

Weidong Bao<sup>1</sup>, Acacia Alcivar-Warren<sup>2,3</sup>

1. Genetics Information Research Institute (GIRI), 20380 Town Center Lane, Suite 240, Cupertino, CA 95014, USA [weidong@girinst.org](mailto:weidong@girinst.org)

2. FUCOBI Foundation, Quito, Ecuador, [www.fucobi.org](http://www.fucobi.org)

3. Environmental Genomics Inc., P. O. Box 196, Southborough, MA 01772, USA [environmentalgenomics.warren@gmail.com](mailto:environmentalgenomics.warren@gmail.com)

Table 1. Repetitive component in the genome of *Crassostrea gigas* isolate BHY1A (GCA\_005518195.2)

Class & Superfamily	Family No. (autonomous)	Copy No.	Genome (% 587Mb)
<b>DNA Transposon</b>			
Mariner/Tc1	103 (55)	15,724	2.00
Helitron	159 (23)	32,020	14.29
Crypton	97 (35)	20,138	3.16
Kolobok	69 (40)	9,583	2.21
ISL2EU	69 (48)	5,765	1.48
hAT	54 (16)	5,871	0.85
Academ	36 (22)	2,383	
Harbinger	39 (22)	1,987	
P	31 (13)	1,213	
Zator	23 (7)	1,163	
IS3EU	18 (6)	1,034	
EnSpm/CACTA	20 (8)	379	
MuDR	20 (13)	319	
piggyBac	12 (7)	146	
Polinton	12 (11)	47	
Sola	21 (18)	151	
Ginger2/TDD	6 (5)	408	
Dada	2 (2)	9	
Merlin	2 (2)	8	
Unclassified	321	60,028	11.70
<b>Subtotal</b>	<b>1115</b>	<b>158,408</b>	<b>40.78</b>
<b>LTR retrotransposon</b>			
Gypsy	434	8,787	2.70
BEL	100	1,319	
Copia	6	68	
Unclassified	12	1,206	
DIRS	83	750	
<b>Subtotal</b>	<b>635</b>	<b>12,131</b>	<b>4.33</b>
<b>Non-LTR retrotransposon</b>			
Tx1	66	1,129	
RTEX	60	998	
CR1	31	462	
RTE	9	90	
L2	6	56	
Proto2	3	32	
R2	1	4	
Penelope	21	831	
SINE/tRNA	4	1,145	
<b>Subtotal</b>	<b>201</b>	<b>4,750</b>	<b>3.41</b>
<b>Unclassified repeats</b>	<b>48</b>	<b>3,284</b>	<b>0.81</b>
<b>Tandem &amp; Simple repeat</b>			<b>1.72 *</b>
<b>Multicopy genes</b>			<b>0.57</b>
<b>Total</b>			<b>51.6</b>

\* Tandem repeats, detected by Tandem Repeat Finder, account for 6.3% of the genome: 1.72% outside of TEs, and 4.6% within TEs.

## Results

Pacific oyster genome is highly enriched in different types of repeats. Consensus or representative sequences of 1999 families of transposons, 48 types of unclassified elements, and several other of satellite sequences and multicopy genes (Table 1) were reconstructed and manually curated from the two genome assemblies of *Crassostrea gigas*: isolate BHY1A (GCA\_005518195.2, 587Mb) and strain 05x7-T-G4-1.051#20 (GCA\_000297895.2, 564 Mb).

At least 51.6% of the BHY1A genome are repetitive sequence (Table 1), exceeding the previous figure (48.32% [1]). ~80% repetitive sequence is consist of a large variety of DNA transposons. Helitron is the major type of DNA transposon (27.7% of the total repeats). LTR retrotransposons and Non-LTR retrotransposons account for only 8.4% and 6.6%, respectively.

A large portion of the transposons are highly active or in the near past (Figure 1), which may be the major cause of the highly polymorphism of oyster genomes.

~73% of tandem repeats and/or simple repeats are located within various DNA transposons (Table 1)

Except for a few families (Table 2), there is no major differences, with regard to the types of repeats and their relative abundance, between the two oyster genomes. Nevertheless, a slightly more TEs were detected in the larger BHY1A genome (Figure 1).

The predominance of highly active DNA transposons inhibit, to some extent, the activity of retrotransposons by frequent insertion into the latter sequence, and can frequently generate haphazard genomic duplication, estimated to account for another 1-2% of the genome.

