

Ancestry Transitions - Adult Hybrids

steph

2024-06-05

```
library(tidyverse)
library(ggpubr)
library(patchwork)
library(propagate)
library(lme4)
theme_set(theme_classic())

tracts<-read_csv("C:/Users/Steph/GitHub_data/bgchm_clines/tracts_ThrushHybridsAdults.240222.csv")

#actual LD-pruned AIMs states for autosomes
aims<-read_table("C:/Users/Steph/GitHub_data/thrushHybrids.f50.240205.LD.states")
#list of all autosomal AIMs
aims2<-read_table("C:/Users/Steph/GitHub_data/hmm_aims_thrushHybrids",
                  col_names=c("scaffold","position","ref","alt"))

scaf_order_raw<-read.csv("C:/Users/Steph/Downloads/scaffold_order.csv")
scaf_order<-scaf_order_raw%>%
  mutate(scaffold=gsub("_","-",Sequence.Name),
        chromosome=paste("chromosome",sprintf("%02d",as.numeric(Assigned.Molecule)),
                          sep="_"))%>%
  mutate(chromosome=if_else(Assigned.Molecule%in%c("Z","W"),
                            paste("chromosome",Assigned.Molecule,sep="_"),chromosome))%>%
  dplyr::select(scaffold,chromosome)

meta_aims<-read.csv("C:/Users/Steph/GitHub_data/AIMs_metaData/AIMs_metadata_240227.csv")
clines<-read.csv("C:/Users/Steph/GitHub_data/bgchm_clines/bgchm_clines_overlappingOutliers.csv")

pop.cols<-c("#202D26","#49654A","#849D74","#ADB593")
hybCol="#849D74"

meta_aims<-meta_aims%>%filter(reference%in%tracts$reference)

aims<-aims%>%rename(scaffold=CHROM)%>%left_join(scaf_order)%>%
  mutate(ChromPos=paste(chromosome,POS,sep="_"))%>%
  dplyr::select(-c(POS,REF,ALT,scaffold))%>%filter(!is.na(chromosome))%>%
  pivot_longer(-c(ChromPos,chromosome),names_to="reference",values_to="genotype")

aims2<-aims2%>%
  left_join(scaf_order)%>%
  mutate(ChromPos=paste(chromosome,sprintf("%010d",position),sep="_"))%>%
```

```

  mutate(ChromPos100kb=paste(chromosome,sprintf("%010d",plyr::round_any(position,100000)),sep="_"))

tracts<-tracts%>%mutate(scaffold=CHROM)%>%
  left_join(scaf_order)%>%
  mutate(ChromPos=paste(chromosome,sprintf("%010d",start),sep="_"))%>%
  mutate(ChromPos100kb=paste(chromosome,sprintf("%010d",plyr::round_any(start,100000)),sep="_"))%>%
  left_join(meta_aims%>%dplyr::select(reference,release_site))%>%
  rename(population=release_site)%>%
  filter(chromosome%in%c("chromosome_01","chromosome_02",
                          "chromosome_03","chromosome_04",
                          "chromosome_05","chromosome_06",
                          "chromosome_07"))

tractSummary100kb<-tracts%>%group_by(ChromPos100kb,chromosome,population)%>%
  summarise(meanLength=mean(tractLength)/100000,
           startCount=n(),
           meanAncestry=mean(ancestry))%>%ungroup()%>%
  left_join(aims2%>%group_by(ChromPos100kb)%>%summarise(aimsCount=n()))%>%
  left_join(meta_aims%>%rename(population=release_site))%>%
    group_by(population)%>%
    summarise(sampleSize=n())%>%
  mutate(startsPerBird=startCount/sampleSize,
         populationNum=as.numeric(as.factor(as.character(population))))%>%
  mutate(population=factor(population,
                           levels=c("Alaska","Pemberton","Hope","Washington")))%>%
  filter(aimsCount<1600) #remove outlier window

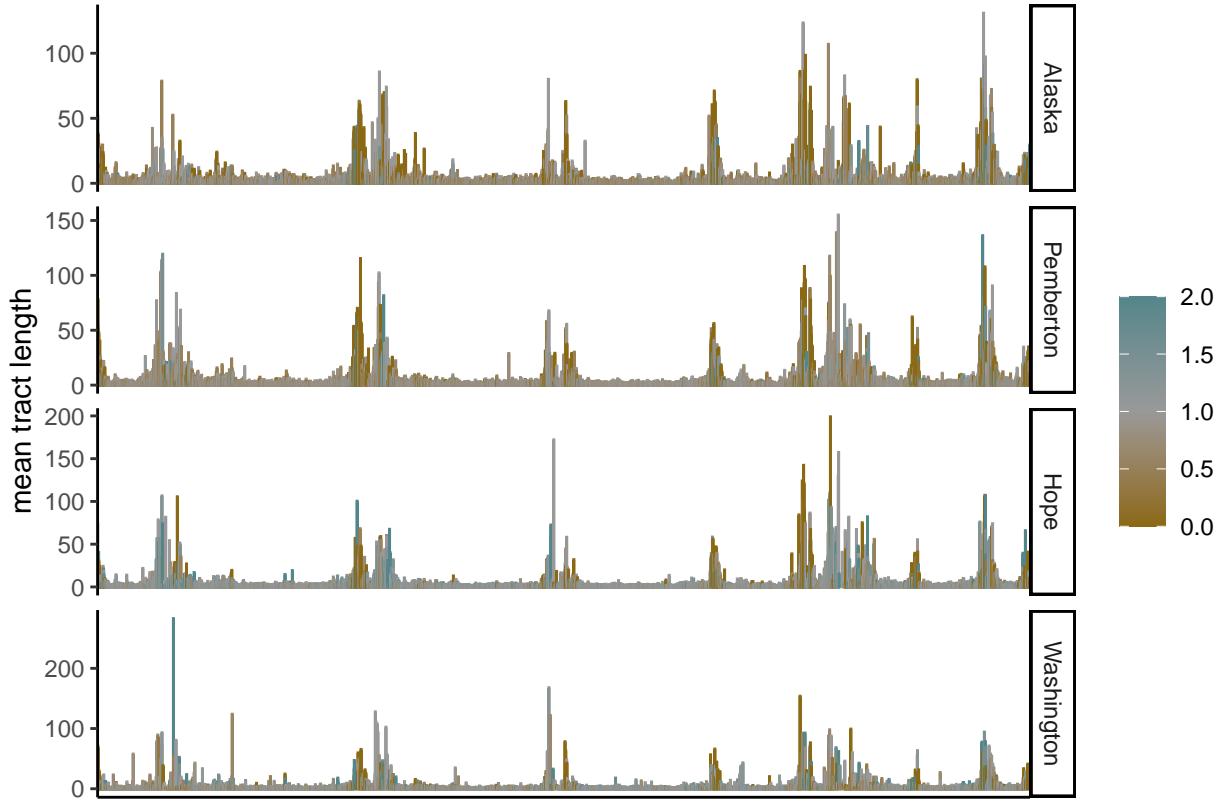
```

