

## SUPPLEMENTARY INFORMATION

### **Interrogating two extensively self-targeting Type I CRISPR-Cas systems in *Xanthomonas albilineans* reveals distinct anti-CRISPR proteins that block DNA degradation**

Franziska Wimmer, Frank Englert\*, Katharina G. Wandera\*, Omer S. Alkhnbashi, Scott P.

Collins, Rolf Backofen, Chase L. Beisel<sup>@</sup>

\*Equal contributions

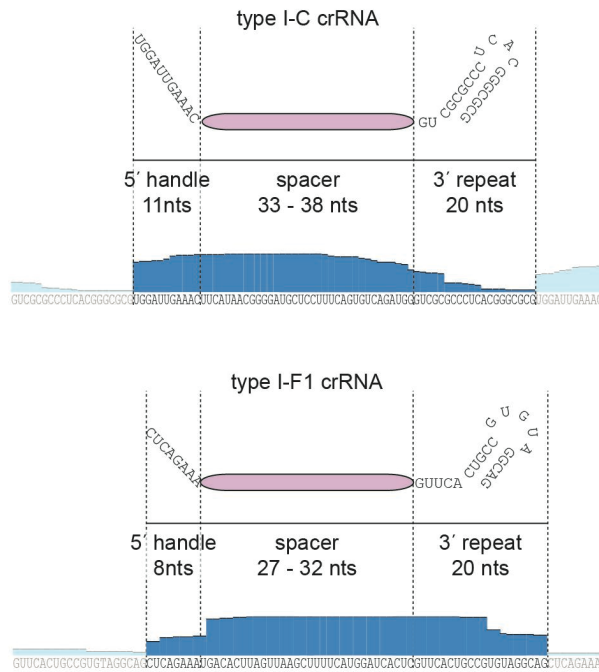
<sup>@</sup>Correspondence: [chase.beisel@helmholtz-hiri.de](mailto:chase.beisel@helmholtz-hiri.de) (to C.L.B.)

## SUPPLEMENTARY FIGURES

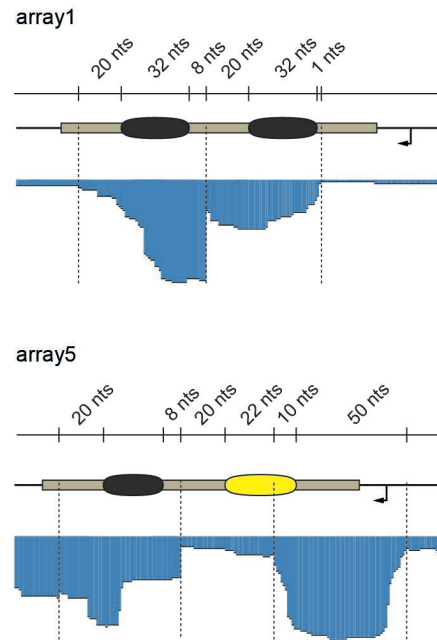
A

| gene ID   | Name           | TPM | gene ID   | Name             | TPM | gene ID   | Name        | TPM |
|-----------|----------------|-----|-----------|------------------|-----|-----------|-------------|-----|
| XALC_2885 | IC <i>cas3</i> | 31  | XALC_3048 | IF <i>cas1</i>   | 10  | XALC_0251 | <i>birA</i> | 93  |
| XALC_2887 | IC <i>cas5</i> | 910 | XALC_3049 | IF <i>cas2-3</i> | 7   | XALC_0807 | <i>mreD</i> | 44  |
| XALC_2888 | IC <i>cas8</i> | 94  | XALC_3050 | IF <i>cas8</i>   | 324 | XALC_0808 | <i>mrdA</i> | 86  |
| XALC_2889 | IC <i>cas7</i> | 146 | XALC_3051 | IF <i>cas5</i>   | 115 | XALC_1005 | <i>nuoG</i> | 964 |
| XALC_2890 | IC <i>cas4</i> | 33  | XALC_3052 | IF <i>cas7</i>   | 87  | XALC_1665 | <i>purB</i> | 55  |
| XALC_2891 | IC <i>cas1</i> | 27  | XALC_3053 | IF <i>cas6</i>   | 118 | XALC_2038 | <i>lpxB</i> | 116 |
| XALC_2892 | IC <i>cas2</i> | 89  |           |                  |     | XALC_2040 | <i>dnaE</i> | 149 |
|           |                |     |           |                  |     | XALC_2768 | <i>thyA</i> | 326 |
|           |                |     |           |                  |     | XALC_2997 | <i>hemB</i> | 130 |
|           |                |     |           |                  |     | XALC_3172 | <i>mpA</i>  | 77  |

