View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Diposit Digital de Documents de la UAB

10506–10519 *Nucleic Acids Research, 2019, Vol. 47, No. 20* *Published online 4 October 2019*
doi: 10.1093/nar/gkz831

SURVEY AND SUMMARY

Synonymous genome recoding: a tool to explore microbial biology and new therapeutic strategies

Miguel Angel Martínez*, Ana Jordan-Paiz, Sandra Franco and Maria Nevot

IrsiCaixa, Hospital Universitari Germans Trias i Pujol, Universitat Autònoma de Barcelona (UAB), Badalona, Spain

Received August 13, 2019; Revised September 12, 2019; Editorial Decision September 16, 2019; Accepted September 30, 2019

ABSTRACT

Synthetic genome recoding is a new means of generating designed organisms with altered phenotypes. Synonymous mutations introduced into the protein coding region tolerate modifications in DNA or mRNA without modifying the encoded proteins. Synonymous genome-wide recoding has allowed the synthetic generation of different small-genome viruses with modified phenotypes and biological properties. Recently, a decreased cost of chemically synthesizing DNA and improved methods for assembling DNA fragments (e.g. lambda red recombination and CRISPR-based editing) have enabled the construction of an *Escherichia coli* variant with a 4-Mb synthetic synonymously recoded genome with a reduced number of sense codons ($n = 59$) encoding the 20 canonical amino acids. Synonymous genome recoding is increasing our knowledge of microbial interactions with innate immune responses, identifying functional genome structures, and strategically ameliorating cis-inhibitory signaling sequences related to splicing, replication (in eukaryotes), and complex microbe functions, unraveling the relevance of codon usage for the temporal regulation of gene expression and the microbe mutant spectrum and adaptability. New biotechnological and therapeutic applications of this methodology can easily be envisaged. In this review, we discuss how synonymous genome recoding may impact our knowledge of microbial biology and the development of new and better therapeutic methodologies.

SYNONYMOUS MUTATIONS AND GENOME CODING

The genetic code is common to all organisms and extensively redundant. Based on the nature of triplet encoding and the fact that there are four different nucleotides, only $4^3 = 64$ different codons are possible. Of these possible 64 codons, 61 are sense codons, coding for 20 amino acids and 3 are stop codons, coding for termination of translation. Eighteen of these amino acids are encoded by more than one synonymous codon. Consequently, a given amino acid might be encoded by one, two, three, four or six different codons (Figure 1). Importantly, the observed ratios of synonymous codons are highly non-random (1). This phenomenon is termed codon usage bias. Foreign protein expression in *Escherichia coli* is well-known to be influenced significantly by the presence or absence of uncommon codons, and heterologous protein expression often requires coding sequence modification (2).

A suggested model of codon usage bias is that abundant codons correlate with the abundances of isoaccepting tRNAs, which correlate with protein production levels (3). However, tRNA availability is not the sole determinant of protein production (4). Synonymous codon mutations can impact many molecular phenotypes, including transcription modifications (5,6), translation initiation (7–9), translation elongation (10,11), translation accuracy (12,13), RNA stability (14), RNA structure and folding (8,15), RNA splicing (16,17), RNA toxicity (18) and cotranslational folding (23). In the case of eukaryotes, synonymous substitutions can affect chromatin organization (19), enhancer functions (20,21) and microRNA targeting (20,22). Due to the impact of synonymous mutations on molecular phenotypes, it is not difficult to assume that they are under selective pressure. Moreover, natural selection has generated unequal codon usage across genes and genomes (24,25). Accordingly, synonymous nucleotide positions have been modelled by genome-wide mutational processes during evolution (26,27). Synonymous codon usage is neither random nor neutral (28–30). Organisms and different genes from the same organism include specific codons at different frequencies (31). Human codon frequencies for synonymous codons are not equal (Figure 1). Synonymous codon frequencies varies among different taxa, with codon usage in

*To whom correspondence should be addressed. Tel: +34 934656374; Fax: +34 934653968; Email: mmartinez@irsicaixa.es

© The Author(s) 2019. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Gly	GGG	16.5	Arg	AGG	12.0	Trp	UGG	13.2	Arg	CGG	11.4
Gly	GGA	16.5	Arg	AGA	12.2	Stop	UGA	1.6	Arg	CGA	6.2
Gly	GGU	10.8	Ser	AGU	12.1	Cys	UGU	10.6	Arg	CGU	4.5
Gly	GGC	22.2	Ser	AGC	19.5	Cys	UGC	12.6	Arg	CGC	10.4
Glu	GAG	39.6	Lys	AAG	31.9	Stop	UAG	0.8	Gln	CAG	34.2
Glu	GAA	29.0	Lys	AAA	24.4	Stop	UAA	1.0	Gln	CAA	12.3
Asp	GAU	21.8	Asn	AAU	17.0	Tyr	UAU	12.2	His	CAU	10.9
Asp	GAC	25.1	Asn	AAC	19.1	Try	UAC	15.3	His	CAC	15.1
Val	GUG	28.1	Met	AUG	22.0	Leu	UUG	12.9	Leu	CUG	39.6
Val	GUA	7.1	Ile	AUA	7.5	Leu	UUA	7.7	Leu	CUA	7.2
Val	GUU	11.0	Ile	AUU	16.0	Phe	UUU	17.6	Leu	CUU	13.2
Val	GUC	14.5	Ile	AUC	20.8	Phe	UUC	20.3	Leu	CUC	19.6
Ala	GCG	7.4	Thr	ACG	6.1	Ser	UCG	4.4	Pro	CCG	6.9
Ala	GCA	15.8	Thr	ACA	15.1	Ser	UCA	12.2	Pro	CCA	16.9
Ala	GCU	18.4	Thr	ACU	13.1	Ser	UCU	15.2	Pro	CCU	17.5
Ala	GCC	27.7	Thr	ACC	18.9	Ser	UCC	17.7	Pro	CCC	19.8

Figure 1. The standard genetic code and codon bias. Each amino acid is encoded by multiple codons, except methionine and tryptophan. On the right are the human frequencies per thousand for each codon. The Kazusa database (http://www.kazusa.or.jp/codon/) was used to compile human codon usage. Codons carrying CpG and UpA dinucleotides are underrepresented.

mammals differing from that in bacteria or insects. Synonymous codons may also play distinct roles at different sites in the genome, and epistatic interactions may occur among codons within and between genes (2). Although this review mainly deals with prokaryotes and viruses, it is worthy to mention that genome-wide association studies and ulterior functional analyses have recently shown that synonymous substitutions can impact complex human diseases, such as cancer (32) and autism (33).

Synonymous genome recoding alters not only codon usage, but also codon-pair usage, and dinucleotide frequency (e.g. CpG/UpA) (reviewed in (34)). Recently, it has been described that efficient mRNA translation is also determined by a triplet-of-triplet genetic code, that is, translation of a given codon is influenced by the two preceding codons and that given codon (56). In addition to codon usage bias, which has been thoroughly studied in many species, codon-pair frequencies (termed codon-pair bias) are different in different organisms. In any given genome, different frequencies of codon-pairs are found than would be expected based on the individual codon usage bias of that genome (reviewed in (31)). Codon-pair usage bias was first described in bacteria (35), and then in all studied organisms (30). An analysis of 14 795 human genes showed that some synonymous codon-pairs are used more or less frequently than expected (36). For example, based on the codon frequencies, the amino acid pair Ala–Glu would be expected to be encoded equally as frequently by GCCGAA and GCAGAG. However, even though the codon-pair GCCGAA contains the most frequent Ala codon, it is strongly underrepresented, as it is used only one-seventh as often as GCAGAG. Analysis of an extended data set of 35 770 human mRNA sequences had similar findings (37). Whether codon-pair usage bias is shaped by selection is controversial. Looking at the underrepresented codon-pair GCCGAA, we easily notice that a CpG dinucleotide is generated between the third base of the first codon and the first base of the accompanying codon. Many of the underrepresented human codon-pairs generate CpG dinucleotides (38). CpG dinucleotides are less frequent than expected in both humans and other mammals (Figure 1) (39). The mechanistic reasons for this underrepresentation are unclear. Compared to other dinucleotides, the mutation rate of the CpG motif is at least 10-fold higher, especially in higher primates (40), likely due to cytosine methylation resulting in deamination to thymine (40). Moreover, 5-methylcytosine in the context of the CpG sequence is an epigenetic mechanism associated with transcriptional repression (41,42). These findings strongly suggest that it is very difficult to distinguish between codon-pair usage bias and dinucleotide frequencies. Because codon-pair bias and dinucleotide frequency are inevitably connected, whether one is generating the other is still in question (37,38,43,44). Interestingly, most mammalian RNA viruses have low frequencies of CpGs (45,46). Furthermore, viruses with high CpG frequencies may be more recognizable by pathogen innate immune sensors (47–50). A broad spectrum antiviral protein as the zinc-finger antiviral protein (ZAP) has been shown to inhibit HIV-1 production by cells infected with CpG-enriched virus genomes (42). Nevertheless, there are other hypotheses to

Table 1. Methods for large-scale genome synonymous recoding

Method	References
Codon bias usage	(68,72,74,77,79,83,85,152)
Codon-pair usage	(64,69,70,71,73,75, 78,80,82,84,88,89,153)
Dinucleotide frequency (e.g. CpG/UpA)	(37,86,130)
Codon-triplet bias usage	(56)
Codons that could generate stop mutations after a single nucleotide substitution	(136)

explain the codon-pair bias. In particular, work with *E. coli* has proposed that the size of the tRNA variable loop determines whether the 3′ nucleotide adjacent to the codon has a context effect (51), but others have claimed that the P-site wobble position within the codon–anticodon interaction, and the A-site anticodon loop and acceptor stem have a decisive effect on the observed genomic codon-pair patterns (52). More recent studies in yeast have shown that the wobble base pairing plays a critical role in whether a codon-pair has an inhibitory effect on translation (53). This work is in agreement with former studies, which would support the role of codon-pair usage bias. In addition, numerous studies have offered evidence of the effect of codon-pair usage bias on translational fidelity and rate (29,54,55). Both codon-pair usage bias and dinucleotide content have been incorporated into algorithms for gene optimization or deoptimization (Table 1) (31). Furthermore, work performed with the *Salmonella* flgM gene, which coordinates flagellar gene expression and flagellum assembly, has shown that mRNA translation is determined by a triplet-of-triplet genetic code (56). That is, and similar to codon-pair usage, the efficiency of translating a particular codon is influenced by the sequence of the immediately adjacent flanking codons. A model explains these codon-context effects by suggesting that codon recognition by elongation factor-bound aminoacyl-tRNA is initiated by hydrogen bond interactions between the first two nucleotides of the codon and anticodon, and then stabilized by base-stacking energy over three successive codons (54). These data suggest that the context of the genome sequence should consider codon triplets in addition to codon-pairs and highlight the challenge of designing genes for maximal expression, whether by natural selection or in the laboratory.

