Received: 30 July 2020 | Revised: 23 December 2020 | Accepted: 29 December 2020

DOI: 10.1002/biot.202000354

Biotechnology Journal

RAPID COMMUNICATION

Oligonucleotide abundance biases aid design of a type IIS synthetic genomics framework with plant virome capacity

Fabio Pasin[1,2]

[1] School of Science, University of Padova, Padova, Italy

[2] Agricultural Biotechnology Research Center, Academia Sinica, Taipei, Taiwan

Correspondence
Fabio Pasin, School of Science, University of Padova, Padova, Italy.
Email: fabio.pasin@studenti.unipd.it

Abstract

Synthetic genomics-driven dematerialization of genetic resources facilitates flexible hypothesis testing and rapid product development. Biological sequences have compositional biases, which, I reasoned, could be exploited for engineering of enhanced synthetic genomics systems. In proof-of-concept assays reported herein, the abundance of random oligonucleotides in viral genomic components was analyzed and used for the rational design of a synthetic genomics framework with plant virome capacity (SynViP). Type IIS endonucleases with low abundance in the plant virome, as well as Golden Gate and No See'm principles were combined with DNA chemical synthesis for seamless viral clone assembly by one-step digestion-ligation. The framework described does not require subcloning steps, is insensitive to insert terminal sequences, and was used with linear and circular DNA molecules. Based on a digital template, DNA fragments were chemically synthesized and assembled by one-step cloning to yield a scar-free infectious clone of a plant virus suitable for *Agrobacterium*-mediated delivery. SynViP allowed rescue of a genuine virus without biological material, and has the potential to greatly accelerate biological characterization and engineering of plant viruses as well as derived biotechnological tools. Finally, computational identification of compositional biases in biological sequences might become a common standard to aid scalable biosystems design and engineering.

KEYWORDS
DNA chemical synthesis, Golden Gate cloning, plant virome, type IIS restriction enzyme, viral infectious clone assembly

1 | INTRODUCTION

Viruses are relatively simple biological entities, with genomic components whose size is compatible both in terms of technical feasibility and costs with current advances in *de novo* DNA synthesis.[1,2] Synthetic genomics is becoming commonplace for the study and engineering of animal viruses, especially those with medical interest and human pandemic potential.[3–7] In plant virology, synthetic genomics can help to demonstrate correctness of genomic sequences, rescue of viruses that might not be physically available (e.g., ancient or environmental samples), as well as to accelerate virus reverse genetics, host-virus interaction studies, biological characterization of emerging viruses, or engineering of biotechnological devices.[8–11] Adoption of this powerful approach for plant virology has nonetheless lagged and is reported only in a handful of studies.[9] The first reports of synthetic genomic approaches in plant virology included the use of oligonucleotides to

assemble a synthetic tobamovirus,[12] and the reconstitution of a full-length tombusvirus genome from synthetic fragments cloned in an intermediate vector.[13]

Successful examples of synthesis and assembly of viral and bacterial genomes, whose sizes are orders of magnitude larger than those of plant viruses, suggest that the field of plant virology has additional constraints beyond DNA synthesis technology. Viral clone or vector delivery to plants requires specialized techniques,[11] one of the most efficient of which is *Agrobacterium*-mediated delivery (agro-infection). Agro-infection requires assembly of viral genomic sequences into binary vectors, which can pose technical challenges.[9] Adoption of synthetic genomic approaches by plant virologists would be greatly facilitated by improved molecular tools and frameworks that enable assembly and plant delivery of chemically-synthesized virus components.

Low-cost and large-scale methods for *de novo* gene synthesis rely on barcode sequences that allow standardized, PCR-mediated assembly of short oligonucleotides.[14,15] Gibson assembly and its variants were used successfully for plant virus clone assembly,[9,16–18] but their applications are limited to linear fragments with terminal homologies. A cloning system compatible with linear or circular DNA molecules and insensitive to specific terminal sequences such as gene synthesis barcodes would facilitate low-cost, rapid engineering of plant virus clones and vectors without biological material requirements.

Restriction enzymes (RE) are robust tools that have been used in molecular cloning for decades. Type IIS REs cleave outside of the recognition sites and can be used to produce cloning junctions with arbitrary sequences; correct junction design allows unidirectional, scar-free assembly.[19,20] Type IIS RE-based strategies such as No See'm cloning allowed recovery of infectious clones of large-sized viruses with no cloning scars.[21–24] Multiple digestion, ligation, purification and/or subcloning steps were nonetheless required. Type IIS REs and DNA ligases can be mixed for one-step digestion-ligation reactions.[25,26] Based on this property, unique junction sites, and different antibiotic selection schemes, Golden Gate cloning allows high throughput transfer of desired fragments from donor to recipient vectors.[26] Golden Gate syntaxes of unique cloning junctions have been used extensively for robust, directional assembly in a single reaction of over 30 standard parts, which are then flanked by 4-nt cloning junctions.[27,28]

The plant virome is composed of highly divergent viruses with polyphyletic origins.[29] Here, I sought to explore compositional biases of viral sequences for rational design of a Synthetic genomics framework to rescue Viruses of Plants (SynViP). A sequence database was analyzed to measure random oligonucleotide abundance in the plant virome. The computational results were used to determine virome cloning constraints of REs and guide the development of a framework that combines Golden Gate and No See'm principles with low-cost DNA chemical synthesis. A new binary vector was generated to incorporate divergent sites of type IIS REs with low frequencies in the plant virome. Appropriate cloning designs allow one-step, seamless assembly of full-length viral genomic components suitable for agro-infection. The SynViP framework has plant virome capacity, can be used with linear and circular DNA molecules, does not require subcloning steps, is insensitive to fragment terminal sequences, and was used successfully to assemble an infectious clone and rescue a plant RNA virus based on a digital template.

2 | RESULTS

2.1 | Abundance of oligonucleotides in a plant virome is non-random and correlates negatively with their GC content

For sequence analysis, I selected a reference plant virome dataset that includes 2044 accessions, which were grouped into 34 taxonomic families with 1 to >600 assigned genomic components (Figure 1A; Table S1). The virome had a guanine + cytosine content (%GC) peak at the 42% bin (Figure 1A), and an average %GC aggregate per family ranging from 31.48% to 58.22% (Figure 1B). To establish virus study systems, genomic components are usually cloned individually into suitable plasmid vectors. Multipartite viruses are inoculated by simultaneous plant delivery of multiple vectors that include each of the genomic components. REs recognize short DNA sequences. Complete non-redundant oligonucleotide (n.r. mer) sets were generated by random union of three to eight DNA mononucleotides, and searched in the virome to comprehensively infer RE constraints in plant virus cloning (Figure 1C). Virome sequence analysis showed that n.r. mer counts had strong negative correlation with oligonucleotide GC content; abundance of an oligonucleotide in the plant virome is, thus, significantly reduced by an increase in its %GC. The result was confirmed for each of the n.r. mer sets, and the Pearson's correlation coefficients ranged between $-0.948 < r < -0.997$ (Figure 1D). The n.r. mer sets were classified by size and %GC, and the subsets obtained were ordered by their virome average counts. The ordered subsets failed to group uniformly by size (Figure 1E), indicating that a size increase alone is not sufficient to reduce virome oligonucleotide abundance. Abundance of the 6-nt subset with 100% GC was lower than that of the 7-nt subsets with 0%, 14.3%, and 28.6% GC; in turn, abundance of the 7-nt subset with 100% GC was lower than that of the 8-nt subsets with 0%, 12.5% and 25.0% GC (Figure 1F). The results revealed that (i) the virome dataset analyzed has a non-random oligonucleotide composition, (ii) the oligonucleotide abundance in the virome is reduced by a concomitant increase in oligonucleotide size and GC content and, by data interpolation, (iii) a RE that recognizes a long DNA sequence with a high %GC would alleviate unwanted targeting and thus cloning constraints of plant virome components.

