Clustered regularly-interspaced short palindromic repeat (CRISPR) sequences cooperate with CRISPR-associated (Cas) proteins to form the basis of CRISPR-Cas adaptive immune systems in prokaryotes. For more than 20 years, these systems were of interest only to specialists, mainly molecular microbiologists, who tried to understand the properties of this unique defense mechanism. In 2012, the potential of CRISPR-Cas systems was uncovered and these were presented as genome-editing tools with an outstanding capacity to trigger targeted genetic modifications that can be applied to virtually any organism. Shortly thereafter, in early 2013, these tools were shown to efficiently drive specific modification of mammalian genomes. This review attempts to summarize, in a comprehensive manner, the key events and milestones that brought CRISPR-Cas technology from prokaryotes to mammals.

Limitations in Mammalian Genome Engineering Are Usually Overcome with Tools Imported from Prokaryotes

When researchers encounter difficulties in modifying the mammalian genome, prokaryotes very often come to the rescue, providing innovative solutions derived directly from nature. Useful reporter genes (such as lacZ) to identify cell types where a gene is expressed [1], robust inducible gene expression (the tetracycline system) [2] and efficient conditional mutagenesis (the cre/loxP system) [3] illustrate how bacteria have often been instrumental in triggering exceptional qualitative advances in our ability to modify the eukaryote genome at will, particularly that of vertebrates and notably, that of mammals.

For more than 30 years, mouse developmental and molecular biologists benefited from random and targeted animal transgenesis techniques [4], from simple methods based on the direct microinjection of DNA constructs into the pronucleus of fertilized eggs, to sophisticated, specific genetic modifications implemented with the help of embryonic stem (ES) cells (see Glossary) and homologous recombination approaches. These and other complementary methods were applied to the genome modification of livestock, fish, and other animal species [5].

For decades, most genome modifications in animals were based on stochastic approaches, whereby the transgene would integrate randomly into the host genome. This often led to chromosomal position effects [6] that resulted in variegation or in unexpected, variable, ectopic, and ultimately, suboptimal transgene expression patterns [7]. Although homologous gene recombination in ES cells provided a suitable solution for undesired random modification of genomes [8], the use of ES cells was limited for many years to mice [9]. It was not until the somatic cell nuclear transfer (SCNT) approach was developed that it became possible to alter
endogenous loci directly in nonrodent species [10]. Both the ES cell and the SCNT approaches nonetheless required laborious, lengthy protocols available only to a handful of research centers and reference laboratories. The need was therefore obvious for better, simpler, more efficient tools to target modification of animal genomes.

The introduction of genome-editing nucleases (at first, the zinc-finger nucleases, ZFN [11], followed by transcription activator-like effector nucleases, TALENs [12], and most recently, clustered regularly interspaced short palindromic repeats (CRISPR)-associated proteins, Cas [13]; see Figure 1) provided an operative solution for targeting any desired genetic modification, for specifically altering genes at will, which was welcomed and rapidly embraced by the scientific community (Box 1).

Initial CRISPR Discoveries

Why then is CRISPR–Cas the most recent and best example of tools imported from prokaryotes for application and benefit to genetic modifications in eukaryotes? What is the origin of the CRISPR-Cas technology? The following paragraphs outline the key discoveries and advances in prokaryote research since the early 1990s that permitted the launch of this disruptive technique in 2012, that is, the unexpected application of a basic research study that resulted in one of the most extraordinary breakthroughs in biotechnology (Table 1 gives a summary of milestones in CRISPR history).

The DNA repeats with dyad symmetry that would later become known as CRISPR were first reported in 1987 by Atsuo Nakata’s group in Japan [14]. These regularly spaced motifs were clustered next to the iap gene, which encodes an aminopeptidase in Escherichia coli K12. A second array was found in the same genome 2 years later, and hybridization assays suggested the presence of similar sequences in very close relatives (Shigella and Salmonella species) [15]. In 1991, interspersed direct repeats (DR) were identified in strains of an evolutionarily distant group of bacteria, the Mycobacterium tuberculosis complex (MTBC) [16]. The DR-intervening sequences, known as spacers, were found to differ among isolates, and hence were harnessed for strain typing [17,18]. The use of DR loci as genetic markers for strain differentiation in MTBC expanded rapidly during the 1990s, and they are still used for this purpose today [19].
Following these first descriptions of CRISPR in Gram-negative bacteria (E. coli) [14,15] and Gram-positive bacteria (MTBC) [16–18], the research group of Francisco Rodríguez-Valera in Spain discovered repeat-spacer clusters in archaea [20]. They detected long stretches of these elements in chromosomal and in resident plasmid regions of strains pertaining to some Haloferax and Haloarcula species [21]. Whereas bacterial systems were not analyzed until the next decade, transcription from repeat loci [20] and the first studies to define a biological role for CRISPR [21] were reported in archaea, in 1993 and 1995, respectively, which described an incompatibility between the chromosome and recombinant multicopy plasmids with identical repeat arrays [21].