Mean lengths and ancestries of tracts starting in 100kb windows

```

ggplot(tractSummary100kb,
       aes(x=ChromPos100kb,y=meanLength,colour=meanAncestry))++
  geom_bar(stat='identity')++
  scale_colour_gradient2(high="cadetblue4",low="goldenrod4",mid="grey60",
                         name=element_blank(),midpoint=1)++
  theme(axis.text.x=element_blank(),axis.ticks.x=element_blank())+_
  ylab("mean tract length")+xlab("")+_
  facet_grid(rows=vars(population),scales='free_y')

```



Histogram of ancestry starts per population

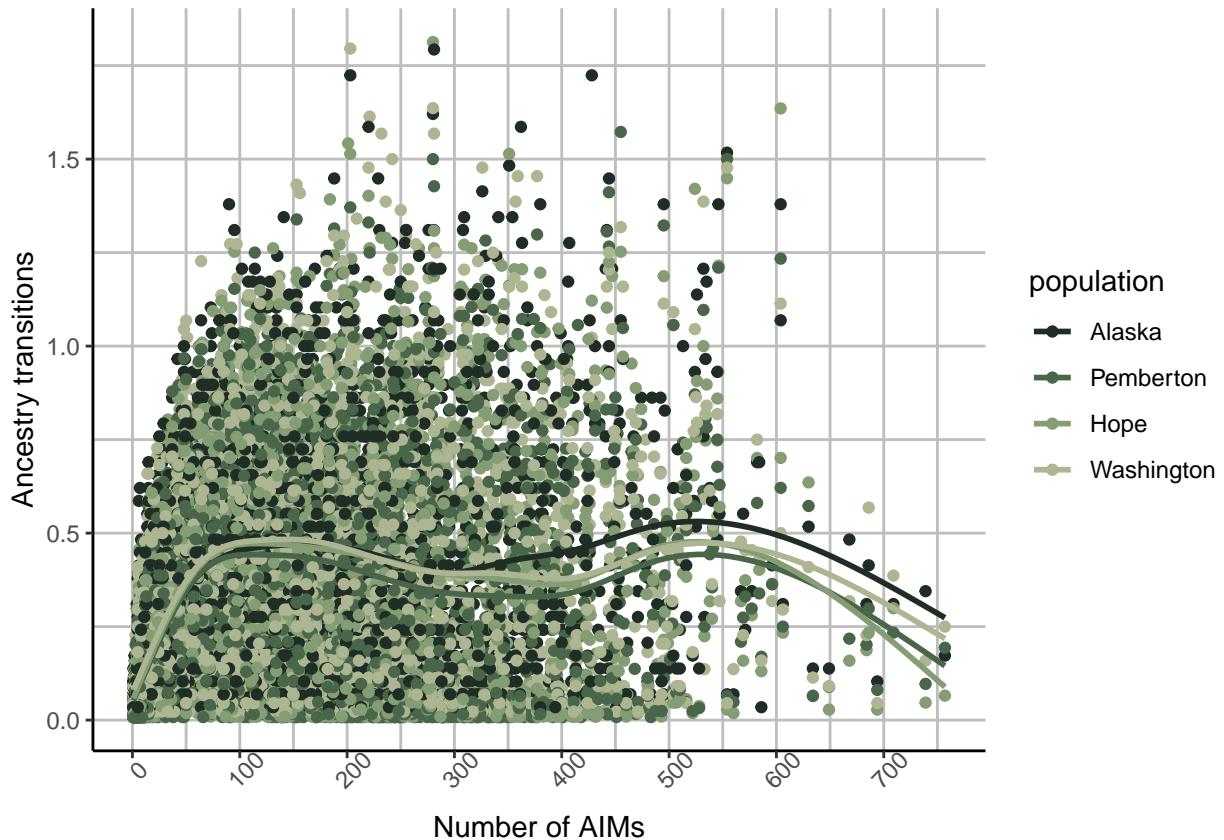
```
tracts<-tracts%>%
  mutate(population=factor(population,levels=c("Alaska","Pemberton","Hope","Washington")))

gg1<-ggplot(tracts%>%mutate(chromosome=gsub("chromosome_","",chromosome)),
  aes(x=start,fill=population))+
  geom_histogram(binwidth=100000)+
  theme_classic()+
  scale_fill_manual(values=pop.cols,name="Population")+
  theme(axis.text.x=element_blank(),axis.ticks.x=element_blank())+
  ylab("Tract starts")+xlab("Genomic position")+
  facet_grid(cols=vars(chromosome),scale="free",space="free")+
  theme(strip.background = element_rect(fill="white"))
ggsave("C:/Users/Steph/GitHub/thrush_hybrids/ancestry_tracts/figures/ancestryTracts.png",gg1,
  width=8,height=2)
```

Number of transitions depends, non-linearly, on the number of windows

```
ggplot(tractSummary100kb,aes(x=aimsCount,y=startsPerBird,colour=population))+
  geom_point()+
  scale_colour_manual(values=pop.cols)+
  geom_smooth(se=F,method='gam')+
  theme(panel.grid.major=element_line(colour = "grey"),
    panel.grid.minor=element_line(colour = "grey"),
    axis.text.x=element_text(angle=45))+
```

```
scale_x_continuous(breaks=seq(0,max(tractSummary100kb$aimsCount),100))+  
xlab("Number of AIMs")+ylab("Ancestry transitions")
```



