Predicting evolution using frequency-dependent selection in bacterial populations

Data analysis and simulations

Taj Azarian and Pamela P. Martinez

Contents

This is the code for reproducing the figures from the paper "Predicting evolution using frequency-dependent selection in bacterial populations" ## add DOI

```
require(ape)
require(car)
require(readr)
require(vegan)
require(gtools)
require(ggtree)
require(readxl)
require(gridSVG)
require(cowplot)
require(deSolve)
require(Metrics)
require(ggrepel)
require(svglite)
require(quadprog)
require(phytools)
require(devtools)
require(tidyverse)
require(wesanderson)
require(RColorBrewer)

#### Quadratic Programming function ####
## X is a matrix with rows = COGs and columns = SCs
## Y is a matrix with rows = COGs and columns = 1
QP <- function(X, Y){
  rinv <- solve(chol(t(X) %*% X)) # M to be minimized in quad. function (Choleski decomp)
```

```
  C <- cbind(rep(1,ncol(X)), diag(ncol(X))) #The contraints to minimize the quandratic function
  b <- c(1,rep(0,ncol(X)))
  d <- t(Y) %*% X  #Vector in the quadratic function to be minimized
  output <- solve.QP(Dmat = rinv, factorized = TRUE, dvec = d,
                     Amat = C, bvec = b, meq = 1)$solution
  output <- round(output, digits = 5)
  return(output)
}
```