These studies illustrate the possible fitness costs of synonymous genome recoding, which can be under strong selection. As suggested previously, their importance in evolution may be under-appreciated (57). Synonymous mutations influence fitness in bacteria (8,11,18,57–67) and viruses (36,68–89). In addition, synonymous rewriting of protein-coding regions allows the encoded amino acid sequences to be retained, but it may remove other genetic information, including hidden control elements embedded within amino acid coding regions (56). We and others hypothesize that synonymous genome recoding offers an opportunity to manipulate the microbial replication capacity and biology.

SYNTHETIC GENOMICS

The ability to read (sequencing) and write (synthesis) DNA sequences has greatly increased our capability to understand and manipulate biological systems. Moreover, DNA sequencing and synthesis are the fundamental technologies in the development of synthetic genomics, which aims to determine the fundamentals and design of a genome and engineer novel biological functions. A genome is defined as synthetic when all building blocks of the final DNA molecule are generated by chemical synthesis (90,91). Synthetic techniques for designing and manufacturing DNA offer the promise of completely controlling an organism's genetic information (92).

In the last decade, sequencing technologies have evolved considerably. Following the development of several novel DNA sequencing methods in the late 1990s and early 2000s (93,94), different technologies have become commercially available for so-called next-generation sequencing (NGS). NGS in all branches of the tree of life has increased our capability to understand biological systems and the interrelated genetic nature of different organisms. Knowledge regarding the genetic diversity of microorganisms, the microbiome, and human disease has also increased due to NGS. Seven years ago, we started to sequence >15 petabases per year (95). This wealth of sequence data also enhanced the discipline of synthetic biology and genomics. In order to better understand DNA sequence information, we need to manipulate it and create new synthetic DNA reads. In contrast to the rapid development of DNA sequencing techniques, the capability to synthesize DNA has not increased at the same pace (90). Writing is much more difficult than reading.

Synthetic genomics starts with short synthetic single-stranded DNA fragments called oligonucleotides. Developments in the 1980s in polymerase chain reaction (PCR) allowed the rapid assembly of overlapping oligonucleotides. Kary Mullis and Michael Smith were awarded with the Nobel Prize in Chemistry in 1993 for the invention of PCR and the fundamental application of synthetic oligonucleotides in DNA manipulation. In addition to other applications in molecular biology, the invention of PCR and the chemical synthesis of DNA oligonucleotides opened the door >30 years ago to synthetic genomics.

DNA chemical synthesis allows the synthesis of oligonucleotides no more than 200 nucleotides in length (reviewed in (90)). This is an intrinsic length limit based on phosphoramidite chemistry coupling efficiency. Yields of full-length oligonucleotides can be affected by spurious depurination of the synthesized oligonucleotide during the acid deprotection steps of the synthesis cycle and becomes especially problematic for longer oligonucleotide sequences. The cost of column-based oligonucleotide synthesis has not decreased substantially in the last decade. Custom commercial oligonucleotide prices range from $0.05 to $0.50 per base depending on the scale of synthesis, the length of the oligonucleotide, and the supplier. Moreover, in contrast to custom pricing for sequencing, oligonucleotide prices have increased in the last few years. Alternatively, the prices have decreased for DNA synthesis via microarray-based oligonucleotide synthesis. Compared to column-based tech-

niques, array oligonucleotide synthesis can be considerably cheaper. The problem with making oligonucleotides from these array platforms for gene synthesis applications is that they produce many more synthesis-related errors than column-synthesized oligonucleotides (96,97). Minimizing synthesis errors is crucial for large-scale genome synthesis. One wrong base out of >1 million in an essential gene renders the genome of *Mycoplasma mycoides* (1.1 Mb) inactive (98). Both the length and price of oligonucleotide synthesis remain the main technical challenges to speeding up synthetic genomics development.

Overlapping single-stranded DNA oligonucleotides 100–200 nucleotides in length are designed to facilitate the assembly and generation of double-stranded DNA. Several oligonucleotides can easily be assembled by PCR or isothermal amplification to generate 1000–2000 base pair (bp) DNA fragments (Figure 2). Using conventional cloning methods, these synthetic DNA fragments can be cloned into a bacterial plasmid vector and individual clones isolated and sequenced by Sanger sequencing. Importantly, both chemical DNA synthesis and PCR are error prone. The probability of any given sequence containing a mutation increases with the length of the DNA, which means that more clones have to be sequenced to obtain a completely correct DNA clone. Five to six overlapping PCR fragments 1000–2 000bp long can be assembled and cloned in a plasmid vector using the Gibson protocol to obtain up to 10–15 kb DNA fragments (99). Via PCR recombination and the isothermal cloning reaction, it has been shown to be possible to synthesize the entire 16.3 kb mouse mitochondrial genome from 600 overlapping 60-nucleotide synthetic oligonucleotides (100). Cloned fragments of this size can be bought from commercial vendors at $0.10 to $0.50 per bp depending on the fragment size and cloning vector. Comparing the prices for custom oligonucleotides and gene synthesis, it may be worthwhile to use commercial DNA fragment synthesis platforms. The Gibson protocol is probably the most commonly used method for assembling multiple pieces of DNA into larger constructs. This method uses a one-pot isothermal technique, which utilizes an enzyme mixture containing a thermostable DNA polymerase, DNA ligase and exonuclease to chew-back, anneal, and repair adjacent overlapping DNA sequences in order to assemble the required DNA construct. More recent work has updated and improved this method to further expand it (101). Alternatively, the introduction of 1557 synonymous TTA/TTG to CTA/CTG leucine codon changes across 200 kb of *Salmonella typhimurium* genomic DNA was performed using stepwise integration of rolling circle amplified segments (SIRCAS) (102). Genome recoding by SIRCAS involves iterative recombination of recoded DNA segments containing a selection marker (typically kanamycin or chloramphenicol resistance cassettes). After each step, one selection marker is gained while the other is lost, providing a readout for successful recoding.

The first full-genome synthetic reconstructions were performed with RNA viruses because of their smaller genomes (∼10 kb). RNA virus genomes can be constructed easily and reasonably inexpensively using the method described above (Figure 2). In 2002, the first complete synthesis of a viral RNA genome was performed (103); *in vitro* transcribed RNA generated from 7.5 kb of synthetic poliovirus DNA representing its genomic sequence yielded infectious virus particles after transfection of HeLa cell-free extract. One year later, Smith and colleagues were able to generate the complete infectious genome of bacteriophage X174 (5386 b) from a single pool of chemically synthesized oligonucleotides in 14 days (104). Recently, additional viral RNA and DNA genomes were synthesized, up to a size of 212 kb with the DNA genome of horsepox virus (105). Horsepox virus is a DNA virus belonging to the genus Orthopoxvirus and is closely related to a group of viruses including variola virus, the etiological agent of smallpox, and vaccinia virus, the vaccine virus used to eradicate smallpox. Variola virus no longer exists in nature, and the last naturally occurring case of smallpox was in 1977. Synthetic horsepox virus was generated by purchasing DNA fragments of ∼30 kb each that collectively represent the entire 212 kb genome with some overlap between the fragments. The most time-consuming part of the project was construction of the requested DNA fragments by the DNA synthesis company (105).

The synthetic 30 kb DNA fragments of horsepox virus were co-transfected into cells already infected with Shope fibroma virus, which served as a helper virus to provide the horsepox virus with the viral proteins it needed to conduct DNA transcription and replication (106). However, we cannot utilize this methodology to construct larger genomes of other microorganisms, such as bacteria. An approach taken by the Venter group was to reduce the genome of *M. mycoides* by approximately 50% (107). Three rounds of design, synthesis, and testing were necessary to produce JCVI-syn3.0 (531 kb), which has a genome smaller than that of any autonomously replicating cell found in nature. Gene-deletion methods allowed reduction of the *E. coli* K-12 genome by 15% (108). Using this reduced *E. coli* genome, Wang, Chin and colleagues developed a method to replace genomic DNA with long (∼100 kb) synthetic DNA through the in vivo excision of double-stranded DNA from an episomal replicon by CRISPR/Cas9, coupled to lambda red-mediated recombination and simultaneous positive and negative selection (Figure 3) (109). Using this approach, this group was the first to complete a 4-Mb synthetic version of the *E. coli* genome (65). Synthetic DNA fragments approximately 10 kb in length were purchased from a DNA synthesis company. Between 9 and 14 of these synthetic fragments were assembled by homologous recombination in *Saccharomyces cerevisiae* using bacterial artificial chromosomes (BACs) to generate 37 fragments 91–136 kb in length. These 37 BAC fragments contained Cas9 cleavage sites that allowed its introduction into the previously reduced *E. coli* K-12 genome.

The declared cost of the synthesis of the 212 kb genome of horsepox virus was $100 000 (105). An international consortium of 21 different institutions is synthetically reconstructing all 16 chromosomes of *S. cerevisiae*. The consortium has already described the synthesis of artificial yeast chromosome synIII (273 kb) (110). Successively, five additional chromosomes have been reported (111–115). Similar to *E. coli*, the redesigned yeast chromosomes had repetitive sequences (tRNA genes, introns and transposons) removed to allow genome reduction. In contrast, CRISPR has not

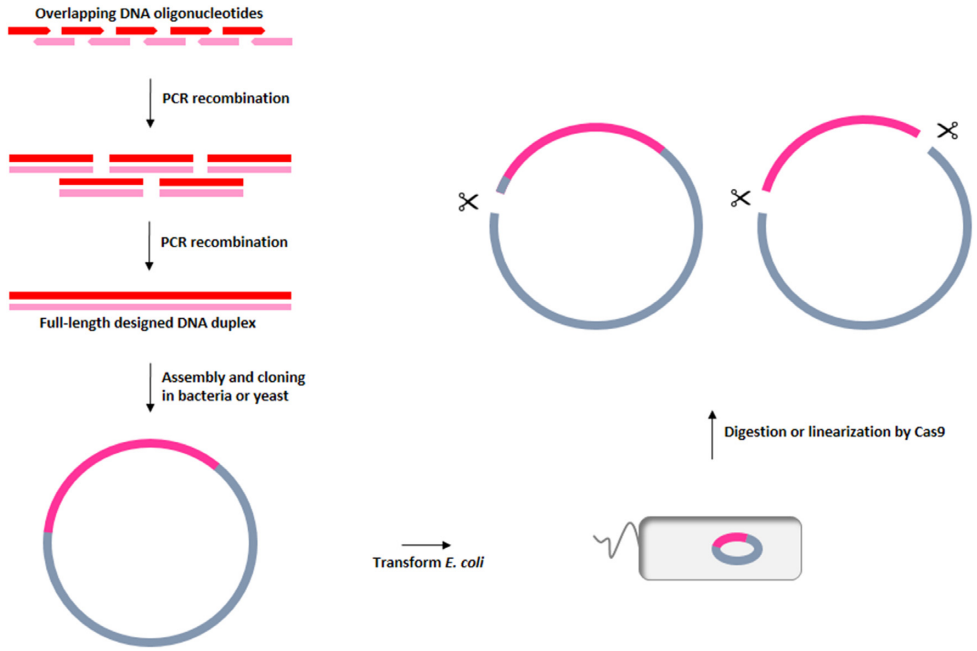

Figure 2. Synthetic DNA oligonucleotide assembly and cloning. Overlapping oligonucleotides encoding a DNA duplex are successively assembled via polymerase chain reaction (PCR) or isothermal amplification. Designed DNA duplex are now cloned in yeast or bacterial plasmid that after *E. coli* transformation can be profusely amplified, sequenced and linearized for further cloning.

been used for progressive genome reduction. At the beginning of this century, the first individual human genome was sequenced (116,117). The Human Genome Project was time-consuming and very expensive, but it accelerated the development of new sequencing methods. Lowering DNA synthesis costs should be one of the major goals in the field of synthetic genomics (118).