2.2 | Abundance of type IIS RE sites in a plant virome depends on site size and GC content

Type IIS REs are compatible with high-throughput and automated construct design and assembly.[19,20,30] A subset of commercially available type IIS REs with ≥6-nt recognition sites and ≥3-nt overhangs was considered (Table 1). Virome average counts for type IIS REs with 6-nt sites ranged from 1.24 to 3.99 (Figure 2A); this variation could not be

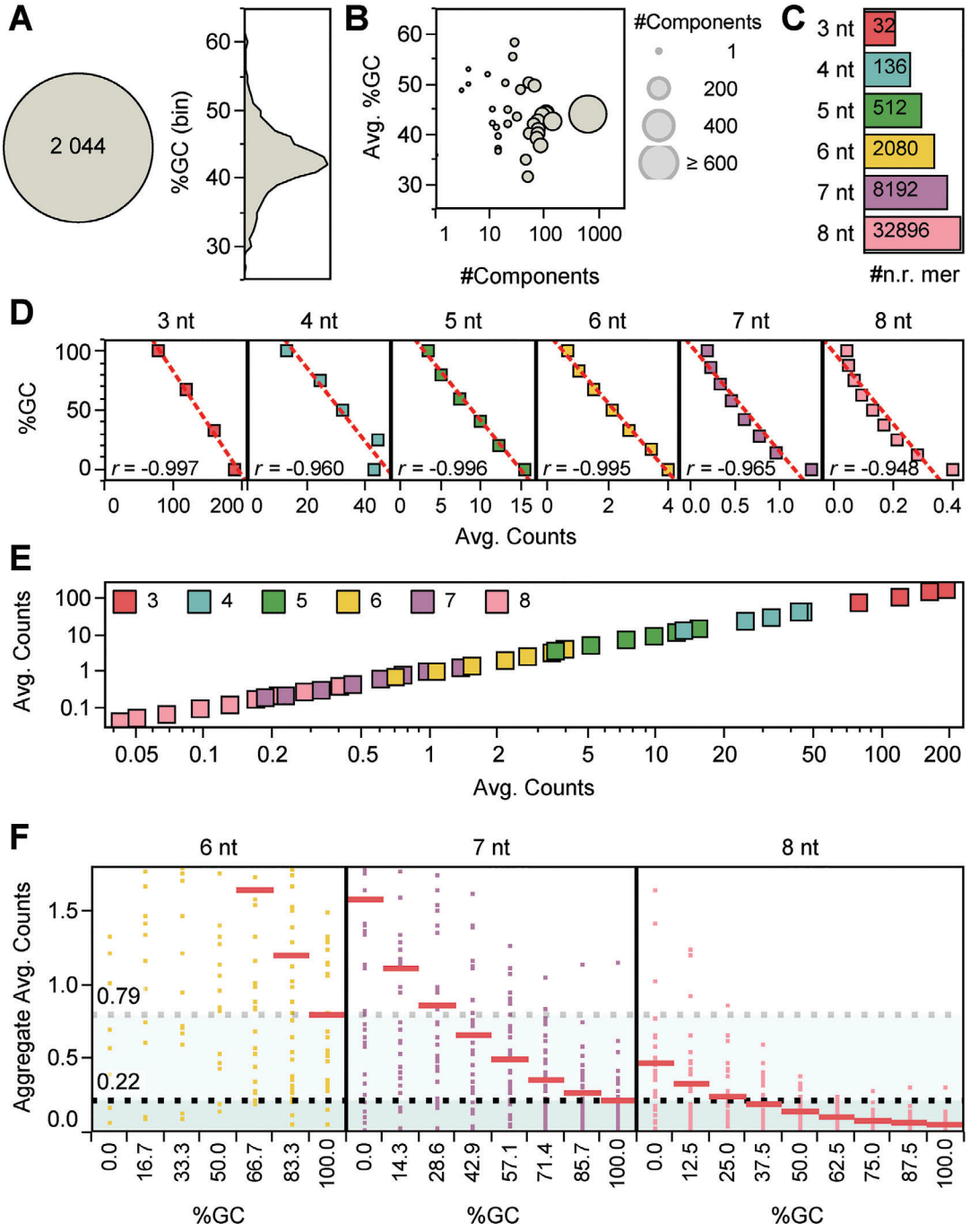

FIGURE 1 Biased abundance of nucleotide oligomers in plant virome sequences. (A) Number of genomic components and guanine-cytosine content (%GC) distribution of the plant virome dataset used. (B) Aggregate component numbers and average %GC per taxonomic family ($n = 34$). (C) Numbers of the random, non-redundant nucleotide oligomers (n.r. mer) used in sequence analysis. (D) The n.r. mers were grouped by size and %GC, and virome average count ($n = 2044$) is shown for each subset; linear correlation values are shown (Pearson's r). (E) Scatter plot shows virome average counts ($n = 2044$) for the subsets in panel E. (F) For each subset, the red line indicates the aggregate average count per taxonomic families ($n = 34$, each point indicates a family); values for the 6-nt (0.79) or 7-nt subsets (0.22) with 100% GC are indicated by dotted lines

explained by recognition site size alone. Abundance of the 6-nt type IIS REs was inversely related to their %GC, as counts for EarI or BsbI (50% GC) were greater than those for high %GC counterparts BsaI, BfuAI, BtgZI or BsmBI (66.67% GC; Figure 2A). Virome counts for the 7-nt type IIS REs were also inversely related to the site %GC, as SapI frequency was 68% greater than AarI (0.74 vs. 0.44) (Figure 2A), which recognizes a sequence with 71.43% GC (Table 1; Figure 2A). The virome dataset comprises diverse taxonomic families with large variation in the number of genomic components (Figure 1B; Table S1). Aggregate counts per family were considered to control overrepresentation of

TABLE 1 The type IIS restriction enzymes considered in the study and their features

		Recognition site			Overhang	
RE	Isoschizomers[a]	Sequence	nt	%GC	nt	References[b]
AarI	n.a.	CACCTGC	7	71.43	4	[60,63–65]
SapI	BspQI, LguI, PciSI	GCTCTTC	7	57.14	3	[63,66,67]
BfuAI	Acc36I, BspMI, BveI	ACCTGC	6	66.67	4	see AarI[c]
BsmBI	Esp3I	CGTCTC	6	66.67	4	[68–71]
BtgZI	n.a.	GCGATG	6	66.67	4	[72]
BsaI	Bso31I, BspTNI, Eco31I	GGTCTC	6	66.67	4	[26,70,73,74]
BbsI	BpiI, BstV2I	GAAGAC	6	50.00	4	[26,75–78]
EarI	Bst6I, Eam1104I	CTCTTC	6	50.00	3	see SapI[c]

