Priming in a permissive type I-C CRISPR-Cas system reveals distinct dynamics of spacer acquisition and loss

Authors:

Chitong Rao[1], Denny Chin[1], Alexander W. Ensminger[1,2,3]*

Affiliations:

[1] Department of Molecular Genetics, University of Toronto. Toronto, Ontario, Canada.

[2] Department of Biochemistry, University of Toronto. Toronto, Ontario, Canada.

[3] Public Health Ontario. Toronto, Ontario, Canada.

Running title: Spacer dynamics in type I-C CRISPR-Cas priming

Keywords: CRISPR-Cas; primed adaptation; next-generation sequencing; spacer loss; Legionella

Abstract

CRISPR-Cas is a bacterial and archaeal adaptive immune system that uses short, invader-derived sequences termed spacers to target invasive nucleic acids. Upon recognition of previously encountered invaders, the system can stimulate secondary spacer acquisitions, a process known as primed adaptation. Previous studies of primed adaptation have been complicated by intrinsically high interference efficiency of most systems against *bona fide* targets. As such, most primed adaptation to date has been studied within the context of imperfect sequence complementarity between spacers and targets. Here, we take advantage of a native type I-C CRISPR-Cas system in *Legionella pneumophila* that displays robust primed adaptation even within the context of a perfectly matched target. Using next-generation sequencing to survey acquired spacers, we observe strand bias and positional preference that are consistent with a 3′ to 5′ translocation of the adaptation machinery. We show that spacer acquisition happens in a wide range of frequencies across the plasmid, including a remarkable hotspot that predominates irrespective of the priming strand. We systematically characterize protospacer sequence constraints in both adaptation and interference and reveal extensive flexibilities regarding the protospacer adjacent motif in both processes. Lastly, in a strain with a genetically truncated CRISPR array, we observe greatly increased interference efficiency coupled with a dramatic shift away from spacer acquisition towards spacer loss. Based on these observations, we propose that the *Legionella* type I-C system represents a powerful model to study primed adaptation and the interplay between CRISPR interference and adaptation.

Introduction

Bacteria and archaea constantly interact with mobile genetic elements including bacteriophages, plasmids, transposons and other conjugative elements (Burrus and Waldor 2004, Frost, Leplae et al. 2005). With their genomes greatly shaped by these mobile elements, these microbes can benefit from

acquisition of foreign DNA, but also suffer detrimental effects from "selfish" elements such as lytic bacteriophages. To combat deleterious horizontal gene transfer, bacteria and archaea harbor multiple resistance mechanisms exemplified by CRISPR-Cas (clustered regularly interspaced short palindromic repeats and CRISPR-associated genes) systems (Labrie, Samson et al. 2010). To date, CRISPR-Cas systems have been identified in about half of genome-sequenced bacteria and archaea and include multiple types that each use distinct protein compositions to function as adaptive immunity against invasive nucleic acids (Horvath and Barrangou 2010, Marraffini and Sontheimer 2010, Makarova, Haft et al. 2011, Makarova, Wolf et al. 2015). A CRISPR array consists of distinct short spacers separated by repeat sequences and is transcribed as a non-coding RNA that undergoes further processing by Cas proteins to form individual repeat-spacer units (crRNAs). These crRNAs are loaded as guide sequences into Cas-crRNA interference complexes that bind to targeted nucleic acids (termed protospacers) with an appropriate protospacer adjacent motif (PAM) and mediate their destruction by recruiting the Cas3 nuclease (in type I systems) or by its intrinsic nuclease activity (in other types), a process known as interference (van der Oost, Westra et al. 2014, Marraffini 2015). One key feature of CRISPR-Cas immunity is its ability to adapt to new threats through the acquisition of new spacers derived from non-productive bacteriophage infection or other encounters with foreign DNA. Spacer acquisition can be either "naïve" (where the invader has not been previously cataloged in the array) or "primed" (where upon recognition of invaders previously targeted by CRISPR-Cas, secondary spacers are acquired in order to enhance protection). Compared with naïve adaptation, primed adaptation is much more efficient and reliant on recruitment of the interference machinery to a pre-existing target (Heler, Marraffini et al. 2014, Vorontsova, Datsenko et al. 2015, Staals, Jackson et al. 2016). When coupled, CRISPR interference and adaptation can effectively protect against evolving invasive elements (Andersson and Banfield 2008, Paez-Espino, Morovic et al. 2013, Paez-Espino, Sharon et al. 2015, van Houte, Ekroth et al. 2016).

Our understanding of primed spacer acquisition is based upon the studies of type I-E, type I-F and type I-B systems in the presence of targeted DNA such as plasmids and bacteriophages (Fineran and

Charpentier 2012, Heler, Marraffini et al. 2014, Amitai and Sorek 2016, Sternberg, Richter et al. 2016, Jackson, McKenzie et al. 2017). During CRISPR adaptation, the conserved proteins Cas1 and Cas2 form a protein complex that plays a key role in pre-spacer capture and insertion into the CRISPR array (Nunez, Kranzusch et al. 2014, Nunez, Harrington et al. 2015, Nunez, Lee et al. 2015, Wang, Li et al. 2015). Regarding the generation of pre-spacers from invasive DNA, characterization of acquired spacers in a priming condition revealed non-conserved patterns in different type I CRISPR-Cas systems. Primed spacer acquisition in the *E. coli* type I-E system showed a clear preference (>90%) from the primed (untargeted) strand and no obvious positional gradient on the plasmid or bacteriophage (Datsenko, Pougach et al. 2012, Savitskaya, Semenova et al. 2013, Fineran, Gerritzen et al. 2014). In the *Haloarcula hispanica* type I-B system, ~70% of new spacers were derived from the primed strand and a moderate preference was seen for the priming-proximal region (Li, Wang et al. 2014). In the *Pectobacterium atrosepticum* type I-F system, ~65% of spacers were acquired from the non-primed (targeted) strand with a clear gradient centered at the priming site (Richter, Dy et al. 2014, Staals, Jackson et al. 2016). A "sliding" model has been proposed to explain these patterns: the spacer acquisition machinery (including the Cas1-Cas2 complex) is recruited to the targeted sequence and subsequently slides away from the priming site in a 3′ to 5′ direction preferentially on one strand and stops at an appropriate PAM site for spacer extraction (Heler, Marraffini et al. 2014). The translocation directionality is consistent with the helicase activity of Cas3 (Mulepati and Bailey 2013, Sinkunas, Gasiunas et al. 2013), indicating that Cas3 may travel in complex with Cas1-Cas2, and this notion is supported by single-molecule imaging (Redding, Sternberg et al. 2015, Wright, Nunez et al. 2016). Besides the sliding model that describes the overall patterns, another model regarding the molecular basis of spacer extraction suggests that double-stranded Cas3 degradation products are preferentially used as donors for Cas1-Cas2 (Swarts, Mosterd et al. 2012, Kunne, Kieper et al. 2016, Severinov, Ispolatov et al. 2016). These two models are not necessarily mutually exclusive, as close interactions between the interference and adaptation machineries are likely involved in robust CRISPR adaptation (Babu, Beloglazova et al. 2011, Richter, Gristwood et al. 2012, Sternberg, Richter et al. 2016). It is possible that, depending on if Cas3 is activated (by some as-yet-

unclear signal) in its nuclease activity, the processing of targeted DNA could contribute to spacer acquisition through either Cas1-Cas2-Cas3 co-sliding (the sliding model) or Cas3 degradation followed by Cas1-Cas2 recycling for protospacer extraction (the alternative model).

Despite the accumulated knowledge on primed adaptation, a number of factors limit direct comparison between most of the previous studies. Specifically, due to the high interference efficiency against *bona fide* targets, most studies used mismatched priming sequences with either a non-canonical PAM or mutations in the seed sequence of a protospacer (Datsenko, Pougach et al. 2012, Savitskaya, Semenova et al. 2013, Fineran, Gerritzen et al. 2014, Li, Wang et al. 2014, Richter, Dy et al. 2014, Staals, Jackson et al. 2016). Such target mismatches not only affect interference (Wiedenheft, van Duijn et al. 2011, Xue, Seetharam et al. 2015), but influence the efficiency of primed adaptation (Fineran, Gerritzen et al. 2014, Li, Wang et al. 2014, Kunne, Kieper et al. 2016, Xue, Whitis et al. 2016), and may pose an impact on how spacers are acquired during priming (Redding, Sternberg et al. 2015, Vorontsova, Datsenko et al. 2015). To prime with a *bona fide* target, others used either inducible expression or anti-CRISPR regulated systems to control interference (Vorontsova, Datsenko et al. 2015, Semenova, Savitskaya et al. 2016, Staals, Jackson et al. 2016). In fact, these *bona fide* targets, despite being cleaved rapidly, were shown to be capable of inducing CRISPR adaptation with an even higher efficiency (Xue, Seetharam et al. 2015, Semenova, Savitskaya et al. 2016, Staals, Jackson et al. 2016). To avoid these complicating factors, the focus of the current study is on a relatively interference-permissive, type I-C CRISPR-Cas system in *L. pneumophila* - a system which is ideally suited to the study of priming with a *bona fide*, perfect-match target. Along with this inherent experimental strength, type I-C systems remain relatively understudied, despite representing the second most abundant type of CRISPR-Cas systems in prokaryotes (Makarova, Haft et al. 2011, Makarova, Wolf et al. 2015).

Results

115 Priming of the permissive *L. pneumophila* type I-C CRISPR-Cas induces robust spacer acquisition.

116 In our previous work, we experimentally showed that a perfectly-targeted plasmid can
117 temporarily co-exist, without detectable mutations (in either plasmid or CRISPR-Cas locus), with the type
118 I-C CRISPR-Cas system in *L. pneumophila* str. Toronto-2005 (Rao, Guyard et al. 2016). These escaped
119 transformants displayed a gradual plasmid loss during non-selective axenic passages and clear spacer
120 acquisition events induced at the end (Rao, Guyard et al. 2016). Here we exploited this robust adaptation
121 system to study spacer acquisition in the type I-C system in depth (Fig. 1A). A targeted plasmid that
122 includes the CRISPR spacer 1 (Sp1) sequence and a canonical TTC PAM (Mojica, Diez-Villasenor et al.
123 2009, Leenay, Maksimchuk et al. 2016) on either the plus strand (pSp1(+)) or minus strand (pSp1(–)) was
124 used to prime spacer acquisition (we refer to protospacer as the target identical to the spacer sequence and
125 PAM as the 5′-3′ sequence upstream of the protospacer on the untargeted strand). These targeted plasmids
126 showed a ~1% relative transformation efficiency compared with untargeted control plasmids, and the
127 escaped transformants were passaged without antibiotic selection for 15 generations to induce spacer
128 acquisition events that we subsequently cataloged by PCR amplification, gel extraction and deep
129 sequencing (Fig. 1A). Around 2 million new spacers were extracted from Illumina raw reads in each
130 priming experiment, and mapped to potential sources including the priming plasmid or the bacterial
131 chromosome. The vast majority (>99.7%) of spacers were derived from the plasmid, with the remaining
132 few from the chromosome or unknown sources (possibly due to chimeric sequences or sequencing errors;
133 Fig. 1B). Collectively these numerous spacer sequences covered all available TTC canonical PAM sites
134 on the plasmid (Table S1), suggesting a sufficient sequencing depth to represent the CRISPR-adapted
135 population.

136 Primed spacer acquisition occurs in a strand-biased fashion but is influenced by local hotspots.

137 Through Sanger sequencing of 23 acquired spacers, we previously observed a moderate
138 preference (74%) of spacers derived from the same strand as the priming protospacer (Rao, Guyard et al.

2016). With a much higher sequencing depth, we comprehensively re-examined the patterns of spacer acquisition from the plasmid. When the priming protospacer is on the plus strand of the plasmid, a majority (83%) of spacers are mapped to the same strand, and an obvious enrichment of acquisitions is seen from the 5′ region proximal to the priming site on both strands (Fig. 1C). When the priming sequence is flipped to the minus strand, the preferred strand of acquisition is also switched, with 64% spacers derived from the minus strand, and as before, more spacers are mapped to the 5′ region of the priming site relative to the 3′ region (Fig. 1D). A clear correlation between the directionality of the priming sequence and the strand preference of acquired spacers is shown from the merged view of pSp1(+) and pSp1(–) mappings (Fig. 1E). These observations are consistent with a strand-specific 3′ to 5′ translocation of Cas3 starting from the priming site.

Besides the strand bias and positional gradient, we observed a wide range of acquisition frequencies across the plasmid. Among all 238 TTC PAM sites, 30 positions each accounted for >1% of all acquisitions in at least one priming experiment, and 62 were acquired at <0.05% frequencies in both priming settings (Table S1). Strikingly, we identified one locus in the coding strand of *repC* that consistently ranked as one of the most frequently acquired spacers regardless of the primed strand (Fig. 1C, D). Interestingly, we did not observe an obvious enrichment of spacers from the origin of plasmid replication (oriV) and open reading frames (*cat*, *mobC*, etc.) - known hotspots in naïve adaptation and primed adaptations in other type I systems (Levy, Goren et al. 2015, Vorontsova, Datsenko et al. 2015, Staals, Jackson et al. 2016).

Lastly, we examined if a mismatched protospacer primes the type I-C system differently from a perfect match. While a single T1A mutation in the seed sequence increased the plasmid relative transformation efficiency from ~1% to ~9%, the overall patterns (strand bias and positional gradient) of acquired spacers were largely unchanged (Fig. 1C, F). These data are consistent with models in which primed spacer acquisition using perfect (interference-driven) or imperfect matches involves shared molecular mechanisms (Semenova, Savitskaya et al. 2016, Staals, Jackson et al. 2016).

Sequence specificity contributes to the acquisition hotspot.

To identify the factors contributing to the high acquisition efficiency of the hotspot, we first examined the hotspot region for the presence of any outstanding feature: PAM density, GC content, origin of replication or predicted small RNA transcription. In the absence of an obvious signal from any of these features, we hypothesized that some other sequence specificity of the hotspot region underlies the high acquisition efficiency. Thus, we generated a set of plasmids carrying mutations upstream, downstream, or within the hotspot while maintaining the *repC* codon to avoid any side-effects due to amino acid changes (Fig. 2A).

We first tested the PAM mutant in pSp1(+) where the TTC PAM of the hotspot is changed to TTT (where coincidentally another TTC motif is made with +1 nt shift). The acquisition efficiency of the mutant sequence decreased by 9-fold (16.7% to 1.8%) (Fig. 2B). We also observed a large reduction of acquisition efficiency at the hotspot by introducing the same mutation in pSp1(–) (Fig. 2C). Importantly, by comparing these hotspot PAM mutants with the *wild type* plasmids, we did not observe a major difference in spacer acquisition patterns due to the elimination of the hotspot, i.e. the imperfect mirroring of strand bias (>80% plus in pSp1(+) and <70% minus in pSp1(–)) is retained. This observation suggests that the local hotspot likely does not affect the initial sliding of the adaptation machinery whose strand preference is partially skewed to the plus strand due to unknown factors.

We next examined the other regions of the hotspot by introducing different sets of mutations in pSp1(+). The acquisition efficiency of the hotspot was dramatically eliminated by internal mutations, slightly reduced by changes upstream, and not reduced at all by the downstream perturbations (Fig. 2D-F). Elimination of the hotspot increases acquisition frequencies of other plasmid loci (Fig. 2B, E), suggesting that its loss modifies the availability of the adaptation machinery to other loci. The major impacts of the PAM mutation and the internal substitutions suggest that some sequence specificity within the hotspot, likely in the 5′ end, contributes to the acquisition preference at this locus.