Fit asymptotic model for relationship between # AIMs and # transitions

Some skew with higher fitted values, but not too bad

```
nlsOut<-read.csv("C:/Users/Steph/GitHub_data/ancestry_tracts/ancestryTransitions.nlsCI.20240605.csv")  
  
tractSummary100kb<-tractSummary100kb%>%  
  left_join(nlsOut%>%dplyr::select(population,ChromPos100kb,nlsPredict,CILwr,CIupr,  
    nlsFitted,nlsResid))%>%  
  mutate(outlierLwr;if_else(startsPerBird<CILwr,1,0))
```

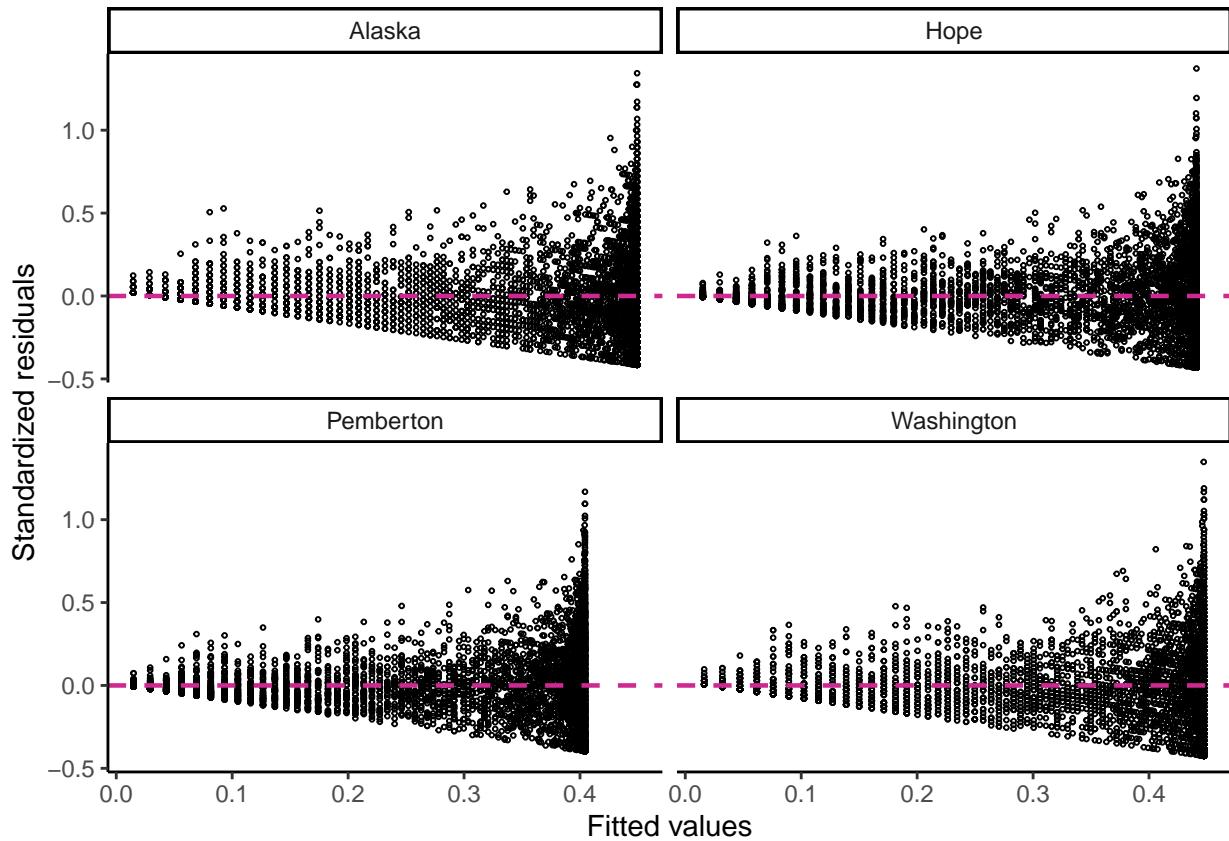
Visualize model fits for each population

```
gg1<-ggplot(tractSummary100kb,aes(x=nlsFitted,y=nlsResid))+  
  geom_point(pch=1,size=0.5)+  
  geom_hline(yintercept = 0,colour="maroon3",linetype=2,size=0.8)+  
  xlab("Fitted values")+ylab("Standardized residuals")  
  facet_wrap(vars(population))
```

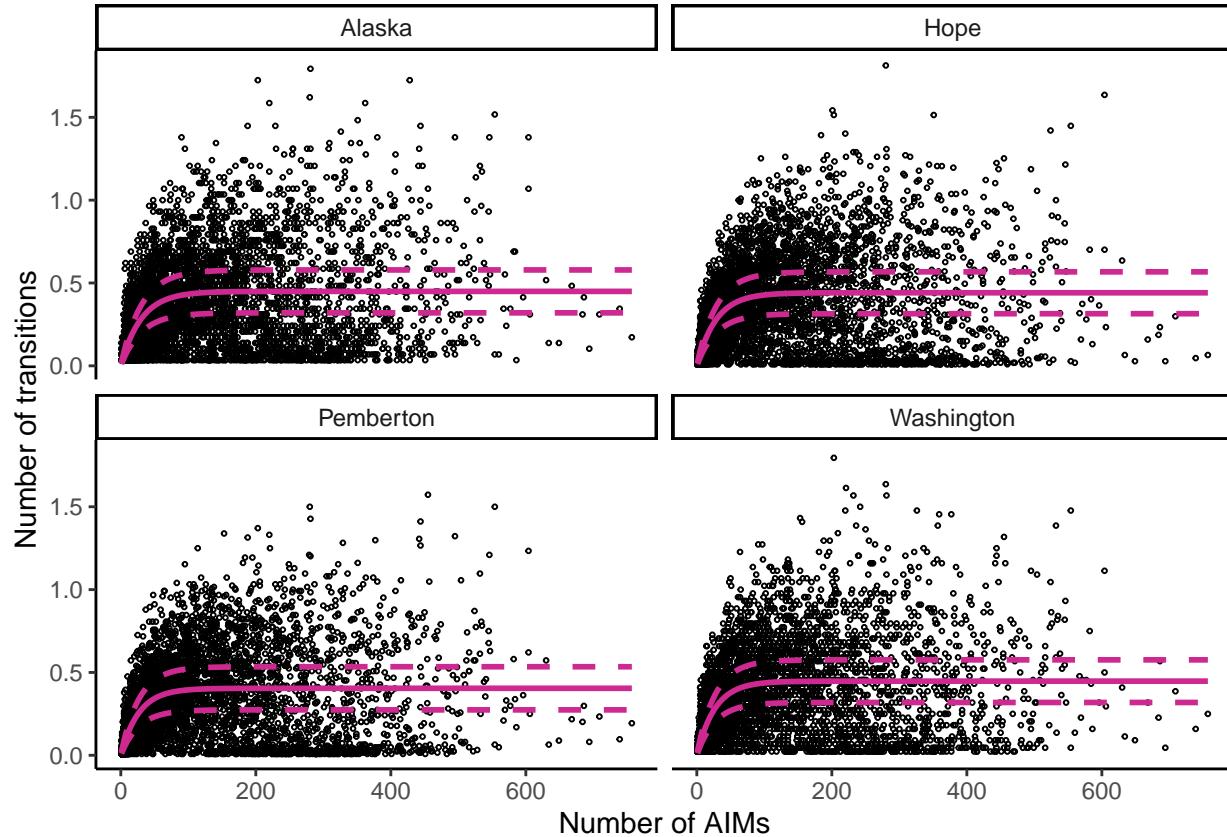
```
gg2<-ggplot(tractSummary100kb,aes(x=aimsCount,y=startsPerBird))+  
  geom_point(pch=1,size=0.5)+
```

```
geom_line(aes(x=aimsCount,y=nlsPredict), colour="maroon3",size=1)+  
geom_line(aes(x=aimsCount,y=CIlwr), colour="maroon3",linetype=2,size=1)+  
geom_line(aes(x=aimsCount,y=CIupr), colour="maroon3",linetype=2,size=1)+  
xlab("Number of AIMs")+ylab("Number of transitions")+  
facet_wrap(vars(population))
```