[a]n.a., non-commercially available isoschizomers (source Rebase[79]).
[b]Representative uses in Golden Gate-based cloning.
[c]BfuAI recognizes the AarI site; EarI recognizes the SapI site.

a specific taxonomic group; results confirmed those of Figure 2A, as per-family abundance of EarI or BsbI was generally greater than that of BsaI, BfuAI, BtgZI or BsmBI (Figure 2B). The aggregate count difference among the 7-nt type IIS REs was significant ($p < 0.01$, $n = 34$), and SapI frequency was 42% greater than AarI (0.75 vs. 0.53) (Figure 2C). Virome counts for type IIS Res; thus, depend on both size and %GC of the recognition sites. It can be concluded that use of a type IIS RE that targets a 7-nt site with high %GC, that is, AarI, would be preferable in the design of a cloning strategy with plant virome capacity.

2.3 | Design of SynViP, a type IIS synthetic genomics framework with plant virome capacity

Despite the low AarI frequency, data inspection identified genomic components with a larger number of AarI instances than SapI (e.g., *Genomoviridae*, Figure 2B). I asked whether a cloning system compatible with both AarI and SapI would confer a technical advantage. Virome abundance of REs with low %GC sites imposes substantial cloning constraints; EarI and BsbI (50% GC) were not considered further. Sequence analysis showed that a hypothetical cloning system based on BsaI, BsmBI, BtgZI, or BfuAI (6-bp targeting REs with 66.67% GC) would allow recovery of clones with no sequence modification for 30.58%-41.93% of the virome; complete virome cloning would nevertheless require targeted insertion of up to 13 or 18 single-nucleotide mutations per component (Figure 2D). SapI- or AarI-based systems would respectively allow recovery of clones with no sequence modification for 56.70% or 69.81% of the virome, and insertion of up to 6 single-nucleotide mutations per component would confer virome capacity. The alternative choice of AarI- or SapI-based designs would allow cloning of faithful sequence copies for >80% of virome components (bottom row, Figure 2D). Similar designs coupled with a targeted insertion of single- or two-nucleotide mutations to remove RE recognition sites would allow recovery of full-length clones of >95% or >99% of virome components, respectively. *In silico* analysis showed that a system compatible with both AarI and SapI is thus superior to those based on an individual type IIS RE.

Guided by these results, I generated pLX-AS, a mini binary vector compatible with Golden Gate and No See'm assembly strategies that employ either AarI or SapI (Figure 2E-G; see Section 4, and Table S2-S3). pLX-AS is based on the pLX backbone,[31] which has been used for *Agrobacterium*-mediated delivery to plants of infectious clones of several RNA and DNA viruses.[16,18] The T-DNA region of pLX-AS comprises a cloning cassette flanked by sequences of the cauliflower mosaic virus (CaMV) 35S promoter and the nopaline synthase (nos) terminator (Figure 2E) for *in planta* transcription regulation of cDNA clones from RNA viruses. The cloning cassette designed includes divergent AarI and SapI sites flanking the *Escherichia coli lacZ* reporter for white-blue screens of recombinant clones (Figure 2F). In RNA virus clone assembly, backbone overhangs generated by AarI or SapI digestion allow seamless fragment insertion downstream of the CaMV 35S promoter, and *in planta* generation of authentic viral genomic 5′ ends (Figure 2G).

2.4 | One-step digestion-ligation assembly of a virus clone using a digital template and uncloned synthetic DNA fragments

An AarI/SapI-based cloning strategy including pLX-AS would (i) be compatible with linear and/or circular DNA molecules, (ii) be insensitive to specific terminal sequences such as gene synthesis barcodes, and (iii) facilitate rapid engineering of plant virus clones without biological material requirements. The family *Solemoviridae* includes RNA viruses with an average genome size of 4208 nt, close to the aggregate median size (4154) of plant virome components (Figure 3A). As proof of the utility of the synthetic genomics framework designed, I focused on turnip rosette virus (TuRV; *Solemoviridae*). The TuRV genomic sequence and its complex organization have recently been revised,[32,33] providing a representative, high-confidence control during method validation. The TuRV accession KC778720.1 used as a

