**Contig validation using exon re-alignment and order checking.**

The idea is to compare the assembled contig sets by comparing their exon content with the reference transcriptome. Two elements are verified, the exon presence and order.

The reference transcriptome is made of all the expressed transcripts (above nil expression level).

Exon presence is checked by aligning the exons of the reference transcripts on the contig set using BLAST (filtered by 80% identity and 80% of coverage). Using the number of hits, each transcript is linked to the corresponding contig: the contig with the highest number of corresponding exons is retained for each transcript. Then the order of the exons on the contig is verified and only the exons of contigs having all their exons respecting the initial order are counted. Gaps in the exon order are not taken into account. Several transcripts and genes can be linked to the same contig. The final figures are calculated for all the genes. The diagram hereunder presents assembly examples with only one gene and two transcripts which have been assembled in two contigs.

Supplemental Fig. S4. Steps of the evaluation process and final marks

The diagram at the top of the figure presents the references gene structure with its exon order. The figure above shows the two corresponding transcript structures. The contig and alignment section present the contigs linked to the transcripts. The first one comprises an exon (exon 5) which does not align on the reference and the second includes a translocation of exon 3 and 4 and the same mis-alignment of exon 5. The table at the bottom presents how the contigs are linked to the transcripts by counting the number of aligned exon (the link is presented in the selected column) and then the mark given to each contig taking into account the correctness of the exon order in the contig.

The final marks of the assembly will be first the number of correctly reconstructed exons divide by the number of exons of the reference transcripts and second number of correctly ordered and reconstructed exons divide by the number of exons of the reference transcripts.