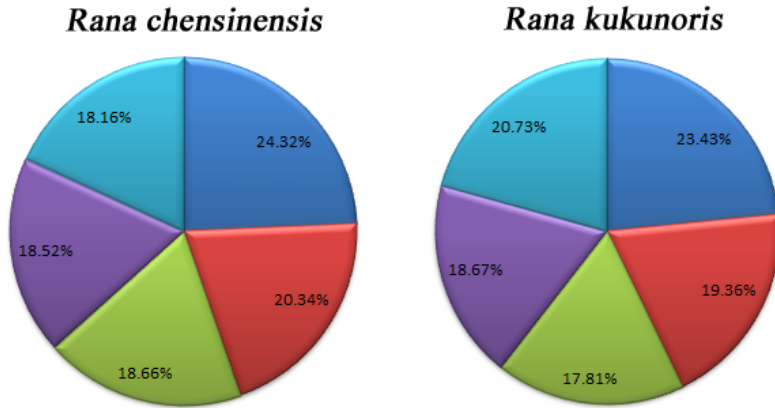


**Additional file 1. Characteristics of gene annotation of assembled transcripts against the reference dataset. A.** E-value distribution of blastx hits for each transcript with a cut-off E-value of 1E-5. **B.** Similarity distribution of blastx hits for each transcripts.

**(A) E-value distribution**

- 1 1E-5 to 1E-20
- 2 1E-20 to 1E-50
- 3 1E-50 to 1E-100
- 4 1E-100 to 0
- 5 0



**(B) Similarity distribution**

- 1 <40%
- 2 40% to 60%
- 3 60% to 80%
- 4 80% to 95%
- 5 95% to 100%