gg1



gg2



```
ggsave("C:/Users/Steph/GitHub/thrush_hybrids/ancestry_tracts/figures/model_fit.png",gg2,
       width=6,height=4,units="in")
```

Compare observed and expected overlap between populations

```
tractSummary100kb %>% filter(outlierLwr==1) %>% pull(ChromPos100kb) %>% unique() %>% length()

## [1] 2791

nLoci=tractSummary100kb %>% pull(ChromPos100kb) %>% unique() %>% length()
popCombos<-t(combn(sort(unique(tractSummary100kb$population)),2))
outOverlap<-data.frame()

for(i in 1:nrow(popCombos)){
  pop1<-popCombos[i,1]
  pop2<-popCombos[i,2]

  out1<-tractSummary100kb %>% filter(outlierLwr&population==pop1) %>% pull(ChromPos100kb)
  out2<-tractSummary100kb %>% filter(outlierLwr&population==pop2) %>% pull(ChromPos100kb)

  overlap1<-sum(out1%in%out2)
  expected1<-length(out1)*length(out2)/nLoci
```

```

outOverlap<-rbind(outOverlap,
                   data.frame(populations=paste(pop1,pop2,sep=" x "),
                               sharedOutliers=overlap1,expectedSharedOutliers=expected1))

}

popCombos3<-t(combn(sort(unique(tractSummary100kb$population)),3))

for(i in 1:nrow(popCombos3)){
  pop1<-popCombos3[i,1]
  pop2<-popCombos3[i,2]
  pop3<-popCombos3[i,3]

  out1<-tractSummary100kb%>%filter(outlierLwr&population==pop1)%>%pull(ChromPos100kb)
  out2<-tractSummary100kb%>%filter(outlierLwr&population==pop2)%>%pull(ChromPos100kb)
  out3<-tractSummary100kb%>%filter(outlierLwr&population==pop3)%>%pull(ChromPos100kb)

  overlap1<-length(intersect(intersect(out1,out2),out3))
  #three way overlap: expected for 2 pops/nLoci * outliers in third pop
  #(5*4*3)/(2^2)==(5*4/2)*3/2
  expected1<-(length(out1)*length(out2)/nLoci)*length(out3)/nLoci

  outOverlap<-rbind(outOverlap,
                     data.frame(populations=paste(pop1,pop2,pop3,sep=" x "),
                               sharedOutliers=overlap1,expectedSharedOutliers=expected1))
}

pop1="Alaska"; pop2="Pemberton"; pop3="Hope"; pop4="Washington"
out1<-tractSummary100kb%>%filter(outlierLwr&population==pop1)%>%pull(ChromPos100kb)
out2<-tractSummary100kb%>%filter(outlierLwr&population==pop2)%>%pull(ChromPos100kb)
out3<-tractSummary100kb%>%filter(outlierLwr&population==pop3)%>%pull(ChromPos100kb)
out4<-tractSummary100kb%>%filter(outlierLwr&population==pop4)%>%pull(ChromPos100kb)

overlap1<-length(intersect(intersect(out1,out2),out3),out4))
#three way overlap: expected for 2 pops/nLoci * outliers in third pop
expected1<-(length(out1)/nLoci)*(length(out2)/nLoci)*(length(out3)/nLoci)*length(out4)

outOverlap<-rbind(outOverlap,data.frame(populations=paste(pop1,pop2,pop3,pop4,sep=" x "),
                                         sharedOutliers=overlap1,expectedSharedOutliers=expected1))

outOverlap<-outOverlap%>%
  mutate(ratio=round(sharedOutliers/expectedSharedOutliers,2))%>%
  mutate(expectedSharedOutliers=round(expectedSharedOutliers,2))%>%
  rename("observed overlap"="sharedOutliers",
         "expected overlap"="expectedSharedOutliers")

outOverlap

##                                     populations observed overlap expected overlap
## 1                           Alaska x Hope           1170      635.59
## 2                           Alaska x Pemberton        1198      624.83
## 3                           Alaska x Washington       1094      583.87
## 4                           Hope x Pemberton        1373      662.07

```

```

## 5          Hope x Washington      1244    618.67
## 6          Pemberton x Washington 1213    608.20
## 7          Alaska x Hope x Pemberton 998    229.82
## 8          Alaska x Hope x Washington 905    214.76
## 9          Alaska x Pemberton x Washington 912    211.12
## 10         Hope x Pemberton x Washington 1035   223.71
## 11         Alaska x Pemberton x Hope x Washington 808    77.65
## ratio
## 1  1.84
## 2  1.92
## 3  1.87
## 4  2.07
## 5  2.01
## 6  1.99
## 7  4.34
## 8  4.21
## 9  4.32
## 10 4.63
## 11 10.41