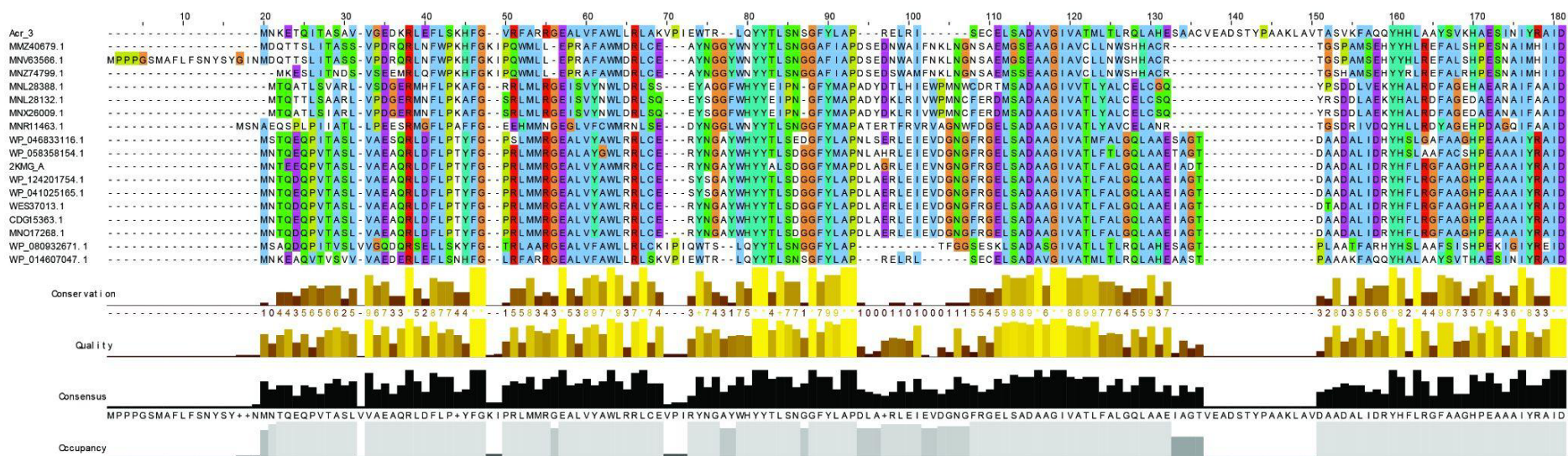
B



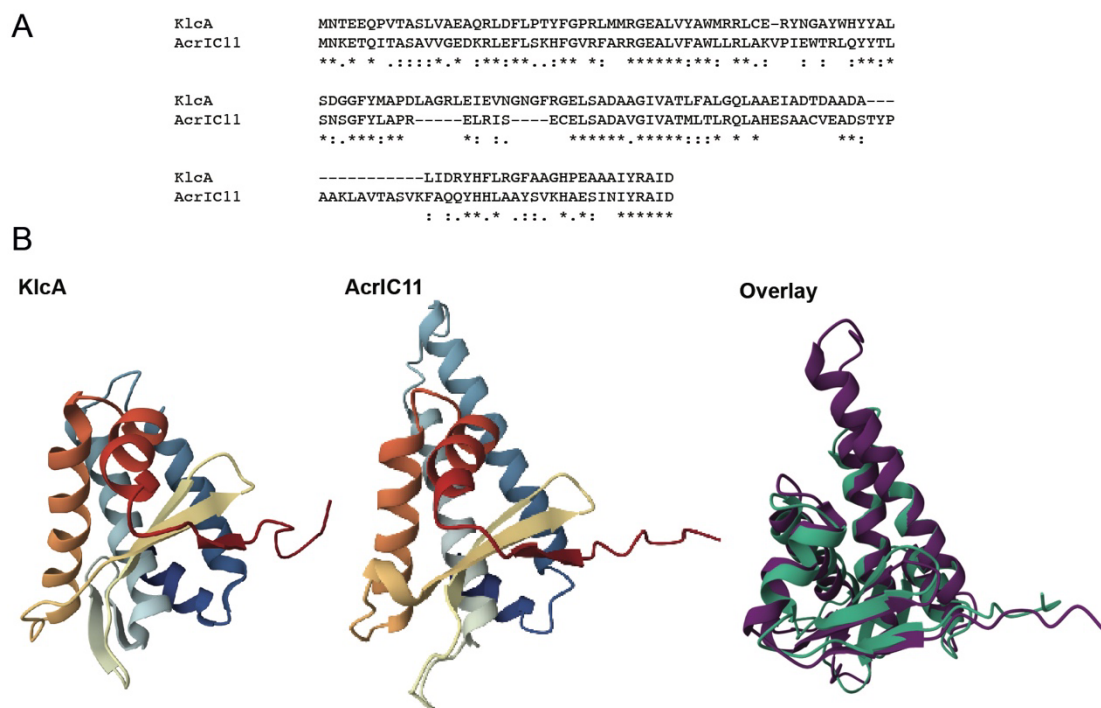
C



**Figure S1:** RNA-Seq reveals transcription of *X. albilineans* *cas* genes and crRNA processing. **(A)** Transcripts per million calculated from total RNA-Seq results. TPM of type I-C and type I-F1 *cas* genes are shown. Right table: essential genes shown for comparison. **(B)** Representative crRNAs undergoing standard processing. Expected mature crRNAs for the *X. albilineans* type I-C and type I-F1 arrays are shown. **(C)** Atypical processing for crRNAs derived from CRISPR array 1 and array 5. Small RNA-Seq coverage of type I-F1 array 1 and array 5 are shown in blue. Self-targeting spacers are depicted in yellow ovals and non self-targeting spacers are represented in black ovals. Dashed lines display the location of the processing event.



**Figure S2:** Amino-acid conservation of identified Acr\_3 homolog.



**Figure S3.** Comparison of the predicted structure of AcrIC11 with the crystal structure of KlcA. **(A)** Amino acid sequence alignment between KlcA and AcrIC11 using MUSCLE. **(B)** Structures of KlcA and AcrIC11. The published structure of KlcA is derived from crystallography (2KMG) (1), while the structure of AcrIC11 was predicted using AlphaFold (2, 3). The overlaid structure was generated using TM-align (4) and indicated strong similarity (RMSD = 1.467 Å, TM = 0.79111).

