B3.1.9—Sequence bias

Spo11Mapper includes a utility script (*SeqBias.pl*) capable of determining per base (A/G/C/T) frequencies flanking a given set of coordinates. *SeqBias* directly samples a user provided FASTA reference genome file (e.g. Cer3H4L2) to pileup and calculate per base frequencies (A/G/C/T) for a given ±bp width, centred on Watson (+) and Crick (-) coordinates listed within a 1bp histogram file generated by *Spo11Mapper*:

perl SeqBias.pl –i <Histogram File> –r <Reference FASTA> –w <Width> –m <Mode> –o <Output File>

For Watson (+) hits, *SeqBias* samples the reference FASTA "as is", without further processing. As FASTA files specify the (+) strand, *SeqBias* reverse complements sequences flanking Crick (-) hits. *SeqBias* provides two analysis modes: (i) *Pos*—under this mode, *SeqBias* calculates biases flanking all specified nucleotides in an unweighted manner (ii) *Freq*—under this mode, *SeqBias* weights biases by the number of hits present at any given nucleotide. Sites with stronger signal will therefore be more heavily represented within the resulting bias. All biases produced throughout this chapter were generated under *Pos* mode. A population averaged bias is calculated proceeding 1bp histogram processing, and provided in a tab delimited file.