```

```

# write.csv(outOverlap,
#           file="C:/Users/Steph/GitHub/thrush_hybrids/ancestry_tracts/outlier_overlap.csv",
#           row.names=F)

```

```

clineOutlierSummary<-read.csv(file="C:/Users/Steph/GitHub/thrush_hybrids/genomic_clines/clines_outlier_
tractOutlierSummary<-tractSummary100kb%>%filter(outlierLwr>0)%>%
  group_by(ChromPos100kb)%>%
  summarise(tractOutlierCount=n())%>%
  rename(ChromPos=ChromPos100kb)

tractOutlierSummary%>%filter(tractOutlierCount==4)%>%
  mutate(chromosome=substr(ChromPos,1,13))%>%group_by(chromosome)%>%
  summarise(n())

```

```

## # A tibble: 7 x 2
##   chromosome  `n()`
##   <chr>        <int>
## 1 chromosome_01    297
## 2 chromosome_02    140
## 3 chromosome_03     91
## 4 chromosome_04     54
## 5 chromosome_05    164
## 6 chromosome_06     24
## 7 chromosome_07     38

```

```
#write.csv(tractOutlierSummary,file="C:/Users/Steph/GitHub/thrush_hybrids/ancestry_tracts/transitions_o
```

```

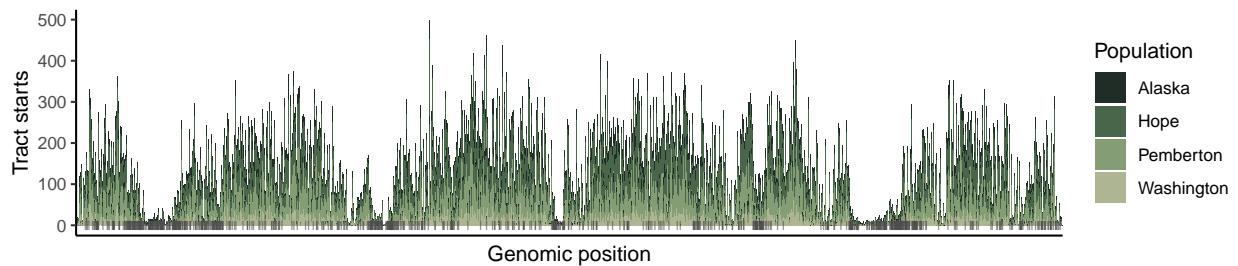
transitionOutliers<-tractSummary100kb%>%filter(outlierLwr==1)%>%
  group_by(ChromPos100kb)%>%
  summarise(countOutlier=n())%>%
  filter(countOutlier==4)%>%
  mutate(yvar=0)

```

```

ggplot(tractSummary100kb,
       aes(x=ChromPos100kb,y=startCount,fill=population))++
  geom_bar(stat='identity')++
  geom_point(data=transitionOutliers,
             mapping=aes(x=ChromPos100kb,y=yvar),
             fill="grey30",colour="grey30",pch="|")+
  scale_fill_manual(values=pop.cols,name="Population")+
  theme(axis.text.x=element_blank(),axis.ticks.x=element_blank())+
  ylab("Tract starts")+xlab("Genomic position")+
  theme(strip.background = element_rect(fill="white"))

```



Look at overlap with interesting loci

SNPs associated with migratory phenotypes

```

gwasSNPs1<-read.csv("C:/Users/Steph/GitHub_data/ancestry_tracts/metaUSAT_SNPs_steph.csv")
gwasSNPs2<-read.csv("C:/Users/Steph/GitHub_data/ancestry_tracts/all_traits_SNPs_plog6.csv")
gwasSNPs<-rbind(gwasSNPs1,gwasSNPs2); rm(gwasSNPs1,gwasSNPs2)

gwasSNPs<-gwasSNPs%>%
  mutate(scaffold=gsub("_","-",chr))%>%
  left_join(scaf_order)%>%
  mutate(pos=as.numeric(as.character(pos)))%>%
  mutate(ChromPos=paste(chromosome,sprintf("%010d",pos),sep="_"))%>%
  mutate(ChromPos100kb=paste(chromosome,sprintf("%010d",plyr::round_any(pos,100000)),sep="_"))%>%
  dplyr::select(chromosome,ChromPos,ChromPos100kb,phenotype)%>%
  filter(!is.na(chromosome))

gwasTransitionsOutliers<-gwasSNPs%>%filter(ChromPos100kb%in%transitionOutliers$ChromPos100kb)
gwasTransitionsOutliers%>%mutate(Chrom=substr(ChromPos100kb,1,13))%>%
  group_by(Chrom,phenotype)%>%
  summarise(count=n())%>%
  arrange(phenotype)

```

```

## # A tibble: 22 x 3
## # Groups:   Chrom [7]
##   Chrom phenotype      count
##   <chr>    <chr>      <int>
## 1 chromosome_02 adult        7
## 2 chromosome_03 adult        5
## 3 chromosome_01 fall_35_radio  2

```

```

##  4 chromosome_05 fall_35_radio      1
##  5 chromosome_05 fall_bear_radio1  35
##  6 chromosome_03 fall_dep          1
##  7 chromosome_02 fall_distance    16
##  8 chromosome_03 fall_distance    3
##  9 chromosome_05 fall_distance    3
## 10 chromosome_01 juvenile         1
## # i 12 more rows

nrow(gwasTransitionsOutliers)

## [1] 121

tractLoci4pops<-tractSummary100kb%>%group_by(ChromPos100kb)%>%
  summarise(count=n())%>%filter(count==4)

nrow(gwasSNPs%>%filter(ChromPos100kb%in%tractLoci4pops$ChromPos100kb))* 
  nrow(transitionOutliers)/nrow(tractLoci4pops)

## [1] 133.6202