188 Analysis of acquired spacers reveals an alternative PAM and extensive acquisition inaccuracies.

189 Of all acquired spacers from the plasmid, those having a TTC PAM account for 92.5% and 90.0%
190 in the pSp1(+) and pSp1(–) priming experiment, respectively (Fig. 3B). We examined the trinucleotide
191 sequence upstream of all acquired spacers from pSp1(+) for the abundance of other PAMs (Fig. S1A).
192 The ~2 million acquired spacers from pSp1(+) are next to 2,978 different PAM loci. While most (90%) of
193 these PAM loci were acquired rarely (with <0.01% frequencies), some PAM sequences other than TTC
194 were oversampled, suggesting one or more alternative PAMs. Based on the frequency rankings of the
195 trinucleotide PAM sequences, the second most frequent PAM is TTT, followed by four TCN motifs (Fig.
196 S1A). As these less frequent PAMs share a 2 nt identity with TTC, which could derive from slipping
197 events where the real PAM is still TTC located nearby (Shmakov, Savitskaya et al. 2014), we first
198 suspected that the TTT and TCN PAMs might be due to –1 nt slips (upstream) and +1 nt slips
199 (downstream), respectively (Fig. 3A). Thus, we separately reanalyzed acquired loci with each of these
200 PAMs with respect to their flanking sequence. Spacers with a TCN PAM showed a major T signal further
201 upstream, consistent with +1 nt slips from TTC (Fig. 3C). However, spacers with a TTT PAM did not
202 show an outstanding signal next to the trinucleotide, suggesting that TTT is an alternative PAM other than
203 TTC (Fig. 3C). Consistent with this interpretation, acquired spacers with a TTT PAM showed
204 independent localizations relative to those with a TTC PAM (Fig. S1B).

205 As we observed +1 nt slips, we wondered if there were other types of acquisition errors. We
206 systematically examined + slips where the acquisition machinery extracts protospacers further
207 downstream at the PAM and – slips where cleavage happens further upstream (Fig. 3A). Indeed, besides
208 +1 nt slips that occurs at a ~2% frequency, other types of slips do happen - though at 0.1%~1%
209 frequencies, with a decreasing trend as the slipping goes further (Fig. 3B). Apart from the aforementioned
210 classes of spacer acquisition where the upstream sequence of the plasmid contains either a TTC or a TTT
211 PAM, ~2% spacers remained unexplained. When we examined the target sequence upstream and
212 downstream of these spacers, we observed a clear GAA signal directly downstream (Fig. 3C). This

downstream GAA signal is consistent with a phenomenon known as "flipping" - where a double-stranded DNA pre-spacer is extracted next to a TTC PAM but subsequently integrated in an opposite direction into the CRISPR array so that the reverse complementary strand is used as spacer (Shmakov, Savitskaya et al. 2014). Indeed, we identified ~1% spacers derived from potential flips with an original TTC PAM, 0.3%~0.4% spacers from a combination of flips and +1 nt slips, and even rarer still - combinations of flips and other types of slips (Fig. 3B). When combined, the canonical TTC PAM, the alternative TTT PAM and slipping and flipping events explain >99.5% of all acquired spacers from the priming plasmid (Fig. 3B).

The native spacers in *L. pneumophila* type I-C CRISPR arrays range from 33 to 37 nt, with 35 nt being the most frequent length. We next asked if acquired spacers had a similar length distribution. Compared with native spacers, laboratory acquired spacers showed a slight shift towards shorter lengths, with 34 nt being the most frequent (Fig. 3D). Compared with type I-E and type I-F systems that acquire spacers mostly (~90%) with a uniform length (Datsenko, Pougach et al. 2012, Savitskaya, Semenova et al. 2013, Fineran, Gerritzen et al. 2014, Richter, Dy et al. 2014, Staals, Jackson et al. 2016), the type I-C system acquired spacers with a broad range of lengths - suggesting a remarkably imprecise molecular ruler in the adaptation machinery. This inaccuracy is largely unattributed to either slipping or flipping events, as spacers with a canonical TTC PAM showed a similarly large distribution of length (Fig. 3E). Interestingly, in +1 nt slips and +2 nt slips, we observed a distinct distribution of spacer length, with an increasing preference for shorter (≤33 nt) ones (Fig. 3E). This is in contrast with the observations in the *P. atrosepticum* type I-F system where – slips instead of + slips correlated with aberrant spacer lengths (Staals, Jackson et al. 2016), indicating another molecular distinction between these two adaptation machineries. Taken together, we identified extensive acquisition inaccuracies in the *L. pneumophila* type I-C system. A representative example of these inaccuracies can also be found at the major spacer acquisition hotspot (Fig. S1C).

Systematic quantification of interference efficiencies confirms a hierarchy of preferred PAMs.

PAM recognition in spacer acquisition is attributed to the adaptation machinery - and this process is likely independent from the Cascade interference complex that by itself recognizes the PAM and binds to target. To examine the possible co-evolution of PAM recognition by these two machineries (Kunne, Kieper et al. 2016), we asked if the Cascade interference complex also recognizes alternative PAMs other than the canonical TTC motif. Indeed, using an *in vivo* positive screen, Leenay *et al*. recently identified TTC, CTC, TCC and TTT, with decreasing preferences, as functional PAMs for interference in the *Bacillus halodurans* type I-C system (Leenay, Maksimchuk et al. 2016). Here, we performed a plasmid-removal based screen to examine functional PAMs for interference in *L. pneumophila* (Fig. 4A). We transformed a plasmid library containing a full spacer 1 match with a randomized trinucleotide PAM into either *L. pneumophila* str. Toronto-2005 *wild type* or Δ*cas3*. By analyzing the PAM abundance in the survived plasmid pools using high-throughput sequencing, we identified PAMs that were depleted to different degrees by the *wild type* type I-C system. Among the 64 PAM sequences, TTC achieved the highest protection efficiency of >99.9%, 6 others (TTT, CTT, CTC, TTA, TTG and TCC) within the range of 95% ~ 99.5%, and 11 more above 50% (Fig. 4B). It is noteworthy that TTT is the second most interference-efficient PAM, consistent with our observation that TTT is also the second most frequent PAM used in spacer acquisition. Many of the less protective PAMs share a 2 nt identity with TTC, suggesting that a 1 nt perturbation of the PAM would still allow some functionality. We confirmed the observed hierarchy of PAM activities using a CFU-based plasmid transformation efficiency assays of 8 selected PAMs (Fig. 4C). Inspection of the escaped transformants of TTT PAM plasmids showed spacer acquisition events similar to the transformants of the *bona fide* target and mismatched protospacer plasmids, suggesting that the alternative PAM primes in a similar manner (Fig. 4D). Together, our assay suggests that the *L. pneumophila* type I-C system possesses a broader range of active PAMs in interference than in adaptation.

Truncation of the type I-C array leads to a dramatic increase in interference and frequent spacer loss.

As *L. pneumophila* type I-C CRISPR-Cas is relatively permissive for interference, we wondered how spacer acquisition efficiency would change if target cleaving by this system is made more efficient. While studying a minimized type I-C array that contains only a single spacer, we made an unexpected observation that allowed us to further explore the relationship between interference efficiency and spacer acquisition. We generated a CRISPR array-minimized strain in which all 43 spacers except the spacer 1 were deleted in *L. pneumophila* str. Toronto-2005. Remarkably, the sole spacer in this strain showed a ~100-fold increased protection efficiency against its matching protospacer as compared to the parental strain that contains a full-length (43 spacers) array (Fig. 5A). When examining the CRISPR loci in the less frequent escaped transformants, no spacer acquisition was observed, in contrast to what we observed for the more permissive, full-length array strain. Instead, these escaped transformants showed clear spacer loss events (Fig. 5A). To test if the modified CRISPR array is adaptable, we next transformed the spacer 1 only strain with a mismatched target plasmid (carrying a T1A seed mutation). Use of this mismatched target led to both decreased interference efficiency and robust spacer acquisition, demonstrating the adaptability of the minimized CRISPR array under certain conditions (Fig. 5A). Consistent with the observations using agarose gel, by quantifying spacer dynamics in the single-spacer strain transformants of the spacer 1 targeted plasmid, we observed 39% spacer loss frequency in the CRISPR loci and a ~100-fold lower spacer acquisition frequency relative to the *wild type* strain transformants (Fig. 5B). As the minimized array was designed to contain two similar but distinct repeat sequences flanking spacer 1, we examined the spacer loss events and found that most loci retained the downstream repeat, consistent with a mechanism of homologous recombination (Fig. 5C). It is also noteworthy that we quantified spacer dynamics in the single-spacer strain transformed with an untargeted plasmid. We observed a detectable <0.1% spacer loss frequency without spacer targeting, suggesting that spacer loss events naturally occur in CRISPR loci at a low frequency and can be enriched under selection (Fig. 5B).

Our observations using the minimized array suggest that when CRISPR interference reaches a threshold of efficiency to no longer tolerate the temporary co-existence between the target and a

functional CRISPR-Cas, the resulting transformants would select for spacer loss rather than spacer acquisition. To further explore the interference efficiency threshold between spacer acquisition and spacer loss, we took advantage of a naturally occuring type I-C system with a greater protection. The type I-C CRISPR-Cas system in *L. pneumophila* str. Toronto-2000 differs from the one in *L. pneumophila* str. Toronto-2005 by three newly acquired spacers (Rao, Guyard et al. 2016). Transformation of a plasmid targeted by the first spacer T1 in this array, resulted in a very low transformation efficiency and correspondingly spacer loss instead of spacer acquisition in the rare escaped transformants (Fig. 5D). Examination of downstream spacers in this system, which provide less efficient protection, suggests that less efficient interference is correlated with stronger spacer acquisition (Fig. 5D). These data, together with the observations in the single-spacer strain, indicate that an interference efficiency of >99.9% by plasmid transformation is a good empirical indicator for an absence of primed adaptation using *bona fide* targets in the *L. pneumophila* type I-C system.

Discussion

A clear interplay between CRISPR interference and adaptation has been established (Sternberg, Richter et al. 2016, Wright, Nunez et al. 2016). For CRISPR-Cas systems that execute the interference very efficiently (interference-strict), a slow or delayed target degradation (by target mismatches or other means) is often necessary to achieve an efficient primed adaptation (Kunne, Kieper et al. 2016, Semenova, Savitskaya et al. 2016). Here we confirmed the observation using a native interference-permissive system. We propose that when the cleaving efficiency of a system allows the temporary coexistence of target and a functional CRISPR-Cas, robust spacer acquisition predominates. In contrast, when this system is made highly efficient in spacer targeting, spacer loss events were selected for and the resulting transformants showed a lack of primed adaptation. The observed low frequency of natural spacer loss in the CRISPR array also provides another perspective of spacer dynamics - that CRISPR-Cas systems can readily update

and diversify its immunological memory through both spacer acquisition and spacer loss, thus providing raw materials of evolution.

Our analyses of spacer acquisition patterns (Fig. 6) are consistent with the sliding model in which the adaptation machinery (which includes Cas1-Cas2 and possibly Cas3) translocates 3′ to 5′ preferentially on the untargeted strand before stopping at an appropriate sequence to extract a protospacer for subsequent integration into the CRISPR array. Several similarities and differences exist between *L. pneumophila* type I-C acquisition and what has been described for other systems in other bacteria (Fig. 6). We observe similar spacer acquisition patterns between the *L. pneumophila* type I-C system and the *H. hispanica* type I-B system (Li, Wang et al. 2014): both systems show a moderate (70~80%) overall bias towards the untargeted strand and a positional preference for the 5′ region on both strands relative to the priming site. In contrast, the *P. atrosepticum* type I-F system prefers spacers on the targeted strand and samples spacers in a narrower distance from the priming sequence (Richter, Dy et al. 2014, Staals, Jackson et al. 2016). Type I-F's opposite strand preference could possibly be due to an opposite spatial organization of the adaptation complex relative to the PAM recognition. The *E. coli* type I-E system, shows robust spacer acquisitions from the untargeted strand, like type I-C and type I-B, but proximity to the priming sequence appears to have little influence on the overall pattern of spacer acquisition (Datsenko, Pougach et al. 2012, Savitskaya, Semenova et al. 2013, Fineran, Gerritzen et al. 2014). To explain the discrepancies of positional preference (sampling distance) in different systems, a variable processivity of Cas3 in different systems was proposed (Redding, Sternberg et al. 2015). This model would suggest that the type I-C and type I-B systems should have an intermediate level of Cas3 processivity (between a highly processive type I-E system and a less processive type I-F system). The different levels of strand bias in these systems, on the other hand, may attribute to different degrees of "PAM-independent processing" in which Cas3 is recruited with the help of Cas1-Cas2 and travels bi-directionally (Redding, Sternberg et al. 2015).

Compared with the earlier study of the *H. hispanica* type I-B system that used Sanger sequencing (Li, Wang et al. 2014), we have achieved a much higher survey depth through next-generation sequencing, thus enabling a more comprehensive examination of spacer acquisition details. In PAM preference, in addition to the canonical TTC PAM (~90% of all acquired spacers), we identified an alternative TTT PAM (2~4%) and extensive slipping and flipping events (6~8%). With respect to spacer size selection, we observed a flexible choice more similar to the type I-B system (Li, Wang et al. 2014) than the highly stringent type I-E (Savitskaya, Semenova et al. 2013, Fineran, Gerritzen et al. 2014) and type I-F (Richter, Dy et al. 2014, Staals, Jackson et al. 2016) systems. These observations suggest that the type I-C Cas1-Cas2 protein complex is relatively promiscuous in spacer extraction from pre-spacer substrates. The similarity between the type I-C and type I-B systems is also consistent with their closer Cas1-based phylogenetic relationship relative to the other two systems (Fig. S3). Further insights into the molecular basis of spacer acquisition stringency may be derived from a detailed structure based comparison of Cas1 and Cas2 from each system.

It is known that different CRISPR-Cas systems, as well as different spacers within one array, often show a wide range of interference efficiencies (Marraffini and Sontheimer 2008, Bikard, Hatoum-Aslan et al. 2012, Cady, Bondy-Denomy et al. 2012, Li, Wang et al. 2014, Xue, Seetharam et al. 2015, Qiu, Wang et al. 2016, Rao, Guyard et al. 2016) (Fig. 5D). Both technical and biological factors could contribute to this variation. On the one hand, transformation methods, plasmid copy number, bacterial culture conditions, etc. could all affect how efficiently invasive DNA is cleaved (Majsec, Bolt et al. 2016, Rao, Guyard et al. 2016, Severinov, Ispolatov et al. 2016). On the other hand, innate factors could also influence interference efficiency – such as expression levels of Cas proteins, transcription and processing efficiencies of individual spacers, and binding affinities between Cascade and crRNA (Xue, Seetharam et al. 2015, Hoyland-Kroghsbo, Paczkowski et al. 2016, Patterson, Jackson et al. 2016, Rao, Guyard et al. 2016). We observed a dramatic increase in interference efficiency for the same spacer (Sp1) when the CRISPR array was minimized. This could be due to a higher abundance of Sp1 crRNA, a relatively

increased availability of Cas proteins for Sp1 (due to lack of competition with other spacers for loading), or a combination of both. Future experiments to examine the crRNA abundance and to over-express each Cas functional group (to determine limiting factors) will be necessary to test these hypotheses.