SYNONYMOUS MUTATIONS AND MICROBE BIOLOGY

Aiming the construction of synthetic genomes will not only enable the development of new technologies, but will also allow an immense increase in knowledge in the biological sciences. Large-scale synonymous recoding has broad applications for modulating diverse biological systems (34). Synonymous recoding of synthetic genomes is an effective alternative to conventional mutational analyses involving a limited segment rather than broad-based genome recoding. Furthermore, genome recoding may allow us to approach biological questions that are very difficult to answer through conventional methodologies (Table 2).

Recently, life was elegantly demonstrated to function with a reduced number of synonymous sense codons (65). An *E. coli* was created in which the entire 4-Mb genome was replaced with synthetic DNA that no longer uses the stop codon UAG (amber stop codon) and the two serine codons UCG and UCA. By recoding 18 2014 codons, an organism was created with 59 codons to encode the 20 amino acids. The previously essential bacterial tRNAs that recognize these three codons can now either be deleted or reassigned to recruit non-canonical amino acids (ncAAs) into proteins synthesized in cells (109). The synthesis of a 57-codon *E. coli* genome is also under way (119). NcAAs are those not naturally encoded or found in the genetic code of any organism. Reassigned tRNAs are being used to expand the genetic code by enabling the co-translational and site-specific incorporation of diverse ncAAs into proteins synthesized in cells (120). To expand the genetic code, an orthogonal aminoacyl-tRNA synthetase–tRNA pair is used to direct the incorporation of an ncAA into proteins through an unassigned codon (e.g. the amber stop codon UAG or the two serine codons UCG and UCA) introduced at the desired site in a gene of interest (121–123). The possibility of integrating ncAAs with diverse structures and properties into proteins in diverse microorganisms will allow proteins to be engineered with alternative structures and functions, and to develop new therapeutic agents (see below) (121).

The combination of DNA synthesis, lambda red recombination (a technique based on homologous recombination), and CRISPR-based editing, which allowed the generation of a full synthetic *E. coli* genome, can be used to explore other aspects of bacterial biology. The introduction in *Caulobacter ethensis* of 133 313 synonymous substitutions that recoded 123 562 codons resulted in the inactivation of 98 genes (124). Interestingly, 27 of the former genes contained non-coding control features embedded within protein-coding sequences. The functional basis of this gene inactivation is still poorly understood. Transcriptional and/or translational control signals at the termini of coding sequences, as well as imprecise genome annotations, may be the underlying cause of gene inactivation. The expression of several hundred synonymous mutants of green fluorescent protein (GFP) in *E. coli* revealed that many synonymous variants reduce bacterial fitness (18). The underly-

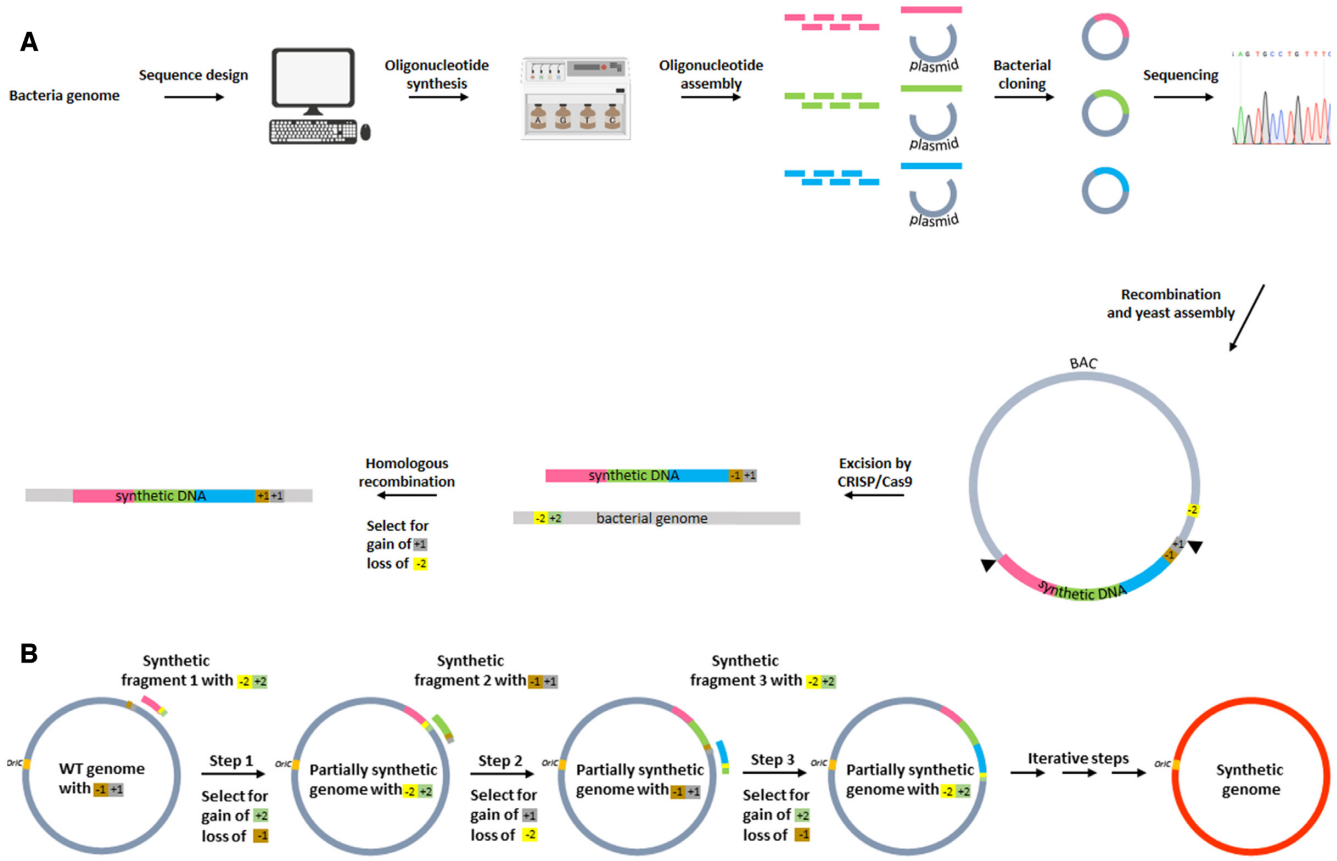

Figure 3. Synthetic DNA generation of a full-length bacterial genome. (**A**) Oligonucleotides (100–200 nucleotides long) are recombined by PCR, cloned in a bacterial plasmid, and sequenced. Synthetic DNA fragments from plasmids of ∼10 kb (pink, green, and blue) are assembled by homologous recombination in *Saccharomyces cerevisiae* using bacterial artificial chromosomes (BACs) to generate a length of ∼100 kb. BAC fragments contain Cas9 cleavage sites (black arrows) that allow them to be liberated and to replace genomic bacterial DNA by homologous recombination. The double-selection cassette (−1, +1) (e.g. *rpsL-kanR*) ensures the integration of the synthetic DNA, and the double-selection cassette (−2, +2) on the genome ensures the removal of the corresponding wild-type DNA. (**B**) Different BAC fragments have homologous overlapping sequences and corresponding selection cassettes for sequentially and iteratively replace the bacterial genome in the precise genome location.

ing mechanism is independent of translation but depends on the production of toxic mRNA molecules, suggesting that translation-independent RNA toxicity may be an unrecognized obstacle in bacterial gene expression.

Due to the shorter genomes of RNA viruses, synthetic genomics has been employed much more often to explore diverse aspects of virus biology. Similar to what has been found in *C. crescentus*, large-scale synonymous recoding has allowed the identification of diverse features encoded within virus genomes (Table 2). Large-scale synonymous recoding of viral genomes has uncovered new antiviral mechanisms within the innate immune response (125). Compared to the host cell preference, the human immunodeficiency virus type 1 (HIV-1) genome contains extremely high frequencies of A nucleotides (between 31.7% and 38%), low G/C content, and suboptimal codon usage, particularly within the *gag* (virus structural proteins) and *pol* (virus enzymes) sequences (126,127). This uncommon rare codon bias favors A/U in the third codon position. Schlafen (SLFN) genes contain a subgroup of interferon-stimulated early-response genes that are directly induced by pathogens via the interferon regulatory factor 3 pathway (54). Human SLFN11 has been described as a novel host restriction factor that suppresses HIV-1 protein synthesis in a codon usage-dependent manner (128). SLFN11 utilizes the unique viral codon bias towards A/U nucleotides. Consequently, wild-type (WT) HIV-1 is affected by SLFN11, but not recoded viruses containing HIV-1 structural *gag* sequences optimized for human cell expression (i.e. without A/U in the third position) (128). Using synonymous mutagenesis, CpG suppression was recently shown to be essential for HIV-1 (50). The effect of CpG dinucleotides is exerted post-transcriptionally and independent of translation efficiency. A zinc-finger antiviral protein (ZAP) has been identified that inhibits virion production by cells infected with CpG-enriched HIV-1. These results suggest that ZAP utilizes host CpG suppression to identify non-self RNA. Similarly, large-scale synonymous recoding of viral genomes allows the identification of viral sequence patterns, including CpG content, detected by innate immune pathways in different viral systems, such as influenza virus (48,129), echovirus (130), foot and mouth disease virus (131), and simian immunodeficiency virus (SIV) (77).

