Avoidance of Self During CRISPR Immunization

Jake L. Weissman[1], Arlin Stoltzfus[2,3], Edze R. Westra[4], and Philip L. F. Johnson[1,*]

[1]Department of Biology, University of Maryland College Park, MD, USA
[2]Office of Data and Informatics, Material Measurement Laboratory, NIST
[3]Institute for Bioscience and Biotechnology Research, 9600 Gudelsky Drive, Rockville, MD 20850
[4]Environment and Sustainability Institute, Centre for Ecology and Conservation, University of Exeter, Biosciences, Penryn, Cornwall, UK
[*]Correspondence: plfj@umd.edu (P.L.F. Johnson)

Keywords: autoimmunity; self-nonself recognition; priming

Abstract

The battle between microbes and their viruses is ancient and ongoing. CRISPR immunity, the first and, to-date, only form of adaptive immunity found in prokaryotes, represents a flexible mechanism to recall past infections while also adapting to a changing pathogenic environment. Critical to the role of CRISPR as an adaptive immune mechanism is its capacity for self versus non-self recognition when acquiring novel immune memories. Yet, CRISPR systems vary widely in both how and to what degree they can distinguish foreign from self-derived genetic material. We document known and hypothesized mechanisms that bias the acquisition of immune memory towards non-self targets. We demonstrate that diversity is the rule, with many widespread but no universal mechanisms for self vs. non-self recognition.

Distinguishing Self from Non-Self During the CRISPR Immune Response

Viruses of microbes severely impact their hosts' population and evolutionary dynamics [1, 2], and, as a result, prokaryotes have evolved a number of anti-viral defense systems, some quite complex [3, 4, 5, 6]. Among the best-studied classes of host defense systems are the CRISPR immune systems, which can acquire novel and highly specific immune "memory" (in the form of short DNA fragments called **"spacers"**; see Glossary) and then use this memory to degrade matching viral genetic material [7, 8]. Typically, immunity proceeds in three steps (1) spacer acquisition (sometimes called 'adaptation' in the literature) [8, 9], (2) biogenesis of short guide RNAs (**crRNAs**) corresponding to the host's spacer repertoire [10, 11, 12], (3) targeting and degradation of the matching sequence on the invading genome (the **"protospacer"**) [8, 10, 11, 12]. During this multi-stage process the host cell must successfuly identify foreign genetic material and distinguish these

potential targets from self genetic material, or else risk costly autoimmunity and inefficient clearance of viral pathogens.

Therefore, CRISPR's capacity for self versus non-self recognition is critical to its role as an adaptive immune mechanism. All immune systems face a fundamental trade-off between pathology induced by the pathogen and pathology associated with autoimmunity. Unlike innate immune systems, the inherent flexibility of adaptive immune systems makes autoimmunity a recurring threat, thus favoring the evolution of continuously acting mechanisms to avoid self-targeting during the lifetime of an organism. In the vertebrate adaptive immune system, numerous mechanisms are well understood to prevent autoimmunity through both biased (i.e., against non-self) acquisition of immunity and biased targeting [13]. Similarly, CRISPR may differentiate self from non-self at multiple stages of immunity. Indeed, non-self recognition in CRISPR immunity has been demonstrated during spacer acquisition (discussed below, e.g., [14]) and target degradation (via mechanisms that prevent cleavage of self targets, e.g., [15]). In principle non-self recognition could also occur during crRNA maturation if self-targeting sequences were not allowed to fully mature (in a process akin to **thymic selection** in vertebrate adaptive immune systems [13]), though to our knowledge this has not been observed. The details of CRISPR immunity, and specific protein machinery involved, are quite variable across systems (see Box 1 for an overview), leading to corresponding variability in the mechanisms of non-self recognition employed by different CRISPR systems.

Box 1 - The unity and diversity of CRISPR defense systems

CRISPR arrays are loci on the host genome where memories (spacers) are stored [7], and the CRISPR-associated (**Cas**) proteins are the machinery responsible for both the acquisition of novel memories and the use of current memories in immune defense [16]. All CRISPR systems share the same core acquisition genes, *cas1* and *cas2*, though the acquisition process may differ in many details between systems (with some systems using additional acquisition proteins [17, 18], and some even acquiring spacers from RNA [19]; see [20, 21] for in-depth reviews of the mechanics of spacer acquisition). In contrast, the Cas targeting machinery, or "effector" module, is highly variable among system types, and is used as the basis for classifying systems [16, 22]. Systems are grouped into two classes on the basis of whether their effector module consists of a single Cas protein (e.g., Cas9 in type II systems or Cas12 in type V systems) or complex of Cas proteins (e.g., the Cascade complex and Cas3 in type I systems). Below the class level, systems can be classified into at least 6 types and 33 subtypes, though the majority of systems belong to types I, II and III, with type I being the most prevalent among sequenced genomes [17, 16, 22]. System types and subtypes have important functional differences (e.g., RNA targeting in type VI systems [23, 24, 25]) that influence their capacity for self vs. non-self recognition (main text).

