Local GC content biases in nucleotide sequences allow rational design of SynViP, a synthetic genomics framework with plant virome capacity

Fabio Pasin [1,2,†,*]

[1] Institute of Molecular and Cellular Biology of Plants (CSIC - UPV), Valencia, Spain

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

[†] Former address: Agricultural Biotechnology Research Center, Academia Sinica, Taipei, Taiwan

* Correspondence: <f.pasin@csic.es>; ORCID ID: 0000-0002-9620-4301

ABSTRACT

Novel methodological concepts for scalable engineering of biosystems are necessary to meet the ever-increasing human needs. Humans directly exploit wild and domesticated plant species for food, energy, biomaterial and biomolecule production. Engineering of novel, useful traits in crops is critical for sustained food security; traditional breeding approaches are nonetheless time and cost consuming.[1,2]

Plant viruses have been used in multiple plant biotechnology and synthetic biology applications as varied as biopharmaceutical production, flowering induction and accelerated plant breeding, transient crop reprogramming, or heritable mutagenesis of crop genomes.[3–5] *Agrobacterium*-mediated inoculation (agro-infection) is the most universal and efficient way to deliver plant viruses and their derived expression vectors to plants. Agro-infection requires T-DNA binary vectors harboring viral genomic sequences whose assembly can pose technical challenges since no standardized cloning frameworks are available.

Biological sequences including those of viruses are known to have non-random compositional tendencies, such as biases in nucleotide contents, dinucleotide frequency, and codon usage.[6–11]

Here, I hypothesized that identification of nucleotide compositional biases in biological sequences could be exploited to guide the rational design of enhanced synthetic genomics systems. Viruses are simple biological entities and, to test my hypothesis at a multispecies level, a reference plant virome that includes 2044 accessions grouped into 34 taxonomic families was used. Global sequence analyses showed a virome guanine + cytosine content (%GC) of 42%; it ranged from 31.48% to 58.22%, per taxonomic family. To identify local compositional biases the virome sequences were searched with complete non-redundant oligonucleotide sets that were generated by random union of 3 to 8 mononucleotides. The oligonucleotide sets were classified by size and %GC, and their abundance was computed. Results showed that, within size classes, abundance of an oligonucleotide in the plant virome is significantly, negatively correlated with its %GC ($-0.948 < r < -0.997$, per size class). The identified local %GC biases could not be anticipated by the global %GC results.

The virome compositional biases were next used to guide the development of SynViP, a synthetic genomics framework with plant virome capacity that combines Golden Gate and No See'm principles with low-cost DNA chemical synthesis. SynViP relies on Type IIS restriction enzymes for unidirectional, scar-free assembly and the pLX series of mini T-DNA binary vectors,[12] which have been used for agro-infection of RNA and DNA viruses to plants.[13,14] SynViP does not require subcloning steps, can be used with linear and circular DNA molecules, is insensitive to fragment terminal sequences, and was used

successfully to rescue a genuine plant virus based on a digital sequence and with no biological template requirements.[15]

SynViP is predicted to greatly facilitate assembly and plant delivery of chemically-synthesized virus components and genomes, and to allow high-throughput biological characterization of plant viruses as well as engineering and prototyping of next-generation viral vectors for crop reprogramming.

Finally, my results show a complementarity of local and global analyses in detecting non-random compositional tendencies of biological sequences. The computational identification of local and global compositional biases in multispecies datasets might become a common step to guide scalable, portable design and engineering of biosystems.

FUNDING

F.P. is supported by a Juan de la Cierva Incorporación contract [grant IJC2019-039970-I] from Ministerio de Ciencia e Innovación (Spain), and was the recipient of a post-doctoral fellowship from Academia Sinica (Taiwan).