RNA virus populations are genetically very heterogeneous. Genetic variation is generated by the accumulation of mutations during replication and their rearrangement by

Table 2. Synonymous recoding as a new tool for exploring microbial biology

Explored features	Microbe	Targeted genome region	References
Reduced number of synonymous sense codons	*E. coli*	Full-genome	(65,121)
Specific mRNA structures required for replication	Poliovirus	P2 and P3	(138,139)
Virus genome cis-inhibitory signal sequences relevant to complex viral functions	Respiratory Syncytial Virus	NS1 and NS2	(153)
Non-coding control features embedded within protein-coding sequences	*C. ethensis*	Full-genome	(124)
Immune innate response	Influenza Virus	Segment 5	(48,50,77,128–130)
	HIV-1	Env and Gag	
	SIV	Gag and Pol	
	Echovirus 7	VP3, 3C, and 3D	
Codon usage and the temporal regulation of viral gene expression	SIV Herpesvirus	Env ORF57	(141)
Genome position in sequence space, evolutionary trajectory, and pathogenicity	Poliovirus	Capsid	(134–137)
	Coxsackie virus	Capsid	
	Influenza Virus	Polymerase and Haemagglutinin	
	HIV-1	Protease	

genetic recombination, and genome segment reassortment in the case of segmented genomes (132). Unlike DNA polymerases, RNA viruses lack proofreading activity, resulting in an estimated error rate during replication of 10^{-3} to 10^{-5} mutations per nucleotide per replication cycle. This high mutation rate generates a virus population composed of closely related mutant spectra or mutant clouds termed viral quasispecies (133). The quasispecies dynamics of RNA viruses are intimately related to both viral disease and antiviral treatment strategies. One unexplored aspect of the genetic architecture of RNA viruses is how codon choice influences population diversity and evolvability. In order to explore this aspect of RNA virus biology, the quasispecies structure of a poliovirus containing 934 synonymous substitutions randomly introduced in the 2643-nucleotide-long capsid coding region (72) was compared to the WT virus (134). As expected, the recoded virus explores a different sequence space and gives rise to a distinct spectrum of mutants. Interestingly, mice infected with the synonymously recoded poliovirus exhibit delayed and decreased mortality compared to mice inoculated with the same doses of WT virus, suggesting that a virus quasispecies structure and mutant spectrum impact viral pathogenesis. This attenuated phenotype exhibited by the recoded virus suggests a possible new strategy for the generation of live attenuated vaccines (34,135). Another strategy explored to attenuate coxsackie B3 and influenza A viruses has been to redirect their evolutionary trajectories towards detrimental regions of sequence space (136). Viral genomes were synonymously recoded to harbor more serine and leucine codons that could generate stop mutations after a single nucleotide substitution (136). These engineered viruses generate more stop codons and result in significant losses in viral fitness. More recently, synonymous virus genome recoding has been used to investigate how the HIV-1 position in sequence space defines its mutant spectrum. To this end, a synthetic virus was generated carrying a reengineered protease sequence including 13% of synonymous mutations (137). WT and recoded viruses were propagated in the presence of HIV-1 protease inhibitors (PIs). Ultradeep sequence clonal analysis revealed that both viruses harbor previously described mutations conferring resistance to PIs. However, the WT and recoded HIV-1 proteases had different resistance variant repertoires, indicating that the HIV-1 protease position in sequence space delineates the evolution of its mutant spectrum. Importantly, in both poliovirus and HIV-1, synonymously recoded viruses did not have a reduced replication capacity compared to the WT virus.

Adenoviruses are DNA viruses broadly employed as vectors for human gene therapy and vaccines. In particular, single-stranded adenovirus-associated virus (AAV) has the unique capacity for site-specific integration into a transcriptionally silent region of the human genome. However, AAV Rep 78 protein has an inhibitory effect on adenovirus replication. In order to explore the inhibitory mechanism of Rep78, synonymous recoding was performed on the 1866-nucleotide-long *Rep* genomic segment (76). The genetically recoded AAV Rep gene permitted the identification of a 135-bp long cis-acting element in the *Rep* sequence. Reengineering this cis-acting element overcame Rep's inhibitory effects on adenovirus replication. Reengineered *Rep* is readily tolerated, facilitating the development of improved adeno-based gene-therapy vectors. Large-scale codon optimization of two DNA viruses, adenovirus and papilloma virus, has also aided in our understanding of how these viruses increase their fitness, immunogenicity, and oncogenic potential (81,83). The utility of synonymous recoding to identify the locations of critical DNA or RNA sequence signals was also demonstrated with poliovirus. The introduction of 2256 synonymous substitutions into the 6 606-nucleotide-long non-structural protein-coding region of poliovirus allowed the discovery of two 75-nucleotide-long unique, functionally redundant RNA elements, which are separated by 150 nucleotides, in the 3′-terminal coding sequence of poliovirus RNA polymerase (138). Similarly, massively recoded poliovirus has allowed the identification of conserved synonymous sites and local secondary RNA that play a role in determining virus packaging (139).

Another example of the utility of large-scale synonymous recoding is the identification of how codon usage is implicated in the temporal regulation of gene expression. A number of viruses are able to maintain life-long infections in their natural host despite strong host immune-surveillance measures. Persisting viruses can be found in the lenti-retroviruses (e.g. HIV-1/SIV), herpes, polyoma, papilloma, parvo, and adenovirus families. Persisting viruses exquisitely regulate their gene expression in a time-based manner. In many cases, such as HIV-1, structural proteins are made late in the viral replication cycle, and expression of these late gene products is typically induced by viral transinducers made earlier in the viral replication cycle (140). Intriguingly, structural gene products of different members of human persisting viruses exhibit a marked bias of codon usage compared to the codon usage of the human exome (141). In HIV-1/SIV, the viral protein Rev acts as a transinducer. Rev protein interacts with a highly structured element in the envelope gene (*env*), referred to as the Rev-responsive element (RRE). In the absence of Rev, mRNAs of the HIV-1 late structural genes (e.g. Env protein) are retained in the nucleus, preventing their translation. *Env* has unusual codon usage that is different from that of the host cell. Synonymous recoding of *env* has demonstrated that Rev induction of Env protein expression depends on its biased codon usage. Rev inducibility was lost when 10.5% of the *env* codons were synonymously replaced (141). Analogous results were obtained when this approach was used with rhesus monkey rhadinovirus (RRV), a member of the herpesvirus family, ORF57 transinducer (141). ORF57 induces the late expression of RRV structural glycoprotein H (gH), but this inducibility was lost when 14.3% the gH codons were replaced with synonymous codons. Interestingly, the transinducibility of Env expression can be regulated by ORF57 instead of Rev by modifying *env* to make it better resemble the codon use of gH. Similarly, the dependence of gH on ORF57 expression can be changed to Rev dependence by changing the nature of its codon usage. These results suggest that persisting viruses use poor codon usage that is biased in a distinctive way to temporally regulate late expression of structural gene products. Coupling synthetic genomics and synonymous recoding is an interesting and effective strategy for modulating diverse biological systems, and a possible alternative to conventional mutational analysis involving a limited number of nucleotide substitutions.

SYNONYMOUS MUTATIONS AND MICROBE THERAPEUTICS

Synthetic genomics is pushing a new era of medicine via the genetic programming of living cells. A recent example is bacteria. An *E. coli* strain containing a synchronized lysis circuit that colonizes tumors and undergoes intratumoral quorum lysis was engineered to locally release an encoded nanobody antagonist of CD47, an anti-phagocytic receptor overexpressed in several human cancers (142). This construct allows for the combined local delivery of an immunotherapeutic agent along with immunestimulatory bacterial lysis adjuvants to stimulate antitumor immunity and promote tumor regression.

An obvious application of synonymous nucleotide recoding is codon optimization to increase protein expression and improve the production of therapeutic antibodies, cytokines, or fusion proteins (reviewed in (143)). However, in mammalian cells, codon optimization is not evident because, as noted previously in this review, not all synonymous codon mutations are neutral.

The approach of reducing expression by intentional introduction of suboptimal synonymous codons has been used extensively with viruses (Table 3). The first complete synthesis of an organism's genome was performed with poliovirus (103). Similarly, the first large-scale synonymous recoding of a genome was carried out with poliovirus. Simultaneously, Burns *et al.* and Mueller *et al.* demonstrated the viability of synonymously recoding the poliovirus genome (72,79). Using the Sabin type 2 oral poliovirus vaccine strain, Burns *et al*. reengineered the capsid coding region of poliovirus by replacing degenerate codons for nine amino acids with corresponding non-preferred synonymous codons. Up to 542 synonymous substitutions were introduced among the 2 555 nucleotides of the virus capsid region. Quantifying virus yields in HeLa cells and plaque areas (visible structures formed within a cell culture by virus cell destruction), these authors demonstrated that the replication capacity of the virus was reduced up to 98%. When recoded viruses were serially propagated in HeLa cells to test their stability, viral evolution was demonstrated by increased virus yields, increased plaque sizes, and mutated genomic sequences (79). Mueller *et al*. were also able to reduce the replication capacity of the Mahoney strain of poliovirus by introducing suboptimal synonymous codon changes in the virus capsid coding region (72). When this suboptimal recoded virus was inoculated in poliovirus CD155 transgenic mice, the variant presented a neuroattenuated phenotype. Attenuation of the classical oral poliovirus vaccine is based on very few point mutations, which can revert to virulence after a few rounds of viral replication (144). These pioneering results obtained with recoded polioviruses suggest that codon-usage in recoded viruses may be much more stable than most RNA virus point mutants, and could possibly enable the development of live attenuated RNA virus vaccines with superior genetic stability. Moreover, synonymous recoding does not affect the viral amino acid sequence and its subsequent antigenicity.

Large-scale synonymous codon usage recoding has also been used to obtain safe live attenuated vaccines for arboviruses, such as Chikungunya virus (74) and tick-borne encephalitis virus (85). An interesting approach to developing new arbovirus vaccines is to codon-optimize the virus to grow in insect cells and deoptimize it to reduce its replication capacity to non-viability in mammalian cells (75). Viruses that are non-viable in mammalian cells could still be grown in insect cells, implying that one might design a vaccine that combines features of live attenuated viruses with features of inactivated viruses. Analogously, work with SIV has shown that codon optimization of *gag* and *pol* results in a 100-fold decrease in virus replicative capacity (77). Re-

Table 3. Examples of microbe attenuation by synonymous genome recoding

Microbe	Targeted genome region	Method	References
Streptococcus pneumoniae	Pneumolysin	Codon-pair usage	(64)
Methylobacterium extorquens	Formaldehyde-activating enzyme	Codon bias usage	(62)
Salmonella enterica	Ribosomal protein S20, L-γ-glutamyl phosphate reductase	Codon bias usage	(57,58)
Escherichia coli	Lac-Z, green florescence protein, Beta-lactamase	Codon bias usage	(11,18,60,61)
Poliovirus	Capsid	Codon bias usage, Codon-pair usage and Dinucleotide frequency	(36,68,72,79)
Echovirus 7	VP1, VP2, 3C, 3D	Dinucleotide frequency	(37,86)
Influenza virus	Polymerase subunit PB1, PB2, PA, nucleoprotein, matrix, nonstructural protein, neuraminidase, hemagglutinin	Codon-pair usage	(71,80,151)
Human immunodeficiency virus type 1	Protease	Codon-pair usage	(69)
Simian immunodeficiency virus	Gag, Pol	Codon bias usage	(77)
Chikungunya virus	Non-structural proteins: nsP1, nsP4 Structural proteins: E2, E1	Codon bias usage	(74)
Dengue virus	E, NS3, NS5	Codon-pair usage	(75)
Zika virus	E, NS1	Codon-pair usage	(89)
Tick-borne encephalitis virus	NS5	Codon bias usage	(85)
Respiratory syncytial virus	NS1, NS2, N, P, M, SH, G, F, L	Codon-pair usage	(87)
Porcine reproductive and respiratory syndrome virus	Envelope GP5	Codon-pair usage	(73)
Vesicular stomatitis virus	L1	Codon-pair usage	(78)
Papillomavirus	Oncogenes E6, E7	Codon bias usage	(83)
Marek's disease virus	UL54/ICP27, UL49/VP22, UL30	Codon-pair usage	(82,84)
Adenovirus	Fiber	Codon bias usage	(81)
Cucumber mosaic virus	Coat protein	Codon bias usage	(70)

markably, the recoded virus has a reduced ability to stimulate type I interferon and possibly attenuated pathogenic potential.

