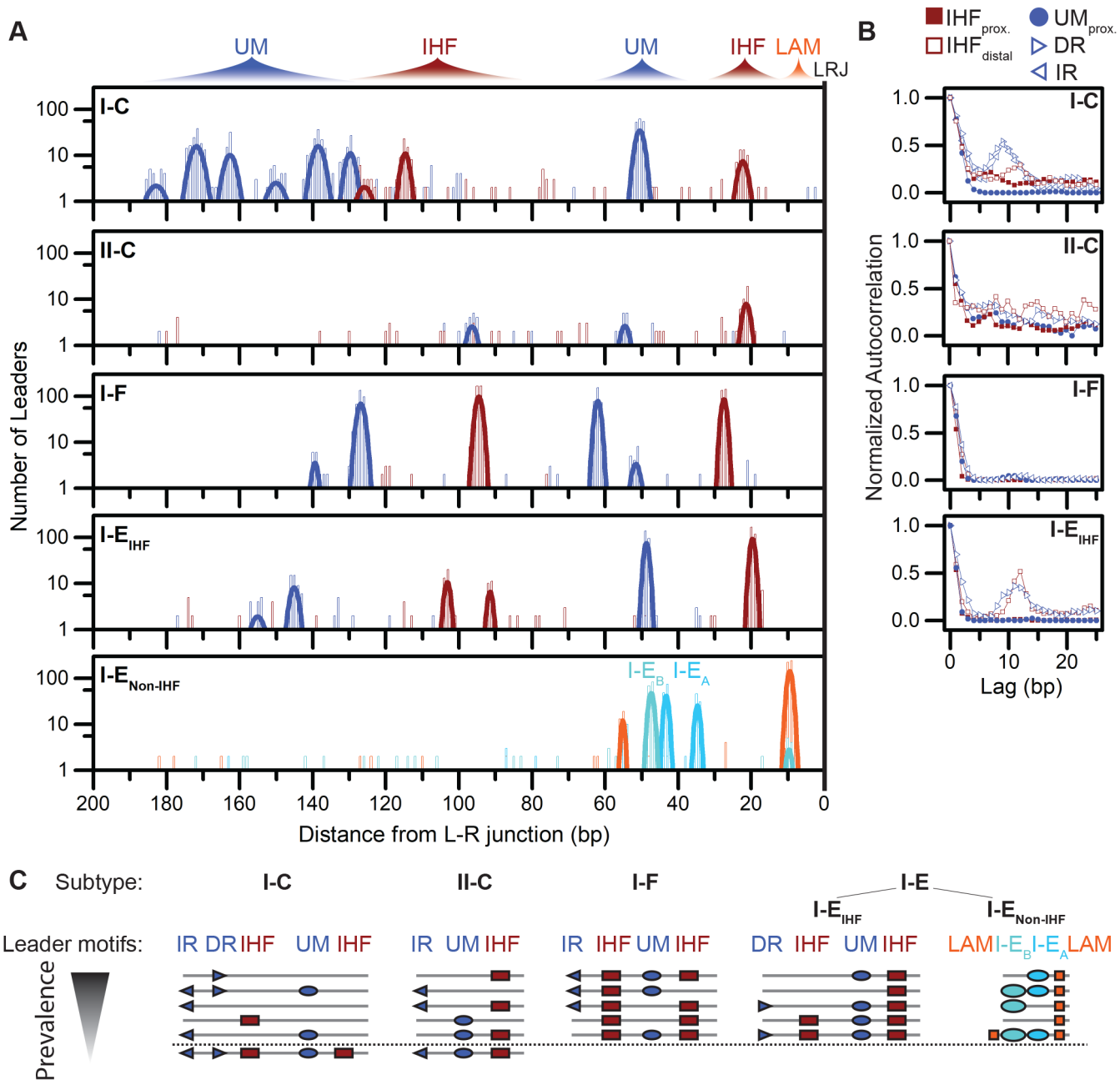


Figure S1. Naturally occurring indels preserve the phase of leader motifs, Related to Figures 2-4 and Table 1. (**A**) Tally of the midpoints of motifs found in I-C, II-C, I-F, I-E_{IHF}, I-$E_{Non\text{-}IHF}$ leaders. Phased motifs are fit to gaussian curves. (**B**) Autocorrelation plots of the midpoints of motifs found in I-C, II-C, I-F, I-E_{IHF}, I-$E_{Non\text{-}IHF}$ leaders. Common indels appear as peaks in the autocorrelation function. Motifs that are frequently shifted by 10-11 bp relative to the LRJ are retained in phase with the DNA helical repeat. (**C**) Schematics of observed leader motif architectures within each CRISPR subtype (**Data S2**). IHF binding sites (IHF) are shown as red rectangles, upstream motifs (UMs) are shown in as blue ovals. The UM motif is often repeated, as a direct repeat (DR) and/or an inverted repeat (IR). The sequences of the UMs and the relative distances between leader motifs varies between leader subtypes. Many I-F leaders possess