Here we focus on mechanisms of self vs. non-self recognition during CRISPR spacer acquisition, as these create a heritable non-self bias passed down through a lineage (though see [26, 27] for examples of recognition during targeting).

To what degree and by what mechanisms does CRISPR distinguish self from non-self during the acquisition of novel immune memories? These questions are not easily answered, as measuring preference for non-self spacer acquisition is challenging in natural, and even many experimental, systems.

Acquisition of self-targeting spacers is typically toxic for individual cells, as it programs the CRISPR system to cleave the self genome [28]. These instances - even if they incur a major cost of carrying the system - are hard to detect due to the strong negative selection that causes these individuals to be rapidly purged from the population (Fig 1). To avoid the confounding effects of selection inherent to population-level studies, much of the experimental work we discuss below estimates the rate of acquisition of self-targeting spacers by tracking engineered or mutant systems that are unable to degrade targets, so that self-targeting carries no cost (e.g., [29, 30]).

We group mechanisms for non-self recognition into two broad categories, (1) those resulting directly from a biased substrate preference by the Cas acquisition machinery and (2) those resulting indirectly from other aspects of the host's ecological or evolutionary dynamics. We demonstrate that diversity is the rule, with many widespread but no universal mechanisms for self vs. non-self recognition during spacer acquisition (Table 1).

Non-Self Recognition Due to Substrate Preference

If the Cas acquisition machinery preferentially associates with foreign genetic material, a strong non-self spacer acquisition bias may result. In order for the Cas machinery to demonstrate this type of substrate preference, there must be some signal recognized by Cas proteins that is enriched in foreign sequences. In cases where no pre-existing spacers targeting the foreign sequence exist ("naive acquisition"), these signals must result from some generic difference between the host genome and the genomes of mobile genetic elements. Alternatively, if the host already has a fully or partially matching spacer towards the foreign sequence, it may leverage this information to acquire additional spacers ("primed acquisition").

Established Mechanisms:	**Description**	**System Types**
Replicon Counting [14, 31, 32]	The spacer acquisition machinery preferentially associates with double-strand breaks, including at collapsed replication forks. Viruses and high-copy plasmids present many more centers of replication than the host genome.	Type I and some type II systems
Synergy with RM Systems [33, 34]	Spacers are acquired from the fragmented byproducts of restriction enzymes. Since RM systems can differentiate self from non-self, CRISPR inherits this bias.	Type II systems (Potentially other types)
Priming [35, 36, 37]	Pre-existing partial or complete matching between a spacer and protospacer leads to a sharp increase in spacer acquisition from sites in the same genome. This allows immunity to be rapidly updated during host-virus coevolution.	Type I and II systems
[a]Induction [38]	The *cas* genes are up-regulated during infection or periods of elevated risk of infection. Induction is particularly relevant when infection is infrequent.	Variable (Depends on genomic background)
Speculative Mechanisms:		
Transcription dependent spacer acquisition [19, 39, 31]	Viral genes are highly expressed during infection. This promotes acquisition in systems that acquire spacers from RNA and also potentially those that acquire spacers from DNA.	Some type III, and possibly type I and type VI systems
Protospacer preference [40, 41]	If host has purged potential sites of spacer acquisition from genome, then self-targeting will be less likely.	Type I and II systems (Potentially other types)
[a]Horizontal transfer of spacers [42, 43, 44, 45]	Recombination occurs between arrays and entire arrays can be transferred horizontally. Presumably self-targeting spacers have already been selected against at this stage.	General to all systems (Depends on rate of horizontal transfer)

[a] Mechanisms arising from features of the host's physiology or ecology rather than any explicit substrate preference of the Cas acquisition machinery.

Table 1: **(Key Table)** Mechanisms of self vs. non-self recognition during spacer acquisition.

Naive Spacer Acquisition

What signals generically distinguish parasitic mobile genetic elements from host sequences? Parasites of all kinds often live and reproduce in large numbers within a given host. Thus, though not perfect signals, sequence multiplicity and replication may serve as indicators of mobile genetic elements. Indeed, some CRISPR systems prefer to acquire spacers from actively replicating sequences within the cell, and this can lead to a strong bias towards non-self acquisition [14, 32].

Working with the *E. coli* type I-E system, Levy et al. [14] demonstrated a preference by CRISPR for free DNA ends during acquisition. Because stalled replication forks frequently produce double-strand breaks in the DNA (i.e., free ends), and because high-copy viruses and plasmids will present many more of these replication forks in the cell than the host genome [14], a strong non-self acquisition bias results [14]. Furthermore, when a break occurs, the RecBCD machinery is recruited and processively degrades the DNA until it reaches a Chi site, producing even more substrate for spacer acquisition. Mobile genetic elements like plasmids and viruses typically lack these Chi sites, meaning that degradation will continue along their genomes, further compounding the resulting non-self bias. Levy et al. [14] estimate a 100- to 1000-fold preference for plasmid over host DNA during acquisition in their system.

Preference for free DNA ends may be a rather general feature of spacer acquisition, and has been experimentally observed in multiple *Streptococcus* type II-A systems [32, 37]. Similarly, the *Pyrococcus furiosus* acquisition module, encoded alongside type I-G and type III-B effector modules, appears to preferentially acquire spacers from regions that are expected to be especially prone to double-strand breaks [31].