Wimmer *et al.* demonstrated that genome recoding using underrepresented codon pairs can also result in poliovirus virus attenuation (36). The introduction of 407 synonymous substitutions into the poliovirus (Mahoney) capsid coding region, which generates underrepresented codon pairs, resulted in decreased rates of virus protein translation, reduced replication capacity in HeLa cells, and attenuation in mice (36). Similar results were obtained with another variant carrying 224 synonymous changes in the same region of the viral genome. These polioviruses were used to immunize mice and provided protective immunity after challenge. Codon-pair deoptimization has been performed in a number of viruses, including negative RNA viruses such as influenza virus (71, 80) and respiratory syncytial virus (RSV) (87), vesicular stomatitis virus (78), positive RNA viruses such as porcine reproductive and respiratory syndrome virus (73), dengue virus (75), zika virus (89), HIV-1 (69), the plant cucumber mosaic virus (70), and DNA viruses such as Marek's disease herpesvirus (82,84). Synonymous codon pair deoptimization has not been restricted to viruses and has been applied to other microorganisms (Table 3) (64).

Synonymously increasing the frequencies of CpG and UpA dinuclotides has also been explored as an alternative method of virus attenuation. Most mammalian RNA and small DNA viruses suppress genomic CpG and UpA dinucleotide frequencies, apparently mimicking host mRNA composition. Increasing the CpG and UpA dinucleotide frequencies attenuated poliovirus, influenza virus and echovirus 7 (another picornavirus like poliovirus) (37,48,68,86). As discussed previously, sometimes increased CpG and UpA dinucleotide frequencies and codon-pair deoptimization are not easily distinguished (43).

Recoding viral genomes by numerous synonymous but suboptimal substitutions can be a new source of live attenuated vaccine candidates. The hypothesis is that large-scale recoded viruses should be genetically stable because deattenuation may require dozens or hundreds of reversions (36). However, their genetic stability under selective pressure is largely unknown. Attempts to evaluate phenotype stability have resulted in interesting results. Codon pair deoptimized versions of RSV were generated and presented an attenuated and temperature-sensitive phenotype (88). When recoded viruses were serially propagated in tissue culture at progressively increasing temperatures, an RSV variant containing 2 692 synonymous mutations in 9 of 11 open reading frames (ORFs) did not lose temperature sensitivity, remained genetically stable, and was restricted at temperatures of 34/35°C and above. However, a variant containing 1 378 synonymous mutations solely in the polymerase L ORF quickly lost substantial attenuation. Sequence analysis of virus populations after tissue culture passages identified many different potentially deattenuating mutations in the L ORF, as well as many others appearing in other ORFs (88). These results emphasize, again, the rapid evolution of RNA viruses and the necessity of stability studies to obtain improved synthetic virus vaccine candidates.

The recent generation of a full synthetic *E. coli* genome (65) has opened up unexplored possibilities for designing a new generation of bacterial live attenuated vaccines. Proof of the concept has been shown with *Streptococcus pneu-*

moniae (64). Codon pair recoding of *S. pneumonia* pneumolysin gene has resulted in strains that are less virulent in mice than the WT. One particular aspect of the new synthetic *E. coli* genome is that it uses only 61 of the 64 available codons in its protein-coding sequences, replacing two serine codons and one stop codon with synonyms (65). These deleted codons can now be reassigned to recruit ncAAs beyond the usual 20 used by most living cells (145). The genome of *S. typhimurium* has also been found to have two fewer codons; 1 557 synonymous leucine replacements were performed to delete two leucine codons (102). Engendered S. typhimurium can be used as a cell-based technology in the human gut. Now ncAAs can take part in chemical reactions that natural proteins cannot perform. An expanded genetic code might engender next-generation biopharmaceuticals (146). An enlarged amino acid repertoire will allow proteins to have new or modified functions, providing ample opportunities for innovative or improved medicines. Current pipelines involve the areas of half-life extension, antibody-drug conjugates (ADCs), bi-specific antibodies, modular chimeric antigen receptor T cell therapy (CAR-T), vaccine development, and expanded applications for diagnosis and bio-compatible materials (146). These new biopharmaceuticals are being explored in cancer and inflammatory, autoimmune, and metabolic diseases (121).

Synonymous codon compression and ncAAs can also be employed to develop viral attenuated vaccines (121). The introduction of amber codons into the genomes of hepatitis D virus, HIV-1 and influenza A has allowed the generation of viral genomes that replicate only in cells that can decode the amber codon (147–149). These viruses can be replicated in cells that contain orthogonal aminoacyl-tRNA synthetase and orthogonal $tRNA_{CUA}$. However, they are attenuated in cells that cannot suppress the amber codon. With the introduction of several amber codons throughout the viral genome, the frequency of revertants can be greatly reduced (148). However, safety studies should be performed because a read-through of stop codons can occur in some cell types (150).

OUTLOOK FOR THE FUTURE

Synthetic genomics and synonymous genome-wide rewriting will decode essential genome functions and allow the design of improved microorganisms for industrial and health applications. Although significant advances have been made in this field, it is still necessary to elucidate the exact mechanistic foundations underlying the contributions of synonymous substitutions to phenotype. Unraveling the fundamental mechanistic impact of synonymous substitutions is essential to support the use of synonymous recoding in an array of biological and clinical contexts. Despite this review focusing on viruses and bacteria, the work can pioneer the study of the impact of synonymous substitution on human diseases. As exemplified by the development of new and massive DNA sequencing methodologies, cheap and accurate DNA synthesis will be necessary to achieve ambitious objectives. The synthesis of relatively small DNA fragments (<10 kb) is currently straightforward. However, new and easier methodologies should be developed to assemble larger pieces of synthetic DNA (e.g. DNA cloning and recombinant protocols). These new methodologies will be necessary if we want to synthetically engineer eukaryotic genomes. Without a doubt, synthetic genomics will greatly impact new diagnostics and therapies, as well as basic research.

ACKNOWLEDGEMENTS

We apologize to those whose works have not been cited in this article due to a lack of space.

FUNDING

Spanish Ministry of Science, Innovation and Universities [SAF2016–75277-R]; Spanish Ministry of Innovation and Universities [BES-2014-069931 to A.J.-P.]; Grifols Grant [miRNA 2018]; Spanish AIDS network (RIS) [RD16/0025/0041 to M.N.]. Funding for open access charge: Spanish Ministry of Science, Innovation and Universities [SAF2016-75277-R].

Conflict of interest statement. None declared.

REFERENCES

1. Grantham,R., Gautier,C. and Gouy,M. (1980) Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. *Nucleic Acids Res.*, **8**, 1893–1912.
2. Quax,T.E., Claassens,N.J., Soll,D. and van der Oost,J. (2015) Codon bias as a means to Fine-Tune gene expression. *Mol. Cell*, **59**, 149–161.
3. Ikemura,T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.*, **2**, 13–34.
4. Brule,C.E. and Grayhack,E.J. (2017) Synonymous Codons: Choose wisely for expression. *Trends Genet.*, **33**, 283–297.
5. Zhou,Z., Dang,Y., Zhou,M., Li,L., Yu,C.H., Fu,J., Chen,S. and Liu,Y. (2016) Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E6117–E6125.
6. Findlay,G.M., Daza,R.M., Martin,B., Zhang,M.D., Leith,A.P., Gasperini,M., Janizek,J.D., Huang,X., Starita,L.M. and Shendure,J. (2018) Accurate classification of BRCA1 variants with saturation genome editing. *Nature*, **562**, 217–222.
7. Goodman,D.B., Church,G.M. and Kosuri,S. (2013) Causes and effects of N-terminal codon bias in bacterial genes. *Science*, **342**, 475–479.
8. Kudla,G., Murray,A.W., Tollervey,D. and Plotkin,J.B. (2009) Coding-sequence determinants of gene expression in Escherichia coli. *Science*, **324**, 255–258.
9. Stergachis,A.B., Haugen,E., Shafer,A., Fu,W., Vernot,B., Reynolds,A., Raubitschek,A., Ziegler,S., LeProust,E.M., Akey,J.M. *et al.* (2013) Exonic transcription factor binding directs codon choice and affects protein evolution. *Science*, **342**, 1367–1372.
10. Boel,G., Letso,R., Neely,H., Price,W.N., Wong,K.H., Su,M., Luff,J., Valecha,M., Everett,J.K., Acton,T.B. *et al.* (2016) Codon influence on protein expression in E. coli correlates with mRNA levels. *Nature*, **529**, 358–363.
11. Sorensen,M.A., Kurland,C.G. and Pedersen,S. (1989) Codon usage determines translation rate in Escherichia coli. *J. Mol. Biol.*, **207**, 365–377.
12. Akashi,H. (1994) Synonymous codon usage in Drosophila melanogaster: natural selection and translational accuracy. *Genetics*, **136**, 927–935.
13. Drummond,D.A. and Wilke,C.O. (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell*, **134**, 341–352.
14. Presnyak,V., Alhusaini,N., Chen,Y.H., Martin,S., Morris,N., Kline,N., Olson,S., Weinberg,D., Baker,K.E., Graveley,B.R. *et al.* (2015) Codon optimality is a major determinant of mRNA stability. *Cell*, **160**, 1111–1124.