From 1996 to 1999, similar repeated elements were found in other archaea and bacteria, and in 2000 these and additional sequences in DNA databases were collected to designate a newly identified type of prokaryotic short repeats that was termed short regularly spaced repeats (SRSR) [22]. The then-rudimentary bioinformatic analyses applied to these SRSR elements indicated that they were partially palindromic and occurred in clusters, regularly interspersed by unique spacer sequences of constant length, and similar to that of the repeats. This very first compilation of a large number of such SRSR in unrelated microorganisms, and the peculiarities deletions (INDELs). NHEJ can also occur in the presence of homologous DNA templates.

**Precursor-crRNA (pre-crRNA):** RNA encoded by a CRISPR locus, usually encompassing the whole repeat array.

**Protospacer adjacent motif (PAM):** a two- to five-nucleotide conserved motif, occurring next to CRISPR targets of most CRISPR-Cas systems, which is necessary for efficient target cleavage.

**Short regularly spaced repeats (SRSR):** original denomination of the family of DNA repeats afterwards renamed as CRISPR.

**Single-guide RNA (sgRNA):** synthetic RNA molecule composed of crRNA and tracrRNA sequences merged through a linker element.

**Somatic cell nuclear transfer (SCNT) approach:** sophisticated reproduction procedure to generate animals from enucleated embryos reconstructed with nuclei from somatic cells.

**Spacer:** CRISPR-intervening sequence, typically of exogenous origin.

**Trans-activating crRNA (tracrRNA):** noncoding RNA of Class II CRISPR-Cas systems required for both the generation of crRNAs and target cleavage.

**Transcription activator-like effector nuclelease (TALEN):** chimeric protein made of a designed transcription activator-like core domain, with specific DNA-binding capacity, fused with the DNA endonuclease domain from the FokI restriction enzyme.

**Zinc-finger nucleases (ZFN):** chimeric protein made of several designed and sequence-specific zinc-finger DNA-binding domains fused with the DNA endonuclease domain from the FokI restriction enzyme.