Nevertheless, Wei et al. [30] working with the *Streptococcus thermophilus* DGCC7710 type II-A CRISPR1 locus found that spacers were acquired as frequently from the host genome as a plasmid, indicating no non-self bias.

This is a particularly confusing result as the type II-A CRISPR3 locus from the same strain was recently shown to have a preference for free DNA ends [37]. It is possible that the CRISPR1 and CRISPR3 loci of *S. thermophilus* are functionally quite different (after all, they do have different acquisition rates [46]). More likely, we think, is that the identity of the substrate used in each experiment influences the outcome. Specifically, the plasmid used by Wei et al. [30] is thought to have relatively low copy number (~ 3 copies per cell [30], in contrast to the high burst-size lytic phages used by others [37]). We would expect only a weak preference for plasmid-derived spacers in this case, because the number of plasmid replicons is similar to the number of host replicons. Following this logic, we predict that the more rapidly a virus or plasmid reproduces inside the cell, the more replicons it will produce, and thus the more prone it will be to spacer acquisition. Thus, we might expect large, low-copy plasmids and lysogenic phage to coexist for a longer period of time with an active CRISPR system than high copy plasmids or lytic viruses. Similarly, rapidly replicating hosts that are effectively polyploid would be more prone to self-targeting than slow-growing hosts [47, 48, 49]. In fact, this could partially explain why CRISPR is more prevalent among organisms we expect to be slower-growing (e.g., extremophiles, some archaea, anaerobes [50, 51]). Related to this point, we might expect CRISPR to be less effective at acquiring immunity towards mobile genetic elements that employ rolling-circle replication (which have only a single replication fork per genome and may reproduce serially) [52]. For example, in a type II-A system spacers were not acquired from staphylococcal phage $\phi 12\gamma 3$ while it underwent rolling-circle replication, but were only acquired during early stages of infection [32]. On the other hand, contrary to our expectation, it seems that in some plasmids rolling-circle replication may promote spacer acquisition, likely due to a dependence on DNA nicking at the origin of replication [31].

CRISPR may also be able to directly leverage expression level as a sig-

nal of growth rate. During infection, many viruses subvert host transcriptional processes so that host genes are down-regulated even as viral genes are transcribed at a high rate [53]. In these cases, systems that acquire spacers directly from RNA [19] might favor non-self protospacers. Acquisition from RNA has only been experimentally observed in certain type III systems where the *cas* acquisition machinery is fused to a reverse transcriptase [19], but bioinformatic evidence suggests that RNA-targeting type VI systems may also acquire spacers directly from RNA [54, 23, 24, 25]. Even in systems that acquire spacers from DNA, spacer-acquisition hot-spots have been observed in highly expressed genes [39, 31]. It has been hypothesized that transcription may make the DNA physically more accessible to the Cas machinery [39], or may cause double-strand breaks [31].

CRISPR's preference for free DNA ends may also bias acquisition towards non-self in an entirely growth-independent manner via a synergy with innate immune systems, specifically restriction-modification (**RM**) systems. These systems degrade mobile genetic elements and may provide substrates for spacer acquisition [33, 34]. RM systems have been shown to increase the rate of spacer acquisition [33] and also tend to co-occur with CRISPR when looking broadly across species [55]. A CRISPR-RM synergy would allow spacer acquisition to benefit from the strong non-self recognition capacity of RM systems (based on methylation patterns), and might be quite general, as the vast majority of prokaryotes encode at least one RM system [56, 55].

Finally, we note that if the Cas acquisition machinery prefers specific motifs present in only some subsets of potential spacers [41], then selection against these sequences on the host genome may lead to a non-self acquisition bias. Under this mechanism, the non-self signal is not specifically enriched in non-self sequences in general (as discussed above), but rather depleted in the host (via the strong selective pressure imposed by self-targeting). Acquisition biases are well documented, with many systems requiring a 2- to 8-bp system-specific protospacer adjacent motif (**PAM**) directly upstream

of the protospacer [57, 40, 58]. Even among protospacers with the appropriate PAM there is evidence for strong acquisition biases on the basis of motifs internal to the protospacer [39, 59, 41], and single mutations in the protospacer can drastically alter these biases [60]. Motif-avoidance in the host genome will not be possible in the case of short or degenerate motifs (i.e., most PAMs), but may be feasible in the case of longer, less abundant motifs (similar to the avoidance of restriction sites seen on some genomes [61]). Even in this case, viruses are also likely to be under strong pressure to purge preferred motifs (e.g., PAM avoidance in viruses [62]), limiting the ability of this mechanism to differentiate non-self sequences. Thus while the principle behind motif-depletion is quite general (any host can evolve in such a way), its non-self biasing effects are likely to be somewhat weaker than the other substrate preferences discussed above.