15. Shabalina,S.A., Ogurtsov,A.Y. and Spiridonov,N.A. (2006) A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Res.*, **34**, 2428–2437.
16. Pagani,F., Raponi,M. and Baralle,F.E. (2005) Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 6368–6372.
17. Takata,M.A., Soll,S.J., Emery,A., Blanco-Melo,D., Swanstrom,R. and Bieniasz,P.D. (2018) Global synonymous mutagenesis identifies cis-acting RNA elements that regulate HIV-1 splicing and replication. *PLoS Pathog.*, **14**, e1006824.
18. Mittal,P., Brindle,J., Stephen,J., Plotkin,J.B. and Kudla,G. (2018) Codon usage influences fitness through RNA toxicity. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, 8639–8644.
19. Warnecke,T., Weber,C.C. and Hurst,L.D. (2009) Why there is more to protein evolution than protein function: splicing, nucleosomes and dual-coding sequence. *Biochem. Soc. Trans.*, **37**, 756–761.
20. Birnbaum,R.Y., Clowney,E.J., Agamy,O., Kim,M.J., Zhao,J., Yamanaka,T., Pappalardo,Z., Clarke,S.L., Wenger,A.M., Nguyen,L. *et al.* (2012) Coding exons function as tissue-specific enhancers of nearby genes. *Genome Res.*, **22**, 1059–1068.
21. Lin,M.F., Kheradpour,P., Washietl,S., Parker,B.J., Pedersen,J.S. and Kellis,M. (2011) Locating protein-coding sequences under selection for additional, overlapping functions in 29 mammalian genomes. *Genome Res.*, **21**, 1916–1928.
22. Brest,P., Lapaquette,P., Souidi,M., Lebrigand,K., Cesaro,A., Vouret-Craviari,V., Mari,B., Barbry,P., Mosnier,J.F., Hebuterne,X. *et al.* (2011) A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. *Nat. Genet.*, **43**, 242–245.
23. Pechmann,S. and Frydman,J. (2013) Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat. Struct. Mol. Biol.*, **20**, 237–243.
24. Plotkin,J.B. and Kudla,G. (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.*, **12**, 32–42.
25. Chamary,J. V, Parmley,J.L. and Hurst,L.D. (2006) Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat. Rev. Genet.*, **7**, 98–108.
26. Chen,S.L., Lee,W., Hottes,A.K., Shapiro,L. and McAdams,H.H. (2004) Codon usage between genomes is constrained by genome-wide mutational processes. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 3480–3485.
27. Plotkin,J.B., Robins,H. and Levine,A.J. (2004) Tissue-specific codon usage and the expression of human genes. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 12588–12591.
28. Hunt,R.C., Simhadri,V.L., Iandoli,M., Sauna,Z.E. and Kimchi-Sarfaty,C. (2014) Exposing synonymous mutations. *Trends Genet.*, **30**, 308–321.
29. Moura,G.R., Pinheiro,M., Freitas,A., Oliveira,J.L., Frommlet,J.C., Carreto,L., Soares,A.R., Bezerra,A.R. and Santos,M.A. (2011) Species-specific codon context rules unveil non-neutrality effects of synonymous mutations. *PLoS One*, **6**, e26817.
30. Tats,A., Tenson,T. and Remm,M. (2008) Preferred and avoided codon pairs in three domains of life. *BMC Genomics*, **9**, 463.
31. Alexaki,A., Kames,J., Holcomb,D.D., Athey,J., Santana-Quintero,L. V, Lam,P.V.N., Hamasaki-Katagiri,N., Osipova,E., Simonyan,V., Bar,H. *et al.* (2019) Codon and Codon-Pair Usage Tables (CoCoPUTs): facilitating genetic variation analyses and recombinant gene design. *J. Mol. Biol.*, **431**, 2434–2441.
32. Biasoli,D., Compston-Garnett,L., Ricketts,S.L., Birand,Z., Courtay-Cahen,C., Fineberg,E., Arendt,M., Boerkamp,K., Melin,M., Koltookian,M. *et al.* (2019) A synonymous germline variant in a gene encoding a cell adhesion molecule is associated with cutaneous mast cell tumour development in Labrador and Golden Retrievers. *PLos Genet.*, **15**, e1007967.
33. Zhou,J., Park,C.Y., Theesfeld,C.L., Wong,A.K., Yuan,Y., Scheckel,C., Fak,J.J., Funk,J., Yao,K., Tajima,Y. *et al.* (2019) Whole-genome deep-learning analysis identifies contribution of noncoding mutations to autism risk. *Nat. Genet.*, **51**, 973–980.
34. Martinez,M.A., Jordan-Paiz,A., Franco,S. and Nevot,M. (2016) Synonymous virus genome recoding as a tool to impact viral fitness. *Trends Microbiol.*, **24**, 134–147.
35. Gutman,G.A. and Hatfield,G.W. (1989) Nonrandom utilization of codon pairs in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 3699–3703.
36. Coleman,J.R., Papamichail,D., Skiena,S., Futcher,B., Wimmer,E. and Mueller,S. (2008) Virus attenuation by genome-scale changes in codon pair bias. *Science*, **320**, 1784–1787.
37. Tulloch,F., Atkinson,N.J., Evans,D.J., Ryan,M.D. and Simmonds,P. (2014) RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. *Elife*, **3**, e04531.
38. Kunec,D. and Osterrieder,N. (2016) Codon pair bias is a direct consequence of dinucleotide bias. *Cell Rep.*, **14**, 55–67.
39. Simmonds,P., Xia,W., Baillie,J. and McKinnon,K. (2013) Modelling mutational and selection pressures on dinucleotides in eukaryotic phyla –selection against CpG and UpA in cytoplasmically expressed RNA and in RNA viruses. *BMC Genomics*, **14**, 610.
40. Hodgkinson,A. and Eyre-Walker,A. (2011) Variation in the mutation rate across mammalian genomes. *Nat. Rev. Genet.*, **12**, 756–766.
41. Jones,P.A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.*, **13**, 484–492.
42. Takahashi,Y., Wu,J., Suzuki,K., Martinez-Redondo,P., Li,M., Liao,H.K., Wu,M.Z., Hernandez-Benitez,R., Hishida,T., Shokhirev,M.N. *et al.* (2017) Integration of CpG-free DNA induces de novo methylation of CpG islands in pluripotent stem cells. *Science*, **356**, 503–508.
43. Futcher,B., Gorbatsevych,O., Shen,S.H., Stauft,C.B., Song,Y., Wang,B., Leatherwood,J., Gardin,J., Yurovsky,A., Mueller,S. *et al.* (2015) Reply to Simmonds et al.: Codon pair and dinucleotide bias have not been functionally distinguished. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E3635–E3636.
44. Simmonds,P., Tulloch,F., Evans,D.J. and Ryan,M.D. (2015) Attenuation of dengue (and other RNA viruses) with codon pair recoding can be explained by increased CpG/UpA dinucleotide frequencies. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E3633–E3634.
45. Rima,B.K. and McFerran,N. V (1997) Dinucleotide and stop codon frequencies in single-stranded RNA viruses. *J. Gen. Virol.*, **78**, 2859–2870.
46. van der Kuyl,A.C. and Berkhout,B. (2012) The biased nucleotide composition of the HIV genome: a constant factor in a highly variable virus. *Retrovirology*, **9**, 92.
47. Kumagai,Y., Takeuchi,O. and Akira,S. (2008) TLR9 as a key receptor for the recognition of DNA. *Adv. Drug. Deliv. Rev.*, **60**, 795–804.
48. Gaunt,E., Wise,H.M., Zhang,H., Lee,L.N., Atkinson,N.J., Nicol,M.Q., Highton,A.J., Klenerman,P., Beard,P.M., Dutia,B.M. *et al.* (2016) Elevation of CpG frequencies in influenza A genome attenuates pathogenicity but enhances host response to infection. *Elife*, **5**, e12735.
49. Zhang,L., Li,H., Hai,Y., Yin,W., Li,W., Zheng,B., Du,X., Li,N., Zhang,Z., Deng,Y. *et al.* (2017) CpG in combination with an inhibitor of notch signaling suppresses Formalin-Inactivated respiratory syncytial Virus-Enhanced airway hyperresponsiveness and inflammation by inhibiting Th17 memory responses and promoting Tissue-Resident memory cells in lung. *J. Virol.*, **91**, e02111-16.
50. Takata,M.A., Goncalves-Carneiro,D., Zang,T.M., Soll,S.J., York,A., Blanco-Melo,D. and Bieniasz,P.D. (2017) CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature*, **550**, 124–127.
51. Curran,J.F., Poole,E.S., Tate,W.P. and Gross,B.L. (1995) Selection of aminoacyl-tRNAs at sense codons: the size of the tRNA variable loop determines whether the immediate 3′ nucleotide to the codon has a context effect. *Nucleic Acids Res.*, **23**, 4104–4108.
52. Buchan,J.R., Aucott,L.S. and Stansfield,I. (2006) tRNA properties help shape codon pair preferences in open reading frames. *Nucleic Acids Res.*, **34**, 1015–1027.
53. Gamble,C.E., Brule,C.E., Dean,K.M., Fields,S. and Grayhack,E.J. (2016) Adjacent codons act in concert to modulate translation efficiency in yeast. *Cell*, **166**, 679–690.
54. Chevance,F.F., Le Guyon,S. and Hughes,K.T. (2014) The effects of codon context on in vivo translation speed. *PLos Genet.*, **10**, e1004392.