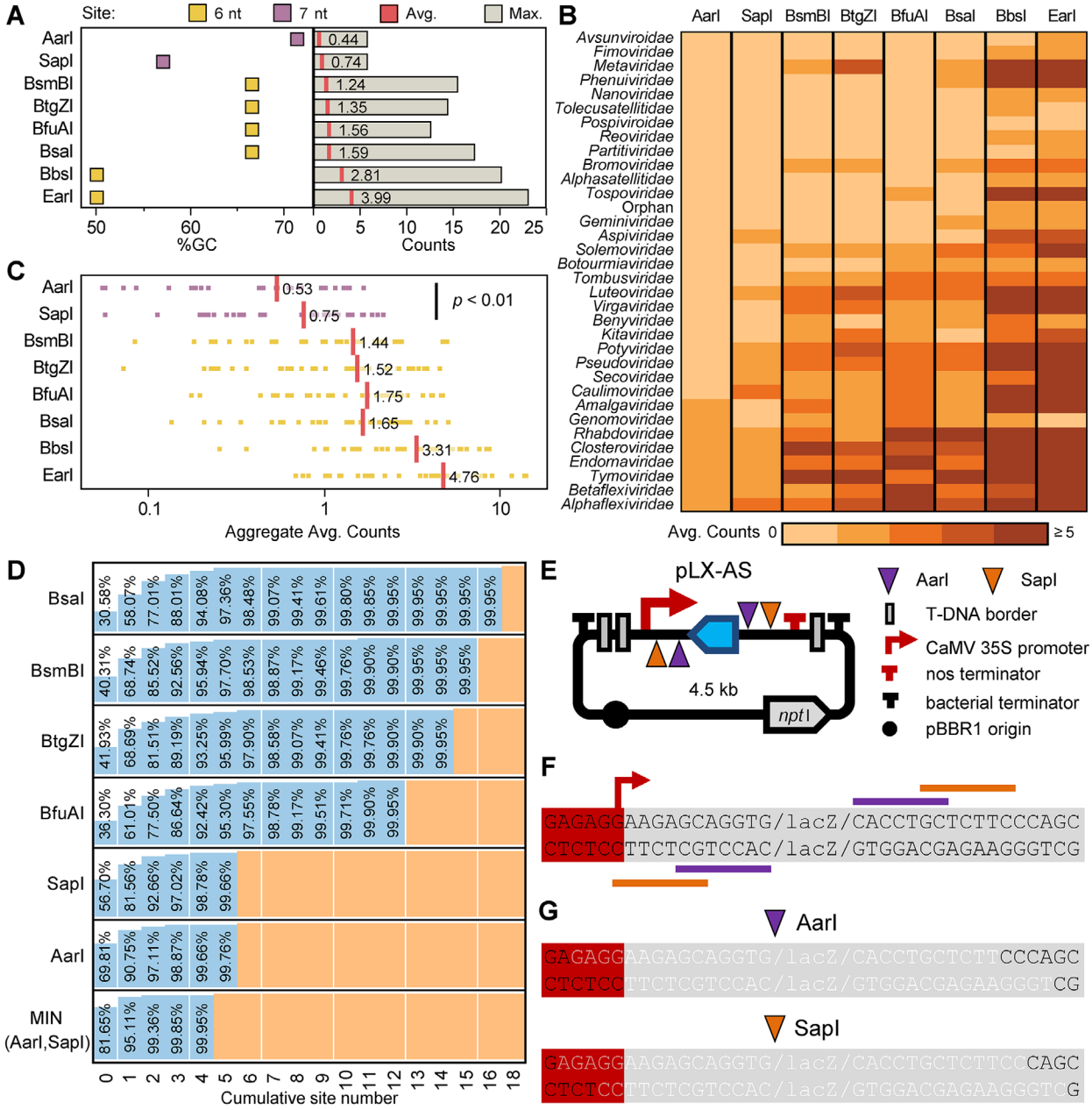

FIGURE 2 Differential abundance of type IIS restriction enzyme sites in the plant virome and design of the SynViP framework. (A) Left, plot shows site size and %GC of reference type IIS restriction enzymes (RE); right, RE site counts in virome components ($n = 2044$; bars, maximum; red lines, mean). (B) Heatmap shows RE site average counts in each taxonomic family; Orphan, species with unassigned family. (C) For each RE, the red line and value indicate the aggregate average count per taxonomic families (each point indicates a family, $n = 34$); p value by Wilcoxon signed-rank test is indicated for the AarI *versus* SapI comparison ($n = 34$). (D) Cumulative component percentages relative to the complete virome are shown for site count numbers; values are shown for each RE, or considering the smallest count number among AarI or SapI, that is, MIN(AarI,SapI); orange bars indicate 100%. (E) Diagram of pLX-AS, a 4.5-kb T-DNA binary vector including divergent AarI and SapI recognition sites. Vector components are indicated (right); blue, *lacZ* reporter gene; *npt*I, kanamycin resistance gene. (F) Detail of the T-DNA cloning cassette of pLX-AS. The 3′ sequence of the CaMV 35S promoter is in red; arrow indicates the nucleotide preceding the transcription initiation site; lines mark AarI (purple) and SapI (orange) sites. (G) Detail of the pLX-AS cloning cassette, after AarI (top) or SapI (bottom) digestions; backbone overhangs are show in black

digital template includes a single SapI and no AarI sites. Three linear DNA fragments ranging from 1.2 to 1.6 kb and spanning the TuRV genome were obtained by chemical synthesis (Table S4). Each synthetic fragment was flanked by convergent AarI sites that would allow removal of synthesis barcodes (Figure 3B), and the generation of orthogonal cloning junctions selected to allow (i) directional fragment assembly in pLX-AS by one-step digestion-ligation, (ii) recovery of a seamless TuRV genomic sequence, and (iii) *in planta* generation of the correct TuRV 5′ end from the CaMV 35S promoter (Figure 3C). A synthetic ribozyme was included in the gene synthesis design for *in planta*

