# Fig 2 for Paper: Build 2A and 2B Glued Together

July 21, 2017

## Contents



## 1 Intro

2017-07-18: Initially we built Fig2A and Fig2B in different scripts (Fig2A-desert-distribution.rnw and nc-snps.rnw, resp.). To streamline paper production and figure-tweaking, I've changed those two scripts to save the figure-relevant data they calculate into two .rda files that are loaded here, so that I can generate both figs together. Most of the documentation about the figures remains in those scripts, but visual parameters (sizes, colors, ...) can all be set here.

## 2 Preliminaries

Load utility R code; do setup:

```
source('../../../R/wlr.R') # load util code; path relative this folder or sibling in scripts/larrys

## Running as: ruzzo @ recycle.cs.washington.edu; SVN Id, I miss you.  $Id: wlr.R  2017-07-21 or later $

setup.my.wd('paperfigs') # set working dir; UPDATE if this file moves, or if COPY/PASTE to new file
setup.my.knitr('Fig2-glue-figs-knitr/') # knitr's "unnamed-chunk-nnn" figures
my.figs.dir <- 'Fig2-glue-figs-mine/'   # my named figures
generic.setup(my.figs.dir)
```

```
# frequently need to add figpath to file name
fpath <- function(base, suffix='.pdf', dir=my.figs.dir){
  return(paste(dir, base, suffix, sep=''))
}
```

## 3 Setup for Fig 2A

```
chr1.len <- genome.length.constants()$chr1.length  ## 3042585
load('Fig2A-data.rda')            # contains the "gap table" all.n.na
load('../../../data/des.rda')     # desert tables from svn+ssh://ceg1.ocean.washington.edu/var/svn/7_strains/trunk
names(des)[[6]] <- 'tp3367'       # override oldschool name
strain.order <- c(7,1,2,5,4,3,6) # Order of isolates, top-to-bottom in fig 2A
```

# 4 Setup for Fig 2B

```
load('../nc-snps/Fig2B-data.rda') # provides snp.rates.blob, needed by snp.rates.plot
```

# 5 Create Plot

```
heights.a.b <- c(2.7, 1.8) # panel heights (inches)
pdf(fpath('Fig2'),width=6.5, height=sum(heights.a.b))
opar <- par(no.readonly=TRUE, oma=c(0,0,0,0), mar=c(3,3,0.4,0.4), tcl=-0.2)
layout(matrix(1:2,nrow=2), heights=heights.a.b)

# Fig 2A:
draw.des.fig(des, all.n.na, width=chr1.len, row.order=strain.order, panel.label='A', chr='Chr1',
             xlab='Chromosome 1 Position (Mb)', gap.col = 'gold', d.col='dodgerblue2',
             min.des=10000, twotone='lightblue', do.par=FALSE)

# Fig 2B:
xx <- snp.rates.plot(snp.rates.blob, des.col='blue3', undes.col='black', yclip=0.01, legend='',
                     xlab='Desert Index', ylab='SNP Density', main='',
                     ylab.sub=list(text='(SNPS / bp)', line=1.1, cex=.75),
                     yticks=list(padj=1.6, at=seq(0,0.01,0.0025),
                                 labels=c('0.000','','0.005','','0.010'), cex=.9))
# panel='B':
text(0.62,0.0092,'B',cex=1.1) # for 2B, coords empirically set to roughly match rel pos in panel 2A

dev.off()

# pdf
#   2

par(opar)
```

Fig 1 shows it. (Surrounding boxes just to make marginal space obvious; change fbox to mbox to remove.)