55. Irwin,B., Heck,J.D. and Hatfield,G.W. (1995) Codon pair utilization biases influence translational elongation step times. *J. Biol. Chem.*, **270**, 22801–22806.
56. Chevance,F.F.V. and Hughes,K.T. (2017) Case for the genetic code as a triplet of triplets. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 4745–4750.
57. Kristofich,J., Morgenthaler,A.B., Kinney,W.R., Ebmeier,C.C., Snyder,D.J., Old,W.M., Cooper,V.S. and Copley,S.D. (2018) Synonymous mutations make dramatic contributions to fitness when growth is limited by a weak-link enzyme. *PLos Genet.*, **14**, e1007615.
58. Knoppel,A., Nasvall,J. and Andersson,D.I. (2016) Compensating the fitness costs of synonymous mutations. *Mol. Biol. Evol.*, **33**, 1461–1477.
59. Raghavan,R., Kelkar,Y.D. and Ochman,H. (2012) A selective force favoring increased G+C content in bacterial genes. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 14504–14507.
60. Ballard,A., Bieniek,S. and Carlini,D.B. (2019) The fitness consequences of synonymous mutations in Escherichia coli: Experimental evidence for a pleiotropic effect of translational selection. *Gene*, **694**, 111–120.
61. Yannai,A., Katz,S. and Hershberg,R. (2018) The codon usage of lowly expressed genes is subject to natural selection. *Genome Biol. Evol.*, **10**, 1237–1246.
62. Agashe,D., Martinez-Gomez,N.C., Drummond,D.A. and Marx,C.J. (2013) Good codons, bad Transcript: Large reductions in gene expression and fitness arising from synonymous mutations in a key enzyme. *Mol. Biol. Evol.*, **30**, 549–560.
63. Brandis,G. and Hughes,D. (2016) The Selective Advantage of Synonymous Codon Usage Bias in Salmonella. *PLos Genet.*, **12**, e1005926.
64. Coleman,J.R., Papamichail,D., Yano,M., Garcia-Suarez Mdel,M. and Pirofski,L.A. (2011) Designed reduction of Streptococcus pneumoniae pathogenicity via synthetic changes in virulence factor codon-pair bias. *J. Infect. Dis.*, **203**, 1264–1273.
65. Fredens,J., Wang,K., de la Torre,D., Funke,L.F.H., Robertson,W.E., Christova,Y., Chia,T., Schmied,W.H., Dunkelmann,D.L., Beránek,V. *et al.* (2019) Total synthesis of Escherichia coli with a recoded genome. *Nature*, **569**, 514–518.
66. Frumkin,I., Schirman,D., Rotman,A., Li,F., Zahavi,L., Mordret,E., Asraf,O., Wu,S., Levy,S.F. and Pilpel,Y. (2017) Gene architectures that minimize cost of gene expression. *Mol. Cell*, **65**, 142–153.
67. Kelsic,E.D., Chung,H., Cohen,N., Park,J., Wang,H.H. and Kishony,R. (2016) RNA structural determinants of optimal codons revealed by MAGE-Seq. *Cell Syst.*, **3**, 563–571.
68. Burns,C.C., Campagnoli,R., Shaw,J., Vincent,A., Jorba,J. and Kew,O. (2009) Genetic inactivation of poliovirus infectivity by increasing the frequencies of CpG and UpA dinucleotides within and across synonymous capsid region codons. *J. Virol.*, **83**, 9957–9969.
69. Martrus,G., Nevot,M., Andres,C., Clotet,B. and Martinez,M. (2013) Changes in codon-pair bias of human immunodeficiency virus type 1 have profound effects on virus replication in cell culture. *Retrovirology*, **10**, 78.
70. Mochizuki,T., Ohara,R. and Roossinck,M.J. (2018) Large-Scale synonymous substitutions in cucumber mosaic virus RNA 3 facilitate amino acid mutations in the coat protein. *J. Virol.*, **92**, e01007-18.
71. Mueller,S., Coleman,J.R., Papamichail,D., Ward,C.B., Nimnual,A., Futcher,B., Skiena,S. and Wimmer,E. (2010) Live attenuated influenza virus vaccines by computer-aided rational design. *Nat. Biotechnol.*, **28**, 723–726.
72. Mueller,S., Papamichail,D., Coleman,J.R., Skiena,S. and Wimmer,E. (2006) Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. *J. Virol.*, **80**, 9687–9696.
73. Ni,Y.-Y., Zhao,Z., Opriessnig,T., Subramaniam,S., Zhou,L., Cao,D., Cao,Q., Yang,H. and Meng,X.-J. (2014) Computer-aided codon-pairs deoptimization of the major envelope GP5 gene attenuates porcine reproductive and respiratory syndrome virus. *Virology*, **450–451**, 132–139.
74. Nougairede,A., De Fabritus,L., Aubry,F., Gould,E.A., Holmes,E.C. and de Lamballerie,X. (2013) Random codon re-encoding induces stable reduction of replicative fitness of Chikungunya virus in primate and mosquito cells. *PLoS Pathog.*, **9**, e1003172.
75. Shen,S.H., Stauft,C.B., Gorbatsevych,O., Song,Y., Ward,C.B., Yurovsky,A., Mueller,S., Futcher,B. and Wimmer,E. (2015) Large-scale recoding of an arbovirus genome to rebalance its insect versus mammalian preference. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 4749–4754.
76. Sitaraman,V., Ward,C. and Skiena,S. (2012) Synthetic sequence design for signal location search. *Lect. Notes Comput. Sci.*, **7262**, 165–179.
77. Vabret,N., Bailly-Bechet,M., Lepelley,A., Najburg,V., Schwartz,O., Verrier,B. and Tangy,F. (2014) Large-scale nucleotide optimization of simian immunodeficiency virus reduces its capacity to stimulate type I interferon in vitro. *J. Virol.*, **88**, 4161–4172.
78. Wang,B., Yang,C., Tekes,G., Mueller,S., Paul,A., Whelan,S.P.J. and Wimmer,E. (2015) Recoding of the vesicular stomatitis virus L gene by Computer-Aided design provides a live, attenuated vaccine candidate. *MBio*, **6**, e00237-15.
79. Burns,C.C., Shaw,J., Campagnoli,R., Jorba,J., Vincent,A., Quay,J. and Kew,O. (2006) Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. *J. Virol.*, **80**, 3259–3272.
80. Yang,C., Skiena,S., Futcher,B., Mueller,S. and Wimmer,E. (2013) Deliberate reduction of hemagglutinin and neuraminidase expression of influenza virus leads to an ultraprotective live vaccine in mice. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 9481–9486.
81. Villanueva,E., Martí-Solano,M. and Fillat,C. (2016) Codon optimization of the adenoviral fiber negatively impacts structural protein expression and viral fitness. *Sci. Rep.*, **6**, 27546.
82. Eschke,K., Trimpert,J., Osterrieder,N. and Kunec,D. (2018) Attenuation of a very virulent Marek's disease herpesvirus (MDV) by codon pair bias deoptimization. *PLoS Pathog.*, **14**, e1006857.
83. Cladel,N.M., Budgeon,L.R., Hu,J., Balogh,K.K. and Christensen,N.D. (2013) Synonymous codon changes in the oncogenes of the cottontail rabbit papillomavirus lead to increased oncogenicity and immunogenicity of the virus. *Virology*, **438**, 70–83.
84. Conrad,S.J., Silva,R.F., Hearn,C.J., Climans,M. and Dunn,J.R. (2018) Attenuation of Marek's disease virus by codon pair deoptimization of a core gene. *Virology*, **516**, 219–226.
85. de Fabritus,L., Nougairede,A., Aubry,F., Gould,E.A. and de Lamballerie,X. (2015) Attenuation of tick-borne encephalitis virus using large-scale random codon re-encoding. *PLoS Pathog.*, **11**, e1004738.
86. Fros,J.J., Dietrich,I., Alshaikhahmed,K., Passchier,T.C., Evans,D.J. and Simmonds,P. (2017) CpG and UpA dinucleotides in both coding and non-coding regions of echovirus 7 inhibit replication initiation post-entry. *Elife*, **6**, e29112.
87. Le Nouen,C., Brock,L.G., Luongo,C., McCarty,T., Yang,L., Mehedi,M., Wimmer,E., Mueller,S., Collins,P.L., Buchholz,U.J. *et al.* (2014) Attenuation of human respiratory syncytial virus by genome-scale codon-pair deoptimization. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 13169–13174.
88. Le Nouen,C., McCarty,T., Brown,M., Smith,M.L., Lleras,R., Dolan,M.A., Mehedi,M., Yang,L., Luongo,C., Liang,B. *et al.* (2017) Genetic stability of genome-scale deoptimized RNA virus vaccine candidates under selective pressure. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E386–E395.
89. Li,P., Ke,X., Wang,T., Tan,Z., Luo,D., Miao,Y., Sun,J., Zhang,Y., Liu,Y., Hu,Q. *et al.* (2018) Zika virus attenuation by codon pair deoptimization induces sterilizing immunity in mouse models. *J. Virol.*, **92**, e00701-18.
90. Hughes,R.A. and Ellington,A.D. (2017) Synthetic DNA synthesis and assembly: putting the synthetic in synthetic biology. *Cold Spring Harb. Perspect. Biol.*, **9**, a023812.
91. Schindler,D., Dai,J. and Cai,Y. (2018) Synthetic genomics: a new venture to dissect genome fundamentals and engineer new functions. *Curr. Opin. Chem. Biol.*, **46**, 56–62.
92. Cameron,D.E., Bashor,C.J. and Collins,J.J. (2014) A brief history of synthetic biology. *Nat. Rev. Microbiol.*, **12**, 381–390.
93. Brenner,S., Johnson,M., Bridgham,J., Golda,G., Lloyd,D.H., Johnson,D., Luo,S., McCurdy,S., Foy,M., Ewan,M. *et al.* (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.*, **18**, 630–634.
94. Ronaghi,M., Karamohamed,S., Pettersson,B., Uhlen,M. and Nyren,P. (1996) Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.*, **242**, 84–89.