Primed Spacer Acquisition

By far the most specific and reliable indicator of a non-self sequence is that the host already has a spacer targeting that sequence (assuming selection has purged all self-targeting spacers from the population, Fig 1). While this specific type of information is useless when the host encounters a completely new mobile element, preexisting immune memory can be extremely useful in the context of an ongoing coevolutionary arms race. For example, viruses frequently coevolve with their hosts to overcome CRISPR immune targeting [63, 64, 65, 66]. A single mutation in the viral protospacer or PAM can be enough to completely prevent CRISPR targeting [63, 40, 58]. How does the host keep up during fast-paced coevolutionary dynamics? Many CRISPR systems, it turns out, are able to quickly update their immune targeting when a foreign sequence encodes a protospacer that has a partial or complete match in the host's CRISPR array [36, 35, 67, 20]. Such "priming" can lead to strongly biased acquisition from already-recognized enemies.

Mechanistically, priming relies on CRISPR's preference for free DNA ends

[14, 32]. DNA fragments produced by CRISPR's immune activity become the substrates for spacer integration by the Cas acquisition machinery [68, 37]. Perfect spacer-protospacer matches stimulate the most efficient primed spacer acquisition [69], but even partial matches may lead to low rates of degradation and stimulate the acquisition of spacers [70, 37].

Priming is a widespread phenomenon, and has been observed experimentally to be acting in type I-B [71, 72], I-C [60], I-E [36, 35, 67], I-F [73], and type II-A [37] CRISPR systems. Bioinformatic evidence has suggested that type II-C systems may also be capable of priming [74]. Type III systems tend to be quite tolerant of mismatches during targeting [75], and thus are less likely to require priming to overcome pathogen coevolution [21], perhaps explaining why priming has not been observed in these systems to-date.

Despite the generality of this mechanism across type I and II CRISPR systems, some important differences exist. There are particular strand and spatial biases of primed acquisition that vary between systems, likely resulting from the fact that the type I endonuclease Cas3 moves along the DNA processively whereas the type II endonuclease Cas9 remains associated with the free ends [37]. These differences are also seen in terms of PAM-dependence, where priming in the type II-A system is reliant on the presence of an intact PAM sequence, which is required for endonuclease activity to produce a fragmented substrate for acquisition [37]. In contrast, PAM-independent priming has been observed in a type I-E system, where recognition of a protospacer target lacking an appropriate PAM leads to recruitment of Cas3 in such a way that endonuclease activity is inhibited. Following recruitment, Cas3 acts as a molecular motor and moves processively along the DNA strand, potentially promoting spacer uptake in regions quite distant from the original protospacer match [76, 77].

Finally, how effective is priming as a mechanism for self versus non-self recognition? In one type I-F study system, priming led to strongly biased acquisition towards non-self (500-fold over naive acquisition), but promiscuous

tolerance of partial matches lead to an elevated number of self-acquisition events, so that the absolute number of self-targeting spacers was approximately the same in naive and primed states [39]. Thus priming may cause strongly non-self biased acquisition, but it may simultaneously not affect, or may even increase, the absolute rate of self-targeting by the spacer acquisition machinery.

Non-Self Biases Related to Host Physiology and Ecology

So far we have discussed a number of ways in which the Cas spacer acquisition machinery may respond preferentially to non-self sequences. Even in the absence of such a preference, environmental cues may lead to non-self biased spacer content in the host CRISPR array. In general, we expect these mechanisms to be weaker than many of the preference-based mechanisms discussed above, but they may still be of ecological or evolutionary importance.

Expression of the *cas* Genes

Though not often discussed explicitly as a means of self vs. non-self recognition, *cas* genes are often up-regulated in response to infection, or under conditions where infection is likely to occur [38]. This amounts to a form of temporal biasing, limiting acquisition events to periods where foreign DNA is likely to be present in the cell. Across systems and host species, though, patterns of expression are variable [38]. The *cas* genes can be up-regulated in response to various stimuli that may correspond to increased infection risk, including nutrient concentrations [78, 79, 80], temperature [81], and host density [82, 83]. Systems may even be up-regulated as a direct response to viral contact or ongoing infection [38, 84]. For a comprehensive discussion of CRISPR regulation, a large and active research area in itself, see Patterson

et al. [38]. How CRISPR is regulated so that the host can dynamically control infection risk is still something of a mystery, but promising new methods to quickly and accurately measure the expression of *cas* genes in a range of genetic backgrounds and ecological scenarios are being developed [80].

We expect the conditions associated with induction to be correlated with the risk of infection, and these indicators likely vary across environments and taxa. Induction will be particularly important for the self vs. non-self recognition when viral (or plasmid) infection is a rare occurrence, since at all other times the only substrate for spacer acquisition will be the host genome. Therefore, if pathogen exposure varies in time, hosts can maximize their capacity for self vs. non-self recognition by employing a strategy that combines induction with various mechanisms to bias the Cas acquisition machinery's substrate preference (discussed earlier). Possibly of note, *cas* genes are typically found as a single operon [17], and often are co-transcribed (e.g., [80]). This implies a temporal coupling of the Cas acquisition and effector complexes, consistent with the idea that at times of increased infection the host will want to both use and add to its spacer repertoire.