Consistent with previous studies of CRISPR adaptation (Paez-Espino, Morovic et al. 2013, Yosef, Shitrit et al. 2013), we observed a great range of spacer acquisition frequencies at different locations of the same element. While this variation could be affected by PAM specificity, strand specificity and strand-specific distance from the priming site, our examination of the major spacer acquisition hotspot points towards other factors that directly contribute to pre-spacer capture by the adaptation machinery. By introducing different mutations in the hotspot neighbourhood, we found that the internal sequence, but not the flanking nucleotides, contributes to the frequent acquisition. Based on these data, we speculate that some DNA motif or ssDNA secondary structure within the hotspot sequence (likely at the PAM-proximal end) could attract the adaptation machinery, and further systematic mutation experiments are required to identify the exact contributor. The intrinsic sequence specificity of the type I-C hotspot stands in stark contrast to a study of the *E. coli* type I-E system that showed a frequently-acquired protospacer was reliant on its upstream and downstream sequences (Yosef, Shitrit et al. 2013). Notably, we also did not observe a detectable enrichment of spacer acquisition from either the origin of plasmid replication or transcriptionally active regions - in contrast to what has been seen in type I-E and type I-F systems (Levy, Goren et al. 2015, Vorontsova, Datsenko et al. 2015, Staals, Jackson et al. 2016). These discrepancies further indicate a mechanistic distinction for pre-spacer capture in different systems and point towards the potential for diverse model systems to inform our understanding of the mechanisms underpinning CRISPR-Cas interference and adaptation.

Going forward, several features make the *L. pneumophila* type I-C system a good model system to study CRISPR-Cas functionality. First, type I-C systems represent one of the most common types of CRISPR-Cas systems yet nevertheless remain relatively understudied (Makarova, Haft et al. 2011, Makarova, Wolf et al. 2015). Second, our earlier comparative genomics data suggest that the system is

naturally adaptable (Rao, Guyard et al. 2016). Third, the relatively permissive interference of the system allows the laboratory study of primed spacer acquisition within the context of perfectly matched target sequences. Lastly, based on our initial characterizations, the system displays several features that distinguish it from type I-E and type I-F systems, the two systems most exhaustively studied to date.

Materials and Methods

Bacterial strains and plasmids

Legionella pneumophila strain Toronto-2005 is a clinical isolate of Sequence Type 222 from Toronto, Canada, with a circularized genome available (Genbank CP012019) (Rao, Guyard et al. 2016). An RpsL^{K43R} streptomycin resistant derivate of the clinical isolate is used as *wild type* in this study. From this RpsL^{K43R} strain, a Δ*cas3* deletion mutant and an array-minimized (Sp1-only) strain were generated by allelic exchange as described (Ensminger, Yassin et al. 2012, Rao, Guyard et al. 2016). Specifically, in the Sp1-only strain, only the first repeat, Sp1 and the last repeat of the original array were retained. A closely-related ST222 strain, Toronto-2000, was also genome sequenced (Rao, Guyard et al. 2016). The priming plasmids were generated by cloning the insert (see Table S2) into the ApaI/PstI-cut pMMB207 backbone (Rao, Benhabib et al. 2013, Rao, Guyard et al. 2016) (see Supplemental File for the full pSp1(+) sequence). Our previous study using Illumina sequencing showed that this plasmid has an average copy number of 7.6 in *L. pneumophila* str. Philadelphila-1 (Rao, Benhabib et al. 2013). Site-directed mutagenesis (QuickChange II) was used to mutate the spacer acquisition hotspot in the original plasmid. Bacterial electroporation and axenic passage were performed as previously described (Rao, Guyard et al. 2016). After axenic passages for 15 generations, the CRISPR adaptation ratio in the bacterial population increased from ~1% to ~24%, as quantified by Illumina sequencing (data not shown). Each priming experiment was performed in two biological replicates and these replicates were largely consistent in spacer mappings (data not shown). Unless specified, data shown are averages of two replicates.

409 PCR of CRISPR loci and preparation of Illumina libraries

410 Roughly 1 OD unit (~1x10^9) bacterial cells from either colony pool (containing at least 50
411 independent colonies) or axenic passage were used for genomic DNA extraction using the NucleoSpin
412 Tissue kit (Machery-Nagel). CRISPR loci were amplified using the Kapa HiFi polymerase (Kapa
413 Biosystems) and primers listed in Table S1. Raw PCR products of 20 amplification cycles were used for
414 library preparation. In addition, to enrich for adapted CRISPR arrays, 30-cycle PCR products were
415 concentrated by ethanol precipitation and separated in 6% acrylamide gel by running at 60V for 3 hours.
416 A ~70 bp higher band than the original array (~350 bp) was extracted and DNA purified from the
417 extraction was subjected to another 10-cycle PCR to increase the yield. These further size-selection steps
418 to enrich for adapted arrays did not introduce significant bias relative to the raw PCR products (data not
419 shown). Purified PCR amplicons were normalized by PicoGreen to 1 ng and processed using the Nextera
420 XT kit (Illumina). Multiplexed libraries were subjected to Illumina NextSeq sequencing at 2 x 150bp read
421 length (CAGEF, University of Toronto).

422 Illumina reads processing and data analyses

423 Paired-end raw reads were first attempted to merge by FLASH (Magoc and Salzberg 2011) using
424 "-m 50 -M 100 -x 0.02" settings. The unassembled single-end reads were quality trimmed by
425 Trimmomatic (Bolger, Lohse et al. 2014) using "SLIDINGWINDOW:3:20 MINLEN:50" settings. These
426 pre-processed reads were combined and processed using a Perl script (available upon request) to annotate
427 the presence of leader sequence (L), CRISPR repeats (R), existing spacers (S), new spacers (X) and
428 downstream sequence (D) in each read. The new spacers were extracted and aligned using blastn to either
429 the priming plasmid or *L. pneumophila* str. Toronto-2005 genome. Blastn results were then summarized
430 into coverages of each nucleotide in the plasmid and subjected to Circos visualization (Krzywinski,
431 Schein et al. 2009). To examine the PAM preference, slipping and flipping of acquired spacers, flanking
432 sequences of acquired spacers were extracted from the plasmid and subjected to Sequence Logo (Crooks,

433 Hon et al. 2004) visualization. To avoid potential redundancy, flipping cases were only examined from
434 spacers without a TTC or TTT PAM in the upstream junction. To quantify spacer acquisition and spacer
435 loss frequencies, the following formulas were used, in which each item denotes the count of reads with
436 the indicated annotation:

437 $$\text{Spacer acquisition ratio} = \text{L-R-X} / (\text{L-R-X} + \text{L-R-S1} + \text{L-R-D})$$

438 $$\text{Sp1 loss ratio} = (\text{L-R-D} + \text{X-R-D}) / (\text{L-R-D} + \text{X-R-D} + \text{S1-R-D})$$

439 Preparation and analyses of PAM plasmids pool

440 Oligos (see Table S2) with a randomized trinucleotide upstream of Sp1 sequence were annealed,
441 digested and ligated into the ApaI/PstI-cut pMMB207 vector. A total of ~3000 *E. coli* colonies were
442 obtained after transformation and combined into a pool. Plasmids were extracted from the *E. coli* pool
443 using the PureYield Plasmid Midiprep kit (Promega), and a control plasmid with a scrambled insert was
444 spiked into the plasmid pool at ~1% ratio. Roughly 1 μg of the pooled plasmids was electroporated into 4
445 OD units of *L. pneumophila* str. Toronto-2005 *wild type* or Δ*cas3* overnight culture. Three biological
446 replicates of electroporation were performed. With 5 μg/ml chloramphenicol selection, over 3000
447 colonies were obtained from each electroporation. Plasmids were then extracted from these *L.*
448 *pneumophila* transformants using the EZ-10 Spin Column Miniprep kit (Biobasic). Without any PCR
449 amplification, these plasmid pools were subjected to the Nextera XT library preparation and Illumina
450 NextSeq sequencing. After quality filtering, reads containing the Sp1 sequence (or the scrambled
451 sequence) were extracted and PAM sequences were identified from these reads. PAM frequencies in *L.*
452 *pneumophila* transformants were normalized to both the scrambled control and the *E. coli* plasmid pool.

453 Data accessibility

454 The NextSeq sequencing data have been deposited in the NCBI Sequence Read Archive under the
455 BioProject PRJNA360289.

Acknowledgments

This work was supported by a Project Grant from the Canadian Institutes of Health Research (PHT-148819), the Connaught Fund (NR-2015-16), and an infrastructure grant from the Canada Foundation for Innovation and the Ontario Research Fund (30364) to AWE. The authors thank members of the Centre for the Analysis of Genome Evolution and Function (CAGEF) at the University of Toronto for performing Illumina sequencing. We thank members of the Ensminger laboratory for helpful discussions and for careful reading of the manuscript.

References

Amitai, G. and R. Sorek (2016). "CRISPR-Cas adaptation: insights into the mechanism of action." Nat Rev Microbiol **14**(2): 67-76.

Andersson, A. F. and J. F. Banfield (2008). "Virus population dynamics and acquired virus resistance in natural microbial communities." Science **320**(5879): 1047-1050.

Babu, M., N. Beloglazova, R. Flick, C. Graham, T. Skarina, B. Nocek, A. Gagarinova, O. Pogoutse, G. Brown, A. Binkowski, S. Phanse, A. Joachimiak, E. V. Koonin, A. Savchenko, A. Emili, J. Greenblatt, A. M. Edwards and A. F. Yakunin (2011). "A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair." Mol Microbiol **79**(2): 484-502.

Bikard, D., A. Hatoum-Aslan, D. Mucida and L. A. Marraffini (2012). "CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection." Cell Host Microbe **12**(2): 177-186.

Bolger, A. M., M. Lohse and B. Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data." Bioinformatics **30**(15): 2114-2120.

Burrus, V. and M. K. Waldor (2004). "Shaping bacterial genomes with integrative and conjugative elements." Res Microbiol **155**(5): 376-386.

Cady, K. C., J. Bondy-Denomy, G. E. Heussler, A. R. Davidson and G. A. O'Toole (2012). "The CRISPR/Cas adaptive immune system of Pseudomonas aeruginosa mediates resistance to naturally occurring and engineered phages." J Bacteriol **194**(21): 5728-5738.

Crooks, G. E., G. Hon, J. M. Chandonia and S. E. Brenner (2004). "WebLogo: a sequence logo generator." Genome Res **14**(6): 1188-1190.

Datsenko, K. A., K. Pougach, A. Tikhonov, B. L. Wanner, K. Severinov and E. Semenova (2012). "Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system." Nat Commun **3**: 945.

Ensminger, A. W., Y. Yassin, A. Miron and R. R. Isberg (2012). "Experimental evolution of Legionella pneumophila in mouse macrophages leads to strains with altered determinants of environmental survival." PLoS Pathog **8**(5): e1002731.

Fineran, P. C. and E. Charpentier (2012). "Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information." Virology **434**(2): 202-209.

Fineran, P. C., M. J. Gerritzen, M. Suarez-Diez, T. Kunne, J. Boekhorst, S. A. van Hijum, R. H. Staals and S. J. Brouns (2014). "Degenerate target sites mediate rapid primed CRISPR adaptation." Proc Natl Acad Sci U S A **111**(16): E1629-1638.

Frost, L. S., R. Leplae, A. O. Summers and A. Toussaint (2005). "Mobile genetic elements: the agents of open source evolution." Nat Rev Microbiol **3**(9): 722-732.

Heler, R., L. A. Marraffini and D. Bikard (2014). "Adapting to new threats: the generation of memory by CRISPR-Cas immune systems." Mol Microbiol **93**(1): 1-9.

Horvath, P. and R. Barrangou (2010). "CRISPR/Cas, the immune system of bacteria and archaea." Science **327**(5962): 167-170.

Hoyland-Kroghsbo, N. M., J. Paczkowski, S. Mukherjee, J. Broniewski, E. Westra, J. Bondy-Denomy and B. L. Bassler (2016). "Quorum sensing controls the Pseudomonas aeruginosa CRISPR-Cas adaptive immune system." Proc Natl Acad Sci U S A.

Jackson, S. A., R. E. McKenzie, R. D. Fagerlund, S. N. Kieper, P. C. Fineran and S. J. Brouns (2017). "CRISPR-Cas: Adapting to change." Science **356**(6333).

Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones and M. A. Marra (2009). "Circos: an information aesthetic for comparative genomics." Genome Res **19**(9): 1639-1645.

Kunne, T., S. N. Kieper, J. W. Bannenberg, A. I. Vogel, W. R. Miellet, M. Klein, M. Depken, M. Suarez-Diez and S. J. Brouns (2016). "Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation." Mol Cell **63**(5): 852-864.

Labrie, S. J., J. E. Samson and S. Moineau (2010). "Bacteriophage resistance mechanisms." Nat Rev Microbiol **8**(5): 317-327.

Leenay, R. T., K. R. Maksimchuk, R. A. Slotkowski, R. N. Agrawal, A. A. Gomaa, A. E. Briner, R. Barrangou and C. L. Beisel (2016). "Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems." Mol Cell **62**(1): 137-147.

Levy, A., M. G. Goren, I. Yosef, O. Auster, M. Manor, G. Amitai, R. Edgar, U. Qimron and R. Sorek (2015). "CRISPR adaptation biases explain preference for acquisition of foreign DNA." Nature **520**(7548): 505-510.

Li, M., R. Wang and H. Xiang (2014). "Haloarcula hispanica CRISPR authenticates PAM of a target sequence to prime discriminative adaptation." Nucleic Acids Res **42**(11): 7226-7235.

Li, M., R. Wang, D. Zhao and H. Xiang (2014). "Adaptation of the Haloarcula hispanica CRISPR-Cas system to a purified virus strictly requires a priming process." Nucleic Acids Res **42**(4): 2483-2492.

Magoc, T. and S. L. Salzberg (2011). "FLASH: fast length adjustment of short reads to improve genome assemblies." Bioinformatics **27**(21): 2957-2963.

Majsec, K., E. L. Bolt and I. Ivancic-Bace (2016). "Cas3 is a limiting factor for CRISPR-Cas immunity in Escherichia coli cells lacking H-NS." BMC Microbiol **16**: 28.

Makarova, K. S., D. H. Haft, R. Barrangou, S. J. Brouns, E. Charpentier, P. Horvath, S. Moineau, F. J. Mojica, Y. I. Wolf, A. F. Yakunin, J. van der Oost and E. V. Koonin (2011). "Evolution and classification of the CRISPR-Cas systems." Nat Rev Microbiol **9**(6): 467-477.

Makarova, K. S., Y. I. Wolf, O. S. Alkhnbashi, F. Costa, S. A. Shah, S. J. Saunders, R. Barrangou, S. J. Brouns, E. Charpentier, D. H. Haft, P. Horvath, S. Moineau, F. J. Mojica, R. M. Terns, M. P. Terns, M. F. White, A. F. Yakunin, R. A. Garrett, J. van der Oost, R. Backofen and E. V. Koonin (2015). "An updated evolutionary classification of CRISPR-Cas systems." Nat Rev Microbiol **13**(11): 722-736.

Marraffini, L. A. (2015). "CRISPR-Cas immunity in prokaryotes." Nature **526**(7571): 55-61.

Marraffini, L. A. and E. J. Sontheimer (2008). "CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA." Science **322**(5909): 1843-1845.