Figure 1: Southwest US Data

```
#### Data mining ####
set.seed(9340)
df <- read_csv("data_southwestUS.csv") ## presence absence dataframe with metadata

dfE1 <- df %>% subset(Epoch1 == "E1") ## E1 prevaccine
vaccineT <- df %>% distinct(BAPS2, PCV7.actual)
vaccineT <- vaccineT %>% subset(PCV7.actual == "VT") %>%
  dplyr::rename(W = PCV7.actual) %>%
  full_join(subset(vaccineT,PCV7.actual == "NVT")) %>%
  unite("vaccine", W:PCV7.actual, na.rm = TRUE)

dfF <- df %>% select(BAPS2, Epoch1) %>% group_by(Epoch1) %>%
  count(BAPS2) %>% mutate(freq = prop.table(n)) %>% ungroup() %>%
  select(Epoch1, BAPS2, freq) %>%
  spread(Epoch1, freq, fill = 0)

zero_E1 <- dfE1 %>% distinct(BAPS2) %>%
  mutate(n = 0, freq = 0)
```

```
#### Replicates  - null expectation Pro rata ####
dfE1_NVT_all <- data.frame(NULL)
replicates <- 10000

for(i in 1:replicates){
  #sub-sampling from each epoch independently with replacemnt - best subsampling strategy
  dfE1_NVT_i <- dfE1 %>% sample_frac(1, replace = TRUE) %>%
    subset(PCV7.actual == "NVT") %>% count(BAPS2) %>%
    mutate(freq = prop.table(n)) %>% bind_rows(zero_E1) %>%
    group_by(BAPS2) %>% summarise(n = sum(n), freq = sum(freq)) %>%
    arrange(BAPS2) %>% ungroup %>% mutate(iter = i)

  dfE1_NVT_all <- bind_rows(dfE1_NVT_all, dfE1_NVT_i)
}

dfE1_NVT_all <- dfE1_NVT_all %>% group_by(BAPS2) %>%
  summarise(expected_E3 = quantile(freq, 0.5), cil = quantile(freq, 0.025),
            ciu = quantile(freq, 0.975)) %>% ungroup

dfF <- left_join(dfF,dfE1_NVT_all) %>% replace(., is.na(.), 0) %>%
  mutate(delta = E3 - E1, deltaExp = expected_E3-E1,
         CI_low = cil-E1, CI_up = ciu-E1) %>%
```

```
  mutate(SC=BAPS2) %>%
  mutate(expectation = "Null exp.", signif = ifelse(delta > CI_up, "pos",NA)) %>%
  mutate(signif = ifelse(delta < CI_low, "neg", signif)) %>%
  mutate(signif = ifelse(E1 == 0 | E3 == 0, NA, signif))

dfF <- left_join(dfF,vaccineT)

#### Plot A: Prevalence by sequence cluster ####
dat1 <- mutate(dfF, signif = ifelse(E3 > CI_up | E3 < CI_low, -0.005,NA), Epoch = "E1")
dat1 <- subset(dat1, select = c(SC, Epoch, signif))

datPlotA <- dfF %>% select(SC, E1, E3) %>%
  pivot_longer(-SC, names_to =  "Epoch", values_to = "Prevalence") %>%
  left_join(dat1) %>% left_join(select(dfF, SC, vaccine)) %>%
  mutate(SC = as.character(SC)) %>% arrange(Epoch,-Prevalence)

rank <- as.character(unique(datPlotA$SC))

datPlotA <- datPlotA %>% mutate(SC = factor(SC, levels = rank))

plot1A <- ggplot(datPlotA, aes(x=SC, y=Prevalence, alpha=Epoch, fill=vaccine)) +
  geom_bar(stat='identity', position='dodge') +
  scale_alpha_manual(values = c(1,0.4),
                     labels = c("Pre-vaccine", "Post-vaccine")) +
  scale_fill_manual(values = c("#143c77","darkred","mediumpurple4"),
                    labels = c("Non-vaccine type", "Vaccine type", "Mixed"),
                    name = "Composition") +
  xlab("Strain (SC)") + theme_classic() +
  scale_y_continuous(expand = c(0, 0), limits = c(0,0.15)) +
  theme(legend.title = element_text(face="bold", size = 10),
        legend.justification = c(1, 1), legend.box = "horizontal",
        legend.position = c(1, 1), # legend.position = c(0.725, 0.85),
        legend.spacing.y = unit(0.2, "cm"),
        legend.text=element_text(size=9),
        legend.background = element_blank(),
        legend.box.background = element_rect(fill = gray(0.96), color = NA))

#### Plot B: Change in prevalence ####
dfF <- dfF %>% mutate(SC = factor(SC, levels = rank))

plot1B <- ggplot() + ylim(c(-0.09,0.08)) +
  geom_hline(yintercept=0, lty="dashed",size=0.65) +
  scale_colour_manual(values = c("#143c77","darkred","mediumpurple4"),
                      labels = c("Non-vaccine type", "Vaccine type", "Mixed"),
                      name = "SC Composition") +
  scale_linetype_manual(values = c(1), name = NULL) +
  scale_shape_manual(values = c("-", "+")) + theme_classic() +
  geom_point(data=dfF, aes(SC, delta, col = vaccine),
             size =4, alpha  = 0.85, show.legend = F) +
  geom_pointrange(data=dfF, aes(x=SC, y=deltaExp,
                                ymin=CI_low, ymax=CI_up, lty=expectation),
                  size=.15, fatten = 6, show.legend = F) +
  geom_point(data=dfF, aes(x=SC, y=-0.09, shape = signif), size=6,
```

```
            col = "lightsteelblue4", show.legend = F) +
  labs(x="Strain (SC)") + labs(y="Change in prevalence") +
  theme(legend.title = element_text(face="bold", size = 10),
        legend.text=element_text(size=9),
        legend.position = c(0.01, 1),
        legend.justification = c(0.01, 1),
        legend.box = "horizontal",
        legend.background = element_blank(),
        legend.box.background = element_rect(fill = gray(0.95), color = NA))

#### Final figure 1 ####
figure1 <- plot_grid(plot1A, plot1B, labels = c("A","B"), ncol = 1)
ggsave("figure1.png", figure1, width = 14,height = 7)
```