## SUPPLEMENTARY TABLES

**Table S1:** List of plasmids used in the work. See the attached Excel file.

**Table S2:** Genomic location of predicted prophage regions. Prophage regions are predicted with VirSorter v1.0.3 (5), Prophage Hunter (6) and PHASTER (7, 8).

| Used tool       | Genomic region      | Comment             |
|-----------------|---------------------|---------------------|
| VirSorter       | 204,842-292,722     | prophage category 5 |
| Prophage Hunter | 217,510-251,216     | active              |
| PHASTER         | 251,524-271,683     | incomplete          |
| VirSorter       | 1,197,275-1,218,686 | prophage category 6 |
| Prophage Hunter | 1,691,999-1,713,262 | ambiguous           |
| VirSorter       | 1,696,720-1,861,289 | prophage category 5 |
| PHASTER         | 1,790,323-1,798,253 | intact              |
| Prophage Hunter | 2,174,815-2,185,313 | ambiguous           |
| Prophage Hunter | 2,830,643-2,860,815 | ambiguous           |
| PHASTER         | 2,837,733-2,874,463 | questionable        |
| PHASTER         | 2,841,444-2,856,796 | incomplete          |
| PHASTER         | 3,098,545-3,132,674 | questionable        |
| Prophage Hunter | 3,102,291-3,128,520 | ambiguous           |
| PHASTER         | 3,107,687-3,129,023 | incomplete          |
| Prophage Hunter | 3,120,278-3,132,701 | ambiguous           |