SVN ID I miss you Id: Fig2.rnw 2017-07-20 or later ruzzo

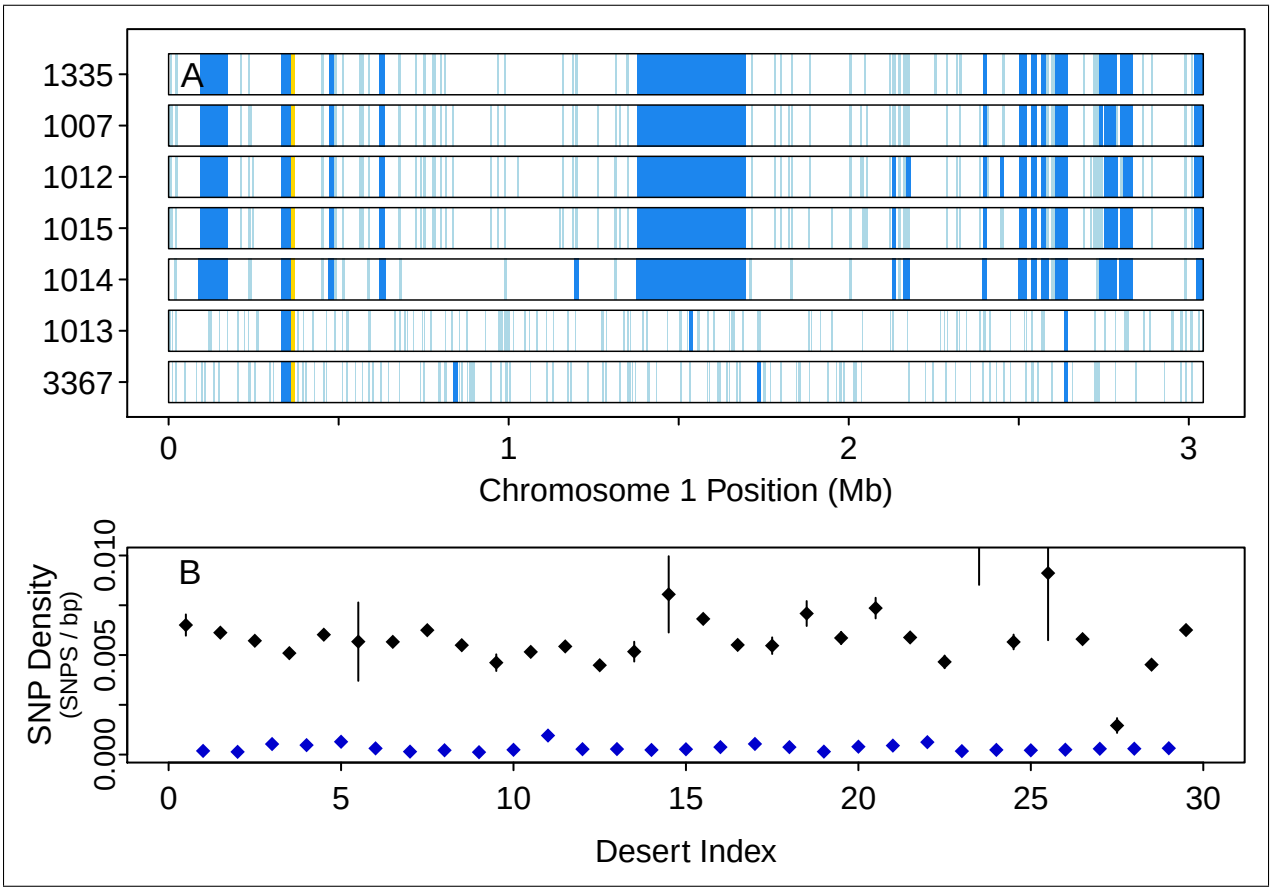


Figure 1: Proposed caption: Attributes of SNP deserts for *T. pseudonana* isolates. A) SNP distributions across the 3 Mb of Chromosome 1 for the seven *T. pseudonana* isolates. Regions in blue have significantly low SNP density ("SNP deserts") based on a negative binomial model (Methods). Pink(???) region is a gap of known size in the reference sequence. The large region centered near 1.5Mb is a 320Kb SNP desert present in all L-isolates but neither H-isolate. B) SNP densities (SNP per base-pair—$\mu \pm 2\sigma$) in the 29 deserts that span at least 50Kb of the CCMP 1335 genome (blue) and the thirty regions surrounding these deserts (including deserts smaller than 50Kb; black).