all four motifs, whereas most I-E leaders possess only the proximal UM and IHF site. I-C and II-C leaders with the full complement of leader motifs are rare, and not among the top five most prevalent leader architectures (below dotted line).

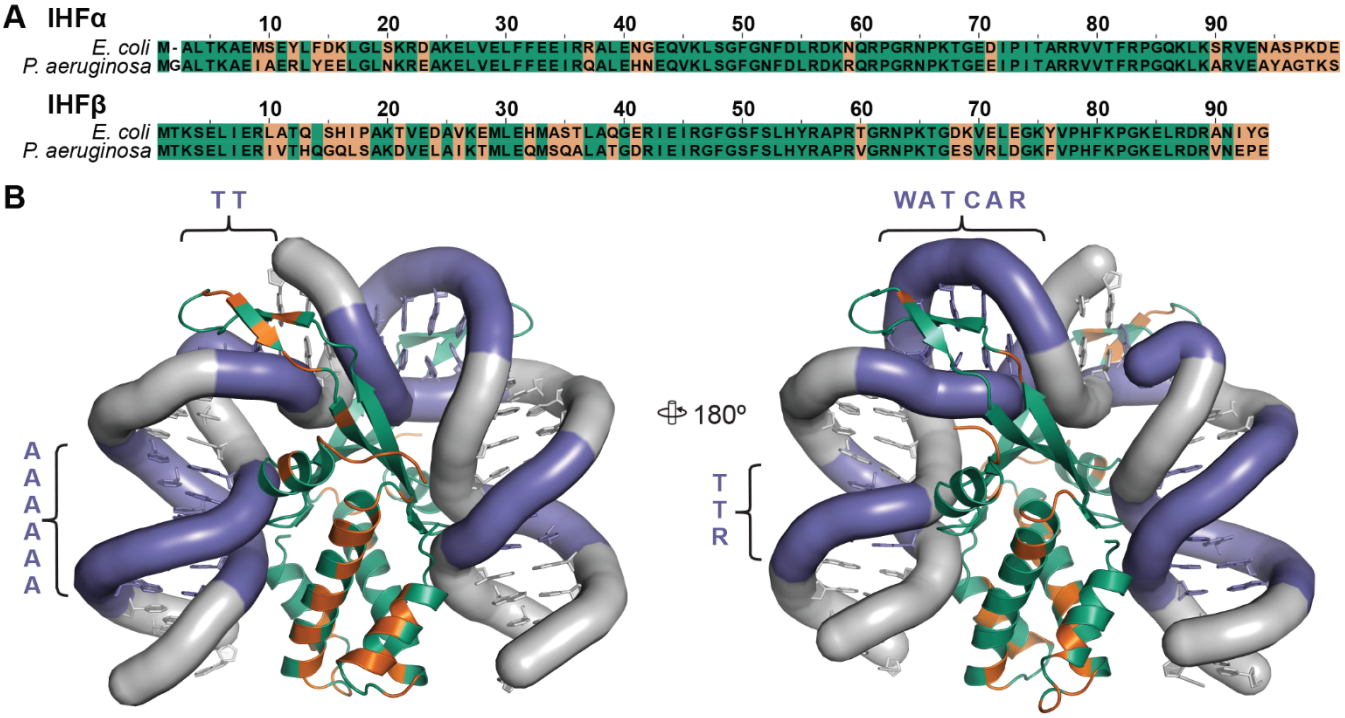


Figure S2. DNA-interacting residues are conserved between *E. coli and P. aeruginosa* IHF proteins. Related to Figure 3. (**A**) Pairwise alignments of *E. coli* and *P. aeruginosa* IHFα (top), and IHFβ (bottom). Identical residues (green) and non-identical residues (orange) are indicated. (**B**) Cartoon representations of *E. coli* IHF bound to cognate DNA[S1] (PDB:1IHF). DNA consensus sites (purple) are labelled. Non-consensus DNA regions are shown in light grey. Both IHF monomers are shown in green. Residues that are not identical in *P. aeruginosa* IHF monomers are shown (orange), none of which are expected to affect clamping of the A-tract, clamping of the TTR motif, kinking of DNA, or recognition of the TT and WATCAR motifs.