Table 2. A few families contrast in copy numbers in oyster BHY1A (GCA\_005518195.2) and 05x7-T-G4-1.051#20 (GCA\_000297895.2)

Family	BHY1A	05x7-T-G4-1.051#20
Academ-6_CGi	14	2
AcademH-11_CGi	10	1
CryptonV-4_CGi	0	1
DNA-30_CGi	58	4
BEL-1_CGi-I	9	2

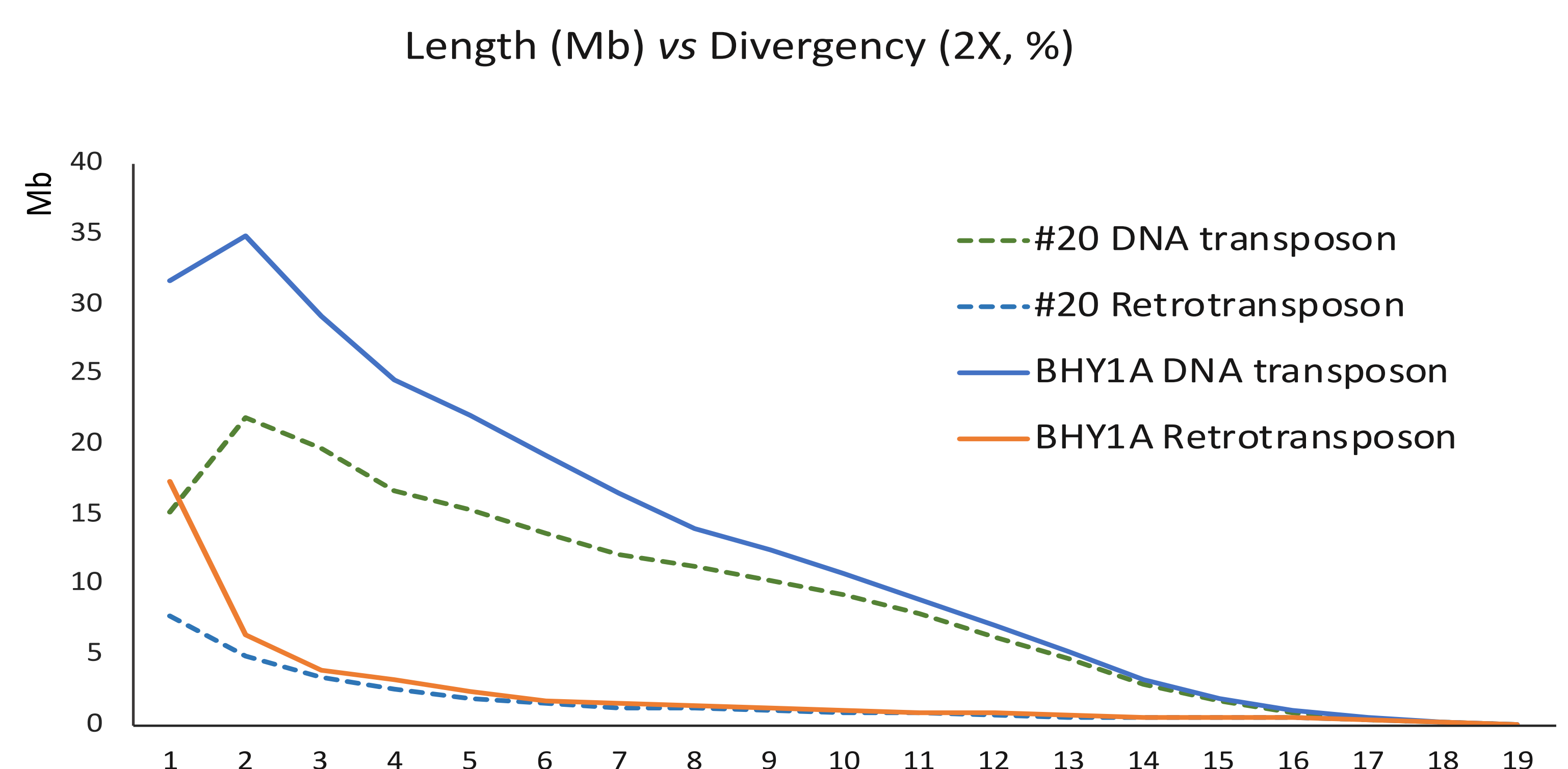


Figure 1. Cumulative lengths of DNA transposons or RNA retrotransposons fragment vs the sequence divergence to the respective family consensus.