 reverse ( <sup>9</sup> S4 <sub>r</sub> )   | CCGTTCTGATTTTTCTTTTCCT<br>TTTACGTTTTGAGGATACCTCA | LA_Cr <sup>7</sup> S2 to Cr <sup>9</sup> S4  |
| LA_Cr <sup>9</sup> S2 forward ( <sup>9</sup> S2 <sub>f</sub> )<br>LA_Cr <sup>10</sup> S4 reverse ( <sup>10</sup> S4 <sub>r</sub> ) | AAAGACAAATCGGTTCAATTGA<br>TTTACGTTTTGAGGATACCTCA | LA_Cr <sup>9</sup> S2 to Cr <sup>10</sup> S4 |

## REFERENCES

- [1] C. Pourcel, M. Touchon, N. Villeriot, J.-P. Vernadet, D. Couvin, C. Toffano-Nioche, G. Vergnaud, *Nucleic acids research*, 48 (2020) D535-D544.
- [2] F. Madeira, Y.M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A.R.N. Tivey, S.C. Potter, R.D. Finn, R. Lopez, *Nucleic acids research*, 47 (2019) W636-W641.
- [3] X. Robert, P. Gouet, *Nucleic acids research*, 42 (2014) W320-W324.

## LEGENDS TO SUPPLEMENTARY FIGURES

**Fig. S1. Analysis of interCRISPR regions at the hypervariable region of serovar Lai.** (A) The architecture of subtype I-B array locus in serovar Lai. CRISPRs identified in hypervariable (the region between *cas2/LA3182* and *cas6/LA3189*) of serovar Lai via the CRISPRCasdb database are drawn in the direction of CRISPR-Cas (5'-3'). CRISPRs and *cas* genes are shown in the architecture to the scale, except for flanking regions. Black and unique color-filled rectangles represent similar repeats and spacer regions, respectively. Length of CRISPR and flanking regions are indicated by vertical dashed lines (grey) and numbers (in bp) given at the apex of the double arrowhead over the architecture. (B) Multiple sequence alignment (MSA) of regions flanked by two adjacent CRISPRs. Nucleotide sequences flanked by LA\_Cr<sup>6</sup> and <sup>7</sup>, LA\_Cr<sup>7</sup> and <sup>8</sup>, LA\_Cr<sup>8</sup> and <sup>9</sup>, LA\_Cr<sup>9</sup> and <sup>10</sup>, and LA\_Cr<sup>11</sup> and <sup>12</sup> were aligned. Grey and unique colored regions at the 3' terminus of alignment represent the first repeat-spacer units (redefined now) of LA\_Cr<sup>7</sup>, Cr<sup>9</sup>, and Cr<sup>11</sup>. Red and yellow colored nucleotides in the alignments represent 100% and more than 70% conservation, respectively, at that position among whole sequences. Consensus sequences of each MSA were shown below the alignment where upper and lower case denotes conserved and semi-conserved nucleotides, respectively. Bold letter code in the alignments indicates consensus nucleotide. Dot in consensus sequences indicates no nucleotide conservation at that particular position.

**Fig. S2. RT-PCR of CRISPR I-B of serovar Lai.** (A) Schematic representation of primer's position used to amplify cDNA of serovar Lai. The primers used in RT-PCR analysis and corresponding annealing regions are indicated over the graphic for clarity. LA\_Cr<sup>6</sup> to LA\_Cr<sup>12</sup>, including CRISPR flanking regions, are drawn to scale in the direction of CRISPR-Cas I-B (5'-3'). Similar repeats and distinct spacers in CRISPRs are represented by black and unique color-filled rectangles, respectively. Vertical dashed lines (grey) and numbers (in bp) given at the apex of the double arrowhead over the architecture indicate DNA regions and corresponding lengths that could be amplified in RT-PCR. (B) Identification of transcription from one CRISPR to another in serovar Lai. The primer pairs aforementioned were used in PCR with cDNA (right panel) or a control template (right panel) prepared using random hexamers. PCR products were resolved on 2% agarose gel. 'M1' denotes the DNA marker used for the size estimation of PCR-amplified fragments.