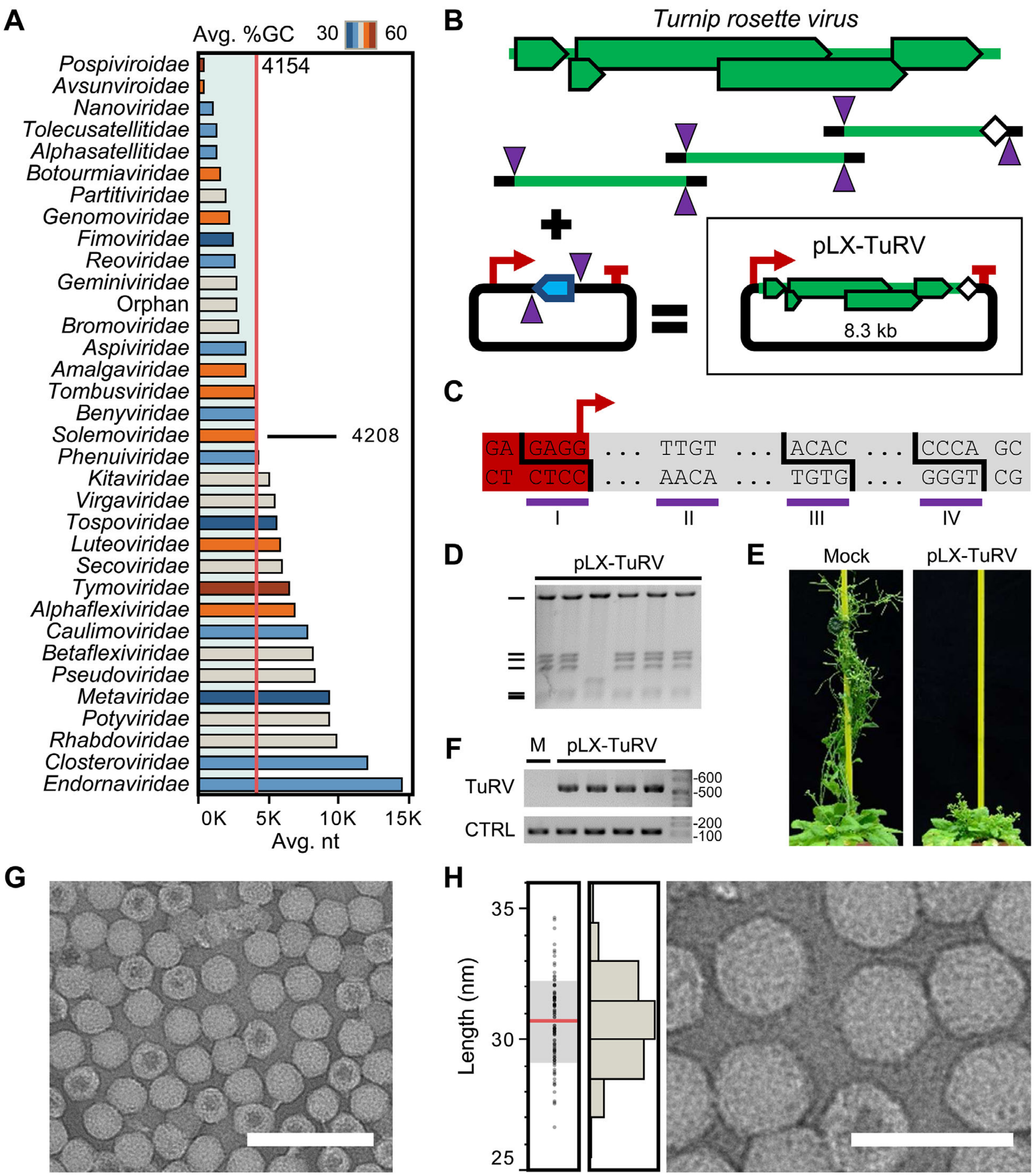

FIGURE 3 Use of SynViP for the one-step digestion-ligation assembly of an infectious clone and the rescue of a genuine virus with no natural template. (A) Bar plot shows average component size and %GC of each taxonomic family; the red line indicates the aggregate virome median size (4154 nt; $n = 34$); *Solemoviridae* average size is shown (4208 nt; $n = 19$). (B) Top, genomic representation of turnip rosette virus (TuRV). DNA fragments spanning the TuRV genome were chemically synthesized; fragments were flanked by synthesis barcodes (black boxes) and convergent AarI sites (triangles); diamond, a synthetic ribozyme. Fragments were assembled into pLX-AS by one-step digestion-ligation. pLX-TuRV is a T-DNA binary vector with a seamless cDNA copy of the TuRV genome flanked by the CaMV 35S promoter (red arrow) and nos terminator (T-shaped mark). (C) Detail of the orthogonal cloning junctions generated during the AarI-based digestion-ligation depicted in panel B; the 3′ end of the CaMV 35S promoter is in red. (D) Restriction profiles are shown of plasmids purified from selected colonies; left, computed restriction pattern. (E) A control vector (Mock) or pLX-TuRV were transformed into *Agrobacterium* and delivered to *Arabidopsis thaliana* by agro-infection. Images show plants at 30 days post agro-infection. (F) Viral RNA detection in uninoculated leaf samples of agro-inoculated plants. RT-PCR reactions are shown using virus-specific primers (TuRV) or actin primers (CTRL); DNA size markers are indicated (right); M, mock sample. (G) Transmission electron micrograph of particles purified from infected plant material; scale bar, 100 nm. (H) Size mean ± SD and distribution ($n = 100$), and micrograph magnification are shown of purified particles; scale bar, 50 nm

removal of non-viral nucleotides from the TuRV 3′ end (Figure 3B, Table S4).

The uncloned synthetic fragments and pLX-AS were mixed and subjected to digestion-ligation using AarI and T4 DNA ligase in AarI buffer supplemented with ATP (see Section 4); the reaction was transformed into *E. coli* and white colonies were selected using a chromogenic substrate to recover recombinant clones. Correct fragment assembly into pLX-AS was confirmed by restriction analysis of plasmid DNA (Figure 3D). A plasmid clone with the correct digestion pattern and sequence was designated pLX-TuRV, that is, an 8.3-kb vector suitable for agro-infection that includes a cDNA copy of the full-length TuRV genome (Figure 3B).

2.5 | Delivery of synthetic infectious clone to plants by agro-infection

Arabidopsis thaliana is a model plant and TuRV host.[32] pLX-TuRV was transformed into *Agrobacterium*, and its infectivity was evaluated in *A. thaliana* by agro-infection. After 30 days, plants agro-inoculated with pLX-TuRV showed severe symptoms and significant growth reduction compared to the mock condition ($p < 0.001$; Figure 3E). Results are consistent with the stunted growth phenotype reported for TuRV-infected *A. thaliana* plants.[32] RNA analysis of upper uninoculated leaves by RT-PCR assays showed TuRV accumulation in pLX-TuRV agro-inoculated plants, but not in the mock condition (Figure 3F). Nanoparticles were purified from infected plant material and analyzed by electron microscopy (Figure 3G); the particles observed had an icosahedral shape and an average size of 30.67 nm (Figure 3H), consistent with those described for TuRV virions.[34] These results confirm the infectivity of the pLX-TuRV binary vector assembled, and thus the suitability of the SynViP framework and its cloning strategy for the generation of plant virus infectious clones and rescue of a genuine plant virus, with no natural template or biological material requirements.

3 | DISCUSSION

Gene synthesis has revolutionized the study of bacteriophages and animal viruses, and spurred major advances in prevention and control of human infectious diseases. First studies in synthetic genomics used viruses as simple model systems to prove the feasibility of recovering functional biological entities solely by sequence data and *in vitro* chemical means.[1,2] More recently, synthetic genomic approaches have been used in basic and applied research,[1,2] including sequence refactoring to facilitate human understanding and manipulation of individual genetic elements,[35] to probe the completeness of our knowledge of biological systems,[36] and for rational vaccine design and rapid vaccine manufacturing.[3–7]

In plant virology, studies have reported the recovery of plant infectious agents in the absence of a physical source of biological materials,[9] although they relied on cloning methods with limited scalability. Instructed by sequence analysis of a plant virome dataset, I describe here SynViP, a synthetic genomics framework with plant virome capacity. The SynViP framework described (i) does not require subcloning steps, (ii) is compatible with DNA fragments generated by low-cost and large-scale chemical synthesis methods and thus has no natural template requirements, (iii) complies with Golden Gate principles to allow directional assembly of multiple DNA fragments in one-step digestion-ligation reactions, (iv) allows recovery of scar-free constructs by following No See'm principles, (v) through its simultaneous compatibility with AarI and SapI, considerably reduces type IIS RE constraints in cloning plant virome components, and (vi) relies on the newly designed pLX-AS binary vector for construct assembly.

pLX-AS includes elements for stable vector maintenance in *E. coli* and *Agrobacterium* cells, and for viral clone delivery to plants by agro-infection. pLX-AS autonomously replicates in *E. coli* and *Agrobacterium*; thus, overcoming a major limitation of the pSoup-pGreen system.[37–41] To promote stable plasmid propagation in bacteria, pLX-AS includes bacterial terminators up- and downstream of T-DNA, and has a lower copy number than pPZP and derivative vectors (e.g., pCAMBIA, pCass4-Rz).[42] The RK2-based pCB301 and its derivatives (e.g., pJL-89, pDIVA) have been used in virus infectious clone assembly;[10,17,43–46] pLX-AS includes the pBBR1 origin, which outperformed the RK2 replicon in transient expression assays.[31]

SynViP was successfully applied for one-step, seamless assembly of an infectious clone of a representative plant RNA virus with no natural template requirements. Uncloned synthetic fragments spanning the TuRV genome were assembled by one-step digestion-ligation into pLX-AS to yield pLX-TuRV, a TuRV infectious clone suitable for agro-infection. The pLX-TuRV infectivity and rescue of a genuine plant virus without biological material requirements was confirmed by visual assessment of host symptoms, and by detection of virus genomes and assembled virus particles in agro-inoculated plant samples. pLX-TuRV is a new entry in our previously reported binary vector collection of *A. thaliana*-infecting viruses,[16] which now covers members of families with disparate genome composition (single-stranded RNA, single- and double-stranded DNA) and organization, including *Caulimoviridae*, *Geminiviridae*, *Potyviridae*, *Virgaviridae*, and *Solemoviridae* (this study). Very short or repetitive sequences can limit efficiency and accuracy of Gibson assembly, but they do not interfere with type IIS-based methods.[19] The approach described herein is, therefore, predicted to be suitable for assembly of tandem genomic repeats needed to rescue viral agents with circular components.[9,47,48] Collectively, it can be anticipated that the framework described in this study will support spread and further development of synthetic genomic approaches for the biological characterization and engineering of plant viruses and derived biotechnologies.