```

Misexpressed genes

```

misexpress<-rbind(readxl::read_excel("C:/Users/Steph/GitHub_data/ancestry_tracts/Misexpression_tables.xlsx",
                                         sheet=2)%>%dplyr::select(-11)%>%mutate(Brain="CNH1"),
                    readxl::read_excel("C:/Users/Steph/GitHub_data/ancestry_tracts/Misexpression_tables.xlsx",
                                         sheet=3)%>%dplyr::select(-11)%>%mutate(Brain="CNM1"),
                    readxl::read_excel("C:/Users/Steph/GitHub_data/ancestry_tracts/Misexpression_tables.xlsx",
                                         sheet=4)%>%dplyr::select(-11)%>%mutate(Brain="HC1"),
                    readxl::read_excel("C:/Users/Steph/GitHub_data/ancestry_tracts/Misexpression_tables.xlsx",
                                         sheet=5)%>%dplyr::select(-11)%>%mutate(Brain="HC2"),
                    readxl::read_excel("C:/Users/Steph/GitHub_data/ancestry_tracts/Misexpression_tables.xlsx",
                                         sheet=6)%>%dplyr::select(-11)%>%mutate(Brain="HT1"))

misexpress<-misexpress%>%filter(Spring.Misexpressed==1|Winter.Misexpressed==1)%>%
  rename(RefSeq.Accn=Chr)%>%
  left_join(scaf_order_raw%>%dplyr::select(Sequence.Name,RefSeq.Accn))%>%
  rename(scaffold=Sequence.Name)%>%
  dplyr::select(scaffold,Str,End,Gene,Summer.Misexpressed,Winter.Misexpressed)%>%
  filter(!is.na(scaffold))%>%
  pivot_longer(cols=c(Summer.Misexpressed,Winter.Misexpressed),names_to="Season")%>%
  mutate(Season=gsub(".Misexpressed","",Season))%>%
  filter(value==1)%>%dplyr::select(-value)%>%
  distinct()%>%
  mutate(scaffold=gsub("_","-",scaffold))%>%left_join(scaf_order)%>%
  mutate(ChromPos100kbStr=paste(chromosome,sprintf("%010d",plyr::round_any(Str,100000)),sep="_"))%>%
  mutate(ChromPos100kbEnd=paste(chromosome,sprintf("%010d",plyr::round_any(End,100000)),sep="_"))

misexpressTransitionOutliers<-misexpress%>%

```

```

filter(ChromPos100kbStr%in%transitionOutliers$ChromPos100kb |
      ChromPos100kbEnd%in%transitionOutliers$ChromPos100kb)

#unique genes
misexpressTransitionOutliers%>%dplyr::select(chromosome, Gene)%>%distinct()

## # A tibble: 118 x 2
##   chromosome     Gene
##   <chr>          <chr>
## 1 chromosome_01 ANKS6
## 2 chromosome_05 ARFIP1
## 3 chromosome_01 AVL9
## 4 chromosome_07 BIN1
## 5 chromosome_04 CBLL1
## 6 chromosome_05 CFAP97
## 7 chromosome_03 CLVS2
## 8 chromosome_03 CNRIP1
## 9 chromosome_01 CREB5
## 10 chromosome_07 DYTN
## # i 108 more rows

misexpressTransitionOutliers%>%dplyr::select(ChromPos100kbStr)%>%distinct()

## # A tibble: 108 x 1
##   ChromPos100kbStr
##   <chr>
## 1 chromosome_01_0096600000
## 2 chromosome_05_0029500000
## 3 chromosome_01_0047700000
## 4 chromosome_07_0024700000
## 5 chromosome_04_0014800000
## 6 chromosome_05_0042600000
## 7 chromosome_03_0065400000
## 8 chromosome_03_0009500000
## 9 chromosome_01_0035600000
## 10 chromosome_07_0012500000
## # i 98 more rows

misexpressTransitionOutliers%>%
  group_by(chromosome)%>%summarise(n())

## # A tibble: 7 x 2
##   chromosome   `n()`
##   <chr>        <int>
## 1 chromosome_01     52
## 2 chromosome_02     16
## 3 chromosome_03     22
## 4 chromosome_04      9
## 5 chromosome_05     31
## 6 chromosome_06      7
## 7 chromosome_07      7

```

```

misexpress %>% filter(ChromPos100kbStr %in% tractLoci4pops$ChromPos100kb) %>%
  pull(ChromPos100kbStr) %>% unique() %>% length() *
  nrow(transitionOutliers) / nrow(tractLoci4pops)

## [1] 89.86498

tractOutlierGenesSummary <- 
  rbind(misexpress %>% dplyr::rename(ChromPos=ChromPos100kbStr) %>%
    left_join(tractOutlierSummary) %>%
    filter(!is.na(tractOutlierCount)) %>%
    mutate(ld_dataset="misexpress") %>%
    dplyr::select(chromosome, ChromPos, tractOutlierCount, ld_dataset) %>%
    distinct(),
  
  gwasSNPs %>% dplyr::select(-ChromPos) %>%
    dplyr::rename(ChromPos=ChromPos100kb) %>%
    left_join(tractOutlierSummary) %>%
    filter(!is.na(tractOutlierCount)) %>%
    mutate(ld_dataset="gwas") %>%
    dplyr::select(chromosome, ChromPos, tractOutlierCount, ld_dataset) %>%
    mutate(chromosome=gsub("chromosome_","",chromosome))

figS3 <- ggplot(tractOutlierGenesSummary,
  aes(x=chromosome)) +
  geom_bar(aes(fill=as.character(tractOutlierCount))) +
  scale_fill_manual(values=c("grey70","lightpink2","lightpink3","lightpink4"),
                    name=element_blank()) +
  facet_wrap(vars(ld_dataset))

ggsave(figS3, filename="C:/Users/Steph/GitHub/thrush_hybrids/ancestry_tracts/figures/FigS3.png", width=8, height=6)

intersect(misexpressTransitionOutliers$ChromPos100kbStr, gwasTransitionsOutliers$ChromPos100kb)

## [1] "chromosome_05_0050000000" "chromosome_02_0031800000"
## [3] "chromosome_03_0073300000" "chromosome_02_0086000000"