95. Schatz,M.C. and Phillippy,A.M. (2012) The rise of a digital immune system. *Gigascience*, **1**, 4.
96. Kosuri,S. and Church,G.M. (2014) Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods*, **11**, 499–507.
97. Wan,W., Li,L., Xu,Q., Wang,Z., Yao,Y., Wang,R., Zhang,J., Liu,H., Gao,X. and Hong,J. (2014) Error removal in microchip-synthesized DNA using immobilized MutS. *Nucleic Acids Res.*, **42**, e102.
98. Gibson,D.G., Glass,J.I., Lartigue,C., Noskov,V.N., Chuang,R.Y., Algire,M.A., Benders,G.A., Montague,M.G., Ma,L., Moodie,M.M. *et al.* (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, **329**, 52–56.
99. Gibson,D.G., Young,L., Chuang,R.Y., Venter,J.C., Hutchison,C.A. and Smith,H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*, **6**, 343–345.
100. Gibson,D.G., Smith,H.O., Hutchison,C.A. 3rd, Venter,J.C. and Merryman,C. (2010) Chemical synthesis of the mouse mitochondrial genome. *Nat. Methods*, **7**, 901–903.
101. Torella,J.P., Lienert,F., Boehm,C.R., Chen,J.H., Way,J.C. and Silver,P.A. (2014) Unique nucleotide sequence-guided assembly of repetitive DNA parts for synthetic biology applications. *Nat. Protoc.*, **9**, 2075–2089.
102. Lau,Y.H., Stirling,F., Kuo,J., Karrenbelt,M.A.P., Chan,Y.A., Riesselman,A., Horton,C.A., Schafer,E., Lips,D., Weinstock,M.T. *et al.* (2017) Large-scale recoding of a bacterial genome by iterative recombineering of synthetic DNA. *Nucleic Acids Res.*, **45**, 6971–6980.
103. Cello,J., Paul,A.V. and Wimmer,E. (2002) Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science*, **297**, 1016–1018.
104. Smith,H.O., Hutchison,C.A. 3rd, Pfannkoch,C. and Venter,J.C. (2003) Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 15440–15445.
105. Koblentz,G.D. (2017) The de novo synthesis of horsepox Virus: Implications for biosecurity and recommendations for preventing the reemergence of smallpox. *Heal. Secur*, **15**, 620–628.
106. Yao,X.D. and Evans,D.H. (2003) High-frequency genetic recombination and reactivation of orthopoxviruses from DNA fragments transfected into leporipoxvirus-infected cells. *J. Virol.*, **77**, 7281–7290.
107. Hutchison,C.A., Chuang,R.Y., Noskov,V.N., Assad-Garcia,N., Deerinck,T.J., Ellisman,M.H., Gill,J., Kannan,K., Karas,B.J., Ma,L. *et al.* (2016) Design and synthesis of a minimal bacterial genome. *Science*, **351**, aad6253.
108. Posfai,G., Plunkett,G., Feher,T., Frisch,D., Keil,G.M., Umenhoffer,K., Kolisnychenko,V., Stahl,B., Sharma,S.S., de Arruda,M. *et al.* (2006) Emergent properties of reduced-genome Escherichia coli. *Science*, **312**, 1044–1046.
109. Wang,K., Fredens,J., Brunner,S.F., Kim,S.H., Chia,T. and Chin,J.W. (2016) Defining synonymous codon compression schemes by genome recoding. *Nature*, **539**, 59–64.
110. Annaluru,N., Muller,H., Mitchell,L.A., Ramalingam,S., Stracquadanio,G., Richardson,S.M., Dymond,J.S., Kuang,Z., Scheifele,L.Z., Cooper,E.M. *et al.* (2014) Total synthesis of a functional designer eukaryotic chromosome. *Science*, **344**, 55–58.
111. Mitchell,L.A., Wang,A., Stracquadanio,G., Kuang,Z., Wang,X., Yang,K., Richardson,S., Martin,J.A., Zhao,Y., Walker,R. *et al.* (2017) Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science*, **355**, eaaf4831.
112. Richardson,S.M., Mitchell,L.A., Stracquadanio,G., Yang,K., Dymond,J.S., DiCarlo,J.E., Lee,D., Huang,C.L., Chandrasegaran,S., Cai,Y. *et al.* (2017) Design of a synthetic yeast genome. *Science*, **355**, 1040–1044.
113. Shen,Y., Wang,Y., Chen,T., Gao,F., Gong,J., Abramczyk,D., Walker,R., Zhao,H., Chen,S., Liu,W. *et al.* (2017) Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science*, **355**, eaaf4791.
114. Wu,Y., Li,B.-Z., Zhao,M., Mitchell,L.A., Xie,Z.-X., Lin,Q.-H., Wang,X., Xiao,W.-H., Wang,Y., Zhou,X. *et al.* (2017) Bug mapping and fitness testing of chemically synthesized chromosome X. *Science*, **355**, eaaf4706.
115. Xie,Z.-X., Li,B.-Z., Mitchell,L.A., Wu,Y., Qi,X., Jin,Z., Jia,B., Wang,X., Zeng,B.-X., Liu,H.-M. *et al.* (2017) "Perfect" designer chromosome V and behavior of a ring derivative. *Science*, **355**, eaaf4704.
116. Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
117. Venter,J.C., Adams,M.D., Myers,E.W., Li,P.W., Mural,R.J., Sutton,G.G., Smith,H.O., Yandell,M., Evans,C.A., Holt,R.A. *et al.* (2001) The sequence of the human genome. *Science*, **291**, 1304–1351.
118. Boeke,J.D., Church,G., Hessel,A., Kelley,N.J., Arkin,A., Cai,Y., Carlson,R., Chakravarti,A., Cornish,V.W., Holt,L. *et al.* (2016) GENOME ENGINEERING. The Genome Project-Write. *Science*, **353**, 126–127.
119. Ostrov,N., Landon,M., Guell,M., Kuznetsov,G., Teramoto,J., Cervantes,N., Zhou,M., Singh,K., Napolitano,M.G., Moosburner,M. *et al.* (2016) Design, synthesis, and testing toward a 57-codon genome. *Science*, **353**, 819–822.
120. Mukai,T., Lajoie,M.J., Englert,M. and Söll,D. (2017) Rewriting the Genetic Code. *Annu. Rev. Microbiol.*, **71**, 557–577.
121. Chin,J.W. (2017) Expanding and reprogramming the genetic code. *Nature*, **550**, 53–60.
122. Chin,J.W. (2014) Expanding and reprogramming the genetic code of cells and animals. *Annu. Rev. Biochem.*, **83**, 379–408.
123. Liu,C.C. and Schultz,P.G. (2010) Adding new chemistries to the genetic code. *Annu. Rev. Biochem.*, **79**, 413–444.
124. Venetz,J.E., Del Medico,L., Wolfle,A., Schachle,P., Bucher,Y., Appert,D., Tschan,F., Flores-Tinoco,C.E., van Kooten,M., Guennoun,R. *et al.* (2019) Chemical synthesis rewriting of a bacterial genome to achieve design flexibility and biological functionality. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 8070–8079.
125. Vabret,N., Bhardwaj,N. and Greenbaum,B.D. (2017) Sequence-Specific sensing of Nucleic Acids. *Trends Immunol.*, **38**, 53–65.
126. Kypr,J. and Mrazek,J. (1987) Unusual codon usage of HIV. *Nature*, **327**, 20.
127. Meintjes,P.L. and Rodrigo,A.G. (2005) Evolution of relative synonymous codon usage in Human Immunodeficiency Virus type-1. *J. Bioinform. Comput. Biol.*, **3**, 157–168.
128. Li,M., Kao,E., Gao,X., Sandig,H., Limmer,K., Pavon-Eternod,M., Jones,T.E., Landry,S., Pan,T., Weitzman,M.D. *et al.* (2012) Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. *Nature*, **491**, 125–128.
129. Jimenez-Baranda,S., Greenbaum,B., Manches,O., Handler,J., Rabadan,R., Levine,A. and Bhardwaj,N. (2011) Oligonucleotide motifs that disappear during the evolution of influenza virus in humans increase alpha interferon secretion by plasmacytoid dendritic cells. *J. Virol.*, **85**, 3893–3904.
130. Atkinson,N.J., Witteveldt,J., Evans,D.J. and Simmonds,P. (2014) The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. *Nucleic Acids Res.*, **42**, 4527–4545.
131. Diaz-San Segundo,F., Medina,G.N., Ramirez-Medina,E., Velazquez-Salinas,L., Koster,M., Grubman,M.J. and de los Santos,T. (2016) Synonymous deoptimization of foot-and-mouth disease virus causes attenuation in vivo while inducing a strong neutralizing antibody response. *J. Virol.*, **90**, 1298–1310.
132. Andino,R. and Domingo,E. (2015) Viral quasispecies. *Virology*, **479–480**, 46–51.
133. Mas,A., Lopez-Galindez,C., Cacho,I., Gomez,J. and Martinez,M.A. (2010) Unfinished stories on viral quasispecies and Darwinian views of evolution. *J. Mol. Biol.*, **397**, 865–877.
134. Lauring,A.S., Acevedo,A., Cooper,S.B. and Andino,R. (2012) Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus. *Cell Host Microbe.*, **12**, 623–632.
135. Lauring,A.S., Frydman,J. and Andino,R. (2013) The role of mutational robustness in RNA virus evolution. *Nat. Rev. Microbiol.*, **11**, 327–336.
136. Moratorio,G., Henningsson,R., Barbezange,C., Carrau,L., Bordería,A. V., Blanc,H., Beaucourt,S., Poirier,E.Z., Vallet,T., Boussier,J. *et al.* (2017) Attenuation of RNA viruses by redirecting their evolution in sequence space. *Nat. Microbiol.*, **2**, 17088.

137. Nevot,M., Jordan-Paiz,A., Martrus,G., Andrés,C., García-Cehic,D., Gregori,J., Franco,S., Quer,J. and Martinez,M.A. (2018) HIV-1 protease evolvability is affected by synonymous nucleotide recoding. *J. Virol.*, **92**, e00777-18.
138. Song,Y., Liu,Y., Ward,C.B., Mueller,S., Futcher,B., Skiena,S., Paul,A.V. and Wimmer,E. (2012) Identification of two functionally redundant RNA elements in the coding sequence of poliovirus using computer-generated design. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 14301–14307.
139. Song,Y., Gorbatsevych,O., Liu,Y., Mugavero,J., Shen,S.H., Ward,C.B., Asare,E., Jiang,P., Paul,A. V, Mueller,S. *et al.* (2017) Limits of variation, specific infectivity, and genome packaging of massively recoded poliovirus genomes. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E8731–E8740.
140. Malim,M.H., Hauber,J., Le,S.Y., Maizel,J.V. and Cullen,B.R. (1989) The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature*, **338**, 254–257.
141. Shin,Y.C., Bischof,G.F., Lauer,W.A. and Desrosiers,R.C. (2015) Importance of codon usage for the temporal regulation of viral gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 14030–14035.
142. Chowdhury,S., Castro,S., Coker,C., Hinchliffe,T.E., Arpaia,N. and Danino,T. (2019) Programmable bacteria induce durable tumor regression and systemic antitumor immunity. *Nat. Med.*, **25**, 1057–1063.
143. Mauro,V.P. and Chappell,S.A. (2014) A critical analysis of codon optimization in human therapeutics. *Trends Mol. Med.*, **20**, 604–613.
144. Minor,P.D. (1993) Attenuation and reversion of the Sabin vaccine strains of poliovirus. *Dev. Biol. Stand.*, **78**, 17–26.
145. Blount,B.A. and Ellis,T. (2019) Construction of an Escherichia coli genome with fewer codons sets records. *Nature*, **569**, 492–494.
146. Kang,M., Lu,Y., Chen,S. and Tian,F. (2018) Harnessing the power of an expanded genetic code toward next-generation biopharmaceuticals. *Curr. Opin. Chem. Biol.*, **46**, 123–129.
147. Lin,S., Yan,H., Li,L., Yang,M., Peng,B., Chen,S., Li,W. and Chen,P.R. (2013) Site-specific engineering of chemical functionalities on the surface of live hepatitis D virus. *Angew. Chem. Int. Ed. Engl.*, **52**, 13970–13974.
148. Si,L., Xu,H., Zhou,X., Zhang,Z., Tian,Z., Wang,Y., Wu,Y., Zhang,B., Niu,Z., Zhang,C. *et al.* (2016) Generation of influenza A viruses as live but replication-incompetent virus vaccines. *Science*, **354**, 1170–1173.
149. Wang,N., Li,Y., Niu,W., Sun,M., Cerny,R., Li,Q. and Guo,J. (2014) Construction of a live-attenuated HIV-1 vaccine through genetic code expansion. *Angew. Chem. Int. Ed. Engl.*, **53**, 4867–4871.
150. Loughran,G., Chou,M.Y., Ivanov,I.P., Jungreis,I., Kellis,M., Kiran,A.M., Baranov,P.V. and Atkins,J.F. (2014) Evidence of efficient stop codon readthrough in four mammalian genes. *Nucleic Acids Res.*, **42**, 8928–8938.
151. Fan,R.L., Valkenburg,S.A., Wong,C.K., Li,O.T., Nicholls,J.M., Rabadan,R., Peiris,J.S. and Poon,L.L. (2015) Generation of live attenuated influenza virus by using codon usage bias. *J. Virol.*, **89**, 10762–10773.
152. Meng,J., Lee,S., Hotard,A.L. and Moore,M.L. (2014) Refining the balance of attenuation and immunogenicity of respiratory syncytial virus by targeted codon deoptimization of virulence genes. *MBio*, **5**, e01704-14.
153. Sitaraman,V., Hearing,P., Ward,C.B., Gnatenko,D. V, Wimmer,E., Mueller,S., Skiena,S. and Bahou,W.F. (2011) Computationally designed adeno-associated virus (AAV) Rep 78 is efficiently maintained within an adenovirus vector. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 14294–14299.