



Research article

Mathematical evaluation of the formation and evolution of hierarchical structures of repetitive DNA sequences

Ziyan Hao¹, Xiaochang Xu² and Xiyin Wang^{1,2,3,*}

¹ College of Mathematics and Science, North China University of Science and Technology, Tangshan, China

² School of Life Science, North China University of Science and Technology, Tangshan, China

³ School of Public Health and Protective Medicine, North China University of Science and Technology, Tangshan, China

* **Correspondence:** Email: wangxiyin@vip.sina.com; Tel: +13603381029.

Supplementary

Table S1. Euclidean distances of mononucleotide composition vectors between hierarchical repeat units under different mutation types and mutation rates. Values are reported as mean $\pm$ standard deviation ($n = 30$ independent simulation runs).

Mutation rate		$r_1 = 5\%, r_2 = 5\%$			$r_1 = 15\%, r_2 = 15\%$		
Point mutation	Inter-order comparison	Z-F	Z-S	F-S	Z-F	Z-S	F-S
	Euclidean distance ($\times 10^{-4}$)	0.71 ± 0.06	1.58 ± 0.06	0.78 ± 0.06	2.61 ± 0.07	3.88 ± 0.12	2.87 ± 0.09
Nucleotide insertion	Inter-order r comparison	Z-F	Z-S	F-S	Z-F	Z-S	F-S
	Euclidean distance ($\times 10^{-4}$)	1.12 ± 0.07	0.99 ± 0.15	0.23 ± 0.03	1.91 ± 0.15	2.27 ± 0.12	0.71 ± 0.09
Nucleotide deletion	Inter-order r comparison	Z-F	Z-S	F-S	Z-F	Z-S	F-S
	Euclidean distance ($\times 10^{-4}$)	1.19 ± 0.07	0.86 ± 0.12	0.24 ± 0.03	0.76 ± 0.15	0.71 ± 0.12	0.32 ± 0.09

Note: Z: zero-order repeat unit, F: first-order repeat unit and S: second-order repeat unit.

Table S2. Euclidean distances of dinucleotide composition vectors between hierarchical repeat units under different mutation types and mutation rates. **Values** are reported as mean $\pm$ standard deviation ($n = 30$ independent simulation runs).

	Mutation rate	$r_1 = 5\%, r_2 = 5\%$			$r_1 = 15\%, r_2 = 15\%$		
	Inter-order comparison	Z-F	Z-S	F-S	Z-F	Z-S	F-S
Point mutation	Euclidean distance ($\times 10^{-4}$)	1.89 ± 0.08	2.66 ± 0.11	1.57 ± 0.09	4.12 ± 0.21	5.23 ± 0.26	2.93 ± 0.17
Nucleotide insertion	Inter-order r comparison	Z-F	Z-S	F-S	Z-F	Z-S	F-S
	Euclidean distance ($\times 10^{-4}$)	1.53 ± 0.17	2.81 ± 0.19	0.56 ± 0.15	2.91 ± 0.23	3.41 ± 0.28	1.72 ± 0.22
Nucleotide deletion	Inter-order r comparison	Z-F	Z-S	F-S	Z-F	Z-S	F-S
	Euclidean distance ($\times 10^{-4}$)	1.71 ± 0.16	1.58 ± 0.19	0.47 ± 0.11	0.94 ± 0.19	1.20 ± 0.25	0.63 ± 0.12

Note: Z: zero-order repeat unit, F: first-order repeat unit and S: second-order repeat unit.

Table S3. Euclidean distances of trinucleotide composition vectors between hierarchical repeat units under different mutation types and mutation rates. Values are reported as mean $\pm$ standard deviation ($n = 30$ independent simulation runs).

Point mutation	Mutation rate	$r_1 = 5\%, r_2 = 5\%$			$r_1 = 15\%, r_2 = 15\%$		
	Inter-order comparison	Z-F	Z-S	F-S	Z-F	Z-S	F-S
	Euclidean distance ($\times 10^{-4}$)	2.21 ± 0.16	3.19 ± 0.19	2.42 ± 0.23	6.71 ± 0.23	6.60 ± 0.34	4.04 ± 0.22
Nucleotide insertion	Inter-order r comparison	Z-F	Z-S	F-S	Z-F	Z-S	F-S
	Euclidean distance ($\times 10^{-4}$)	2.36 ± 0.23	2.15 ± 0.24	0.72 ± 0.19	2.52 ± 0.33	3.97 ± 0.38	1.38 ± 0.27
Nucleotide deletion	Inter-order r comparison	Z-F	Z-S	F-S	Z-F	Z-S	F-S
	Euclidean distance ($\times 10^{-4}$)	1.19 ± 0.23	1.51 ± 0.28	0.35 ± 0.14	0.99 ± 0.29	1.33 ± 0.35	0.44 ± 0.18

Note: Z: zero-order repeat unit, F: first-order repeat unit, and S: second-order repeat unit.