539 Marraffini, L. A. and E. J. Sontheimer (2010). "CRISPR interference: RNA-directed adaptive immunity in
540 bacteria and archaea." Nat Rev Genet **11**(3): 181-190.
541 Mojica, F. J., C. Diez-Villasenor, J. Garcia-Martinez and C. Almendros (2009). "Short motif sequences
542 determine the targets of the prokaryotic CRISPR defence system." Microbiology **155**(Pt 3): 733-740.
543 Mulepati, S. and S. Bailey (2013). "In vitro reconstitution of an Escherichia coli RNA-guided immune
544 system reveals unidirectional, ATP-dependent degradation of DNA target." J Biol Chem **288**(31): 22184-
545 22192.
546 Nunez, J. K., L. B. Harrington, P. J. Kranzusch, A. N. Engelman and J. A. Doudna (2015). "Foreign DNA
547 capture during CRISPR-Cas adaptive immunity." Nature **527**(7579): 535-538.
548 Nunez, J. K., P. J. Kranzusch, J. Noeske, A. V. Wright, C. W. Davies and J. A. Doudna (2014). "Cas1-Cas2
549 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity." Nat Struct Mol
550 Biol **21**(6): 528-534.
551 Nunez, J. K., A. S. Lee, A. Engelman and J. A. Doudna (2015). "Integrase-mediated spacer acquisition
552 during CRISPR-Cas adaptive immunity." Nature **519**(7542): 193-198.
553 Paez-Espino, D., W. Morovic, C. L. Sun, B. C. Thomas, K. Ueda, B. Stahl, R. Barrangou and J. F. Banfield
554 (2013). "Strong bias in the bacterial CRISPR elements that confer immunity to phage." Nat Commun **4**:
555 1430.
556 Paez-Espino, D., I. Sharon, W. Morovic, B. Stahl, B. C. Thomas, R. Barrangou and J. F. Banfield (2015).
557 "CRISPR immunity drives rapid phage genome evolution in Streptococcus thermophilus." MBio **6**(2).
558 Patterson, A. G., S. A. Jackson, C. Taylor, G. B. Evans, G. P. Salmond, R. Przybilski, R. H. Staals and P. C.
559 Fineran (2016). "Quorum Sensing Controls Adaptive Immunity through the Regulation of Multiple
560 CRISPR-Cas Systems." Mol Cell **64**(6): 1102-1108.
561 Qiu, Y., S. Wang, Z. Chen, Y. Guo and Y. Song (2016). "An Active Type I-E CRISPR-Cas System Identified in
562 Streptomyces avermitilis." PLoS One **11**(2): e0149533.
563 Rao, C., H. Benhabib and A. W. Ensminger (2013). "Phylogenetic reconstruction of the Legionella
564 pneumophila Philadelphia-1 laboratory strains through comparative genomics." PLoS One **8**(5): e64129.
565 Rao, C., C. Guyard, C. Pelaz, J. Wasserscheid, J. Bondy-Denomy, K. Dewar and A. W. Ensminger (2016).
566 "Active and adaptive Legionella CRISPR-Cas reveals a recurrent challenge to the pathogen." Cell
567 Microbiol **18**(10): 1319-1338.
568 Redding, S., S. H. Sternberg, M. Marshall, B. Gibb, P. Bhat, C. K. Guegler, B. Wiedenheft, J. A. Doudna and
569 E. C. Greene (2015). "Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas
570 System." Cell **163**(4): 854-865.
571 Richter, C., R. L. Dy, R. E. McKenzie, B. N. Watson, C. Taylor, J. T. Chang, M. B. McNeil, R. H. Staals and P.
572 C. Fineran (2014). "Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer
573 acquisition, bi-directionally from the primed protospacer." Nucleic Acids Res **42**(13): 8516-8526.
574 Richter, C., T. Gristwood, J. S. Clulow and P. C. Fineran (2012). "In vivo protein interactions and complex
575 formation in the Pectobacterium atrosepticum subtype I-F CRISPR/Cas System." PLoS One **7**(12): e49549.
576 Savitskaya, E., E. Semenova, V. Dedkov, A. Metlitskaya and K. Severinov (2013). "High-throughput
577 analysis of type I-E CRISPR/Cas spacer acquisition in E. coli." RNA Biol **10**(5): 716-725.
578 Semenova, E., E. Savitskaya, O. Musharova, A. Strotskaya, D. Vorontsova, K. A. Datsenko, M. D.
579 Logacheva and K. Severinov (2016). "Highly efficient primed spacer acquisition from targets destroyed
580 by the Escherichia coli type I-E CRISPR-Cas interfering complex." Proc Natl Acad Sci U S A **113**(27): 7626-
581 7631.
582 Severinov, K., I. Ispolatov and E. Semenova (2016). "The Influence of Copy-Number of Targeted
583 Extrachromosomal Genetic Elements on the Outcome of CRISPR-Cas Defense." Front Mol Biosci **3**: 45.
584 Shmakov, S., E. Savitskaya, E. Semenova, M. D. Logacheva, K. A. Datsenko and K. Severinov (2014).
585 "Pervasive generation of oppositely oriented spacers during CRISPR adaptation." Nucleic Acids Res **42**(9):
586 5907-5916.

Sinkunas, T., G. Gasiunas, S. P. Waghmare, M. J. Dickman, R. Barrangou, P. Horvath and V. Siksnys (2013). "In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus." EMBO J **32**(3): 385-394.

Staals, R. H., S. A. Jackson, A. Biswas, S. J. Brouns, C. M. Brown and P. C. Fineran (2016). "Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system." Nat Commun **7**: 12853.

Sternberg, S. H., H. Richter, E. Charpentier and U. Qimron (2016). "Adaptation in CRISPR-Cas Systems." Mol Cell **61**(6): 797-808.

Swarts, D. C., C. Mosterd, M. W. van Passel and S. J. Brouns (2012). "CRISPR interference directs strand specific spacer acquisition." PLoS One **7**(4): e35888.

Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar (2013). "MEGA6: Molecular Evolutionary Genetics Analysis version 6.0." Mol Biol Evol **30**(12): 2725-2729.

van der Oost, J., E. R. Westra, R. N. Jackson and B. Wiedenheft (2014). "Unravelling the structural and mechanistic basis of CRISPR-Cas systems." Nat Rev Microbiol **12**(7): 479-492.

van Houte, S., A. K. Ekroth, J. M. Broniewski, H. Chabas, B. Ashby, J. Bondy-Denomy, S. Gandon, M. Boots, S. Paterson, A. Buckling and E. R. Westra (2016). "The diversity-generating benefits of a prokaryotic adaptive immune system." Nature **532**(7599): 385-388.

Vorontsova, D., K. A. Datsenko, S. Medvedeva, J. Bondy-Denomy, E. E. Savitskaya, K. Pougach, M. Logacheva, B. Wiedenheft, A. R. Davidson, K. Severinov and E. Semenova (2015). "Foreign DNA acquisition by the I-F CRISPR-Cas system requires all components of the interference machinery." Nucleic Acids Res **43**(22): 10848-10860.

Wang, J., J. Li, H. Zhao, G. Sheng, M. Wang, M. Yin and Y. Wang (2015). "Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems." Cell **163**(4): 840-853.

Wiedenheft, B., E. van Duijn, J. B. Bultema, S. P. Waghmare, K. Zhou, A. Barendregt, W. Westphal, A. J. Heck, E. J. Boekema, M. J. Dickman and J. A. Doudna (2011). "RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions." Proc Natl Acad Sci U S A **108**(25): 10092-10097.

Wright, A. V., J. K. Nunez and J. A. Doudna (2016). "Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering." Cell **164**(1-2): 29-44.

Xue, C., A. S. Seetharam, O. Musharova, K. Severinov, S. J. Brouns, A. J. Severin and D. G. Sashital (2015). "CRISPR interference and priming varies with individual spacer sequences." Nucleic Acids Res **43**(22): 10831-10847.

Xue, C., N. R. Whitis and D. G. Sashital (2016). "Conformational Control of Cascade Interference and Priming Activities in CRISPR Immunity." Mol Cell **64**(4): 826-834.

Yosef, I., D. Shitrit, M. G. Goren, D. Burstein, T. Pupko and U. Qimron (2013). "DNA motifs determining the efficiency of adaptation into the Escherichia coli CRISPR array." Proc Natl Acad Sci U S A **110**(35): 14396-14401.

Figure Legends

Figure 1: Primed spacer acquisition by *L. pneumophila* type I-C CRISPR-Cas occurs in a strand-biased manner.

628 **A.** Schematic workflow to characterize primed spacer acquisition. Escaped transformants of targeted
629 plasmids were passaged for 15 generations without antibiotic selection to enrich for spacer acquisition.
630 CRISPR loci were PCR amplified and adapted arrays were further isolated through gel size selection.
631 Amplicons were subjected to Illumina sequencing, and acquired spacers were extracted from raw reads
632 and mapped to either the plasmid or the bacterial chromosome. **B.** The vast majority of acquired spacers
633 during priming were derived from the plasmid instead of the chromosome. **C-D.** Circos plots of acquired
634 spacers mapped to the pSp1 priming plasmid where the priming protospacer (identical to the spacer 1
635 sequence from the type I-C system) is either on the plus (+) strand (**C**) or on the minus (–) strand (**D**). In
636 the strand-specific mappings, bars protruding inside and outside of plasmid circle represent spacers
637 matching the minus and plus strand of the plasmid, respectively, and the height of bars indicates the
638 number of spacers mapped to indicated positions. Note that a secondary scale was used for plasmid loci
639 acquired at a frequency of over 10% of all spacers. The frequency of the major spacer acquisition hotspot
640 is indicated. To numerically represent the overall spacer acquisition patterns, the plasmid is divided into
641 four geographic fractions relative to the priming protospacer (denoted by the colored rectangle): the 5′
642 half (Left) and the 3′ half (Right) on the + strand, and the 3′ half (Left) and the 5′ half (Right) on the –
643 strand. **E.** A merged view of the two mappings was created where overlapped coverages were shown in
644 cyan. **F.** Priming by an imperfect target with a seed mismatch showed similar overall patterns of spacer
645 acquisition as priming using a *bona fide* target. Each Circos plot in the figure represents the average of
646 two independent biological replicates.

647 Figure 2: The spacer acquisition hotspot is reliant on its internal sequence.

648 **A.** Mutations were introduced, with *repC* codons maintained, at the upstream, PAM, internal or
649 downstream sequences of the major spacer acquisition hotspot to examine factors contributing to the high
650 acquisition frequency. **B-C.** The mutation at the PAM dramatically reduced the acquisition frequency at
651 the hotspot, while the overall patterns of spacer acquisition remained largely unaffected. Note that the
652 mutation does not eliminate available PAM, but shifted the PAM 1 nt away. **D-F.** Mutations within, but

not flanking, the hotspot also largely decreased the acquisition frequency at the hotspot. Each Circos plot in the figure represents the average of two independent biological replicates.

Figure 3: PAM preference and acquisition inaccuracies in primed adaptation.

A. Schematic representation of spacer selection by the adaptation machinery. In most cases, the machinery extracts the double-stranded sequence immediately downstream of the PAM, with an inexact molecular ruler at the PAM-distal end. Less frequently, the machinery shifts a few nucleotides downstream (+ slip) or upstream (– slip) at the PAM-proximal end, causing slipping events. Flipping events were also observed where the double-stranded DNA substrates were incorporated in an opposite orientation into the CRISPR array. **B.** Based on the PAM localization within the upstream or downstream junction, spacer acquisition events were categorized into different types, with their frequencies shown for both pSp1(+) and pSp1(–) priming. Note that for the alternative TTT PAM, spacers with the first nucleotide being C were excluded as these were classified as potential -1 nt slipping events. **C.** Sequence Logo of the upstream and downstream 20 nt junctions of indicated categories of spacer acquisitions from pSp1(+). **D.** Length distribution of all acquired spacers primed by pSp1(+) and pSp1(–), compared with the native spacers from type I-C CRISPR loci in *L. pneumophila* ST222 strains. **E.** Length distribution of acquired spacers from each slipping category.

Figure 4: PAM preference for the *L. pneumophila* type I-C CRISPR-Cas interference.

A. Schematic workflow to characterize functional PAMs for CRISPR interference. A pool of plasmids containing the spacer 1 sequence and a random trinucleotide PAM was generated and transformed into either *L. pneumophila* str. Toronto-2005 *wild type* or Δ*cas3*, and the abundance of each PAM sequence in the pool was quantified through Illumina sequencing. **B.** Pooled abundances were derived by normalizing the ratio of each PAM in the *wild type* transformant pool to that in the Δ*cas3* pool. These relative abundances categorized PAM sequences into different preferences for CRISPR interference. **C.** Individual plasmid transformation efficiency assay confirmed, with a lower sensitivity, the observations

in the Illumina-based pooled assay. Plasmids containing either spacer 1 and an indicated PAM or a scrambled control sequence were electroporated into *L. pneumophila* str. Toronto-2005 *wild type*. The relative transformation efficiency is calculated by normalizing transformation efficiency of the spacer 1 plasmids to that of the control plasmid. Error bars represent the SEM of three biological replicates. **D.** Spacer acquisition was also observed in escaped transformants of targeted plasmids with the alternative TTT PAM.

Figure 5: Highly-efficient interference leads to spacer loss rather than spacer acquisition.

A. Spacer loss, rather than spacer acquisition, was seen in escaped transformants when the array-minimized (Sp1-only) CRISPR-Cas system highly efficiently (>99.9% by relative transformation efficiency) protects against targeted plasmids. Plasmid transformation efficiency assay was performed to measure interference efficiencies of the modified CRISPR-Cas system (compared with those of the original system, shown by the upper lines). The resulting transformants were examined by PCR amplification for the dynamics of CRISPR loci. **B.** Quantification of spacer acquisition and spacer loss frequencies in plasmid transformants (colonies pool without further passages) of *L. pneumophila* str. Toronto-2005 *wild type* or Sp1-only. Spacer loss frequency is not determined for *wild type* transformants because only the leader-end of the CRISPR array was surveyed by PCR. **C.** Most spacer loss events retained the downstream non-consensus repeat (R2), shown in the bar graph, consistent with a mechanism of homologous recombination between the two flanking repeats (R1 and R2). The few Rx repeats contain mismatches to both R1 and R2 and may derive from sequencing errors. **D.** Spacer loss was also observed in another native type I-C CRISPR-Cas system in *L. pneumophila* str. Toronto-2000 where the first spacer (SpT1) is highly efficient in interference. Plasmids containing a targeted sequence for one of the three indicated spacers showed different relative transformation efficiencies in *L. pneumophila* str. Toronto-2000. The resulting transformants were tested for spacer acquisition or spacer loss by PCR using indicated primers. Error bars represent the SEM of three biological replicates.

Figure 6: Schematic summary of primed spacer acquisition in type I CRISPR-Cas.

Primed spacer sampling is separated into three steps: 1) priming initiation where the Cascade-crRNA complex binds to the targeted DNA and recruits the adaptation machinery; 2) the adaptation complex surveying the plasmid, consistent with a 3′ to 5′ sliding with variable strand specificity; 3) spacer selection from another plasmid locus by the adaptation complex upon recognition of an appropriate PAM sequence. Comparisons of each step in the type I-C system versus type I-B (Li, Wang et al. 2014), type I-E (Datsenko, Pougach et al. 2012, Savitskaya, Semenova et al. 2013, Fineran, Gerritzen et al. 2014) and type I-F systems (Richter, Dy et al. 2014, Staals, Jackson et al. 2016), show similarities and distinctions in molecular mechanisms. Note that most previous studies used an imperfectly-targeted priming sequence with mutations in either the PAM or the protospacer sequence.