### Table 1. Milestones, Discoveries, and Achievements in the History of CRISPR-Cas Technology (1987–2013)

<table>
<thead>
<tr>
<th>Year</th>
<th>Milestone</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1987</td>
<td>First report of CRISPR arrays in Gram-negative bacteria</td>
<td>[14]</td>
</tr>
<tr>
<td>1991</td>
<td>First report of CRISPR arrays in Gram-positive bacteria</td>
<td>[16]</td>
</tr>
<tr>
<td>1993</td>
<td>First report of CRISPR arrays in archaea</td>
<td>[23]</td>
</tr>
<tr>
<td>1995</td>
<td>First insight on CRISPR functionality</td>
<td>[21]</td>
</tr>
<tr>
<td>2000</td>
<td>Large number of regularly spaced repeats are found in bacteria and archaea, suggesting a relevant function</td>
<td>[22]</td>
</tr>
<tr>
<td>2002</td>
<td>Regularly spaced repeats of bacteria and archaea are termed with the acronym CRISPR</td>
<td>[23]</td>
</tr>
<tr>
<td>2002</td>
<td>First identification of CRISPR-associated (cas) genes</td>
<td>[23]</td>
</tr>
<tr>
<td>2005</td>
<td>First identification of CRISPR spacers as homologous to sequences in bacteriophages and plasmids</td>
<td>[24–26]</td>
</tr>
<tr>
<td>2005</td>
<td>First suggestion that CRISPR-Cas systems would represent a bacterial defense mechanism</td>
<td>[24]</td>
</tr>
<tr>
<td>2007</td>
<td>First experimental demonstration that CRISPR-Cas systems are involved in acquired immunity against bacteriophages</td>
<td>[31]</td>
</tr>
<tr>
<td>2008</td>
<td>First experimental demonstration that CRISPR-Cas systems interfere with plasmid horizontal transfer, by targeting DNA</td>
<td>[34]</td>
</tr>
<tr>
<td>2008</td>
<td>First description of the role of CRISPR small RNAs (crRNA) as the guides for CRISPR interference</td>
<td>[33]</td>
</tr>
<tr>
<td>2008</td>
<td>It is anticipated that conserved sequences next to protospacers are important for CRISPR-mediated phage resistance</td>
<td>[37]</td>
</tr>
<tr>
<td>2010</td>
<td>First description of the CRISPR-Cas interference mechanism through Cas proteins cutting target DNA at precise sites</td>
<td>[38]</td>
</tr>
<tr>
<td>2011</td>
<td>Identification of trans-activating crRNAs (tracrRNAs)</td>
<td>[41]</td>
</tr>
<tr>
<td>2011</td>
<td>First successful transfer of a CRISPR-Cas system between two evolutionary distant organisms: from Streptococcus thermophilus to Escherichia coli</td>
<td>[80]</td>
</tr>
<tr>
<td>2012</td>
<td>First reports documenting functional CRISPR-Cas systems reconstructed in vitro and suggesting their potential application as RNA-programmable genome editing tools</td>
<td>[42,48]</td>
</tr>
<tr>
<td>2013</td>
<td>First reports demonstrating the use of CRISPR-Cas tools for efficient genome editing in mammalian cells</td>
<td>[52,53]</td>
</tr>
<tr>
<td>2013</td>
<td>First reports showing efficient genome engineering at multiple loci in mice, through the use of CRISPR-Cas tools</td>
<td>[59–61]</td>
</tr>
</tbody>
</table>
of their genomic structure suggested an underlying, probably important but yet-unknown biological function [22].

By examining SRSR loci in many archaea and bacteria, it was possible to detect a set of four genes in their vicinity (cas1–cas4) that encoded proteins possibly associated with the clustered repeats [23]. This 2002 publication [23], from a team of microbiologists in the Netherlands, credited a proposal by Francisco J.M. Mojica’s group to unify the diversity of names and labels used in the literature for these DNA repeated elements under the concept “clustered
regularly-interspaced short palindromic repeats”, with the acronym CRISPR. This name rapidly became popular and was readily accepted in the still-incipient field.

**CRISPR-Cas Systems Involved in Bacterial Immunity**

Until 2005 there was no clear evidence of the biological function or the underlying mechanism associated with the CRISPR arrays. The first suggestion of a putative link between CRISPR and prokaryotic immunity arose from the discovery that some of the spacers were homologous to DNA sequences from bacteriophages or plasmids [24]. Indeed, some bacteria carrying CRISPR arrays were known to be resistant to phages, which were shown to bear these sequences [24]. This fundamental observation, independently confirmed by other laboratories [25,26], strongly supported the idea that CRISPR elements were associated with some sort of a bacterial defense system. A tentative mechanism of action was proposed [24], involving target recognition by CRISPR-RNA molecules [20,27], reminiscent of the eukaryotic interference RNA [24,28]. One of these 2005 studies [26] was also the first to observe a short stretch of conserved DNA next to the protospacers (the original sequence in viruses and other foreign genetic elements from which the spacers were derived [29]), which would later be termed the protospacer adjacent motif (PAM) [30].

CRISPR element involvement in bacterial immunity was confirmed experimentally in 2007 by Rodolphe Barrangou, Philippe Horvath and collaborators [31]. This landmark publication demonstrated that resistance to bacteriophage infection can be developed by a sensitive bacterial strain through acquisition of spacers that match the viral genome [31]. CRISPR arrays in prokaryotes had been previously found to produce numerous small RNAs [20,27,32]. In 2008, Brouns et al. [33] showed that these CRISPR RNAs (crRNAs) have a crucial role in driving antiviral defense. Moreover, their results suggested that DNA was the target of CRISPR action. Indeed, Luciano Marraffini and Erik Sontheimer subsequently reported that CRISPR could efficiently interfere with the horizontal transfer of plasmid sequences through DNA targeting [34]. The observation that there is a dynamic interplay between the spacer content of CRISPR arrays and potential targets in natural microbial communities further supported a connection between CRISPR genotype and host immunity [35]. These findings corroborated CRISPR-Cas as a general genetic barrier to horizontally transferred DNA and an efficient adaptive immune system in prokaryotes [36].