Horizontal Transfer of Immune Memory

Horizontally transferred spacers, if coming from a closely related strain, are likely to target non-self. This conclusion follows from the assumption that the standing spacer diversity in a population has already experienced strong selection against self-targeting spacers (Fig 1). This line of logic also suggests that spacers acquired via horizontal transfer will be particularly beneficial to their hosts (Box 2). Such a mechanism will only be relevant to individuals if horizontal transfer of immunity is very frequent, which appears to be the case. CRISPR arrays are extremely labile [85, 43], and spacers can be transferred via recombination between arrays [42]. Homology between spacers and viral genomes may actually help these arrays propagate themselves via transduction [45]. In fact, it has even been proposed that repeats are

highly conserved across systems specifically to aid in the horizontal transfer of spacers between arrays through homologous recombination [44]. Clearly, these spacers will only be useful if they come from individuals that share viral pathogens (typically in the same species), though in general we expect horizontal transfer to be most common among closely related organisms (e.g., [86]).

Box 2 - The fitness of acquired spacers

CRISPR immunity is often referred to as **"Lamarkian"** [87], but this is an anachronistic and controversial term [88], with no clear translation into contemporary molecular biology. It is clear, all else being equal, that spacer acquisition will favor locally abundant mobile genetic elements, as there will be many opportunities for acquisition from these sequences. This abundance-bias, independent of any non-self bias, may prove to be either adaptive or maladaptive depending on the mobile element concerned. In the case of phage, acquisition from locally-abundant pathogens is likely to represent a fitness benefit. At the same time, we expect beneficial plasmids or beneficial genes on those plasmids specific to an environment to be locally enriched in that environment (due to selection; [89]), meaning that CRISPR may be more likely to target these sequences, ultimately leading to a loss in relative fitness as compared to CRISPR-lacking strains (e.g., [90]). Thus a preference for spacer acquisition from locally abundant mobile genetic elements does not necessarily lead to a consistent change in fitness, but may amplify preexisting costs or benefits of CRISPR immunity. This is further complicated by the fact that CRISPR does not necessarily prevent horizontal gene transfer over longer timescales [91].

A slightly different line of logic applies to spacers gained via horizontal transfer. Beneficial spacers are likely to have undergone positive selection, and costly spacers will have been selected against. Thus we expect beneficial spacers to be enriched in the population, and therefore more likely to be transferred than costly ones. Since spacers themselves have been "pre-screened" in this case, we expect horizontal transfer to yield spacers that are not only strongly biased towards non-self (i.e., are not harmful), but also that specifically target the most common pathogens in a given environment (i.e., confer the greatest fitness benefit).

Concluding Remarks

CRISPR systems employ a diverse set of mechanisms for non-self recognition during spacer acquisition, and some of these mechanisms are quite widespread. No mechanism, though, is universal (Table 1), and even those that are widespread show a great deal of variability in their details across systems. Included in this diversity are some organisms that are able to circumvent the issue of self-targeting induced mortality entirely. In certain highly polyploid archaea, the presence of many chromosomal copies appears to allow for rapid template-based repair, and this in turn abolishes the cost of self-targeting spacers under natural conditions [92]. Even so, an inability to recognize non-self could still negatively impact the efficiency with which infections are cleared.

Despite the enormous diversity of CRISPR systems, there are some commonalities across mechanisms for non-self recognition, specifically that many rely on CRISPR's preference for free DNA ends (Fig 2). This dependency is obvious in some cases, such as CRISPR's synergy with RM systems and in the context of certain priming mechanisms, but free ends may also contribute to transcription-dependent spacer acquisition. This suggests that DNA ends are a universal signal of infection that can promote recognition of non-self DNA across host taxonomic domains and across classes of mobile genetic elements (e.g., plasmids, viruses). If this is true, we might expect other infection-response mechanisms to also specifically target free DNA ends, including mechanisms controlling the induction or targeting activity of CRISPR immune systems, as well as response mechanisms found in completely distinct classes of prokaryotic antiviral defense systems.

Glossary

- **Cas:** The CRISPR-associated protein machinery that is involved in acquisition of novel spacers, crRNA processing, and immune targeting.

- **CRISPR Array:** The genomic location at which CRISPR immune memories (spacers) are stored.

- **crRNA:** A short RNA produced from a transcribed and processed CRISPR array. The crRNAs guide the Cas effector proteins to a specific target.

- **Thymic Selection:** A key step during T-cell maturation in the vertebrate thymus that promotes functional immunity while reducing autoimmunity. In order to be retained, developing T-cells must show at least minimal binding to an MHC molecule (promoting immunity) but not excessive binding to MHC-presented self-antigens (reducing autoimmunity).

- **PAM:** A protospacer adjacent motif is found directly upstream of the protospacer in many systems; typical length is 2-8nt.

- **Protospacer:** The target sequence matching a spacer from which that spacer was originally derived (e.g., the target sequence on a viral genome).

- **RM:** Restriction-modification systems are a nearly ubiquitous class of innate immune systems in prokaryotes that differentiate self from non-self using DNA methylation patterns.

- **Spacer:** An individual CRISPR immune memory. Typically, spacers are about 30 bp corresponding to some matching target on a viral or plasmid genome.