Supplemental Information

Supplemental Figure 1: Analyses of acquired spacers from pSp1(+), related to Fig. 3.

A. PAM frequencies of all acquired spacers derived from pSp1(+). Note that the canonical TTC and alternative TTT PAMs are the two most frequent motifs, followed by TCN motifs that are likely due to +1 nt slips. **B.** Spacers with the alternative TTT PAM (red) showed independent localizations relative to those with the canonical TTC PAM (grey). Note that the plot scale is not continuous (disrupted by grey rings) in order to fully represent a wide range of spacer acquisition efficiencies. **C.** The major spacer acquisition hotspot exemplifies imprecise size selection, slipping (blue) and flipping (green) events. Unique spacers mapped to either strand of the region were categorized and counted regarding their frequencies.

Supplemental Figure 2: Cas1-based phylogenies of select type I systems.

723 Cas1 protein sequences were retrieved from genomes of *Legionella pneumophila* str. Toronto-2005
724 (Genbank CP012019), *Haloarcula hispanica* str. ATCC 33960 (Genbank CP002922), *Escherichia coli* str.
725 K-12 (Genbank NC_000913) and *Pectobacterium atrosepticum* str. SCRI1043 (Genbank BX950851).
726 These sequences were aligned using the ClustalW option and subjected to the Maximum Likelihood
727 phylogenetic tree construction using the LG model with 500 bootstrap iterations in MEGA v6.0(Tamura,
728 Stecher et al. 2013).

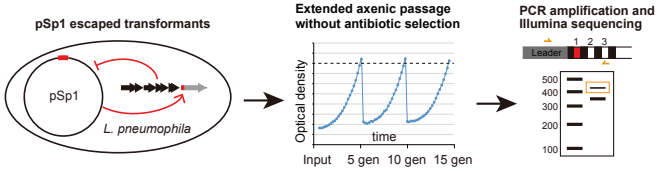
729 Table S1: Spacer acquisition frequencies at TTC PAM sites on the pSp1 priming plasmid.

730 Table S2: Oligos used in this study.

731 Supplemental File: Full nucleotide sequence of pSp1(+).

Figure 1: Primed spacer acquisition by *L. pneumophila* type I-C CRISPR-Cas occurs in a strand-biased manner.

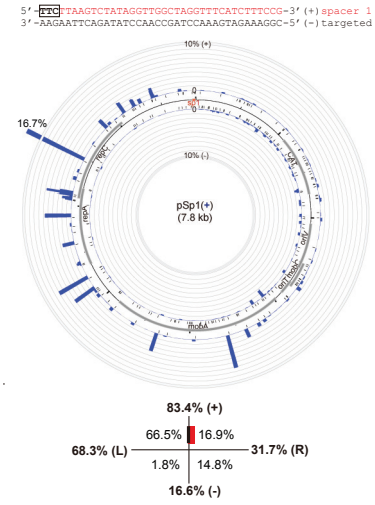
A



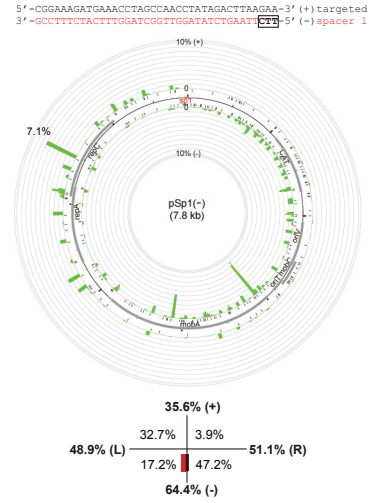
B

Toronto-2005 transformants	pSp1(+)	pSp1(-)	pSp1(+) seed T1A
Total acquired spacers	1,964,679	2,363,914	2,312,824
Spacers targeting Plasmid	99.74%	99.81%	99.79%
Spacers targeting Chromosome	0.01%	0.01%	0.01%
Spacers matching Unknown	0.25%	0.18%	0.21%

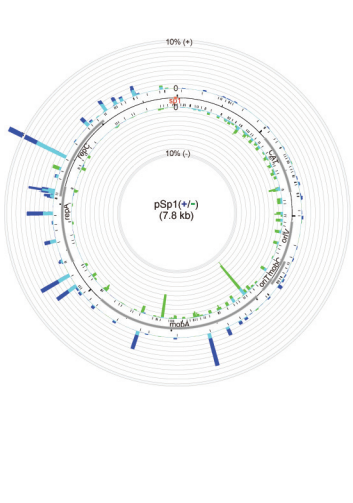
C



D



E



F

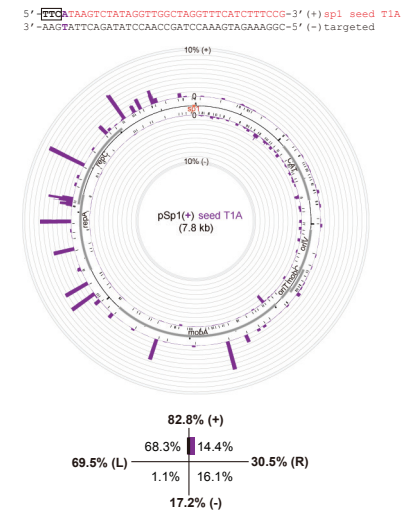


Figure 2: The spacer acquisition hotspot is reliant on its internal sequence.

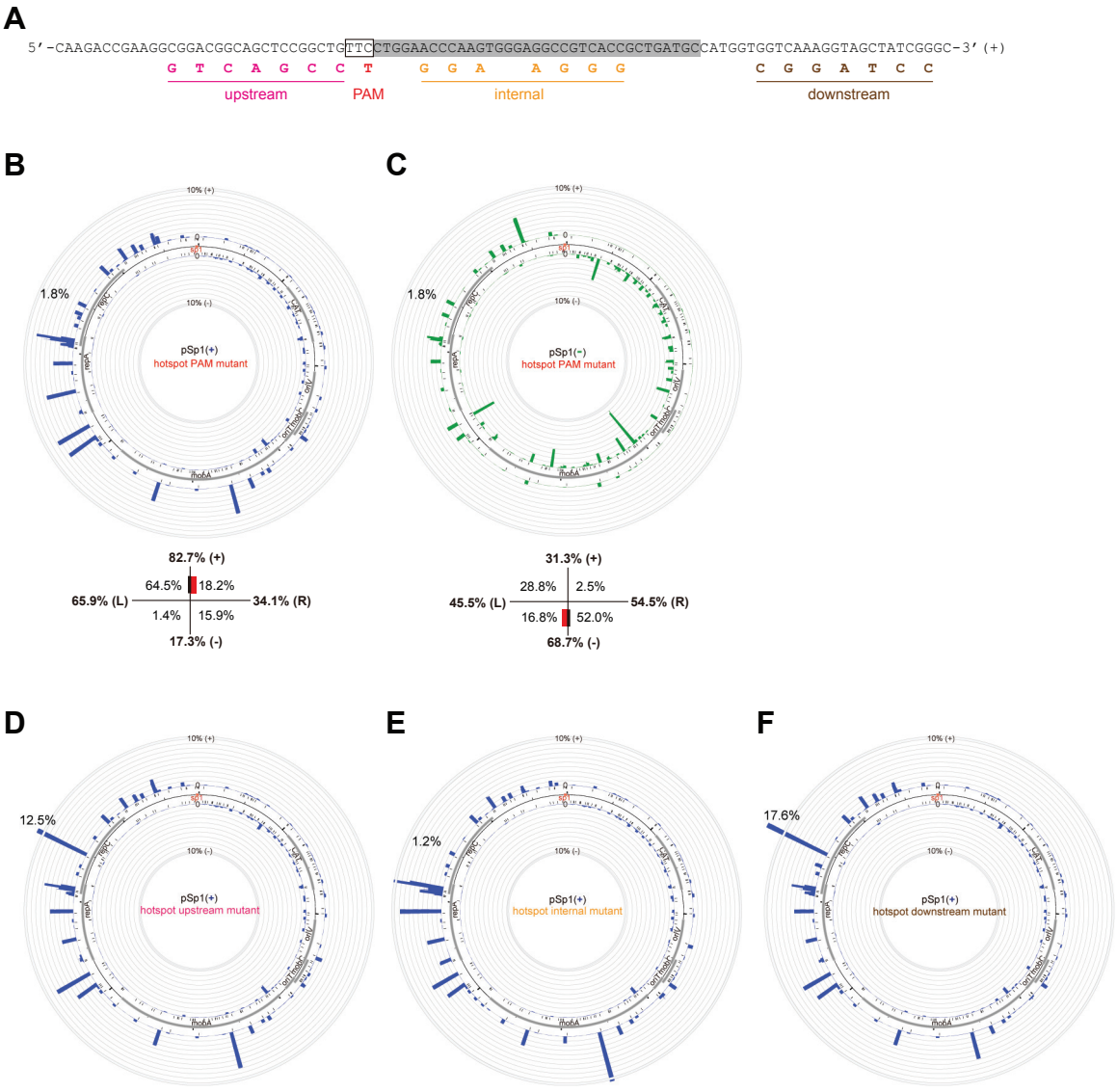


Figure 3: PAM preference and acquisition inaccuracies in primed adaptation.

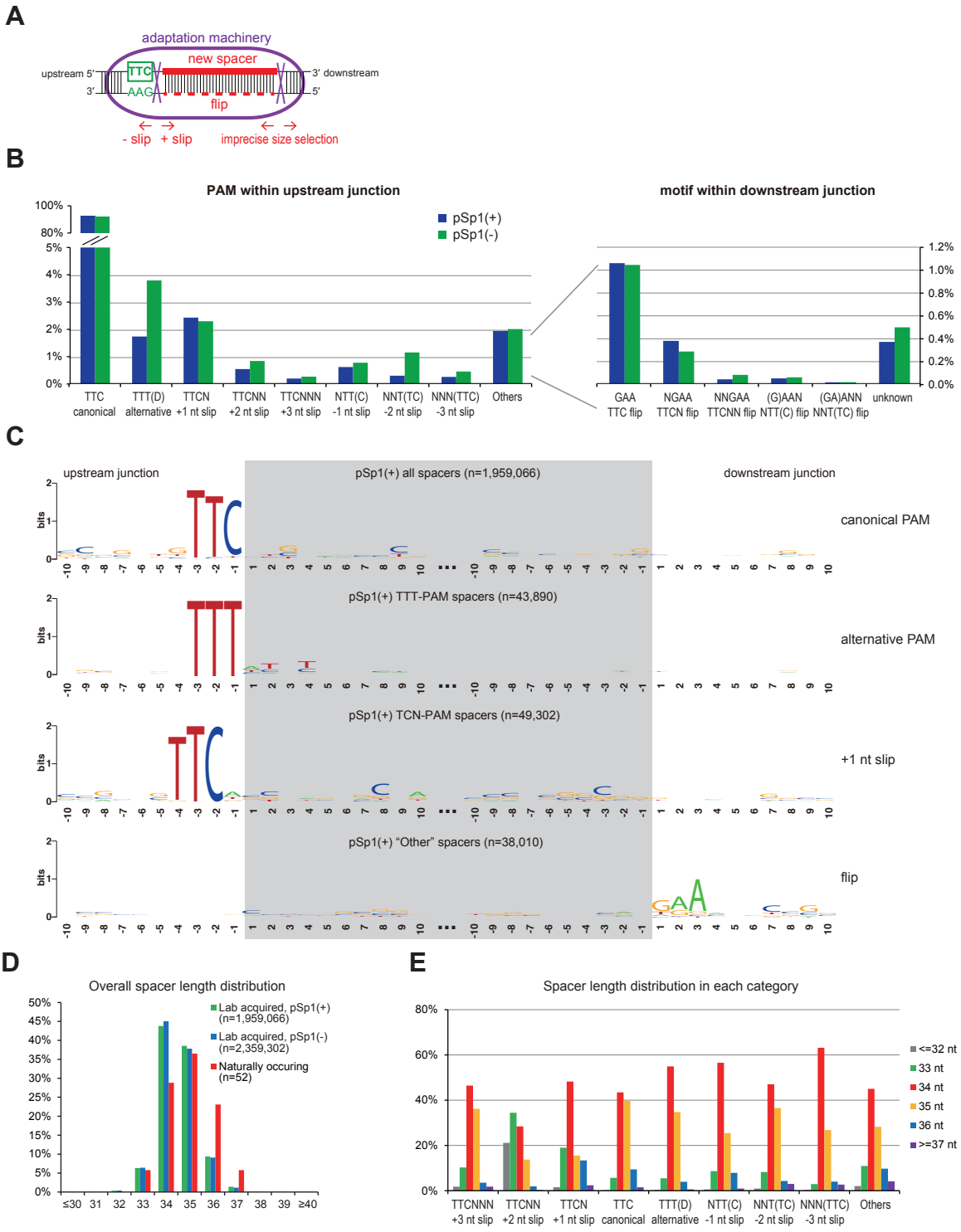


Figure 4: PAM preference for *L. pneumophila* type I-C CRISPR-Cas interference.

