Multicolored De Bruijn graph visualization of first, second, and third generation RNA-seq offers deeper insight into transcript isoforms in Picea abies

Warren W. Kretzschmar[1], Shirin Akhter[2], Veronika Nordal[2], Nicolas Delhomme[2], Nathaniel R. Street[2], Ove Nilsson[3], Olof Emanuelsson[1], Jens F. Sundström[2]

KTH Royal Institute of Technology, Stockholm, Sweden[1], Swedish University of Agricultural Sciences, Sweden[2], Umeå University, Umeå, Sweden[3]

Third-generation sequencing technologies offer cheap, long reads, but with higher error rates than reads from first or second generation technologies. In transcriptome assembly, these technologies are complementary. However, useful approaches to combining data from all these technologies are needed. De Bruijn graphs, a staple of transcriptome assemblers, offer a natural way of combining low-error reads into a graph. Multicolored De Bruijn graphs allow the joint representation of read information from multiple sequencing technologies.

We present a visualization of Picea abies transcript isoforms assembled from Sanger, Illumina, and Pac-Bio IsoSeq reads based on a multicolored De Bruijn graph. This visualization allows close assessment of read support for transcript isoforms from all three technologies simultaneously.