Over the following years, these seminal discoveries allowed further detailed characterization of the CRISPR mechanism of action. PAM sequences appeared to be important for interference [37], and these motifs emerged as a common feature in many systems [29,30], further supporting their functional relevance. The interference/defense mechanism was then defined as Cas protein cleavage of target DNA next to a PAM [38]. However, some systems were documented to cleave RNA instead, not requiring a specific sequence motif next to the target [39,40]. Another crucial piece of information of CRISPR-Cas was uncovered in 2011 [41], the existence in particular systems of an additional small RNA molecule, the trans-activating crRNA (tracrRNA), which was needed to generate mature crRNA molecules. TracrRNAs are exclusive of some Class 2 systems (see below), where, in addition to their implication in crRNA maturation, they also bridge crRNA and the Cas protein responsible for target cleavage [42]. By that time, many CRISPR-Cas systems had been identified [43] and partially characterized in archaea and bacteria [44] (Box 2). This large amount of information on CRISPR-Cas systems led to the first attempt to classify them, from an evolutionary perspective, into distinct functional and structural types (Type I, II, and III) and subtypes [45]. A top level classification, the Class category (that is, Class 1, comprising types I, III, and IV, and Class 2, including types II and V), has recently been adopted [46]. In contrast to Class 1, Class 2 systems require only one Cas protein (Cas9 in the case of Type II systems, instead of a multiprotein complex as in Class 1) for target recognition and cleavage, producing single DNA cuts [47]. These properties of Class 2 systems explain why
Type II were chosen among characterized CRISPR-Cas systems for development of future applications based on target cleavage.

CRISPR-Cas Introduced as New Genome-Editing Tools

In the summer of 2012 two independent teams reported the biochemical properties of partially reconstituted CRISPR-Cas systems in vitro, and went beyond the state-of-the-art knowledge by suggesting that its elements could be used as genome-editing tools [42,48]. In a combined effort, the laboratories of Jennifer Doudna in the USA and Emmanuelle Charpentier in Sweden reconstituted in vitro and demonstrated the function of three of the six elements of the Streptococcus pyogenes CRISPR-Cas9 system (tracrRNA, crRNA and Cas9 protein; see Figure 1A), which target a double-strand break (DSB) and cut a specific DNA sequence homologous to the crRNA spacer region [42]. The same study also showed how tracrRNA and crRNA could be fused into a sgRNA, a chimeric synthetic RNA molecule that retains the full properties of the two original small RNAs. This remarkable achievement further simplified an already very simple RNA-directed DNA endonuclease mechanism for use as a tool in programmable genome editing [42].

In parallel to the Doudna and Charpentier study [42,49], the Siksnys group collaborated with Barrangou and Horvath to assess the function of the Streptococcus thermophilus Type II system in vitro [48]. Their results were similar, and demonstrated the crucial role of crRNA and Cas9 complexes in directing DSB in crRNA-targeted DNA sequences. They also understood the relevance of these findings, and proposed that universal programmable RNA-guided DNA endonucleases could be engineered as unique molecular tools for RNA-directed DNA surgery [48]. These two inspiring publications triggered a few laboratories to assess the conjectured genome-editing capacities of these newly characterized bacteria-derived, RNA-programmable DNA endonucleases [50].

It took less than 6 months to experimentally confirm the predictions of these founder publications of the nascent field of CRISPR-Cas9 technology. By January 2013, three independent US teams, one led by Luciano Marraffini [51], another by Feng Zhang [52], in collaboration with Luciano Marraffini, and a third one by George Church [53], communicated the successful editing of bacterial [51] and mammalian genomes [52,53] using Cas9. CRISPR-Cas9 tools derived from S. pyogenes were improved, and the cas9 gene adapted to mammals by human codon-optimization, for efficient genome modification of various mammalian cell types from mice and humans [52,53], including pluripotent cells. Later the same month, an independent study also reported the formation of DSB at a specific locus, in human cells, using CRISPR-Cas9 methods [54]. The current CRISPR excitement had begun and the rest of the scientific community, including many who had probably not noticed the two 2012 publications on in vitro studies, learned of these far-reaching tools for genome editing in eukaryotes [55–57] and prokaryotes [57,58].

