**Supplementary table 2:** A summary of the assembly post-processing stages for generating the reference transcriptome assembly of *Striga hermonthica*.

|  |  |
| --- | --- |
| **Post Assembly analysisa** | **Number of Transcriptsb** |
| Merged Trinity primary transcripts from all *S. hermonthica* developmental stages (tissue-specific). | 2,583,765 |
| Merged transcripts with predicted coding regions (PlantTribes *AssemblyPostProcesso*r pipeline). | 575,971 |
| Generated non-redundant transcripts with predicted coding regions (PlantTribes *AssemblyPostProcessor* pipeline). | 316,095 |
| Cleaned transcripts determined by excluding non-embryophyte contaminant transcripts after searching against the NCBI non-redundant protein databases, followed by taxonomic assignment using the NCBI taxonomy databases (custom scripts). | 274,008 |
| Created gene expression clusters using tissue-specific RNA-Seq data. Excludes all functionally unannotated transcripts that have no expression across all tissue-specific RNA-Seq libraries (AHRD, InterProScan, and Corset pipelines). | 194,969 |
| Selected reference transcripts from expression clusters as described by Yang et al., 2014 (Corset expression clustering and PlantTribes GeneFamilyClassifier pipelines). | 59,243 |

aIndividual stages of the post-assembly analysis

bTranscript counts for each stage