Additional data file

On our website at [43] we make available the 208 highest-scoring candidates from miRseeker analysis of the sequenced Drosophilid genomes. Here we provide a detailed description of the miRseeker output data available at [43], including the folded structures and evaluation of the top 208 miRseeker candidates:

The left column displays the rank position (given in parentheses) and information on its homologs in other species. miR-2b-1 and miR-100 scored very highly but failed conservation filters; they have been included in the list according to their score but are not assigned an overall rank. The middle column displays information on its score, physical location on the chromosome, and alignment with *D. pseudoobscura*; the latter two are linked to Drosophila genome browser plots and VISTA plots, respectively. The 'super-region' notation is an active link that will generate a report on all folds of all regions within the selected super-region. Color-coding for the left and middle columns is as follows: green—reference set miRNAs; blue—novel miRNAs experimentally verified by Northern blot; orange—miRNA candidates with genuine Anopheles homologs whose expression was not verified; grey-contains a nucleotide repeat (and is thus a poor candidate). The right column shows the region sequence that was selected as the highest-scoring fold for that super-region, the alignment and folded structures of homologous Drosophilid sequences. The *Dm/Dp* alignment is color-coded as follows: blue—loop, red—helical arms, yellow—potential miRNA encoding sequences (defined as >22 nt long blocks of perfectly conserved sequence located <10 nt from the end of either side of the terminal loop). Where relevant, equivalent analysis of the top scoring hit in Anopheles is shown. Note that only a subset of Anopheles hits actually represent homologous miRNA sequence and that many sequences produce no BLAST hit. Candidates that passed all conservation filters are colored green (thus miR-2b-1 and miR-100 are white). The last three candidates on the listed (miR-125, miR-287 and miR-87) scored poorly, but are included here so that their structures may be viewable.