- **Lamarkian Inheritance:** A theory of inheritance attributed to Jean-Baptiste Lamark that proposed that organisms pass on physical changes acquired during their lifetime to their offspring. The precise definition of "Lamarkism" and its relevance (if any) to modern biology have been hotly debated.

Acknowledgments

JLW was supported in part by NSF award DGE-1632976. ERW acknowledges funding from the Natural Environment Research Council (NE/M018350/1). The identification of any specific commercial products is for the purpose of specifying a protocol, and does not imply a recommendation or endorsement by the National Institute of Standards and Technology.

References

[1] Suttle, C.A. (2007) Marine viruses-major players in the global ecosystem. *Nat. Rev. Microbiol.* 5, 801

[2] Marston, M.F. *et al.* (2012) Rapid diversification of coevolving marine Synechococcus and a virus. *Proceedings of the National Academy of Sciences* 109, 4544–4549

[3] Goldfarb, T. *et al.* (2015) BREX is a novel phage resistance system widespread in microbial genomes. *The Embo Journal* 34, 169–183

[4] Koonin, E.V. *et al.* (2017) Evolutionary genomics of defense systems in archaea and bacteria. *Annu. Rev. Microbiol.* 71, 233–261

[5] Doron, S. *et al.* (2018) Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* , eaar4120

[6] Ofir, G. *et al.* (2018) Disarm is a widespread bacterial defence system with broad anti-phage activities. *Nature Microbiology* 3, 90

[7] Mojica, F.J.M. *et al.* (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* 60, 174–182

[8] Barrangou, R. *et al.* (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712

[9] Wei, Y. *et al.* (2015) Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in Streptococcus thermophilus. *Nucleic Acids Res.* 43, 1749–1758

[10] Hale, C. *et al.* (2008) Prokaryotic silencing (psi) RNAs in Pyrococcus furiosus. *RNA* 14, 2572–2579

[11] Carte, J. *et al.* (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes & development* 22, 3489–3496

[12] Brouns, S.J. *et al.* (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321, 960–964

[13] Jr, C.A.J. *et al.* (2001) *Immunobiology.* Garland Science, 5th edition

[14] Levy, A. *et al.* (2015) CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* 520, 505–510

[15] Westra, E.R. *et al.* (2013) Type I-E CRISPR-Cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. *PLos Genet.* 9, e1003742

[16] Makarova, K.S. *et al.* (2018) Classification and nomenclature of CRISPR-Cas systems: where from here? *The CRISPR journal* 1, 325–336

[17] Makarova, K.S. *et al.* (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13, 722–736

[18] Shiimori, M. *et al.* (2018) Cas4 nucleases define the PAM, length, and orientation of DNA fragments integrated at CRISPR loci. *Mol. Cell* 70, 814–824

[19] Silas, S. *et al.* (2016) Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase–Cas1 fusion protein. *Science* 351, aad4234

[20] Jackson, S.A. *et al.* (2017) CRISPR-Cas: Adapting to change. *Science* 356, eaal5056

[21] McGinn, J. and Marraffini, L.A. (2019) Molecular mechanisms of CRISPR–Cas spacer acquisition. *Nat. Rev. Microbiol.* 17, 7–12

[22] Makarova, K.S. *et al.* (2019) Evolutionary classification of crispr–cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* , 1–17

[23] Abudayyeh, O.O. *et al.* (2016) C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* 353, aaf5573

[24] Smargon, A.A. *et al.* (2017) Cas13b is a type VI-B CRISPR-associated RNA-guided rnase differentially regulated by accessory proteins Csx27 and Csx28. *Mol. Cell* 65, 618–630

[25] Meeske, A.J. *et al.* (2019) Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. *Nature* , 1

[26] Westra, E.R. *et al.* (2013) CRISPR-Cas systems preferentially target the leading regions of mobf conjugative plasmids. *RNA Biol.* 10, 749–761

[27] Goldberg, G.W. *et al.* (2014) Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature* 514, 633–637

[28] Stern, A. *et al.* (2010) Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet.* 26, 335–340

[29] Yosef, I. *et al.* (2012) Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res.* , gks216

[30] Wei, Y. *et al.* (2015) Cas9 function and host genome sampling in type II-A CRISPR–cas adaptation. *Genes & Development* 29, 356–361

[31] Shiimori, M. *et al.* (2017) Role of free DNA ends and protospacer adjacent motifs for CRISPR DNA uptake in Pyrococcus furiosus. *Nucleic Acids Res.* 45, 11281–11294

[32] Modell, J.W. *et al.* (2017) CRISPR-Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. *Nature* 544, 101–104

[33] Dupuis, M.È. *et al.* (2013) CRISPR-Cas and restriction–modification systems are compatible and increase phage resistance. *Nat. Commun.* 4, 2087

[34] Hynes, A.P. *et al.* (2014) Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages. *Nat. Commun.* 5, 4399

[35] Datsenko, K.A. *et al.* (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* 3, 945

[36] Swarts, D.C. *et al.* (2012) CRISPR interference directs strand specific spacer acquisition. *PLoS One* 7, e35888