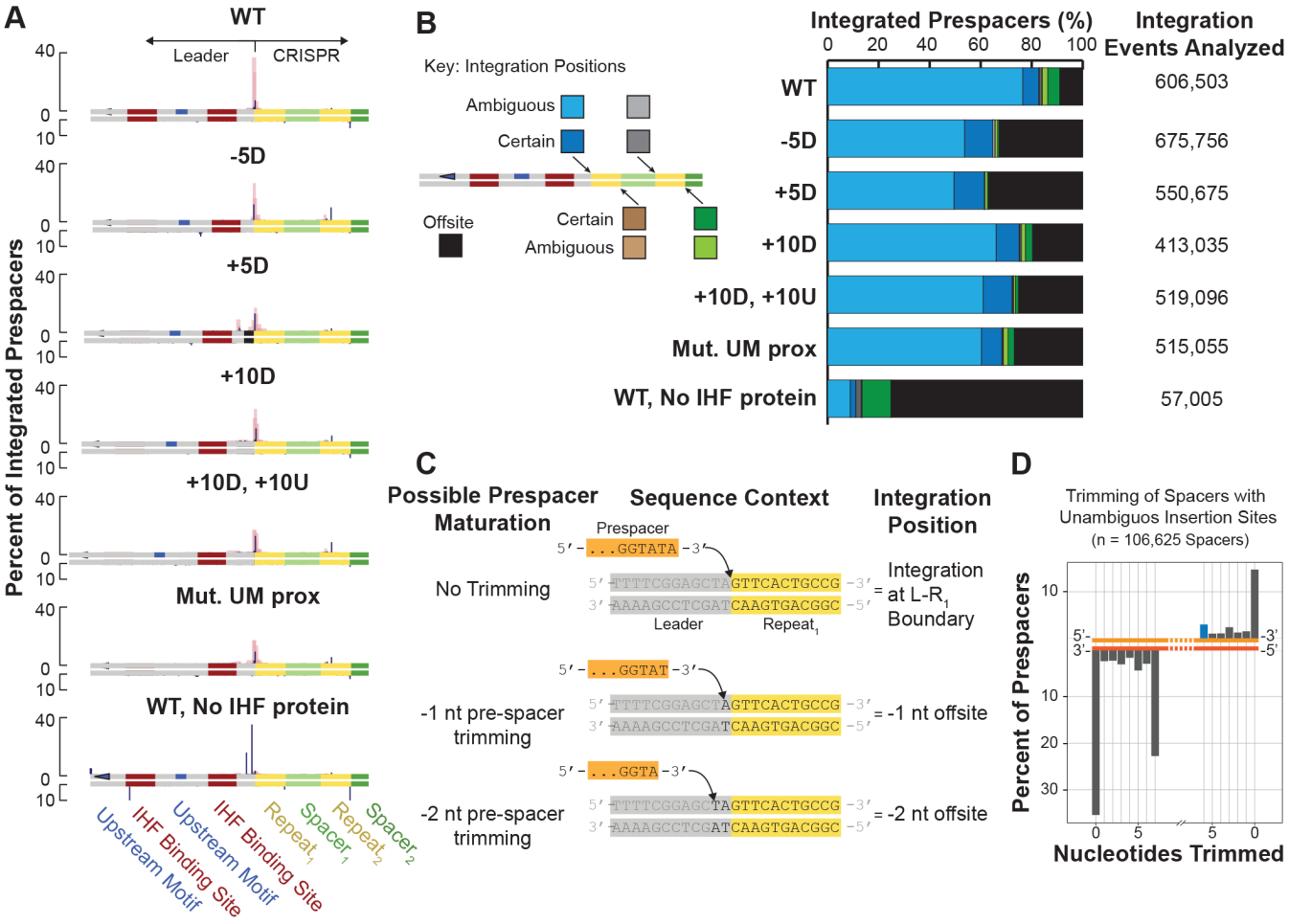


Figure S3. High-throughput sequencing and quantification of prespacer integration events. Related to Figures 3 and 4. (**A**) Graphical representation of Cas1-2/3-mediated integration events of a 40 bp dsDNA prespacer into a wildtype fragment of the CRISPR2 locus from *P. aeruginosa* PA14, in the absence of IHF (WT, No IHF protein), or in the presence of IHF (WT). Integration into variants containing leader mutations was also analyzed. "-5D" signifies deletion of 5 bp downstream of IHF proximal site. "+5D" signifies the addition of 5 bp downstream of IHF proximal site. "+10D" signifies the addition of 10 bp downstream of IHF proximal site "+10D,+10U" signifies the addition of 10 bp both downstream and upstream of the IHF proximal site. "Mut. UM prox" is used to describe a CRISPR containing a scrambled proximal upstream motif. Prespacers with confirmed integration positions (blue) represent the minority of mapped reads, since most integration positions cover two to three nucleotides that could have come from either the prespacer or the CRISPR substrate (red boxes). Based on this uncertainty, we refer to these insertions as "ambiguous". (**B**) Data from panel A is quantified (right). Reads were counted as "ambiguous" insertion events at leader-repeat or spacer-repeat boundaries if the window of possible integration positions for that read covers one of the boundaries shown (left). (**C**) Ambiguous integration events arise because prespacers can be trimmed by Cas1-2/3 or contaminating nucleases prior to integration[S2,S3]. An example of an ambiguous integration event at the leader-repeat$_1$ boundary is shown. (**D**) Quantification of trimming events in prespacers confirmed integration positions into WT CRISPR substrate. Half of the prespacers analyzed are trimmed (49.8%), but there is no preference for cleavage at a PAM-spacer boundary (position shown in blue) indicating that under the conditions tested, prespacer maturation may be stochastic.