Figure 2: Simulations

```
#### Funct. rootfun ####
rfun <- function(t, state, pars){
  dstate <- unlist(repEq(t, state, pars)) # rate of change vector
  return(sum(abs(dstate)) - 1e-4)
}

#### Funct. checkFeas ####
checkFeas <- function(e, g){
  g <- as.data.frame(g)
  temp <- sapply(g, function(x)(length(unique(x))))
  id <- which(temp==1)
  e[id] <- as.numeric(unique(g[,id]))
  return(e)
}

#### Funct. replicator ####
repEq <- function(t, Nf, pars){
  with(as.list(c(Nf, pars)), {
    f <- NULL
    dfdt <- rep(0, nSC)
    xifi <- rep(0, nSC)

    ## loci frequencies ##
    for(k in 1:nCOG){ f[k] <- sum(Nf*genot[,k]) }

    for(k in 1:nSC){
      for(l in 1:nCOG){
        xifi[k] <- xifi[k] + Nf[k]*(genot[k,l]*(eqbm[l] - f[l]))
      }
    }

    ## dfdt ##
    for(k in 1:nSC){ dfdt[k] <- xifi[k] - Nf[k]*sum(xifi) }

    return(list(dfdt))
```

```
  })
}


####################################

#### Pre-intervation ####
#### Parameters ####
nCOG <- 10
nSC <- 8
eqbm <- c(0.5677,0.5138,0.4050,0.4388,0.4981,
          0.5065,0.5725,0.4513,0.5811,0.4034)
vacT <- c(2,3,5)
timeSteps <- 0.5
posCom <- 2^nCOG
genot <- data.frame(permutations(n=2,r=nCOG,v=c(0,1), repeats.allowed = T))
colnames(genot) <- as.character(1:nCOG)
pres <- c(7,193,320,337,340,621,674,842)
genotF <- c(0.2342,0.1511,0.0033,0.1248, 0.1750,0.0067,0.2219,0.083)
genotP <- genot[pres,]
times <- seq(from=0, to=1000, by=timeSteps)
pars <- list(eqbm = eqbm, nCOG = nCOG, nSC = nSC, genot = genotP)

#### Simulations ####
out1 <- out1P <- as.data.frame(lsodar(func=repEq,y=genotF,times=times,parms=pars,rootfun=rfun))
out1 <- out1 %>% pivot_longer(-time, names_to = "genotype", values_to = "frequency") %>%
  mutate(genotype = paste('G', genotype, sep = "")) %>% subset(frequency > 0)

out1_pre <- out1 %>% mutate(time = time/3)
E1 <- round(as.numeric(out1P[nrow(out1P),-1]), digits = 5) ### last time in the data frame

#### Intervation ####
#### Parameters ####
E2 <- E1
E2[vacT] <- 0 ### remove genotypes G2 (001) and G6 (101) 'vaccine types'
E2 <- E2/sum(E2)
idZ <- which(E2 > 0)
timeS <- 40
genotPV <- genotP[idZ,]
eqbmPV <- checkFeas(eqbm, genotPV)
pars$eqbm <- eqbmPV
pars <- list(eqbm = eqbm, nCOG = nCOG, nSC = nSC, genot = genotP)

rfun <- function(t, state, pars){
  dstate <- unlist(repEq(t, state, pars)) # rate of change vector
  return(sum(abs(dstate)) - 1e-6)
}

### Simulations ###
out2 <- out2P <- as.data.frame(lsodar(func=repEq,y=E2,times=times,parms=pars,rootfun=rfun))
out2 <- out2 %>% mutate(time = time + timeS) %>%
  pivot_longer(-time, names_to = "genotype", values_to = "frequency") %>%
  mutate(genotype = paste('G', genotype, sep = "")) %>% subset(time <= 80)
```

```
outF <- rbind(out1_pre, out2)
E3 <- round(as.numeric(out2P[nrow(out2P),-1]), digits = 5)

lineT <- data.frame(linetype = "Non-vaccine type",
                    genotype = paste('G', 1:8, sep = ""), index = 1:8)
lineT <- lineT %>% mutate(linetype = ifelse(index %in% vacT, "Vaccine type",
                                            "Non-vaccine type"), index = NULL)

outF <- outF %>% full_join(lineT) %>%
  mutate(frequency = ifelse(frequency == 0,
                            runif(1, min = 0.00001, max = 0.00002), frequency))

E3 <- outF %>% subset(time == max(outF$time)) %>% arrange(-frequency)
alphaT <- data.frame(genotype = E3$genotype,
                     alpha = c(seq(from=1,to=0.9,length.out =5),1,1,1))
outF <- full_join(outF, alphaT)

#### Figure A ####
plot2A <- ggplot(outF, aes(time, frequency, group = genotype, alpha = alpha,
                           colour = factor(linetype), linetype = factor(linetype))) +
  annotate("rect", xmin = 31.5, xmax = 37.5, ymin = -Inf,
           ymax = 0.28, fill = "gray93", colour = NA) +
  annotate("rect", xmin = 69, xmax = 75, ymin = -Inf,
