

Supplementary Table 9. Statistic of repeat sequence in *C. goeringii*.

	RepBase TEs		TE Proteins		<i>De novo</i>		Combined TEs	
	Length (bp)	%in Genome	Length (bp)	% in Genome	Length (bp)	% in Genome	Length (bp)	% in Genome
DNA	116,219,426	2.83427	12,083,795	0.29469	180,406,646	4.399618	289,365,130	7.056813
LINE	284,219,896	6.931335	133,072,674	3.245274	488,630,560	11.916345	670,468,422	16.35087
SINE	134,833	0.003288	0	0	450,639	0.01099	585,472	0.014278
LTR	290,467,898	7.083707	332,400,204	8.10632	2,479,874,487	60.477264	2,543,168,258	62.02082
Other	37,638	0.000918	0	0	0	0	37,638	0.000918
Unknown	0	0	0	0	20,087,787	0.489885	20,087,787	0.489885
Total	615,367,667	15.00711	477,510,519	11.64515	2,981,789,187	72.717573	3,107,397,891	75.78082

*LINE, long interspersed nuclear element; SINE, short interspersed element; LTR, long terminal repeat;

*Denovo+Repbase denotes transposable elements identified by RepeatMasker (<http://www.repeatmasker.org>) with default options after

RepeatModeler/RepeatScout/Piler/LTR_finder software use with RepBase database prediction.

*TE proteins were transposable elements identified in the genome through the annotation of Repeat ProteinMask software using the RepBase database.

*Combined TEs involved a combination of the above two methods.

*Unknown repeat sequences could not be clustered by Repeat Masker.