Finally, biological systems are often engineered through *ad hoc* strategies that can be applied to a limited number of related species;[49] novel theoretical approaches and methodological concepts for biosystems engineering are necessary to meet the ever-increasing human needs.[50–54] Non-random compositional tendencies, dinucleotide frequency, and codon usage biases have been reported in genomic sequences, including those of viruses.[55–59] The proposed computational identification of compositional biases in biological sequences

might become a common standard to aid scalable biosystems design and engineering.

4 | MATERIALS AND METHODS

Methods and details of the sequence analysis, digestion-ligation,[60] virus inoculation and purification[16,61,62] are provided in the Supporting Information. The vector sequences from this study can be found at NCBI under the GenBank accession numbers MW281334 (pLX-AS) and MW281335 (pLX-TuRV).

ACKNOWLEDGMENTS

The author is grateful to D. Endy, K. Gandall, and the BioBricks Foundation for the supply of materials, to D. San León for bioinformatics advice and scripting, to X.-A. Tseng for technical support, and to C. Mark for editorial assistance. Resources from Y.R. Chen, W. Schmidt, S.-T.D. Hsu, P. Draczkowski, and Academia Sinica Cryo-EM Center (ASCEM) are acknowledged. F.P. was the recipient of a post-doctoral fellowship from Academia Sinica (Taiwan).

CONFLICT OF INTEREST

The author declares no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

ORCID

Fabio Pasin https://orcid.org/0000-0002-9620-4301

REFERENCES

1. Schindler, D., Dai, J., & Cai, Y. (2018). Synthetic genomics: A new venture to dissect genome fundamentals and engineer new functions. *Curr. Opin. Chem. Biol.*, *46*, 56.
2. Zhang, W., Mitchell, L. A., Bader, J. S., & Boeke, J. D. (2020). Synthetic genomes. *Annu. Rev. Biochem.*, *89*, 77.
3. Coleman, J. R., Papamichail, D., Skiena, S., Futcher, B., Wimmer, E., & Mueller, S. (2008). Virus attenuation by genome-scale changes in codon pair bias. *Science*, *320*, 1784.
4. Mueller, S., Coleman, J. R., Papamichail, D., Ward, C. B., Nimnual, A., Futcher, B., … Wimmer, E. (2010). Live attenuated influenza virus vaccines by computer-aided rational design. *Nat. Biotechnol.*, *28*, 723.
5. Dormitzer, P. R., Suphaphiphat, P., Gibson, D. G., Wentworth, D. E., Stockwell, T. B., Algire, M. A., … Venter, J. C. (2013). Synthetic generation of influenza vaccine viruses for rapid response to pandemics. *Sci. Transl. Med.*, *5*, 185ra68.
6. Moratorio, G., Henningsson, R., Barbezange, C., Carrau, L., Bordería, A. V., Blanc, H., … Vignuzzi, M. (2017). Attenuation of RNA viruses by redirecting their evolution in sequence space. *Nat. Microbiol*, *2*, 17088.
7. Rourke, M. F., Phelan, A., & Lawson, C. (2020). Access and benefit-sharing following the synthesis of horsepox virus. *Nat. Biotechnol.*, *38*, 537.
8. Massart, S., Candresse, T., Gil, J., Lacomme, C., Predajna, L., Ravnikar, M., … Wetzel, T. (2017). A framework for the evaluation of biosecurity, commercial, regulatory, and scientific impacts of plant viruses and viroids identified by NGS technologies. *Front. Microbiol.*, *8*, 45.
9. Pasin, F., Menzel, W., & Daròs, J.-A. (2019). Harnessed viruses in the age of metagenomics and synthetic biology: An update on infectious clone assembly and biotechnologies of plant viruses. *Plant Biotechnol. J.*, *17*, 1010.
10. Feng, M., Cheng, R., Chen, M., Guo, R., Li, L., Feng, Z., … Tao, X. (2020). Rescue of tomato spotted wilt virus entirely from complementary DNA clones. *Proc. Natl. Acad. Sci. USA*, *117*, 1181.
11. Abrahamian, P., Hammond, R. W., & Hammond, J. (2020). Plant virus-derived vectors: Applications in agricultural and medical biotechnology. *Annu. Rev. Virol.*, *7*, 513.
12. Cooper, B. (2014). Proof by synthesis of *Tobacco mosaic* virus. *Genome Biol.*, *15*, R67.
13. Lovato, A., Faoro, F., Gambino, G., Maffi, D., Bracale, M., Polverari, A., & Santi, L. (2014). Construction of a synthetic infectious cDNA clone of *Grapevine Algerian latent virus* (GALV-Nf) and its biological activity in *Nicotiana benthamiana* and grapevine plants. *Virol. J.*, *11*, 186.
14. Kosuri, S., & Church, G. M. (2014). Large-scale *de novo* DNA synthesis: Technologies and applications. *Nat. Methods*, *11*, 499.
15. Hughes, R. A., & Ellington, A. D. (2017). Synthetic DNA synthesis and assembly: Putting the synthetic in synthetic biology. *Cold Spring Harb. Perspect. Biol.*, *9*, a023812.
16. Pasin, F., Tseng, X.-A., Bedoya, L. C., Heydarnejad, J., Deng, T.-C., García, J. A., & Chen, Y.-R. (2018). Streamlined generation of plant virus infectious clones using the pLX mini binary vectors. *J. Virol. Methods*, *262*, 48.
17. Blawid, R., & Nagata, T. (2015). Construction of an infectious clone of a plant RNA virus in a binary vector using one-step Gibson Assembly. *J. Virol. Methods*, *222*, 11.
18. Zhao, M., García, B., Gallo, A., Tzanetakis, I. E., Simón-Mateo, C., García, J. A., & Pasin, F. (2020). Home-made enzymatic premix and Illumina sequencing allow for one-step Gibson assembly and verification of virus infectious clones. *Phytopathol. Res.*, *2*, 36.
19. Casini, A., Storch, M., Baldwin, G. S., & Ellis, T. (2015). Bricks and blueprints: Methods and standards for DNA assembly. *Nat. Rev. Mol. Cell Biol.*, *16*, 568.
20. Young, R., Haines, M., Storch, M., & Freemont, P. S. (2020). Combinatorial metabolic pathway assembly approaches and toolkits for modular assembly. *Metab. Eng.*, https://doi.org/10.1016/j.ymben.2020.12.001.
21. Yount, B., Denison, M. R., Weiss, S. R., & Baric, R. S. (2002). Systematic assembly of a full-length infectious cDNA of mouse hepatitis virus strain A59. *J. Virol.*, *76*, 11065.
22. Cockrell, A. S., Beall, A., Yount, B., & Baric, R. (2017). Efficient reverse genetic systems for rapid genetic manipulation of emergent and pre-emergent infectious coronaviruses. In D. R. Perez (Ed.), Reverse Genetics of RNA Viruses (pp. 59–81). New York, NY: Springer New York.
23. Ma, Z., Li, Z., Dong, L., Yang, T., & Xiao, S. (2020). Reverse genetic systems: Rational design of coronavirus live attenuated vaccines with immune sequelae. *Adv. Virus Res.*, *107*, 383.
24. Xie, X., Muruato, A., Lokugamage, K. G., Narayanan, K., Zhang, X., Zou, J., … Shi, P.-Y. (2020). An infectious cDNA clone of SARS-CoV-2. *Cell Host Microbe*, *27*, 841.
25. Kotera, I., & Nagai, T. (2008). A high-throughput and single-tube recombination of crude PCR products using a DNA polymerase inhibitor and type IIS restriction enzyme. *J. Biotechnol.*, *137*, 1.
26. Marillonnet, S., & Grützner, R. (2020). Synthetic DNA assembly using Golden Gate cloning and the hierarchical modular cloning pipeline. *Curr. Protoc. Mol. Biol.*, *130*, e115.
27. Patron, N. J., Orzaez, D., Marillonnet, S., Warzecha, H., Matthewman, C., Youles, M., … Haseloff, J. (2015). Standards for plant synthetic biology: A common syntax for exchange of DNA parts. *New Phytol*, *208*, 13.
28. Pryor, J. M., Potapov, V., Kucera, R. B., Bilotti, K., Cantor, E. J., & Lohman, G. J. S. (2020). Enabling one-pot Golden Gate assemblies of unprecedented complexity using data-optimized assembly design. *PLoS One*, *15*, e0238592.