           ymax = 0.28, fill = "gray93", colour = NA) +
  annotate("rect", xmin = 40, xmax = 42, ymin = -Inf,
           ymax = 0.28, fill = "darkslategray4", colour = NA, alpha = 0.3) +
  geom_line(size = 1) + theme_classic() +
  scale_colour_manual(values = c('#143c77', "darkred")) + xlim(0,75) + #899DA4 #C93312
  #scale_linetype_manual(values = c("solid", "dashed")) +
  geom_segment(x = 38.2, y = 0.3175, xend = 38.2, yend = 0.265,
               arrow = arrow(length = unit(0.25, "cm")),
               colour = 'black', show.legend = F) +
  annotate("text", x = 38.2, y = 0.33, label="Vaccine introduction",
           fontface="bold", color="black", size = 3.7) +
  #geom_segment(aes(x = 65, y = 0.29, xend = 74, yend = 0.29), colour = 'black') +
  annotate("text", x = 34.5, y = 0.295, label="Pre-vaccine\nequilibrium",
           color="dimgrey", fontface="bold", size = 2.65) +
  annotate("text", x = 72, y = 0.295, label="Post-vaccine\nequilibrium",
           color="dimgrey", fontface="bold", size = 2.65) +
  annotate("text", x = 41, y = 0.295, label="Predicted\nFitness",
           color="darkslategray", fontface="bold", size = 2.5) +
  ylab("Strain Prevalence") + xlab("Time") +
  scale_alpha(range = c(0.15, 1), guide = 'none') +
  theme(legend.justification=c(0,0),legend.position=c(0.02,0.85),
        legend.title=element_blank(),
        axis.text = element_text(colour = "black"),
        legend.text=element_text(size=8),
        plot.title = element_text(hjust = 0, size=10, face="bold"))

###################################
COGs <- 2371 ## 2371
SCs <- 35
dataOutF <- data.frame(NULL)
```

```r
replicates <- 10
VTselect <- 3

for(int in 1:replicates){
  el0 <- runif(COGs, min = 0.05, max = 0.95)
  dat <- as.matrix(replicate(SCs, sample(c(0,1), COGs, replace = TRUE)))
  dat <- unique(dat, MARGIN = 2)
  VT <- sample(1:SCs, VTselect)
  NVT <- (1:SCs)[-VT]

  if(ncol(dat) < SCs){
    cat("run again\n")
    break
  }

  #### E1 = pre-vaccine frequencies ####
  x1 <- QP(dat, as.matrix(el0))

  ### re-calculate el ###
  el <- dat %*% as.matrix(x1)

  #### E2 - frequencies just after vaccine intro ####
  x2 <- x1[-VT]
  x2 <- round(x2/sum(x2), digits = 5)
  dat2 <- dat[,-VT]
  fl <- dat2 %*% as.matrix(x2)

  #### E3 - frequencies long-term post-vaccine ####
  x3 <- QP(dat2, el)
  el3 <- dat2 %*% as.matrix(x3)

  #### fitness function "omega" just after vaccine intro ####
  whole <- as.numeric(el - fl) ### this is similar to (el - fl) and thus fitness.
  deltaE <- x3 - x2
  omega <- as.vector(t(dat2) %*% whole) ## length SCs, similar to FFS

  phi <- sum(x2 * omega) ## average fitness
  rateOfChange <- omega - phi ## "rate of change": omega_g - phi

  ################### Relative fitness ###################
  dataOut <- data.frame(riskDif = deltaE, r = rateOfChange)
  dataOut <- subset(dataOut, riskDif != 0)
  dataOut <- mutate(dataOut, change = ifelse(riskDif < 0, "Decreased", "Increased"), rep = int)
  dataOutF <- rbind(dataOutF, dataOut)
}

dataOutF <- dataOutF %>% mutate(col = ifelse(sign(riskDif) == sign(r), "same", "diff"))

#### Figure B ####
plot2B <- ggplot(dataOutF, aes(x=change, y=r,
                               group=factor(rep), colour = col)) +
  geom_point(position = position_dodge(width = 1), size = 2.5, alpha  = 0.6) +
  geom_hline(yintercept = 0, linetype = 3) + theme_classic() +
```

```
  xlab("Observed prevalence change") +
  ylab("Standardized Fitness") +
  theme(legend.position="none", axis.text = element_text(colour = "black"),
        plot.title = element_text(hjust = 0, size=10, face="bold")) +
  scale_colour_manual(values = c("gray70", "darkslategray4"))

figure2 <- plot_grid(plot2A, plot2B, labels = c("A", "B"), rel_widths = c(1.75,1))
ggsave("figure2.png", figure2, width = 14, height = 4.5)
```