**Table S3:** Genomic location of putative Acrs and their amino acid sequence. TPM: transcripts per million based on RNA-seq analyses.

| Acr cand. | Genomic location                        | TPM     | AA sequence   | Comment   |
|-----------|---|---------|---|---|
| Acr_1     | 265340-265717                           | 316     | MNYMKKWIREHVAEVIKANELSRWVDDS<br>DMKFAMYVVECGQGAQLAQDVGREIGN<br>ETIVAIAQTVIDTIDEVSRGGTPRTRSRRKI<br>TDKQRHVLAVVLEKEYGTARGIAAAGWG<br>LTDEEIDNADV*                            | Identified by HMM profile search                                      |
| Acr_2     | 2852041-<br>2852406;3116105-<br>3116470 | 0;<br>0 | MPRKAPTPCRHPGCGKLVSDGSGYCAD<br>HQRDKVGWHKDRRNAHQRGYGATWQK<br>LRAFVMQRDQGLCQPCQSGRLTPAVA<br>VDHIVPKSQGGTDHPNNCQAICHRCHVL<br>KTAQESHQGREGA*                                 | Identified by HMM profile search                                      |
| Acr_3     | plasmII: 16827-<br>17273                | 55      | MNKETQITASAVVGEDKRLEFLSKHFGV<br>RFARRGEALVFAWLLRLAKVPIEWTRLQ<br>YYTLSNSGFYLAPRELRISECELSADAVGI<br>VATMLTLRQLAHESAACVEADSTYPAAK<br>LAVTASVKFAQQYHHLAAYSVKHAESINI<br>YRAID* | Identified by guilt-by-association                                    |
| Acr_4     | plasmIII: 7731-<br>7997                 | 0       | MSTLTVTTRGQVTFRAEVLQHLGIKPGEK<br>IEVYLMPDGRAELKAAKPKGSFRELRGIL<br>KHKTNGARLSIEEINDAIAEAGDAAGTGN<br>T*   | Identified by HMM profile search                                      |
| Acr_5     | 2852137-2852406;<br>3116201-3116470     | 0;<br>0 | MVGWHKDRRNAHQRGYGATWQKLRAF<br>VMQRDQGLCQPCQSGRLTPAVAVDHI<br>VPKSQGGTDHPNNCQAICHRCHVLKTAQ<br>ESHQGREGA*  | Identified by HMM profile search,<br>N-terminal M added; part of Acr2 |
| Acr_6     | 2852659-2852865                         | 4       | MQSQKTARPLNFSRVQNEKLFLDRTVSE<br>LSVARDYKADLAQIEQIDATPWTAAASHAD<br>MTSELKTYARS*  | Identified by HMM profile search                                      |
| Acr_7     | 2852734-2852865                         | 0       | MVSELSVARDYKADLAQIEQIDATPWTAA<br>SHADMTSELKTYARS*   | Identified by guilt-by-association, N-terminal M added; part of Acr6  |