29. Lefeuvre, P., Martin, D. P., Elena, S. F., Shepherd, D. N., Roumagnac, P., & Varsani, A. (2019). Evolution and ecology of plant viruses. *Nat. Rev. Microbiol*., *17*, 632.
30. Smanski, M. J., Bhatia, S., Zhao, D., Park, Y., Woodruff, L. B. A., Giannoukos, G., … Voigt, C. A. (2014). Functional optimization of gene clusters by combinatorial design and assembly. *Nat. Biotechnol*., *32*, 1241.
31. Pasin, F., Bedoya, L. C., Bernabé-Orts, J. M., Gallo, A., Simón-Mateo, C., Orzaez, D., & García, J. A. (2017). Multiple T-DNA delivery to plants using novel mini binary vectors with compatible replication origins. *ACS Synth. Biol*, *6*, 1962.
32. Ling, R., Pate, A. E., Carr, J. P., & Firth, A. E. (2013). An essential fifth coding ORF in the sobemoviruses. *Virology*, *446*, 397.
33. Sõmera, M., & Truve, E. (2013). The genome organization of lucerne transient streak and turnip rosette sobemoviruses revisited. *Arch. Virol., 158*, 673.
34. Horne, R. W., Harnden, J. M., & Hull, R. (1977). The *in vitro* crystalline formations of turnip rosette virus I. Electron microscopy of two- and three-dimensional arrays. *Virology*, *82*, 150.
35. Chan, L. Y., Kosuri, S., & Endy, D. (2005). Refactoring bacteriophage T7. *Mol. Syst. Biol*., *1*, 2005.0018.
36. Jaschke, P. R., Dotson, G. A., Hung, K. S., Liu, D., & Endy, D. (2019). Definitive demonstration by synthesis of genome annotation completeness. *Proc. Natl. Acad. Sci. USA*, *116*, 24206.
37. Bouton, C., King, R. C., Chen, H., Azhakanandam, K., Bieri, S., Hammond-Kosack, K. E., & Kanyuka, K. (2018). *Foxtail mosaic virus*: A viral vector for protein expression in cereals. *Plant Physiol*., *177*, 1352.
38. Hellens, R. P., Edwards, E. A., Leyland, N. R., Bean, S., & Mullineaux, P. M. (2000). pGreen: A versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. *Plant Mol. Biol*., *42*, 819.
39. Lacomme, C., & Chapman, S. (2008). Use of potato virus X (PVX)-based vectors for gene expression and virus-induced gene silencing (VIGS). *Curr. Protoc. Microbiol*., *Chapter 16*, Unit 16I.1.
40. Wrzesińska, B., Wieczorek, P., & Obrępalska-Stęplowska, A. (2016). Recombination-based generation of the agroinfectious clones of *Peanut stunt virus*. *J. Virol. Methods*, *237*, 179.
41. Majer, E., Llorente, B., Rodríguez-Concepción, M., & Daròs, J.-A. (2017). Rewiring carotenoid biosynthesis in plants using a viral vector. *Sci. Rep*., *7*, 41645.
42. Annamalai, P., & Rao, A. L. N. (2005). Replication-independent expression of genome components and capsid protein of brome mosaic virus *in planta*: A functional role for viral replicase in RNA packaging. *Virology*, *338*, 96.
43. Xiang, C., Han, P., Lutziger, I., Wang, K., & Oliver, D. J. (1999). A mini binary vector series for plant transformation. *Plant Mol. Biol*., *40*, 711.
44. Lindbo, J. A. (2007). TRBO: A high-efficiency tobacco mosaic virus RNA-based overexpression vector. *Plant Physiol*., *145*, 1232.
45. Wang, Q., Ma, X., Qian, S., Zhou, X., Sun, K., Chen, X., … Li, Z. (2015). Rescue of a plant negative-strand RNA virus from cloned cDNA: Insights into enveloped plant virus movement and morphogenesis. *PLoS Pathog*., *11*, e1005223.
46. Laufer, M., Mohammad, H., Maiss, E., Richert-Pöggeler, K., Dall'Ara, M., Ratti, C., … Varrelmann, M. (2018). Biological properties of *Beet soilborne mosaic virus* and *Beet necrotic yellow vein virus* cDNA clones produced by isothermal *in vitro* recombination: Insights for reassortant appearance. *Virology*, *518*, 25.
47. Marquez-Molins, J., Navarro, J. A., Pallas, V., & Gomez, G. (2019). Highly efficient construction of infectious viroid-derived clones. *Plant Methods*, *15*, 87.
48. Krupovic, M., Varsani, A., Kazlauskas, D., Breitbart, M., Delwart, E., Rosario, K., … Koonin, E. V. (2020). *Cressdnaviricota*: A virus phylum unifying seven families of Rep-encoding viruses with single-stranded, circular DNA genomes. *J. Virol*., *94*, e00582.
49. Lv, J., Yu, K., Wei, J., Gui, H., Liu, C., Liang, D., … Kelliher, T. (2020). Generation of paternal haploids in wheat by genome editing of the centromeric histone CENH3. *Nat. Biotechnol*., *38*, 1397–1401. https://doi.org/10.1038/s41587-020-0728-4.
50. Nielsen, A. A. K., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E. A., … Voigt, C. A. (2016). Genetic circuit design automation. *Science*, *352*, aac7341.
51. HamediRad, M., Chao, R., Weisberg, S., Lian, J., Sinha, S., & Zhao, H. (2019). Towards a fully automated algorithm driven platform for biosystems design. *Nat. Commun*., *10*, 5150.
52. Radivojević, T., Costello, Z., Workman, K., & Martin, H. G. (2020). A machine learning Automated Recommendation Tool for synthetic biology. *Nat. Commun*., *11*, 4879.
53. Bartley, B. A., Beal, J., Karr, J. R., & Strychalski, E. A. (2020). Organizing genome engineering for the gigabase scale. *Nat. Commun*., *11*, 689.
54. Yang, X., Medford, J. I., Markel, K., Shih, P. M., De Paoli, H. C., Trinh, C. T., … Tuskan, G. A. (2020). Plant biosystems design research roadmap 1.0. *BioDesign Res*, *2020*, 8051764.
55. Burge, C., Campbell, A. M., & Karlin, S. (1992). Over- and underrepresentation of short oligonucleotides in DNA sequences. *Proc. Natl. Acad. Sci*, *89*, 1358.
56. Karlin, S., & Burge, C. (1995). Dinucleotide relative abundance extremes: A genomic signature. *Trends Genet*., *11*, 283.
57. Cheng, X., Virk, N., Chen, W., Ji, S., Ji, S., Sun, Y., & Wu, X. (2013). CpG usage in RNA viruses: Data and hypotheses. *PLoS One*, *8*, e74109.
58. Kunec, D., & Osterrieder, N. (2016). Codon pair bias is a direct consequence of dinucleotide bias. *Cell Rep*., *14*, 55.
59. Di Giallonardo, F., Schlub, T. E., Shi, M., & Holmes, E. C. (2017). Dinucleotide composition in animal RNA viruses is shaped more by virus family than by host species. *J. Virol*., *91*, e02381.
60. van Dolleweerd, C. J., Kessans, S. A., Van de Bittner, K. C., Bustamante, L. Y., Bundela, R., Scott, B., … Parker, E. J. (2018). MIDAS: A modular DNA assembly system for synthetic biology. *ACS Synth. Biol*, *7*, 1018.
61. Denloye, A. O., Homer, R. B., & Hull, R. (1978). Circular dichroism studies on turnip rosette virus. *J. Gen. Virol*., *41*, 77.
62. Pasin, F., Shan, H., García, B., Müller, M., San León, D., Ludman, M., … García, J. A. (2020). Abscisic acid connects phytohormone signaling with RNA metabolic pathways and promotes an antiviral response that is evaded by a self-controlled RNA virus. *Plant Commun*, *1*, 100099.
63. Čermák, T., Curtin, S. J., Gil-Humanes, J., Čegan, R., Kono, T. J. Y., Konečná, E., … Voytas, D. F. (2017). A multipurpose toolkit to enable advanced genome engineering in plants. *Plant Cell*, *29*, 1196.
64. Andreou, A. I., & Nakayama, N. (2018). Mobius assembly: A versatile Golden-Gate framework towards universal DNA assembly. *PLoS One*, *13*, e0189892.
65. Hussey, S. G., Grima-Pettenati, J., Myburg, A. A., Mizrachi, E., Brady, S. M., Yoshikuni, Y., & Deutsch, S. (2019). A standardized synthetic *Eucalyptus* transcription factor and promoter panel for re-engineering secondary cell wall regulation in biomass and bioenergy crops. *ACS Synth. Biol*, *8*, 463.
66. Pollak, B., Cerda, A., Delmans, M., Álamos, S., Moyano, T., West, A., … Haseloff, J. (2019). Loop assembly: A simple and open system for recursive fabrication of DNA circuits. *New Phytol*, *222*, 628.
67. Pollak, B., Matute, T., Nuñez, I., Cerda, A., Lopez, C., Vargas, V., … Federici, F. (2020). Universal loop assembly: Open, efficient and cross-kingdom DNA fabrication. *Synth. Biol*., *5*, ysaa001.
68. Moore, S. J., Lai, H.-E., Kelwick, R. J. R., Chee, S. M., Bell, D. J., Polizzi, K. M., & Freemont, P. S. (2016). EcoFlex: A multifunctional MoClo Kit for *E. coli* synthetic biology. *ACS Synth. Biol*., *5*, 1059.
69. Martella, A., Matjusaitis, M., Auxillos, J., Pollard, S. M., & Cai, Y. (2017). EMMA: An extensible mammalian modular assembly toolkit for the rapid design and production of diverse expression vectors. *ACS Synth. Biol*., *6*, 1380.