REFERENCES

(1) Yang, X.; Medford, J. I.; Markel, K.; Shih, P. M.; De Paoli, H. C.; Trinh, C. T.; McCormick, A. J.; Ployet, R.; Hussey, S. G.; Myburg, A. A.; Jensen, P. E.; Hassan, M. M.; Zhang, J.; Muchero, W.; Kalluri, U. C.; Yin, H.; Zhuo, R.; Abraham, P. E.; Chen, J.-G.; Weston, D. J.; Yang, Y.; Liu, D.; Li, Y.; Labbe, J.; Yang, B.; Lee, J. H.; Cottingham, R. W.; Martin, S.; Lu, M.; Tschaplinski, T. J.; Yuan, G.; Lu, H.; Ranjan, P.; Mitchell, J. C.; Wullschleger, S. D.; Tuskan, G. A. *BioDesign Res.* **2020**, *2020*, 8051764. https://doi.org/10.34133/2020/8051764.

(2) Steinwand, M. A.; Ronald, P. C. *Nat. Food* **2020**, *1* (5), 273–283. https://doi.org/10.1038/s43016-020-0072-3.

(3) Pasin, F.; Menzel, W.; Daròs, J.-A. *Plant Biotechnol. J.* **2019**, *17* (6), 1010–1026. https://doi.org/10.1111/pbi.13084.

(4) Torti, S.; Schlesier, R.; Thümmler, A.; Bartels, D.; Römer, P.; Koch, B.; Werner, S.; Panwar, V.; Kanyuka, K.; Wirén, N. von; Jones, J. D. G.; Hause, G.; Giritch, A.; Gleba, Y. *Nat. Plants* **2021**, *7* (2), 159–171. https://doi.org/10.1038/s41477-021-00851-y.

(5) Khakhar, A.; Voytas, D. F. *Front. Plant Sci.* **2021**, *12*, 668580. https://doi.org/10.3389/fpls.2021.668580.

(6) Burge, C.; Campbell, A. M.; Karlin, S. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (4), 1358–1362. https://doi.org/10.1073/pnas.89.4.1358.

(7) Karlin, S.; Burge, C. *Trends Genet.* **1995**, *11* (7), 283–290. https://doi.org/10.1016/s0168-9525(00)89076-9.

(8) Cheng, X.; Virk, N.; Chen, W.; Ji, S.; Ji, S.; Sun, Y.; Wu, X. *PLoS ONE* **2013**, *8* (9), e74109. https://doi.org/10.1371/journal.pone.0074109.

(9) Kunec, D.; Osterrieder, N. *Cell Rep.* **2016**, *14* (1), 55–67. https://doi.org/10.1016/j.celrep.2015.12.011.

(10) Di Giallonardo, F.; Schlub, T. E.; Shi, M.; Holmes, E. C. *J. Virol.* **2017**, *91* (8), e02381-16, e02381-16. https://doi.org/10.1128/JVI.02381-16.

(11) González de Prádena, A.; Sánchez Jimenez, A.; San León, D.; Simmonds, P.; García, J. A.; Valli, A. A. *mBio* **2020**, *11* (1), e02818-19. https://doi.org/10.1128/mBio.02818-19.

(12) Pasin, F.; Bedoya, L. C.; Bernabé-Orts, J. M.; Gallo, A.; Simón-Mateo, C.; Orzaez, D.; García, J. A. *ACS Synth. Biol.* **2017**, *6* (10), 1962–1968. https://doi.org/10.1021/acssynbio.6b00354.

(13) Pasin, F.; Tseng, X.-A.; Bedoya, L. C.; Heydarnejad, J.; Deng, T.-C.; García, J. A.; Chen, Y.-R. *J. Virol. Methods* **2018**, *262*, 48–55. https://doi.org/10.1016/j.jviromet.2018.09.007.

(14) Zhao, M.; García, B.; Gallo, A.; Tzanetakis, I. E.; Simón-Mateo, C.; García, J. A.; Pasin, F. *Phytopathol. Res.* **2020**, *2* (1), 36. https://doi.org/10.1186/s42483-020-00077-4.

(15) Pasin, F. *Biotechnol. J.* **2021**, *16* (5), e2000354. https://doi.org/10.1002/biot.202000354.

Correspondence: f.pasin@csic.es