```

Relationship to genome structure

```

rho <- read.csv("C:/Users/Steph/GitHub/thrush_hybrids/genomeArchitecture.100kb.20240808.csv")
rho <- rho %>%
  mutate(ChromPos100kb=paste(chromosome, sprintf("%010d", plyr::round_any(startpos, 100000)), sep="_")) %>%
  mutate(chromosomeType=case_when(substr(chromosome, 1, 12) == "chromosome_Z" ~ "sex",
                                 scaffoldLength_Mb < 20 ~ "micro",
                                 scaffoldLength_Mb > 40 ~ "macro"))

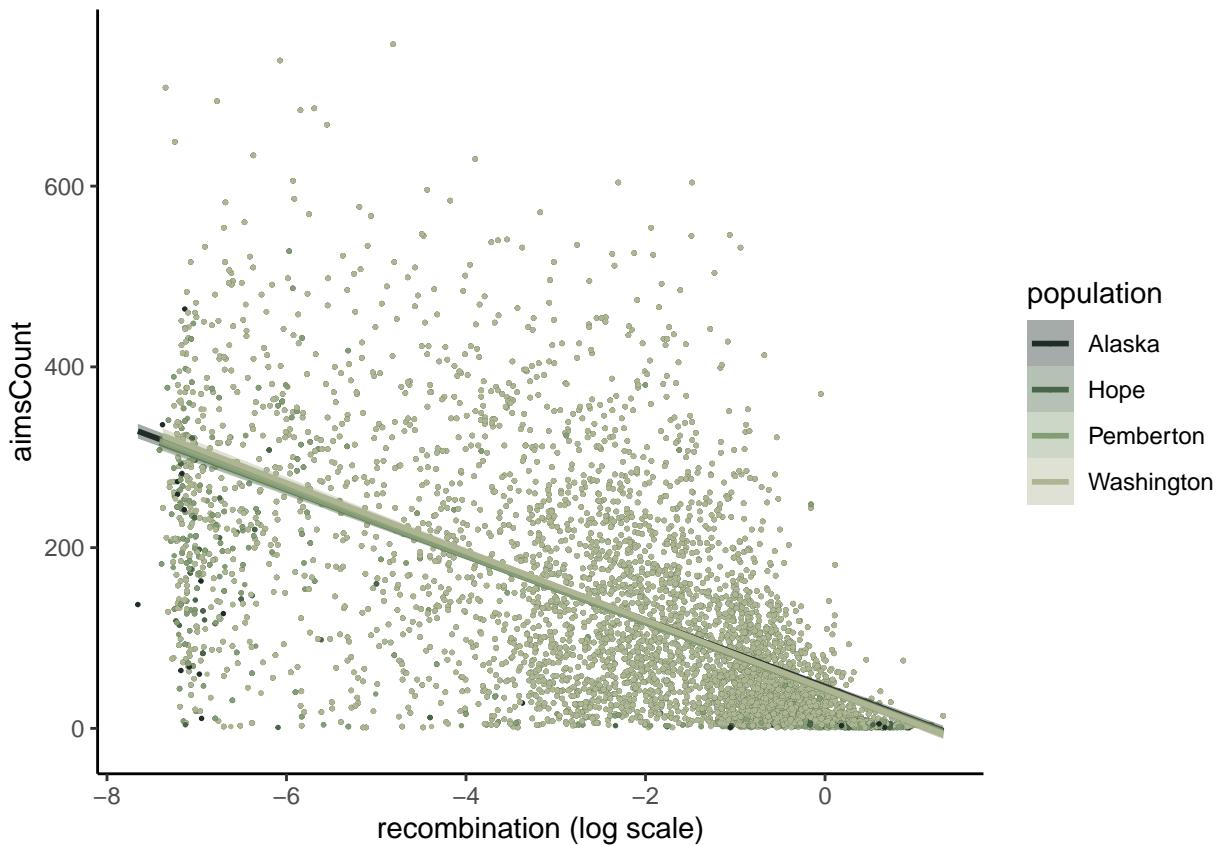
tractSummary100kb <- tractSummary100kb %>%
  left_join(rho %>%
    dplyr::select(ChromPos100kb, mean_cMMb, gene_density_100kb, scaffoldLength_Mb, Fst, chromosome))

# write.csv(tractSummary100kb,

```

```
#           file="C:/Users/Steph/GitHub/thrush_hybrids/genomicFeatures/tractsGenomeArchitecture.csv",
#           row.names=F)
```

```
ggplot(tractSummary100kb,
       aes(x=log(mean_cMMb), y=aimsCount, colour=population, fill=population)) +
  geom_point(size=0.3) +
  geom_smooth(method="lm") +
  scale_colour_manual(values=pop.cols) +
  scale_fill_manual(values=pop.cols) +
  xlab("recombination (log scale)")
```



```
tractSummary100kb$population<-factor(tractSummary100kb$population, levels=c("Alaska", "Pemberton", "Hope"))

p1<-ggplot(tractSummary100kb%>%filter(aimsCount>50&mean_cMMb<3),
            aes(x=log(mean_cMMb), y=startsPerBird, colour=population, fill=population)) +
  geom_point(size=0.3) +
  geom_smooth(method="lm") +
  scale_colour_manual(values=pop.cols) +
  scale_fill_manual(values=pop.cols) +
  xlab("Recombination (log scale)") +
  ylab("Ancestry transitions")
```

```
p2<-ggplot(tractSummary100kb%>%filter(aimsCount>50),
            aes(x=Fst, y=startsPerBird, colour=population, fill=population)) +
```

```

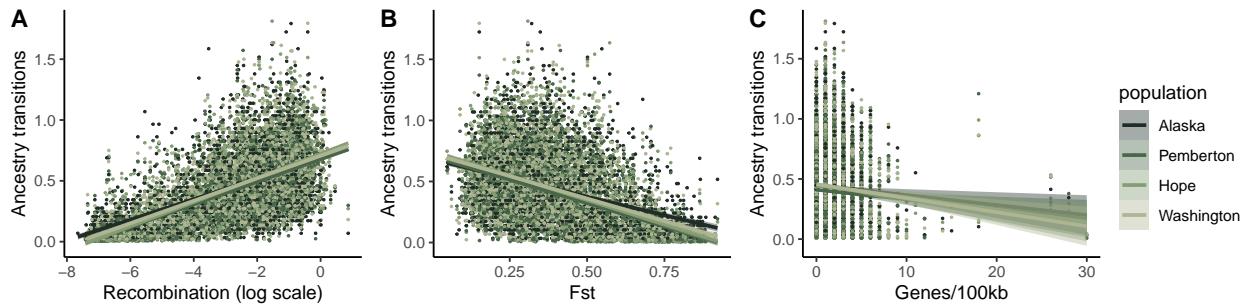
geom_point(size=0.3)+
geom_smooth(method="lm")+
scale_colour_manual(values=pop.cols)+
scale_fill_manual(values=pop.cols)+
xlab("Fst")+
ylab("Ancestry transitions")

p3<-ggplot(tractSummary100kb%>%filter(aimsCount>50),
            aes(x=gene_density_100kb,y=startsPerBird,colour=population,fill=population))++
  geom_point(size=0.3)+
  geom_smooth(method="lm")+
  scale_colour_manual(values=pop.cols)+
  scale_fill_manual(values=pop.cols)+
  xlab("Genes/100kb")+
  ylab("Ancestry transitions")

#leave out scaffold length - macrochromosomes only

gg1<-ggarrange(p1,p2,p3,common.legend=T,nrow=1,legend="right",labels=LETTERS)
gg1