It took a few more months, still in 2013, before the first publications appeared that reported similar in vivo findings in vertebrates. A team in China presented preliminary results on genome
Box 3. The Successful Application of CRISPR-Cas Tools for Mammalian Genome Editing

In mammals, CRISPR-Cas9 tools have now been applied successfully in numerous projects:
- To functionally assess mutations in coding [53,84] and noncoding [85,86] mouse genomic DNA sequences.
- To generate genome-edited nonhuman primates [87].
- To explore genome editing in nonviable human embryos [88].
- To alter epigenetic markers leading to the activation of specific genes [89].
- To generate genome-edited livestock for biomedical [90] and agricultural applications [91].
- To inactivate multiple retroviral genomic insertions simultaneously in porcine cells, for xenotransplantation purposes [92].
- To assess gene and cell therapy approaches for human diseases [93].
- To develop lentiviral knockout libraries that facilitate forward genetics in human cells [94–96].
- To produce chromosomal rearrangements, large inversions, and translocations observed in patients that were extremely difficult to model in human cells [97] and in mice [98].
- To apply rapid somatic genome-editing approaches for cancer modeling [99].
- To correct the genome in patient-derived induced pluripotent cells (iPS) [100,101].
- To develop in vivo genome-editing protocols [67] and promising somatic gene therapy approaches for devastating degenerative [102–104] or rare [105,106] diseases in animal models.

Concluding Remarks

Researchers have been astonished to confirm the ease, efficiency, and apparently unlimited number of applications that arise from the use of CRISPR-Cas components. It is important, however, that we not forget the origin of these tools. CRISPR-Cas systems have been evolving in bacteria and archaea subject to strong selection by infectious genetic elements for billions of years. It should hence not be a surprise to discover that, after this long period of optimization of tools meant to cleave intruder DNA, these elements would also perform most efficiently for genome editing outside their natural context. That which today benefits mammals is derived from what once evolved in archaea and bacteria. Likewise, the reagents we now apply for genome editing in mammals could not be understood without the systematic, fundamental, and often underestimated contribution of the many microbiologists who discovered, dissected and described the functional compounds of the native CRISPR-Cas systems [47,64].

The characterization of other known and to be discovered CRISPR-Cas systems from different prokaryotes [46,65,66] ensures a regular flow of new reagents with slightly different and useful properties [67,68]. Notably, this growing list includes Cpf1 [69], a biochemically validated nuclease of Type V CRISPR-Cas systems [46]. Compared to Cas9, Cpf1 requires a different, T-rich PAM sequence at the 5’ location, and interacts with a single shorter RNA molecule, producing PAM-distal protruding ends upon cutting the DNA [69]. Together with the generation of improved mutant variants of currently known Cas proteins [70–72], through the structure-guided rational design of Cas9 PAM variants [73,74], and with the description of similar gene-editing properties from unrelated immune systems of prokaryotes (i.e., prokaryotic Argonaute,
pAgo [75,76]), all these approaches will further widen our capacity to edit complex genomes at will (see Outstanding Questions).

Acknowledgments
The authors wish to thank Luciano Marraffini and Davide Serugia for critical reading of this manuscript and Catherine Mark for editorial assistance. CRISPR-Cas research in FJMM and LM laboratories is funded through the Spanish Ministry of Economy and Competitiveness (MINECO) [BIO2012-39890 and BIO2015-70787-R] to LM, and [BIO2014-33028-P] to FJMM, the Centre for Networked Biomedical Research on Rare Diseases (CIBERER-ISCIII) [ACCI 2015] to LM, and the Biomedical and Biological Sciences (ISBMBS) European Cooperation in Science and Technology (COST) action [BM1308 SALAAM] to LM. FJMM and LM conceived, discussed and wrote the manuscript.

References
25. Poutrel, C. et al. (2005) CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology 151, 653-663
26. Botkin, A. et al. (2005) Clustered regularly interspaced short palindromic repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151, 2551-2561
47. van der Oost, J. et al. (2014) Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nat. Rev. Microbiol. 12, 479–492
49. Lander, E.S. (2016) The Heroes of CRISPR. Cell 164, 18–28
94. Wang, T. et al. (2014) Genetic screens in human cells using the CRISPR-Cas9 system. Science 343, 80–84
95. Shalem, O. et al. (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 343, 84–87
100. Huang, X. et al. (2015) Production of gene-corrected adult beta globin protein in human erythrocytes differentiated from patient iPSCs after genome editing of the sickle point mutation. Stem Cells 33, 1470–1479