Hi Hongxu,

Here are the commands I use with Bowtie2, in order to make it easier when I show you what to do this afternoon.

To build the index (only need to do this once):

bowtie2-build Sbicolor\_255\_v2.0.hardmasked.fa Sbicolor\_2-0

To align tags:

bowtie2 -f -x Sbicolor\_2-0 -U HapMap.fas.txt -S output.sam -D 20 -R 3 -N 1 -L 18 -i S,1,0.50 --local

Here is part of an R script that I used for extracting the alignment information from the SAM file:

#### marker alignment

# read sam file generated by Bowtie, aligning HapMap fasta to Sorghum

mySam <- readLines("HapMap11-4\_Alignments141230.sam")

mySam <- mySam[-(1:1571)] # get rid of header rows

# find which rows have successful alignments

aligned <- c(grep("Chr",mySam), grep("super",mySam))

length(aligned)  # 34372

mySam[aligned[1:20]]

alignedC <- grep("Chr",mySam)

length(alignedC) # 34320

# convert into table

temp <- strsplit(mySam[alignedC], "\t")

postable <- data.frame(marker=sapply(temp,

                           function(x) strsplit(x[1], "\_")[[1]][1]),

                       allele=sapply(temp,

                           function(x) strsplit(x[1], "\_")[[1]][2]),

                       reference=sapply(temp, function(x) x[3]),

                       position=as.integer(sapply(temp, function(x) x[4])))

dim(postable)

postable[1:20,]

# get markers for which both alleles align, and to the same position

markers <- unique(postable$marker)

toKeep <- rep(FALSE, length(markers))

names(toKeep) <- markers

for(m in markers){

    if(!m %in% locToKeep) next # removes paralogs from DH lines

    myRows <- postable$marker==m

    if(sum(myRows)!=2) next

    if(!identical(as.factor(c("query","hit")), postable$allele[myRows])) next

    myChr <- postable$reference[myRows]

    if(myChr[1] != myChr[2]) next

    myPos <- postable$position[myRows]

    if(myPos[1] == myPos[2]) toKeep[m] <- TRUE

}

sum(toKeep) # 13823 markers

markersToKeep <- as.character(markers[toKeep])

chrToKeep <- postable$reference[match(markersToKeep, postable$marker)]

posToKeep <- postable$position[match(markersToKeep, postable$marker)]