Comparing methods of RNAseq analysis for species without a reference genome

Lilian Padilha, Eveline Teixeira Caixeta, Felipe Rodrigues da Silva
Embrapa Café, Embrapa Café, Embrapa Informática Agropecuária

New sequencing technologies brought deep probing of transcriptomes (so-called RNAseq) to the reach of individual researches. Analysis of RNAseq sequences, however, depend on the alignment of reads to a reference genome. Some approaches have been proposed to allow de novo assembly of the transcripts, therefore allowing RNAseq to be used on organisms lacking a reference genome. The efficiency of those approaches, however, were tested only with diploid species. Here we propose a methodology to evaluate the results of de novo assembly of allotetraploid Coffea arabica RNAseq reads obtained from libraries of coffee leaves infected by Hemileia vastatrix. Trinity was able to assemble longer transcripts when compared to ABySS, SOAPdenovo and Oases. Moreover, the transcripts assembled by Trinity where more similar to the ones obtained using Cufflinks after aligning the RNAseq to 10 Coffea sp. publically available BAC sequences. Comparison of the most abundant transcripts with several Coffea sp. single copy described genes, though, shows that all assemblers failed to perfectly reconstruct the transcripts. Supported by: CNPq, Finep, INCT-Café, CBP&D Café
Keywords: RNAseq, transcriptome, Coffea, EST
Concentration area: Transcrip Proteomic