[37] Nussenzweig, P.M. *et al.* (2019) Cas9 cleavage of viral genomes primes the acquisition of new immunological memories. *Cell host & microbe*

[38] Patterson, A.G. *et al.* (2017) Regulation of crispr–cas adaptive immune systems. *Curr. Opin. Microbiol.* 37, 1–7

[39] Staals, R.H.J. *et al.* (2016) Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nat. Commun.* 7, 12853

[40] Mojica, F.J.M. *et al.* (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155, 733–740

[41] Heler, R. *et al.* (2019) Spacer acquisition rates determine the immunological diversity of the type II CRISPR-Cas immune response. *Cell Host & Microbe* 25, 242–249

[42] Held, N.L. *et al.* (2010) CRISPR associated diversity within a population of Sulfolobus islandicus. *PLoS One* 5, e12988

[43] Puigbò, P. *et al.* (2017) Reconstruction of the evolution of microbial defense systems. *BMC Evol. Biol.* 17, 94

[44] Yair, Y. and Gophna, U. (2019) Repeat modularity as a beneficial property of multiple CRISPR-Cas systems. *RNA Biol.* 16, 585–587

[45] Varble, A. *et al.* (2019) Recombination between phages and CRISPR-Cas loci facilitates horizontal gene transfer in staphylococci. *Nature Microbiology* 4, 956

[46] Paez-Espino, D. *et al.* (2015) CRISPR immunity drives rapid phage genome evolution in Streptococcus thermophilus. *mBio* 6, e00262–15

[47] Akerlund, T. *et al.* (1995) Analysis of cell size and DNA content in exponentially growing and stationary-phase batch cultures of Escherichia coli. *J. Bacteriol.* 177, 6791–6797

[48] Nielsen, H.J. *et al.* (2007) Dynamics of Escherichia coli chromosome segregation during multifork replication. *J. Bacteriol.* 189, 8660–8666

[49] Sun, L. *et al.* (2018) Effective polyploidy causes phenotypic delay and influences bacterial evolvability. *PLoS Biol.* 16, e2004644

[50] Weinberger, A.D. *et al.* (2012) Viral diversity threshold for adaptive immunity in prokaryotes. *mBio* 3, e00456–12

[51] Weissman, J.L. *et al.* (2019) Visualization and prediction of CRISPR incidence in microbial trait-space to identify drivers of antiviral immune strategy. *The ISME Journal*

[52] Wawrzyniak, P. *et al.* (2017) The different faces of rolling-circle replication and its multifunctional initiator proteins. *Front. Microbiol.* 8, 2353

[53] Nechaev, S. and Severinov, K. (2003) Bacteriophage-induced modifications of host RNA polymerase. *Annual Reviews in Microbiology* 57, 301–322

[54] Toro, N. *et al.* (2019) Recruitment of reverse transcriptase-Cas1 fusion proteins by type VI-A CRISPR-Cas systems. *Front. Microbiol.* 10, 2160

[55] Oliveira, P.H. *et al.* (2014) The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.* 42, 10618–10631

[56] Roberts, R.J. *et al.* (2010) REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* 38, D234–D236

[57] Bolotin, A. *et al.* (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151, 2551–2561

[58] Shah, S.A. *et al.* (2013) Protospacer recognition motifs. *RNA Biol.* 10, 891–899

[59] Paez-Espino, D. *et al.* (2013) Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nat. Commun.* 4, 1430

[60] Rao, C. *et al.* (2017) Priming in a permissive type IC CRISPR–Cas system reveals distinct dynamics of spacer acquisition and loss. *RNA* 23, 1525–1538

[61] Rusinov, I. *et al.* (2015) Lifespan of restriction-modification systems critically affects avoidance of their recognition sites in host genomes. *BMC genomics* 16, 1084

[62] Kupczok, A. and Bollback, J.P. (2014) Motif depletion in bacteriophages infecting hosts with CRISPR systems. *BMC Genomics* 15

[63] Deveau, H. *et al.* (2008) Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J. Bacteriol.* 190, 1390–1400

[64] Andersson, A.F. and Banfield, J.F. (2008) Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* 320, 1047–1050

[65] Laanto, E. *et al.* (2017) Long-term genomic coevolution of host-parasite interaction in the natural environment. *Nat. Commun.* 8, 111

[66] Common, J. *et al.* (2019) CRISPR-Cas immunity leads to a coevolutionary arms race between Streptococcus thermophilus and lytic phage. *Philosophical Transactions of the Royal Society B* 374, 20180098

[67] Fineran, P.C. *et al.* (2014) Degenerate target sites mediate rapid primed CRISPR adaptation. *Proceedings of the National Academy of Sciences* 111, E1629–E1638

[68] Künne, T. *et al.* (2016) Cas3-derived target DNA degradation fragments fuel primed CRISPR adaptation. *Mol. Cell* 63, 852–864

[69] Semenova, E. *et al.* (2016) Highly efficient primed spacer acquisition from targets destroyed by the Escherichia coli type IE CRISPR-Cas interfering complex. *Proceedings of the National Academy of Sciences* 113, 7626–7631