70. Vazquez-Vilar, M., Gandía, M., García-Carpintero, V., Marqués, E., Sarrion-Perdigones, A., Yenush, L., … Orzaez, D. (2020). Multigene engineering by GoldenBraid cloning: From plants to filamentous fungi and beyond. *Curr. Protoc. Mol. Biol.*, *130*, e116.
71. Dusek, J., Plchova, H., Cerovska, N., Poborilova, Z., Navratil, O., Kratochvilova, K., … Moravec, T. (2020). Extended set of GoldenBraid compatible vectors for fast assembly of multigenic constructs and their use to create geminiviral expression vectors. *Front. Plant Sci.*, *11*, 522059.
72. Sarrion-Perdigones, A., Vazquez-Vilar, M., Palací, J., Castelijns, B., Forment, J., Ziarsolo, P., … Orzaez, D. (2013). GoldenBraid 2.0: A comprehensive DNA assembly framework for plant synthetic biology. *Plant Physiol.*, *162*, 1618.
73. Engler, C., Kandzia, R., & Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PLoS One*, *3*, e3647.
74. Agmon, N., Mitchell, L. A., Cai, Y., Ikushima, S., Chuang, J., Zheng, A., … Boeke, J. D. (2015). Yeast Golden Gate (yGG) for the efficient assembly of *S. cerevisiae* transcription units. *ACS Synth. Biol.*, *4*, 853.
75. Engler, C., Youles, M., Gruetzner, R., Ehnert, T.-M., Werner, S., Jones, J. D., … Marillonnet, S. (2014). A Golden Gate modular cloning toolbox for plants. *ACS Synth. Biol*, *3*(11), 839–843.
76. Vasudevan, R., Gale, G. A. R., Schiavon, A. A., Puzorjov, A., Malin, J., Gillespie, M. D., … McCormick, A. J. (2019). CyanoGate: A modular cloning suite for engineering cyanobacteria based on the plant MoClo syntax. *Plant Physiol.*, *180*, 39.
77. Chiasson, D., Giménez-Oya, V., Bircheneder, M., Bachmaier, S., Studtrucker, T., Ryan, J., … Parniske, M. (2019). A unified multikingdom Golden Gate cloning platform. *Sci. Rep.*, *9*, 10131.
78. Occhialini, A., Piatek, A. A., Pfotenhauer, A. C., Frazier, T. P., Stewart, C. N., & Lenaghan, S. C. (2019). MoChlo: A versatile, modular cloning toolbox for chloroplast biotechnology. *Plant Physiol.*, *179*, 943.
79. Roberts, R. J., Vincze, T., Posfai, J., & Macelis, D. (2015). REBASE—a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Res.*, *43*, D298.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Pasin, F. (2021). Oligonucleotide abundance biases aid design of a type IIS synthetic genomics framework with plant virome capacity. *Biotechnol. J*, *16*:e2000354. https://doi.org/10.1002/biot.202000354