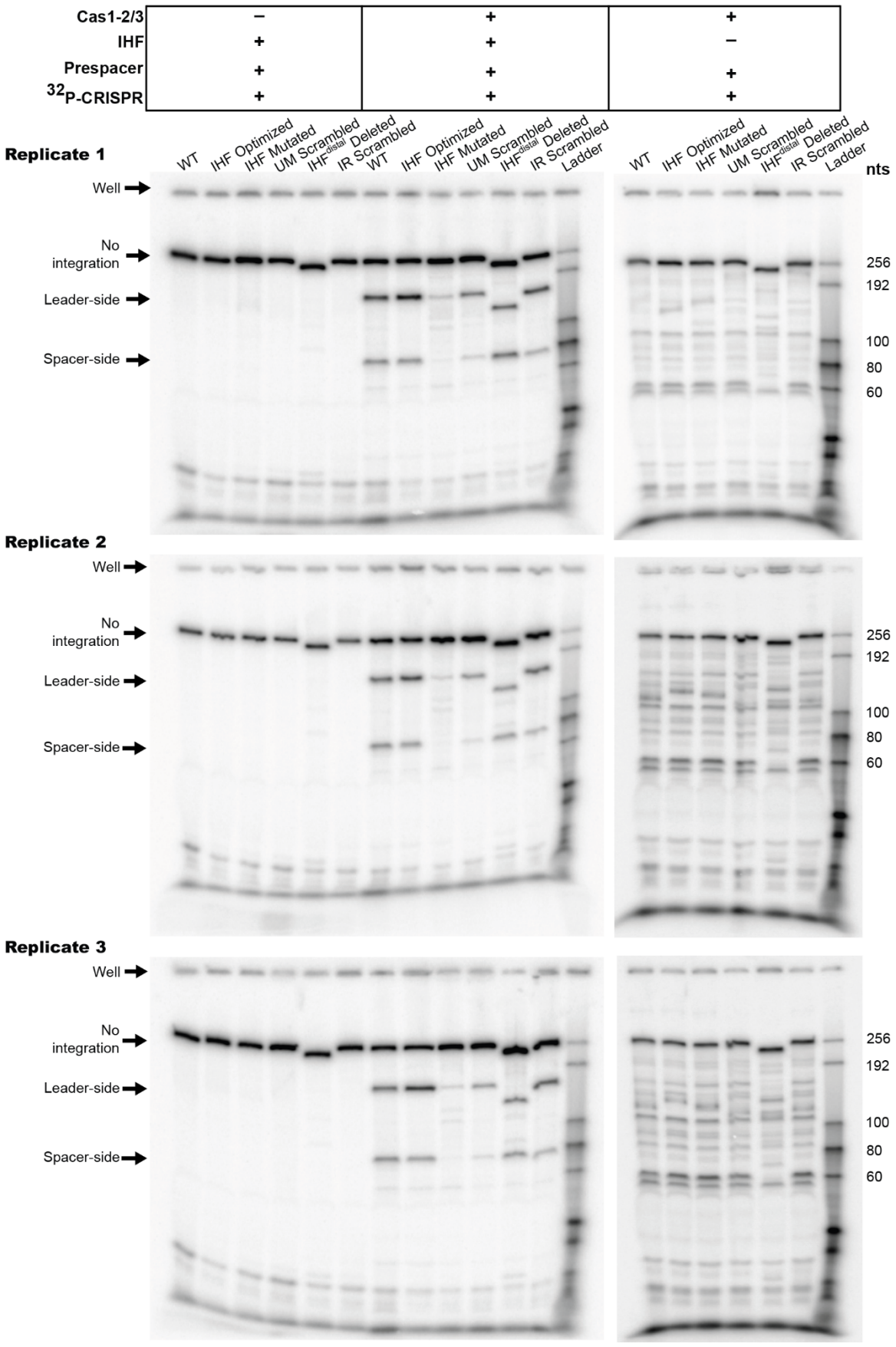
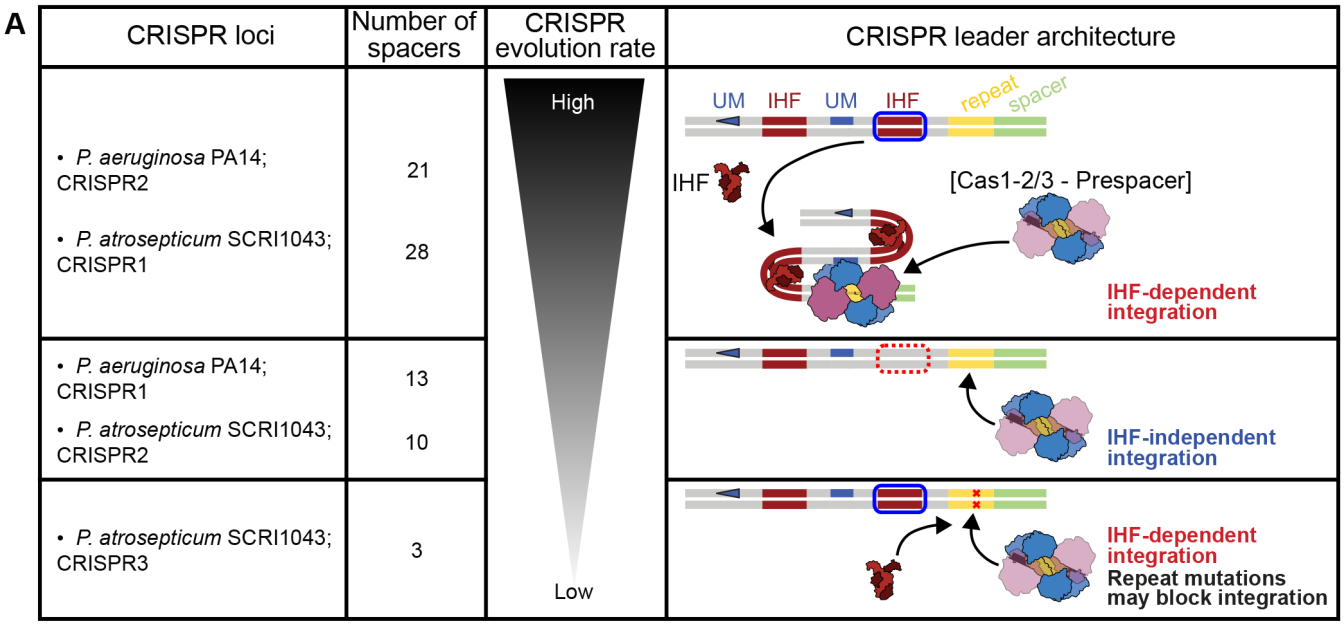


Figure S4. IHF binds I-F leaders to recruit an upstream motif needed for efficient prespacer integration. Related to Figure 3. Uncropped images of denaturing gel electrophoresis of integration reactions into a ^{32}P-labelled CRISPR fragment substrate. Reactions were performed in the absence of Cas1-2/3 (left), in the presence of all components (middle), or in the absence of IHF (right). Each row of gels depicts a replicate.