[70] Severinov, K. *et al.* (2016) The influence of copy-number of targeted extrachromosomal genetic elements on the outcome of CRISPR-Cas defense. *Frontiers in Molecular Biosciences* 3, 45

[71] Li, M. *et al.* (2013) Adaptation of the haloarcula hispanica CRISPR-Cas system to a purified virus strictly requires a priming process. *Nucleic Acids Res.* 42, 2483–2492

[72] Li, M. *et al.* (2014) Haloarcula hispanica CRISPR authenticates PAM of a target sequence to prime discriminative adaptation. *Nucleic Acids Res.* 42, 7226–7235

[73] Richter, C. *et al.* (2014) Priming in the Type IF CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res.* 42, 8516–8526

[74] Nicholson, T.J. *et al.* (2019) Bioinformatic evidence of widespread priming in type I and II CRISPR-Cas systems. *RNA Biol.* 16, 566–576

[75] Pyenson, N.C. *et al.* (2017) Broad targeting specificity during bacterial type III CRISPR-Cas immunity constrains viral escape. *Cell host & microbe* 22, 343–353

[76] Redding, S. *et al.* (2015) Surveillance and processing of foreign DNA by the Escherichia coli CRISPR-Cas system. *Cell* 163, 854–865

[77] Dillard, K.E. *et al.* (2018) Assembly and translocation of a CRISPR-Cas primed acquisition complex. *Cell* 175, 934–946

[78] Yang, C.D. *et al.* (2014) CRP represses the CRISPR/Cas system in Escherichia coli: evidence that endogenous CRISPR spacers impede phage p1 replication. *Mol. Microbiol.* 92, 1072–1091

[79] Patterson, A.G. *et al.* (2015) Regulation of the Type IF CRISPR-Cas system by CRP-cAMP and galm controls spacer acquisition and interference. *Nucleic Acids Res.* 43, 6038–6048

[80] Hampton, H.G. *et al.* (2019) Galk limits type IF CRISPR-Cas expression in a CRP-dependent manner. *FEMS Microbiol. Lett.*

[81] Høyland-Kroghsbo, N.M. *et al.* (2018) Temperature, by controlling growth rate, regulates CRISPR-Cas activity in Pseudomonas aeruginosa. *mBio* 9, e02184–18

[82] Høyland-Kroghsbo, N.M. *et al.* (2016) Quorum sensing controls the Pseudomonas aeruginosa CRISPR-Cas adaptive immune system. *Proceedings of the National Academy of Sciences* , 201617415

[83] Patterson, A.G. *et al.* (2016) Quorum sensing controls adaptive immunity through the regulation of multiple CRISPR-Cas systems. *Mol. Cell* 64, 1102–1108

[84] Ratner, H.K. *et al.* (2015) I can see CRISPR now, even when phage are gone: a view on alternative CRISPR-Cas functions from the prokaryotic envelope. *Current opinion in infectious diseases* 28, 267

[85] Makarova, K.S. *et al.* (2013) The basic building blocks and evolution of CRISPR–cas systems. *Biochem. Soc. Trans.* 41, 1392–1400

[86] Popa, O. *et al.* (2011) Directed networks reveal genomic barriers and dna repair bypasses to lateral gene transfer among prokaryotes. *Genome Res.* 21, 599–609

[87] Koonin, E.V. and Wolf, Y.I. (2016) Just how Lamarckian is CRISPR-Cas immunity: the continuum of evolvability mechanisms. *Biology Direct* 11, 9

[88] Wideman, J.G. *et al.* (2019) Mutationism, not Lamarckism, captures the novelty of CRISPR-Cas. *Biology & Philosophy* 34, 12

[89] Koonin, E.V. and Wolf, Y.I. (2009) Is evolution Darwinian or/and Lamarckian? *Biology Direct* 4, 42

[90] Jiang, W. *et al.* (2013) Dealing with the evolutionary downside of CRISPR immunity: Bacteria and beneficial plasmids. *PLos Genet.* 9, e1003844

[91] Gophna, U. *et al.* (2015) No evidence of inhibition of horizontal gene transfer by CRISPR–cas on evolutionary timescales. *The ISME Journal* 9, 2021–2027

[92] Stachler, A.E. *et al.* (2017) High tolerance to self-targeting of the genome by the endogenous CRISPR-Cas system in an archaeon. *Nucleic Acids Res.*

Figure Legends

Figure 1: Observed frequencies of self-targeting spacers can lead to underestimates of the actual rate of autoimmunity. When acquisition is unbiased, strong selection against self-targeting spacers will purge them from the population. When acquisition is biased, self-targeting spacers will not be acquired in the first place. In both cases, the population will end up with very few self-targeting spacers. Thus, even CRISPR systems that lack a mechanism for self vs. non-self recognition may appear to prefer non-self spacers on the basis of population-level immune diversity.

Figure 2: Multiple mechanisms for non-self recognition may rely on the production of excess free DNA ends by mobile genetic elements. Drawn is a schematic of a host cell infected by multiple plasmids. Regions expected to experience a high rate of double-strand break formation are indicated by red rectangles.