```



```

ggsave("C:/Users/Steph/GitHub/thrush_hybrids/ancestry_tracts/figures/tracts_genomeStructure.png",gg1,
       width=10,height=2.5,units="in",bg='white')

```

```

library(lme4)
library(car)

lmm1<-lmer(startsPerBird~gene_density_100kb+(1|population),data=tractSummary100kb%>%filter(aimsCount>50))
summary(lmm1)

## Linear mixed model fit by REML ['lmerMod']
## Formula: startsPerBird ~ gene_density_100kb + (1 | population)
##   Data: tractSummary100kb %>% filter(aimsCount > 50)
##
## REML criterion at convergence: 4492.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max 
## -1.4806 -0.8021 -0.0915  0.6433  4.6650 
## 
```

```

## Random effects:
## Groups      Name      Variance Std.Dev.
## population (Intercept) 0.0003241 0.018
## Residual          0.0870542 0.295
## Number of obs: 11260, groups: population, 4
##
## Fixed effects:
##                   Estimate Std. Error t value
## (Intercept)      0.441044   0.009712 45.414
## gene_density_100kb -0.009509   0.001448 -6.567
##
## Correlation of Fixed Effects:
##             (Intr)
## gn_dnst_100 -0.243

```

```
confint(lmm1)
```

```

##                  2.5 %      97.5 %
## .sig01        0.007453846  0.039617316
## .sigma         0.291223910  0.298932801
## (Intercept)    0.419998454  0.462101662
## gene_density_100kb -0.012347940 -0.006671452

```

```
Anova(lmm1)
```

```

## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: startsPerBird
##              Chisq Df Pr(>Chisq)
## gene_density_100kb 43.126  1  5.133e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

lmm2<-lmer(startsPerBird~log(mean_cMMb)+(1|population),data=tractSummary100kb%>%filter(aimsCount>50))
summary(lmm2)

```

```

## Linear mixed model fit by REML ['lmerMod']
## Formula: startsPerBird ~ log(mean_cMMb) + (1 | population)
##   Data: tractSummary100kb %>% filter(aimsCount > 50)
##
## REML criterion at convergence: -1231.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1531 -0.6716 -0.1080  0.5204  5.4743
##
## Random effects:
## Groups      Name      Variance Std.Dev.
## population (Intercept) 0.0002248 0.01499
## Residual          0.0523537 0.22881
## Number of obs: 11260, groups: population, 4
##

```

```

## Fixed effects:
##                               Estimate Std. Error t value
## (Intercept)      0.698390   0.008410  83.04
## log(mean_cMMb)  0.092245   0.001063  86.79
##
## Correlation of Fixed Effects:
##             (Intr)
## lg(mn_cMMb)  0.374
confint(lmm2)

##                   2.5 %     97.5 %
## .sig01        0.006379718 0.03284404
## .sigma        0.225842603 0.23182081
## (Intercept)   0.680514754 0.71625606
## log(mean_cMMb) 0.090164219 0.09433061

Anova(lmm2)

## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: startsPerBird
##           Chisq Df Pr(>Chisq)
## log(mean_cMMb) 7533  1 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ',' 1

lmm3<-lmer(startsPerBird~Fst+(1|population),data=tractSummary100kb%>%filter(aimsCount>50))
summary(lmm3)

## Linear mixed model fit by REML ['lmerMod']
## Formula: startsPerBird ~ Fst + (1 | population)
## Data: tractSummary100kb %>% filter(aimsCount > 50)
##
## REML criterion at convergence: 1953.2
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -2.3536 -0.7024 -0.1903  0.5768  5.0583
##
## Random effects:
## Groups   Name       Variance Std.Dev.
## population (Intercept) 0.0002695 0.01642
## Residual            0.0695020 0.26363
## Number of obs: 11260, groups: population, 4
##
## Fixed effects:
##           Estimate Std. Error t value
## (Intercept)  0.71431   0.01012   70.61
## Fst         -0.72502   0.01347  -53.82
##
## Correlation of Fixed Effects:
##           (Intr)
## Fst -0.530

```

```
confint(lmm3)

##                2.5 %      97.5 %
## .sig01      0.006854895  0.03607677
## .sigma       0.260214038  0.26710208
## (Intercept) 0.693539241  0.73506956
## Fst        -0.751439398 -0.69863498
```

```
Anova(lmm3)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: startsPerBird
##      Chisq Df Pr(>Chisq)
## Fst  2897  1 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lmm4<-lmer(startsPerBird~abs(logRatioMu)+(1|population),data=tractSummary100kb%>%filter(aimsCount>50))
summary(lmm4)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: startsPerBird ~ abs(logRatioMu) + (1 | population)
##   Data: tractSummary100kb %>% filter(aimsCount > 50)
##
## REML criterion at convergence: 3509
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -1.7847 -0.7695 -0.1042  0.6338  4.6887
##
## Random effects:
##   Groups   Name        Variance Std.Dev.
##   population (Intercept) 0.0003063 0.0175
##   Residual             0.0820764 0.2865
## Number of obs: 10316, groups: population, 4
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)      0.519931  0.009739  53.38
## abs(logRatioMu) -0.089528  0.003434 -26.07
##
## Correlation of Fixed Effects:
##          (Intr)
## abs(logRtM) -0.330
```

```
confint(lmm4)
```

```
##                2.5 %      97.5 %
## .sig01      0.007111515  0.03863057
## .sigma       0.282610092  0.29043049
## (Intercept) 0.499077601  0.54078309
## abs(logRatioMu) -0.096265847 -0.08280214
```

Anova(lmm4)

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: startsPerBird
##          Chisq Df Pr(>Chisq)
## abs(logRatioMu) 679.5  1 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```