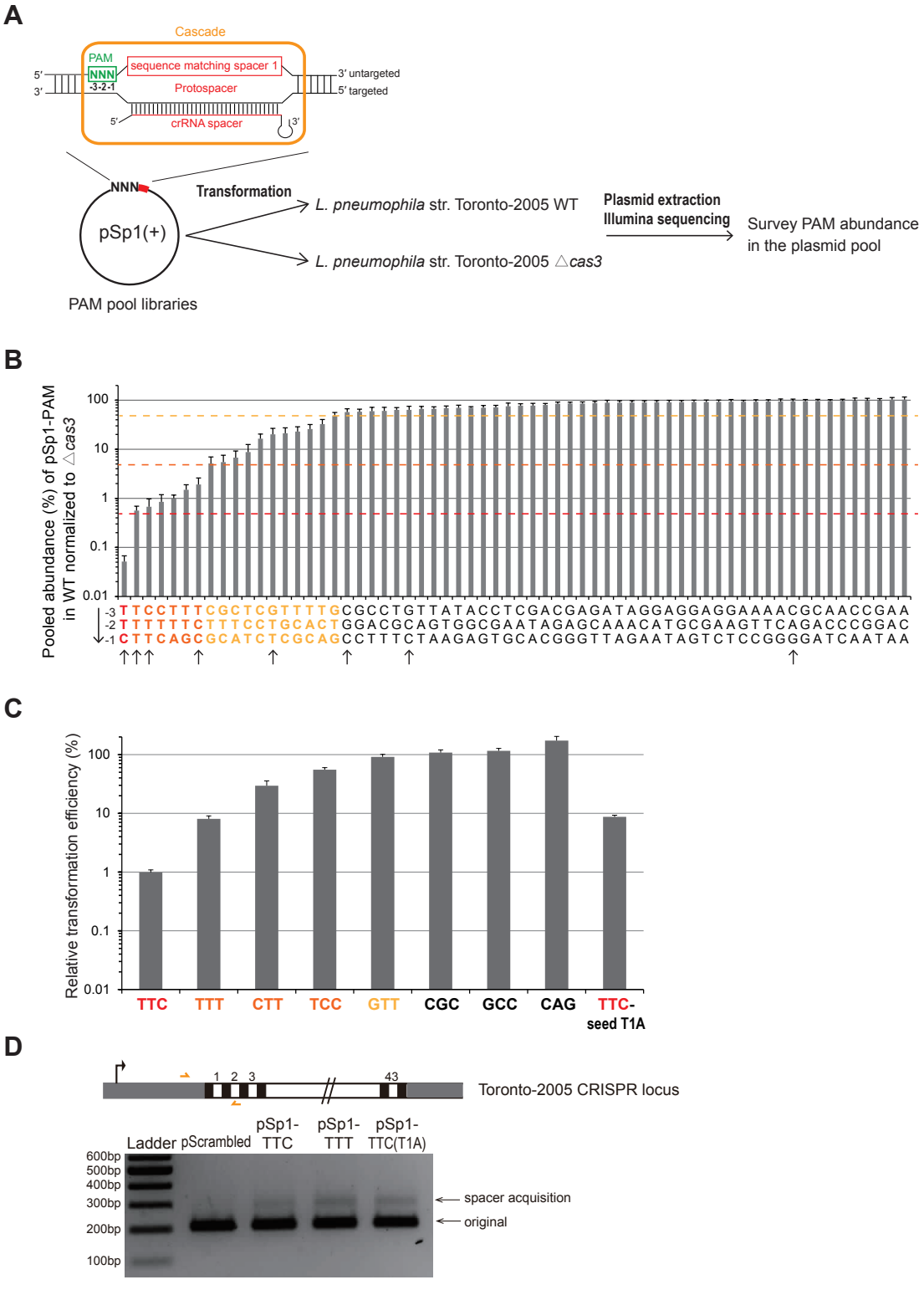
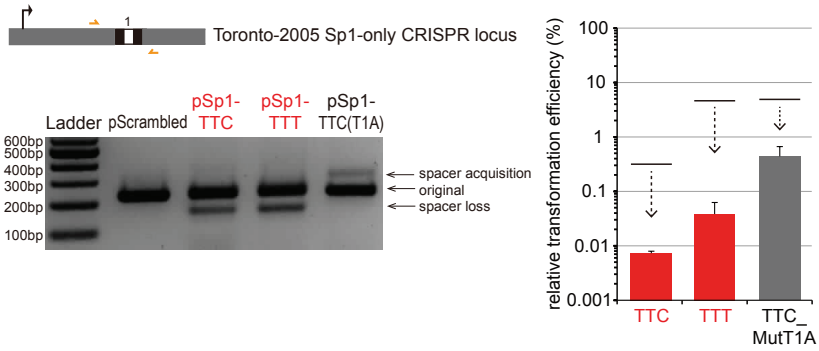


Figure 5: Highly efficient interference leads to spacer loss rather than spacer acquisition.

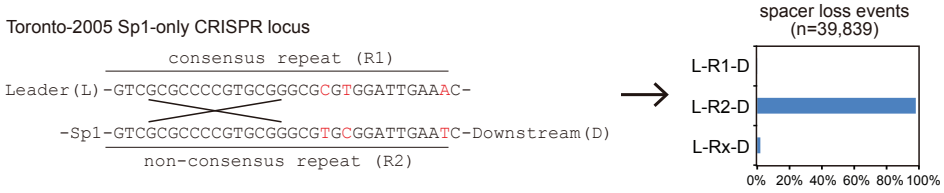
A



B

transformed colony pool	Toronto-2005 WT pSp1(+)	Toronto-2005 Sp1-only pScrambled	Toronto-2005 Sp1-only pSp1(+)
spacer acquisition frequency	1.05%	0.03%	0.01%
spacer loss frequency	ND	0.07%	39.14%

C



D

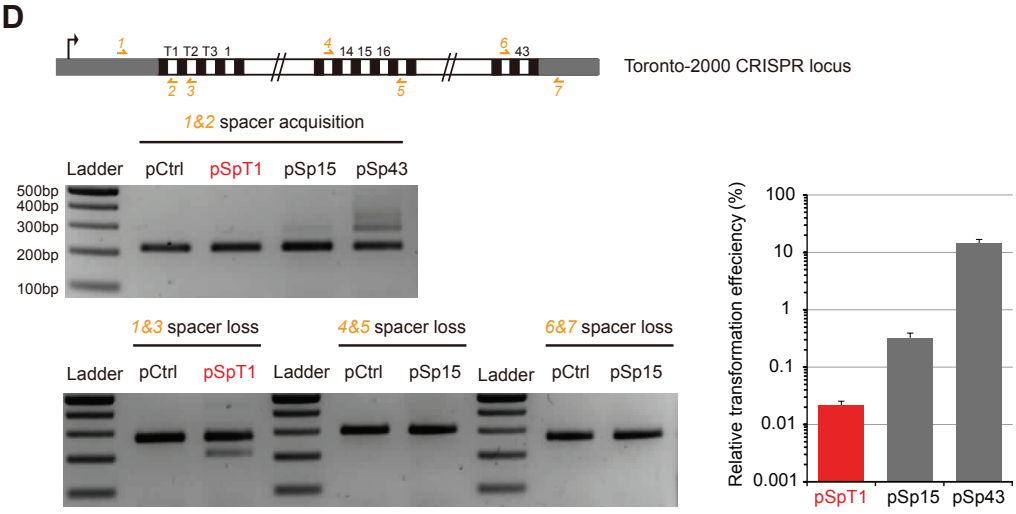
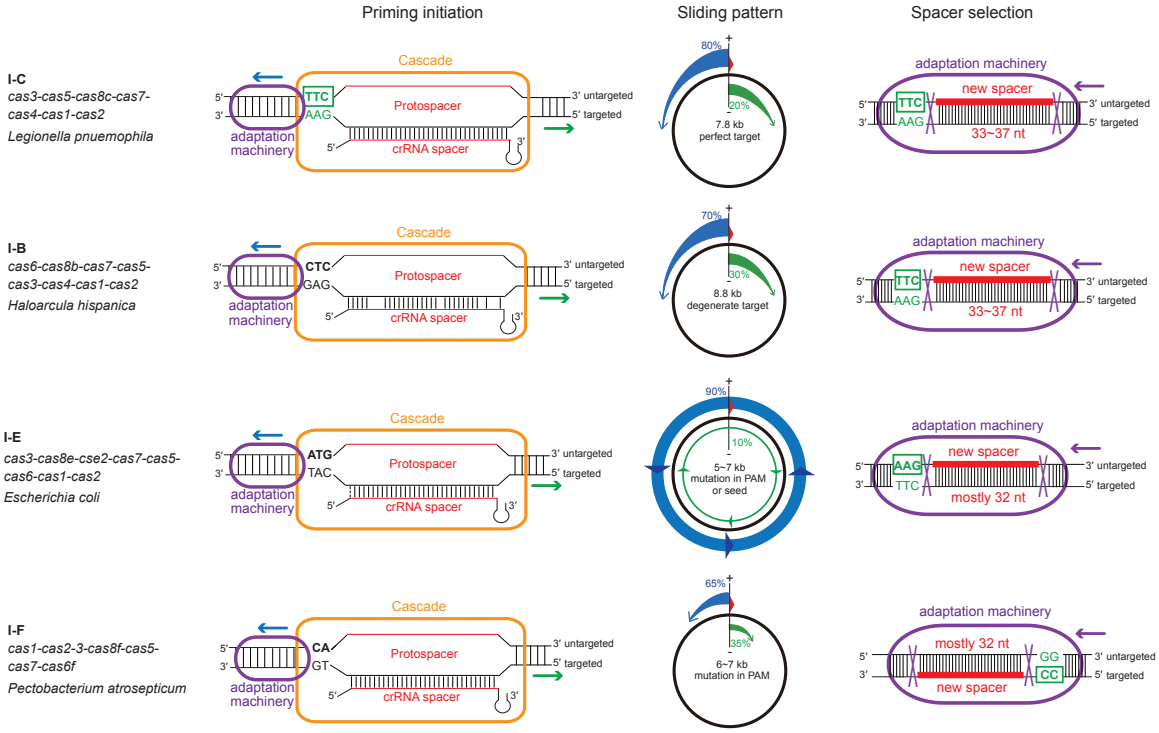
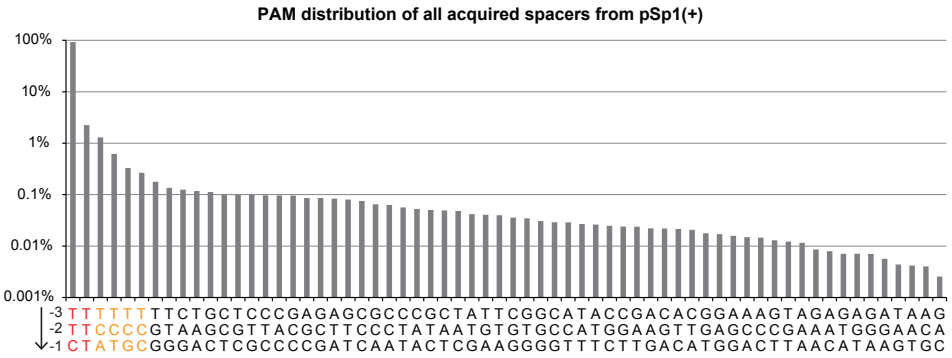


Figure 6: Schematic summary of primed spacer acquisition in type I CRISPR-Cas.

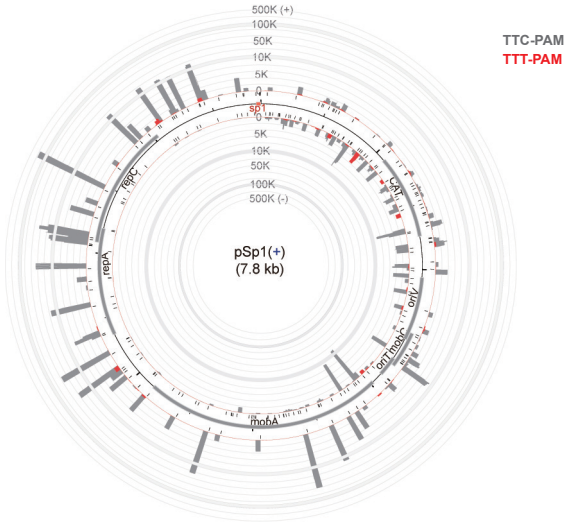


Supplemental Figure 1：Analyses of acquired spacers from pSp1(+), related to Fig. 3.

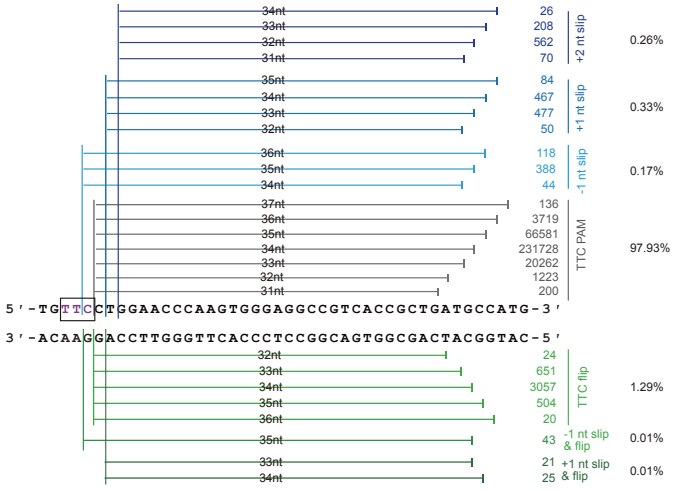
A



B



C



Supplemental Figure 2: Cas1-based phylogenies of select type I systems.

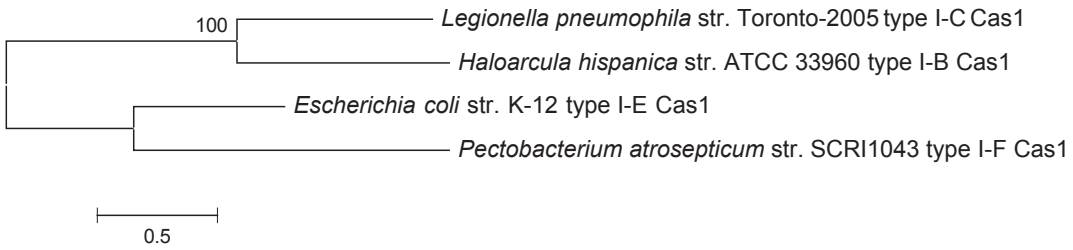


Table S1. Spacer acquisition frequencies at TTC PAM sites on the pSp1 priming plasmid.

pSp1(+) priming			**pSp1(-) priming**		
strand	TTC_position	acquisition frequency	strand	TTC_position	acquisition frequency
plus	6470	16.5184%	minus	3087	7.8981%
plus	3596	7.0230%	plus	6470	7.0235%
plus	5234	6.6901%	minus	4133	5.0959%
plus	5884	5.6131%	plus	5234	2.5909%
plus	5108	4.8968%	plus	5884	2.1271%
plus	4321	3.5812%	plus	5108	1.9950%
plus	6112	2.8217%	plus	6341	1.8248%
plus	6097	2.7199%	plus	7094	1.6486%
plus	7094	2.5139%	minus	4380	1.5024%
plus	5606	2.4831%	minus	1699	1.4929%
plus	6862	2.1105%	minus	799	1.4809%
plus	3400	1.7272%	minus	3239	1.4792%
plus	7417	1.5994%	minus	2073	1.4621%
minus	3087	1.5726%	plus	6112	1.4550%
plus	7213	1.4986%	minus	2212	1.3178%
plus	7282	1.0367%	minus	2699	1.3157%
minus	3239	0.9269%	minus	1989	1.3087%
plus	3262	0.8125%	plus	5606	1.2709%
plus	6055	0.7555%	plus	6097	1.1952%
plus	6260	0.7463%	minus	1015	1.1087%
plus	6052	0.7435%	minus	1823	1.0745%
plus	2731	0.7151%	plus	3596	1.0684%
minus	799	0.6275%	plus	6862	1.0425%
plus	5022	0.6254%	minus	2349	1.0158%
plus	6064	0.6225%	minus	1858	1.0002%
plus	4837	0.6219%	minus	1138	0.9722%
plus	3193	0.6024%	minus	4893	0.9552%
plus	6910	0.5510%	minus	6472	0.9377%
plus	5395	0.5420%	minus	1500	0.9307%
minus	1699	0.5343%	minus	1688	0.9254%
minus	2699	0.4983%	minus	5121	0.9198%
plus	7422	0.4944%	minus	3060	0.9019%
minus	2073	0.4906%	minus	3326	0.8664%
minus	1989	0.4516%	plus	7417	0.8643%
plus	7695	0.4365%	minus	3744	0.8510%
plus	4579	0.4227%	minus	1402	0.7839%
minus	1015	0.4224%	plus	4321	0.7791%
minus	2212	0.4076%	plus	7213	0.7652%
minus	1138	0.3946%	plus	7695	0.7573%
minus	1823	0.3847%	plus	7282	0.7129%
plus	5074	0.3840%	minus	671	0.6986%
plus	6341	0.3778%	minus	3553	0.6526%
minus	1688	0.3760%	minus	3977	0.6384%
plus	1998	0.3717%	minus	1294	0.6295%
minus	1858	0.3701%	minus	6186	0.6272%
plus	3940	0.3497%	minus	3702	0.5875%
plus	7455	0.3054%	minus	3425	0.5684%
plus	5791	0.3052%	minus	3447	0.5320%
minus	671	0.3045%	plus	6260	0.5101%
minus	1500	0.3039%	minus	522	0.4916%
plus	2663	0.2957%	minus	6630	0.4497%
plus	5384	0.2904%	minus	1086	0.4365%
minus	2349	0.2850%	minus	2999	0.4254%
minus	1402	0.2758%	minus	399	0.4251%
plus	1656	0.2692%	minus	2011	0.4198%
plus	261	0.2468%	minus	4908	0.4023%
minus	399	0.2306%	minus	1465	0.3963%
minus	522	0.2285%	minus	942	0.3752%
plus	2887	0.2237%	plus	6055	0.3720%
minus	282	0.2125%	plus	3400	0.3667%
plus	595	0.2090%	minus	772	0.3647%
plus	2612	0.2038%	plus	7422	0.3613%
plus	1802	0.1897%	minus	7677	0.3580%
plus	4232	0.1891%	minus	2117	0.3548%
minus	3060	0.1878%	minus	282	0.3318%
plus	2434	0.1824%	minus	5278	0.3256%
minus	1294	0.1800%	plus	5395	0.3239%
plus	6657	0.1788%	minus	5076	0.3210%
plus	7200	0.1726%	minus	716	0.3126%
plus	2734	0.1687%	plus	6064	0.3022%
minus	230	0.1643%	minus	3726	0.2900%
minus	206	0.1631%	plus	6910	0.2823%
minus	4133	0.1613%	minus	2949	0.2789%
plus	6989	0.1564%	minus	3662	0.2755%
plus	1595	0.1507%	minus	7713	0.2387%
minus	611	0.1392%	minus	611	0.2367%
minus	1465	0.1328%	minus	2529	0.2324%
plus	514	0.1317%	minus	3878	0.2308%