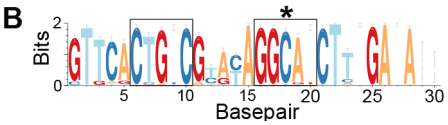


Figure S5. Summary of differential evolution rates between co-existing CRISPR loci. Related to Figure 3. (**A**) CRISPR loci found in *P. aeruginosa* PA14[S4,S5] and *P. atrosepticum* SCRI1043[S6] fall into three categories based on the number of spacers, and rates of evolution. These categories correspond with three different CRISPR architectures. Rapidly evolving CRISPRs from these bacteria possess leaders with a proximal IHF site (midpoint 28 bp from the L-R junction), and a conserved upstream motif (midpoint 63 bp from the L-R junction) (top). IHF folds the leaders of these loci, to recruit an upstream motif to a specific surface of Cas1-2/3, facilitating efficient integration. Slower evolving CRISPRs do not have recognizable IHF binding sites in their leaders but do retain a correctly positioned upstream motif (middle). Integration at these loci is less efficient, and proceeds via an IHF-independent pathway. The slowest evolving CRISPR between these model organisms has the correct motifs in its leader, but integration may be blocked by a rare mutation within the inverted repeat of the first CRISPR repeat (bottom). Mutations in these regions of the CRISPR repeat are known to strongly inhibit integration[S7–S9]. (**B**) Inverted repeat within I-F CRISPR repeat is highly conserved. Sequence logo of the first repeat in I-F CRISPRs. The highly conserved inverted repeats, necessary for efficient spacer integration, are boxed. The position of the C18T mutation that occurs in the first repeat of the *P. atrosepticum* CRISPR3 locus is indicated with an asterisk.