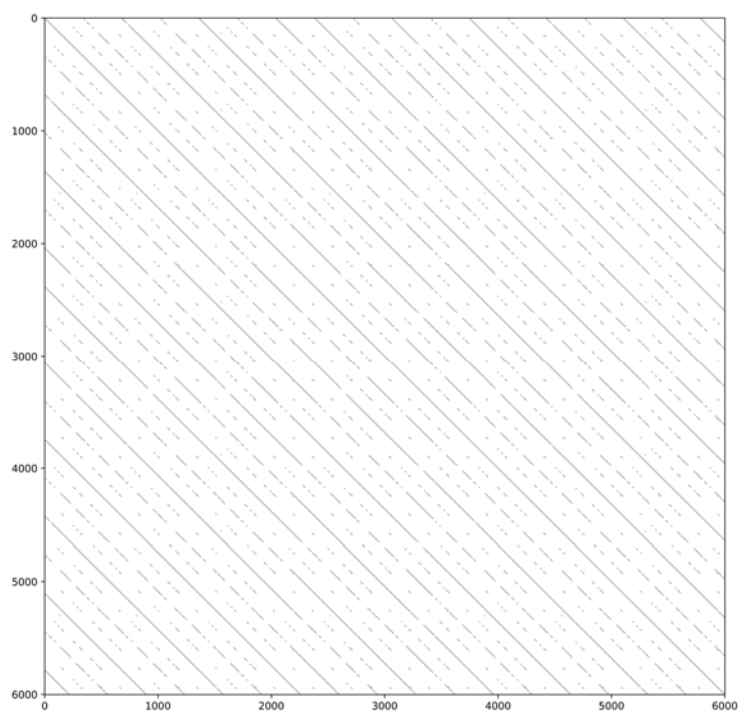


Figure S1. DNA similarity of representative segmental sequences in rice.

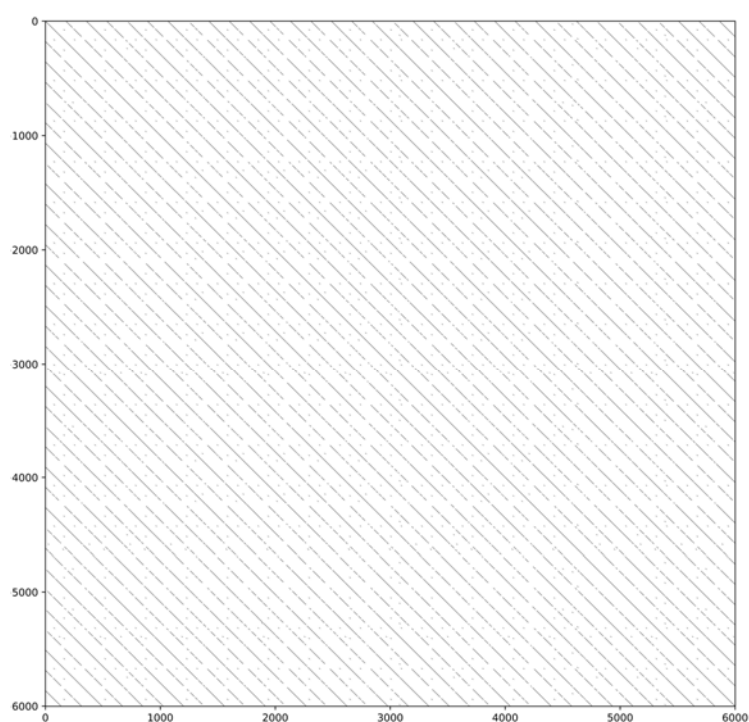


Figure S2. DNA similarity of representative segmental sequences in *Arabidopsis thaliana*.