minus	772	0.1309%	minus	2888	0.2299%
plus	1008	0.1277%	minus	874	0.2256%
plus	2307	0.1263%	plus	7751	0.2196%
plus	1797	0.1254%	minus	1311	0.2164%
plus	470	0.1205%	minus	4224	0.2114%
minus	2117	0.1120%	plus	5022	0.2021%
plus	49	0.1090%	plus	7455	0.1990%
plus	3127	0.1084%	minus	6195	0.1985%
minus	1086	0.1049%	minus	230	0.1935%
plus	1773	0.0932%	minus	5275	0.1925%
plus	7751	0.0926%	minus	965	0.1881%
plus	7229	0.0914%	minus	3168	0.1839%
plus	1945	0.0859%	minus	4466	0.1838%
plus	2659	0.0851%	minus	4034	0.1828%
plus	1650	0.0847%	minus	1378	0.1813%
minus	2949	0.0843%	minus	206	0.1802%
minus	1311	0.0839%	plus	6989	0.1798%
minus	965	0.0830%	minus	6018	0.1748%
minus	874	0.0826%	minus	1281	0.1664%
minus	2999	0.0820%	plus	3262	0.1652%
minus	2529	0.0815%	plus	6052	0.1578%
minus	2011	0.0794%	minus	892	0.1501%
plus	1263	0.0790%	plus	5791	0.1448%
minus	716	0.0760%	minus	1227	0.1446%
plus	1767	0.0754%	minus	638	0.1399%
plus	2557	0.0745%	plus	1431	0.1397%
minus	1072	0.0732%	plus	3193	0.1397%
minus	892	0.0711%	plus	1656	0.1363%
minus	638	0.0628%	minus	2392	0.1361%
plus	1390	0.0626%	minus	7113	0.1335%
minus	3553	0.0604%	plus	5384	0.1323%
plus	547	0.0598%	minus	4010	0.1311%
minus	1227	0.0589%	minus	1072	0.1308%
minus	3326	0.0577%	plus	5074	0.1302%
plus	1431	0.0567%	minus	3747	0.1241%
minus	69	0.0564%	minus	793	0.1222%
minus	3447	0.0554%	minus	7151	0.1219%
minus	1281	0.0544%	plus	4837	0.1168%
minus	622	0.0519%	plus	4579	0.1168%
minus	95	0.0512%	minus	622	0.1121%
minus	2392	0.0509%	minus	6430	0.1071%
minus	1378	0.0503%	minus	3784	0.1066%
minus	793	0.0503%	minus	3102	0.1060%
plus	2707	0.0490%	minus	4647	0.0978%
minus	2888	0.0487%	minus	5646	0.0900%
plus	459	0.0476%	minus	5799	0.0886%
plus	5728	0.0465%	minus	95	0.0849%
minus	942	0.0464%	plus	2663	0.0832%
minus	324	0.0449%	minus	3984	0.0823%
plus	1269	0.0447%	minus	3621	0.0805%
minus	4380	0.0446%	plus	6657	0.0784%
minus	7130	0.0434%	plus	3940	0.0741%
plus	2914	0.0432%	minus	7331	0.0730%
minus	243	0.0425%	plus	7200	0.0725%
plus	2683	0.0418%	minus	7358	0.0710%
minus	3702	0.0416%	minus	7761	0.0688%
plus	1367	0.0414%	minus	7130	0.0679%
minus	333	0.0387%	minus	324	0.0674%
minus	3425	0.0386%	minus	4914	0.0652%
minus	150	0.0377%	plus	7080	0.0650%
plus	712	0.0359%	minus	4616	0.0641%
minus	3168	0.0345%	minus	4167	0.0614%
plus	4929	0.0335%	plus	7229	0.0607%
plus	2882	0.0331%	minus	2442	0.0563%
minus	3977	0.0309%	minus	243	0.0555%
minus	3605	0.0305%	minus	719	0.0555%
minus	3726	0.0288%	minus	3875	0.0539%
plus	7565	0.0265%	minus	333	0.0519%
plus	1541	0.0264%	minus	5248	0.0519%
minus	719	0.0259%	minus	5802	0.0516%
plus	1679	0.0243%	minus	3820	0.0462%
minus	4893	0.0236%	plus	1194	0.0405%
plus	2718	0.0228%	plus	1367	0.0398%
minus	3102	0.0218%	minus	6541	0.0396%
plus	2496	0.0216%	plus	2731	0.0343%
plus	905	0.0213%	minus	4181	0.0341%
minus	3662	0.0209%	minus	155	0.0330%
minus	3744	0.0208%	plus	2496	0.0326%
plus	1194	0.0207%	minus	529	0.0317%
minus	155	0.0199%	minus	7512	0.0303%
plus	1000	0.0196%	minus	907	0.0275%

plus	571	0.0195%	minus	7820	0.0274%
minus	5121	0.0188%	minus	7235	0.0273%
minus	7235	0.0188%	minus	6442	0.0269%
plus	1444	0.0180%	minus	3201	0.0262%
minus	4908	0.0159%	plus	7565	0.0256%
plus	435	0.0158%	minus	5547	0.0255%
plus	7080	0.0137%	plus	1263	0.0242%
minus	529	0.0135%	plus	1998	0.0240%
minus	907	0.0134%	plus	4232	0.0230%
minus	7358	0.0132%	minus	150	0.0220%
plus	1347	0.0127%	plus	5728	0.0219%
minus	2442	0.0126%	minus	6609	0.0207%
plus	1514	0.0116%	minus	3605	0.0179%
minus	3878	0.0114%	plus	3127	0.0178%
plus	1063	0.0110%	plus	595	0.0171%
plus	1484	0.0109%	minus	69	0.0170%
plus	1158	0.0104%	minus	3440	0.0162%
plus	7745	0.0101%	minus	4178	0.0160%
minus	7713	0.0093%	plus	2887	0.0160%
minus	7113	0.0092%	plus	7745	0.0158%
minus	3747	0.0091%	minus	5583	0.0156%
plus	2513	0.0090%	plus	2734	0.0132%
plus	1421	0.0074%	plus	2914	0.0126%
minus	5275	0.0073%	minus	143	0.0125%
minus	5076	0.0073%	plus	2612	0.0114%
minus	3621	0.0068%	plus	4929	0.0113%
minus	76	0.0066%	minus	5202	0.0105%
minus	3201	0.0064%	minus	2187	0.0082%
minus	143	0.0064%	plus	1797	0.0081%
minus	4034	0.0064%	plus	2434	0.0081%
minus	6630	0.0058%	minus	6711	0.0080%
minus	4224	0.0057%	plus	1802	0.0073%
minus	4466	0.0057%	plus	261	0.0071%
minus	4010	0.0056%	plus	49	0.0064%
plus	1580	0.0050%	plus	1945	0.0059%
minus	3784	0.0050%	plus	470	0.0057%
minus	7677	0.0049%	plus	2307	0.0056%
minus	38	0.0049%	plus	1595	0.0055%
minus	6472	0.0045%	plus	1767	0.0054%
minus	7761	0.0044%	plus	2882	0.0053%
minus	6186	0.0042%	plus	514	0.0052%
minus	6018	0.0041%	plus	1008	0.0050%
minus	5278	0.0039%	plus	1269	0.0050%
minus	3875	0.0037%	minus	76	0.0047%
minus	4647	0.0036%	minus	6793	0.0047%
minus	5799	0.0033%	plus	1390	0.0039%
minus	7331	0.0033%	plus	1347	0.0037%
minus	6195	0.0032%	plus	2557	0.0036%
minus	2187	0.0032%	plus	1650	0.0034%
minus	3440	0.0025%	plus	2659	0.0031%
minus	3820	0.0025%	plus	435	0.0030%
minus	7151	0.0025%	plus	2707	0.0028%
minus	5646	0.0024%	plus	1773	0.0027%
minus	4616	0.0023%	plus	2683	0.0025%
minus	3984	0.0021%	minus	4491	0.0024%
minus	4167	0.0020%	minus	4707	0.0023%
plus	5092	0.0018%	plus	905	0.0022%
minus	4181	0.0016%	plus	7325	0.0020%
plus	7325	0.0016%	plus	459	0.0017%
minus	5248	0.0015%	plus	571	0.0016%
minus	7820	0.0014%	plus	5092	0.0014%
plus	1583	0.0013%	plus	1541	0.0014%
minus	4914	0.0013%	plus	547	0.0014%
minus	5802	0.0011%	minus	3999	0.0014%
plus	555	0.0008%	plus	712	0.0013%
minus	6442	0.0008%	minus	38	0.0011%
minus	6609	0.0007%	plus	1580	0.0011%
minus	6541	0.0007%	plus	2513	0.0011%
minus	4178	0.0006%	plus	1679	0.0009%
minus	6430	0.0006%	plus	2718	0.0009%
minus	5547	0.0005%	plus	1514	0.0007%
minus	5202	0.0005%	plus	1000	0.0007%
minus	7512	0.0005%	plus	1421	0.0007%
minus	5583	0.0002%	plus	1063	0.0005%
minus	4491	0.0002%	plus	1444	0.0005%
minus	4707	0.0001%	plus	1158	0.0003%
minus	6711	0.0001%	plus	1484	0.0003%
minus	6793	0.0001%	plus	555	0.0003%
minus	3999	0.0000%	plus	1583	0.0000%

Table S2. Oligos used in this study.

Index	Oligo	Sequence (5' to 3')	Note	Figure
1	Tor2005-IC_Illumina-F	TTTACCGGTCAATTATCGGATT	amplify ~350bp product for Illumina sequencing	Fig. 1,2,5D
2	Tor2005-IC_Illumina-R	CAGATGATCAGGTGATTAAAAACG		Fig. 1,2
3	Tor2005-IC-Sp1(+)_F	cTTCTTAAGTCTATAGGTTGGCTAGGTTTCATCTTTCCGctgca	to clone Sp1(+) into pMMB207 vector	Fig. 1
4	Tor2005-IC-Sp1(+)_R	gCGGAAAGATGAAACCTAGCCAACCTATAGACTTAAGAAgggcc		Fig. 1
5	Tor2005-IC-Sp1(-)_F	cCGGAAAGATGAAACCTAGCCAACCTATAGACTTAAGAActgca	to clone Sp1(-) into pMMB207 vector	Fig. 1
6	Tor2005-IC-Sp1(-)_R	gTTCTTAAGTCTATAGGTTGGCTAGGTTTCATCTTTCCGgggcc		Fig. 1
7	Tor2005-IC-Sp1(+)MutT1A_F	cTTCATAAGTCTATAGGTTGGCTAGGTTTCATCTTTCCGctgca	to clone Sp1(+) MutT1A into pMMB207 vector	Fig. 1F
8	Tor2005-IC-Sp1(+)MutT1A_R	gCGGAAAGATGAAACCTAGCCAACCTATAGACTTATGAAgggcc		Fig. 1F
9	pSp1-hotspot-PAM-sdm_F	CGGCAGCTCCGGCTGTTTCTGGAACCCA	to mutate the hotspot PAM to AAA	Fig. 2B,C
10	pSp1-hotspot-PAM-sdm_R	TGGGTTCCAGAAACAGCCGGAGCTGCCG		Fig. 2B,C
11	pSp1-hotspot-Upstream-sdm_F	ACCCAAGACCGAAGGGGGTCGCCAACTGCGCCTCTTCCTGGAACCCAAG	to mutate 7 nt upstream of the hotspot	Fig. 2D
12	pSp1-hotspot-Upstream-sdm_R	CTTGGGTTCCAGGAAGAGGCGCAGTTGGCGACCCCCTTCGGTCTTGGGT		Fig. 2D
13	pSp1-hotspot-Internal-sdm_F	CCATGGCATCAGCCGTCACCGCTTCCCATTTCGGCTCCAGGAACAGCCGG	to mutate 7nt within the hotspot	Fig. 2E
14	pSp1-hotspot-Internal-sdm_R	CCGGCTGTTCCTGGAGCCGAAATGGGAAGCGGTGACGGCTGATGCCATGG		Fig. 2E
15	pSp1-hotspot-Downstream-sdm_F	CTTTGCCAGCGCGCGGTAACTTCCCTTCACGACCATGGCATCAGCGGTGACGGC	to mutate 7nt downstream of the hotspot	Fig. 2F
16	pSp1-hotspot-Downstream-sdm_R	GCCGTCACCGCTGATGCCATGGTCGTGAAGGGAAGTTACCGCGCGCTGGCAAAG		Fig. 2F
17	Tor2005-IC-Sp1(+)PAMpool_F	agggcccNNNTTAAGTCTATAGGTTGGCTAGGTTTCATCTTTCCGctgcagca	to generate random PAM library in pSp1	Fig. 4
18	Tor2005-IC-Sp1(+)PAMpool_R	tgctgcagCGGAAAGATG		Fig. 4
19	Tor2005-IC-Leader_sdsF	TGTGTGCTTATCAAGCTAATCAAT	to examine spacer dynamics on agarose gel	Fig. 4D
20	Tor2005-IC-Sp1_sdsR	CTATCACCGCGAGATGGTTT		Fig. 4D
21	Scrambled-ctrl_F	cTTCCTTGTCCGTTGATTCTATCGTTGCGACATTGATTActgca	to clone 35nt scrambled control into pMMB207	Fig 4,5
22	Scrambled-ctrl_R	gTAATCAATGTCGCAACGATAGAATCAACGGACAAGGAAgggcc		Fig 4,5
23	Tor2005-IC-downstream_sdsR	AAAGACAAAGAGCTTCTGGCTAAA	reverse primer downstream of I-C array	Fig. 5A,D
24	Tor2000-IC-SpT1_F	cTTCGTCTGATATATGTTCTTTTATTTCAAAATAGGTGAActgca	to clone Tor2000-SpT1 into pMMB207	Fig. 5D
25	Tor2000-IC-SpT1_R	gTTCACCTATTTTGAAATAAAAGAACATATATCAGACGAAgggcc		Fig. 5D
26	Tor2005-IC-Sp15_F	cTTCTAATACATTAATAATCTTGGCAGGGGCTTTTAGCGAGAAactgca	to clone Tor2005-Sp15 into pMMB207	Fig. 5D
27	Tor2005-IC-Sp15_R	gTTCTCGCTAAAAGCCCCTGCCAAGATTATTAATGTATTAGAAgggcc		Fig. 5D
28	Tor2005-IC-Sp43_F	cTTCAGGAATAGCAATTGTGTCAAATAGAAAAGTAGACGGAActgca	to clone Tor2005-Sp43 into pMMB207	Fig. 5D
29	Tor2005-IC-Sp43_R	gTTCCGTCTACTTTTCTATTTGACACAATTGCTATTCCTGAAgggcc		Fig. 5D
30	Tor2000-IC-SpT1_sdsR	GCGCGACACCTATTTTGAAA	oligo #2, to examine spacer dynamics (together w/ o1)	Fig. 5D
31	Tor2000-IC-SpT2_sdsR	CAAAGACGGTTACATCAAGAGGT	oligo #3, to examine spacer dynamics (together w/ o1)	Fig. 5D
32	Tor2005-IC-Sp13-sdsF	TGTATAATGATATTTGTCTGTGAGGGA	oligo #4, to examine spacer dynamics	Fig. 5D
33	Tor2005-IC-Sp17-sdsR	ACTCGACTTGGCCTTATCCA	oligo #5, to examine spacer dynamics	Fig. 5D
34	Tor2005-IC-Sp42-sdsF	TGGTGATTAGGTCGTCAATGC	oligo #6, to examine spacer dynamics (together w/ o23)	Fig. 5D