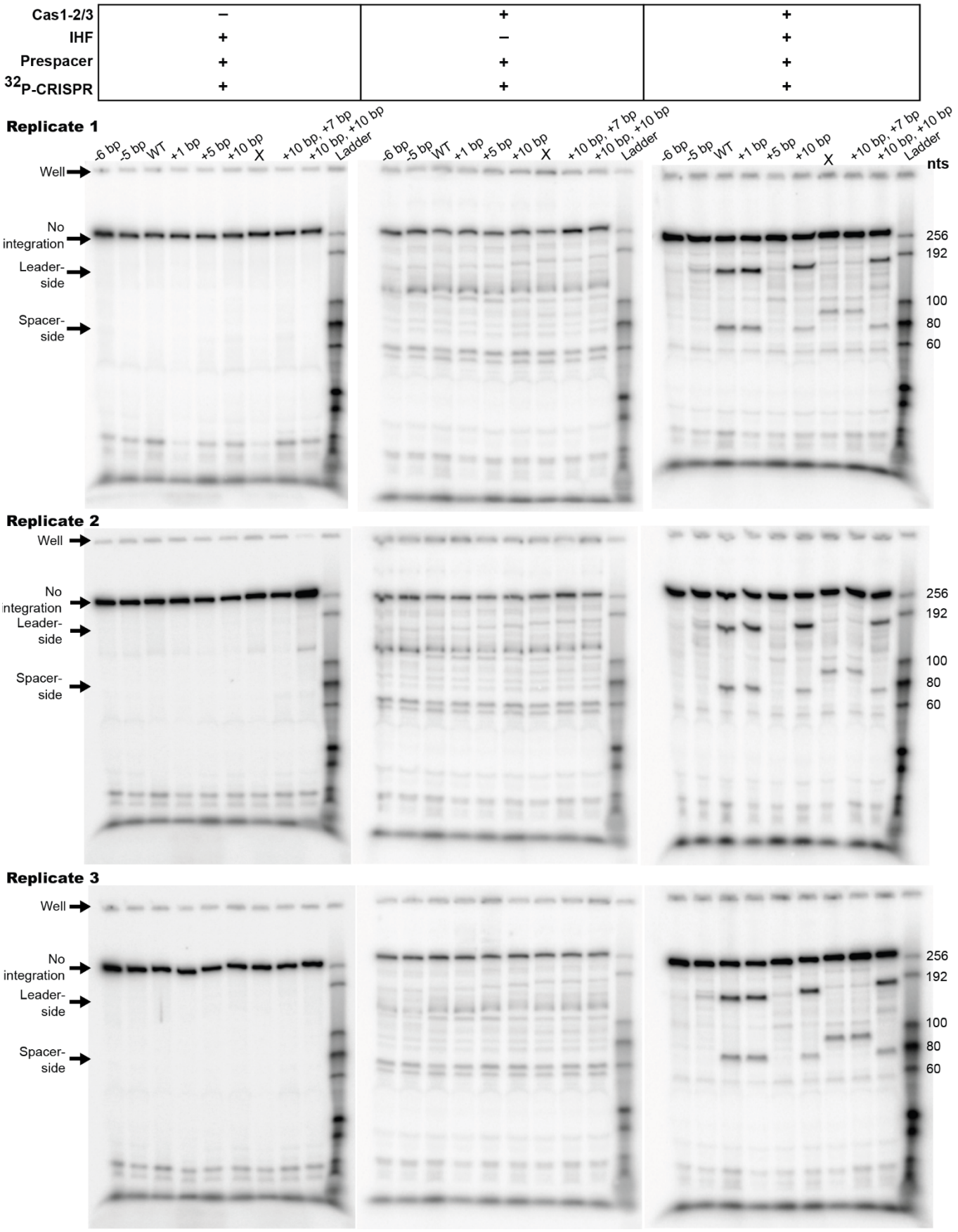


Figure S6. Leader motifs involved in CRISPR evolution are phase sensitive. Related to Figure 4. Uncropped images of denaturing gel electrophoresis of integration reactions into a ^{32}P-labelled CRISPR fragment substrate. Reactions were performed in the absence of Cas1-2/3 (left), in the absence of IHF (middle), or in the presence of all components (right).

“X” marks lanes containing reactions that were not examined further and are not discussed in this paper. Each row of gels depicts a replicate.

Supplemental References

S1. Rice PA, Yang S, Mizuuchi K, Nash HA. Crystal Structure of an IHF-DNA Complex: A Protein-Induced DNA U-Turn. Cell [Internet]. 1996 Dec;87(7):1295–306. Available from: http://dx.doi.org/10.1016/S0092-8674(00)81824-3

S2. Fagerlund RD, Wilkinson ME, Klykov O, Barendregt A, Pearce FG, Kieper SN, Maxwell HWR, Capolupo A, Heck AJR, Krause KL, et al. Spacer capture and integration by a type I-F Cas1–Cas2-3 CRISPR adaptation complex. Proc Natl Acad Sci [Internet]. 2017 Jun 13 [cited 2018 Jan 3];114(26):201618421. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28611213

S3. Kim S, Loeff L, Colombo S, Jergic S, Brouns SJJ, Joo C. Selective loading and processing of prespacers for precise CRISPR adaptation. Nature [Internet]. 2020;579(7797):141–5. Available from: http://dx.doi.org/10.1038/s41586-020-2018-1

S4. Westra ER, van Houte S, Oyesiku-Blakemore S, Makin B, Broniewski JM, Best A, Bondy-Denomy J, Davidson A, Boots M, Buckling A. Parasite Exposure Drives Selective Evolution of Constitutive versus Inducible Defense. Curr Biol [Internet]. 2015 Apr;25(8):1043–9. Available from: http://dx.doi.org/10.1016/j.cub.2015.01.065

S5. Heussler GE, Miller JL, Price CE, Collins AJ, O'Toole GA. Requirements for Pseudomonas aeruginosa type I-F CRISPR-Cas adaptation determined using a biofilm enrichment assay. J Bacteriol. 2016;198(22):3080–90.

S6. Richter C, Dy RL, McKenzie RE, Watson BNJ, Taylor C, Chang JT, McNeil MB, Staals RHJ, Fineran PC. Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. Nucleic Acids Res. 2014;42(13):8516–26.

S7. Goren MG, Doron S, Globus R, Amitai G, Sorek R, Qimron U. Repeat Size Determination by Two Molecular Rulers in the Type I-E CRISPR Array. Cell Rep [Internet]. 2016;16(11):2811–8. Available from: http://dx.doi.org/10.1016/j.celrep.2016.08.043

S8. Wright A V., Liu J-J, Knott GJ, Doxzen KW, Nogales E, Doudna JA. Structures of the CRISPR genome integration complex. Science (80-) [Internet]. 2017 Sep 15;357(6356):1113–8. Available from: http://www.sciencemag.org/lookup/doi/10.1126/science.aao0679

S9. Arslan Z, Hermanns V, Wurm R, Wagner R, Pul Ü. Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. Nucleic Acids Res. 2014;42(12):7884–93.