|        |   |                                 |   |   |
|--------|---|---------------------------------|---|---|
| Acr_8  | plasmIII: 8405-8809   | 45                              | MNKEAQVTVSVVVAEDERLEFLSNHFGL<br>RFARRGEALVFAWLLRLSKVPIEWTRLQ<br>YYTLSNSGFYLAPRELRLSECELSADAVG<br>IVATMLTLRQLAHEAASTPAAAKFAQQYH<br>ALAAYSVTHAESINIYRAID* | Identified by<br>HMM profile<br>search  |
| Acr_9  | plasmIII: 8893-9117   | 0                               | MRVFNIAEIEFAINYWRTRIVPDDGALMC<br>APALSLLQLYGHMIFDRIEAVPESELDAE<br>QGVALSVALYQHELPL*   | Identified by<br>HMM profile<br>search  |
| Acr_10 | 3116786-3116941   | 188                             | LEKSVSELSAARDYKADLAQIGQIDVTPW<br>TAAAHADMTPAEPVELEPYARS*  | Identified by<br>guilt-by-<br>association   |
| Acr_11 | 2852137-2852406;<br>3116201-3116470   | 0;<br>0                         | VGWHKDRRNAHQRGYGATWQKLRAFV<br>MQRDQGLCQPCQSGRLTPAVAVDHIV<br>PKSQGGTDHPNNCQAICHRCHVLKTAQE<br>SHQGREGA*   | Identified by<br>HMM profile<br>search,<br>Acr5 without N-<br>terminal M; part<br>of Acr2 |
| Acr_12 | 2845745-2845960   | 0                               | MVDAKHAAAALRLPYWFSQAMRNKY<br>RIPHYLLGGLVRYRLSELSAWAARSTLVQ<br>RSDTSNVGTSTEEAE*  | Identified by<br>HMM profile<br>search  |
| Acr_13 | 2845431-2845706   | 1                               | MNLITSLRHKLSYLYGEHLPNEIHYHRAD<br>GQHVVVALQDATVDQLAFAIQTINTESVA<br>LSRHRNALEELHTEVRKRSACGADRIAD<br>VAWDN*  | Identified by<br>HMM profile<br>search  |
| Acr_14 | 2844724-2844861   | 2                               | MADGSAPLPSLTTLPPRDHAMRSLDEFV<br>RVDDGRNHKPAHKSRHT*  | Identified by<br>guilt-by-<br>association   |
| Acr_15 | 1430699-1431022;<br>1698203-1698526;<br>1917087-1917410;<br>2197118-2197441;<br>2828433-2828756;<br>3118350-3118673 | 0;<br>2;<br>0;<br>6;<br>2;<br>3 | MSKSNKFSPEVRERAVRMVQEQRGEYQ<br>SLWAAIESIAPKIGCVPQTLNEWVKRAEV<br>DAGAREGVTSSAQRMKELEREVKELR<br>RANEILKLASAFFAQAELDRRLKS*                                 | Identified by<br>HMM profile<br>search  |
| Acr_16 | 2845054-2845317   | 1                               | MTTRLPATQIGQLCESKDPGSTTRIALDE<br>SELAARWGLSVKTLRRWRQEQLGPVFC<br>KLGARVTYLICDVEAFEQRVSRYSTFARA<br>YP*  | Identified by<br>HMM profile<br>search  |
| Acr_17 | 2773986-2774279;<br>3111383-3111676   | 0;<br>0                         | MQRITRRRYTDDFKAQAIALAESVGLAKA<br>ARQLGMSVKTLANWLGASRGGQPLSSP<br>SRKPVSEMESELARLRAENATLKMEREIL<br>KKATAFFARESK*  | Identified by<br>HMM profile<br>search  |

**Table S4:** Features of the self-targeting spacers. See the attached Excel file.

**Table S5:** Mapped read counts for the two biological replicates for the expressed crRNAs. See the attached Excel file.

**Table S6:** Average mapped read counts for all genes in the major predicted prophage regions. See the attached Excel file.



## SUPPLEMENTARY REFERENCES

1. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
2. Mirdita,M., Schütze,K., Moriwaki,Y., Heo,L., Ovchinnikov,S. and Steinegger,M. (2022) ColabFold: making protein folding accessible to all. *Nat. Methods*, **19**, 679–682.
3. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Židek,A., Potapenko,A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.
4. Zhang,Y. and Skolnick,J. (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.*, **33**, 2302–2309.
5. Roux,S., Enault,F., Hurwitz,B.L. and Sullivan,M.B. (2015) VirSorter: mining viral signal from microbial genomic data. *PeerJ*, **3**, e985.
6. Song,W., Sun,H.-X., Zhang,C., Cheng,L., Peng,Y., Deng,Z., Wang,D., Wang,Y., Hu,M., Liu,W., *et al.* (2019) Prophage Hunter: an integrative hunting tool for active prophages. *Nucleic Acids Res.*, **47**, W74–W80.
7. Arndt,D., Grant,J.R., Marcu,A., Sajed,T., Pon,A., Liang,Y. and Wishart,D.S. (2016) PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.*, **44**, W16–21.
8. Zhou,Y., Liang,Y., Lynch,K.H., Dennis,J.J. and Wishart,D.S. (2011) PHAST: a fast phage search tool. *Nucleic Acids Res.*, **39**, W347–52.