>pSp1(+) full sequence

CTGCAGGCATGCAAGCTTGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAA
CAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGG
TCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTC
GGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCC
ATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTTTGTTTATTTTTCTAAATAC
ATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTG
TCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGG
GTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTGATGTCCGGCGGTGCTTTT
GCCGTTACGCACCACCCCGTCAGTAGCTGAACAGGAGGGACAGCTGATAGAAACAGAAGCCACTGGCCTCAAAAACACCATCATACACTAAAT
CAGTAAGTTGGCAGCATCACCCGACGCACTTTGCGCCGAATAAATACCTGTGACGGAAGATCACTTCGCAGAATAAATAAATCCTGGTGTCCC
TGTTGATACCGGGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAG
ATCACTACCGGGCGTATTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGA
TATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGC
CTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTACG
TATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCT
CTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAA
AGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTT
CGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTTTGTGATGG
CTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGTCTTCA
AATTCCCGTTGCACATAGCCCGGCAATTCCTTTCCCTGCTCTGCCATAAGCGCAGCGAATGCCGGGTAATACTCGTCAACGATCTGATAGAGA
AGGGTTTGCTCGGGTCGGTGGCTCTGGTAACGACCAGTATCCCGATCCCGGCTGGCCGTCCTGGCCGCCACATGAGGCATGTTCCGCGTCCTT
GCAATACTGTGTTTACATACAGTCTATCGCTTAGCGGAAAGTTCTTTTACCCTCAGCCGAAATGCCTGCCGTTGCTAGACATTGCCAGCCAGT
GCCCGTCACTCCCGTACTAACTGTCACGAACCCCTGCAATAACTGTCACGCCCCCCTGCAATAACTGTCACGAACCCCTGCAATAACTGTCAC
GCCCCCAAACCTGCAAACCCAGCAGGGGCGGGGGCTGGCGGGGTGTTGGAAAAATCCATCCATGATTATCTAAGAATAATCCACTAGGCGCGG
TTATCAGCGCCCTTGTGGGGCGCTGCTGCCCTTGCCCAATATGCCCGGCCAGAGGCCGGATAGCTGGTCTATTCGCTGCGCTAGGCTACACAC
CGCCCCACCGCTGCGCGGCAGGGGGAAAGGCGGGCAAAGCCCGCTAAACCCCACACCAAACCCCGCAGAAATACGCTGGAGCGCTTTTAGCCG
CTTTAGCGGCCTTTCCCCCTACCCGAAGGGTGGGGGCGCGTGTGCAGCCCCGCAGGGCCTGTCTCGGTCGATCATTCAGCCCGGCTCATCCTT
CTGGCGTGGCGGCAGACCGAACAAGGCGCGGTCGTGGTCGCGTTCAAGGTACGCATCCATTGCCGCCATGAGCCGATCCTCCGGCCACTCGCT
GCTGTTCACCTTGGCCAAAATCATGGCCCCCACCAGCACCTTGCGCCTTGTTTCGTTCTTGCGCTCTTGCTGCTGTTCCCTTGCCCGCACCCG
CTGAATTTCGGCATTGATTCGCGCTCGTTGTTCTTCGAGCTTGGCCAGCCGATCCGCCGCCTTGTTGCTCCCCTTAACCATCTTGACACCCCA
TTGTTAATGTGCTGTCTCGTAGGCTATCATGGAGGCACAGCGGCGGCAATCCCGACCCTACTTTGTAGGGGAGGGCGCACTTACCGGTTTCTC
TTCGAGAAACTGGCCTAACGGCCACCCTTCGGGCGGTGCGCTCTCCGAGGGCCATTGCATGGAGCCGAAAAGCAAAAGCAACAGCGAGGCAGC
ATGGCGATTTATCACCTTACGGCGAAAACCGGCAGCAGGTCGGGCGGCCAATCGGCCAGGGCCAAGGCCGACTACATCCAGCGCGAAGGCAAG
TATGCCCGCGACATGGATGAAGTCTTGCACGCCGAATCCGGGCACATGCCGGAGTTCGTCGAGCGGCCCGCCGACTACTGGGATGCTGCCGAC
CTGTATGAACGCGCCAATGGGCGGCTGTTCAAGGAGGTCGAATTTGCCCTGCCGGTCGAGCTGACCCTCGACCAGCAGAAGGCGCTGGCGTCC
GAGTTCGCCCAGCACCTGACCGGTGCCGAGCGCCTGCCGTATACGCTGGCCATCCATGCCGGTGGCGGCGAGAACCCGCACTGCCACCTGATG
ATCTCCGAGCGGATCAATGACGGCATCGAGCGGCCCGCCGCTCAGTGGTTCAAGCGGTACAACGGCAAGACCCCGGAGAAGGGCGGGGCACAG
AAGACCGAAGCGCTCAAGCCCAAGGCATGGCTTGAGCAGACCCGCGAGGCATGGGCCGACCATGCCAACCGGGCATTAGAGCGGGCTGGCCAC
GACGCCCGCATTGACCACAGAACACTTGAGGCGCAGGGCATCGAGCGCCTGCCCGGTGTTCACCTGGGGCCGAACGTGGTGGAGATGGAAGGC
CGGGGCATCCGCACCGACCGGGCAGACGTGGCCCTGAACATCGACACCGCCAACGCCCAGATCATCGACTTACAGGAATACCGGGAGGCAATA
GACCATGAACGCAATCGACAGAGTGAAGAAATCCAGAGGCATCAACGAGTTAGCGGAGCAGATCGAACCGCTGGCCCAGAGCATGGCGACACT
GGCCGACGAAGCCCGGCAGGTCATCTCGAGCGGCTCGATCTGGCTGCGCTTGGTGGCCCGATGAAGAACGACAGGACTTTGCAGGCCATAGGC
CGACAGCTCAAGGCCATGGGCTGTGAGCGCTTCGATATCGGCGTCAGGGACGCCACCACCGGCCAGATGATGAACCGGGAATGGTCAGCCGCC
GAAGTGCTCCAGAACACGCCATGGCTCAAGCGGATGAATGCCCAGGGCAATGACGTGTATATCAGGCCCGCCGAGCAGGAGCGGCATGGTCTG
GTGCTGGTGGACGACCTCAGCGAGTTTGACCTGGATGACATGAAAGCCGAGGGCCGGGAGCCTGCCCTGGTAGTGGAAACCAGCCCGAAGAAC
TATCAGGCATGGGTCAAGGTGGCCGACGCCGCAGGCGGTGAACTTCGGGGGCAGATTGCCCGGACGCTGGCCAGCGAGTACGACGCCGACCCG
GCCAGCGCCGACAGCCGCCACTATGGCCGCTTGGCGGGCTTCACCAACCGCAAGGACAAGCACACCACCCGCGCCGGTTATCAGCCGTGGGTG
CTGCTGCGTGAATCCAAGGGCAAGACCGCCACCGCTGGCCCGGCGCTGGTGCAGCAGGCTGGCCAGCAGATCGAGCAGGCCCAGCGGCAGCAG
GAGAAGGCCCGCAGGCTGGCCAGCCTCGAACTGCCCGAGCGGCAGCTTAGCCGCCACCGGCGCACGGCGCTGGACGAGTACCGCAGCGAGATG
GCCGGGCTGGTCAAGCGCTTCGGTGATGACCTCAGCAAGTGCGACTTTATCGCCGCGCAGAAGCTGGCCAGCCGGGGCCGCAGTGCCGAGGAA
ATCGGCAAGGCCATGGCCGAGGCCAGCCCAGCGCTGGCAGAGCGCAAGCCCGGCCACGAAGCGGATTACATCGAGCGCACCGTCAGCAAGGTC
ATGGGTCTGCCCAGCGTCCAGCTTGCGCGGGCCGAGCTGGCACGGGCACCGGCACCCCGCCAGCGAGGCATGGACAGGGGCGGGCCAGATTTC
AGCATGTAGTGCTTGCGTTGGTACTCACGCCTGTTATACTATGAGTACTCACGCACAGAAGGGGGTTTTATGGAATACGAAAAAAGCGCTTCA
GGGTCGGTCTACCTGATCAAAAGTGACAAGGGCTATTGGTTGCCCGGTGGCTTTGGTTATACGTCAAACAAGGCCGAGGCTGGCCGCTTTTCA
GTCGCTGATATGGCCAGCCTTAACCTTGACGGCTGCACCTTGTCCTTGTTCCGCGAAGACAAGCCTTTCGGCCCCGGCAAGTTTCTCGGTGAC
TGATATGAAAGACCAAAAGGACAAGCAGACCGGCGACCTGCTGGCCAGCCCTGACGCTGTACGCCAAGCGCGATATGCCGAGCGCATGAAGGC
CAAAGGGATGCGTCAGCGCAAGTTCTGGCTGACCGACGAATACGAGGCGCTGCGCGAGTGCCTGGAAGAACTCAGAGCGGCGCAGGGCGG
GGGTAGTGACCCCGCCAGCGCCTAACCACCAACTGCCTGCAAAGGAGGCAATCAATGGCTACCCATAAGCCTATCAATATTCTGGAGGCGTTC
GCAGCAGCGCCGCCACCGCTGGACTACGTTTTGCCCAACATGGTGGCCGGTACGGTCGGGGCGCTGGTGTCGCCCGGTGGTGCCGGTAAATCC
ATGCTGGCCCTGCAACTGGCCGCACAGATTGCAGGCGGGCCGGATCTGCTGGAGGTGGGCGAACTGCCCACCGGCCCGGTGATCTACCTGCCC
GCCGAAGACCCGCCCACCGCCATTCATCACCGCCTGCACGCCCTTGGGGCGCACCTCAGCGCCGAGGAACGGCAAGCCGTGGCTGACGGCCTG
CTGATCCAGCCGCTGATCGGCAGCCTGCCCAACATCATGGCCCCGGAGTGGTTCGACGGCCTCAAGCGCGCCGCCGAGGGCCGCCGCCTGATG
GTGCTGGACACGCTGCGCCGGTTCCACATCGAGGAAGAAAACGCCAGCGGCCCCATGGCCCAGGTCATCGGTCGCATGGAGGCCATCGCCGCC
GATACCGGGTGCTCTATCGTGTTCCTGCACCATGCCAGCAAGGGCGCGGCCATGATGGGCGCAGGCGACCAGCAGCAGGCCAGCCGGGGCAGC

TCGGTACTGGTCGATAACATCCGCTGGCAGTCCTACCTGTCGAGCATGACCAGCGCCGAGGCCGAGGAATGGGGTGTGGACGACGACCAGCGC
CGGTTCTTCGTCCGCTTCGGTGTGAGCAAGGCCAACTATGGCGCACCGTTCGCTGATCGGTGGTTCAGGCGGCATGACGGCGGGGTGCTCAAG
CCCGCCGTGCTGGAGAGGCAGCGCAAGAGCAAGGGGGTGCCCCGTGGTGAAGCCTAAGAACAAGCACAGCCTCAGCCACGTCCGGCACGACCC
GGCGCACTGTCTGGCCCCCGGCCTGTTCCGTGCCCTCAAGCGGGGCGAGCGCAAGCGCAGCAAGCTGGACGTGACGTATGACTACGGCGACGG
CAAGCGGATCGAGTTCAGCGGCCCGGAGCCGCTGGGCGCTGATGATCTGCGCATCCTGCAAGGGCTGGTGGCCATGGCTGGGCCTAATGGCCT
AGTGCTTGGCCCGGAACCCAAGACCGAAGGCGGACGGCAGCTCCGGCTGTTCCTGGAACCCAAGTGGGAGGCCGTCACCGCTGATGCCATGGT
GGTCAAAGGTAGCTATCGGGCGCTGGCAAAGGAAATCGGGGCAGAGGTCGATAGTGGTGGGGCGCTCAAGCACATACAGGACTGCATCGAGCG
CCTTTGGAAGGTATCCATCATCGCCCAGAATGGCCGCAAGCGGCAGGGGTTTCGGCTGCTGTCGGAGTACGCCAGCGACGAGGCGGACGGGCG
CCTGTACGTGGCCCTGAACCCCTTGATCGCGCAGGCCGTCATGGGTGGCGGCCAGCATGTGCGCATCAGCATGGACGAGGTGCGGGCGCTGGA
CAGCGAAACCGCCCGCCTGCTGCACCAGCGGCTGTGTGGCTGGATCGACCCCGGCAAAACCGGCAAGGCTTCCATAGATACCTTGTGCGGCTA
TGTCTGGCCGTCAGAGGCCAGTGGTTCGACCATGCGCAAGCGCCGCCAGCGGGTGCGCGAGGCGTTGCCGGAGCTGGTCGCGCTGGGCTGGAC
GGTAACCGAGTTCGCGGCGGGCAAGTACGACATCACCCGGCCCAAGGCGGCAGGCTGACCCCCCCCACTCTATTGTAAACAAGACATTTTTAT
CTTTTATATTCAATGGCTTATTTTCCTGCTAATTGGTAATACCATGAAAAATACCATGCTCAGAAAAGGCTTAACAATATTTTGAAAAATTGC
CTACTGAGCGCTGCCGCACAGCTCCATAGGCCGCTTTCCTGGCTTTGCTTCCAGATGTATGCTCTTCTGCTCCCGAACGCCAGCAAGACGTAG
CCCAGCGCGTCGGCCAGCTTGCAATTCGCGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCA
GCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACA
GCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGG
TTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGG
CGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAAC
CGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCG
CCGAGACAGAACTTAATGGGCCC**TTC**TTAAGTCTATAGGTTGGCTAGGTTTCATCTTTCCG