Figure 3: Prediction (Southwest US data)

```
#### Data ####
dfFVT <- df %>% select(BAPS2,PCV7.actual, Epoch1) %>%
  group_by(Epoch1) %>%  count(BAPS2,PCV7.actual) %>%
  mutate(freq = round(prop.table(n), digits = 3)) %>% ungroup() %>%
  select(Epoch1, BAPS2, PCV7.actual, freq) %>%
  spread(Epoch1, freq, fill = 0) %>% arrange(BAPS2,PCV7.actual)

#### Present at E1 ####
SCE1 <- dfFVT %>% subset(E1 > 0) %>% ## & BAPS2 !="27"
  select(BAPS2, PCV7.actual) %>%
  mutate(Epoch1 = "E1")

#### NVTs Present at E1 ####
SCE2 <- SCE1 %>% subset(PCV7.actual == "NVT"  & BAPS2 !="27")


SC_freq_df <- df %>% select(BAPS2, PCV7.actual, Epoch1,
                            HMPREF0837_12128:HMPREF0837_10616) %>%
  arrange(BAPS2) %>% group_by(BAPS2,PCV7.actual,Epoch1) %>%
  mutate(SC_n = n()) %>% ungroup() %>%
  group_by(BAPS2,PCV7.actual,Epoch1,SC_n) %>%
  summarise_at(vars(HMPREF0837_12128:HMPREF0837_10616),mean) %>%
  ungroup()

### Get the matrix and the SC for the pre-vaccine epch "E1"
df_preV <- SCE1 %>% left_join(SC_freq_df)
SC_freq_preV <- as.matrix(df_preV %>% mutate(SC_freq=SC_n/sum(SC_n)) %>% select(SC_freq))
SC_COG_preV <- as.matrix(t(df_preV %>% select(HMPREF0837_12128:HMPREF0837_10616)))

#### Get e_l ####
el <- SC_COG_preV %*% SC_freq_preV

#### Figure A ####
dfImputed <- dfFVT %>% subset(PCV7.actual == "NVT" & E1 == 0) %>%
  select(BAPS2,PCV7.actual) %>% mutate(Epoch1 = "E1", n=1) %>%
  select(Epoch1,BAPS2,PCV7.actual,n)

dfFVTImputed <- df %>% select(BAPS2,PCV7.actual, Epoch1) %>%
  group_by(Epoch1) %>%  count(BAPS2,PCV7.actual) %>%
  ungroup() %>% bind_rows(dfImputed) %>% group_by(Epoch1) %>%
```

```
  mutate(freq = round(prop.table(n), digits = 3)) %>% ungroup() %>%
  select(Epoch1, BAPS2, PCV7.actual, freq) %>%
  spread(Epoch1, freq, fill = 0) %>% arrange(BAPS2,PCV7.actual)

dfFNVTImputed <- dfFVTImputed %>%
  subset(PCV7.actual == "NVT" & BAPS2 !="27") %>%
  mutate(deltaE = E3 - E1) %>% arrange(BAPS2)

#### E2 - frequencies just after vaccine intro ####
x_imputed <- dfFNVTImputed$E1
x_imputed <- as.matrix(round(x_imputed/sum(x_imputed), digits = 5))

dat2_imputed <- dfImputed %>% mutate(Epoch1 = "E2") %>%
  select(BAPS2,PCV7.actual,Epoch1) %>% bind_rows(SCE2) %>%
  arrange(BAPS2) %>% left_join(SC_freq_df) %>%
  select(HMPREF0837_12128:HMPREF0837_10616)
dat2_imputed <- as.matrix(dat2_imputed)

fl_imp <- t(dat2_imputed) %*% x_imputed

#### fitness function "omega" just after vaccine intro ####
whole <- as.numeric(el - fl_imp) ### this is similar to (el - fl) and thus fitness.
deltaE <- dfFNVTImputed$deltaE
omega <- as.vector(dat2_imputed %*% whole) ## length SCs, similar to FFS

phi <- sum(x_imputed * omega) ## average fitness
rateOfChange <- omega - phi ## "rate of change": omega_g - phi

dat3AB <- dfFNVTImputed %>% select(BAPS2, deltaE) %>%
  mutate(r = rateOfChange) %>% subset(deltaE != 0) %>%
  mutate(change = ifelse(deltaE < 0, "Decreased", "Increased")) %>%
  mutate(col = ifelse(sign(deltaE) == sign(r), "same", "diff")) %>%
  left_join(vaccineT)

highlight <- dat3AB %>% filter(col == "diff")

lm3 <- lm(r~deltaE,dat3AB)
summary(lm3)
```

Call: lm(formula = r ~ deltaE, data = dat3AB)

Residuals: Min 1Q Median 3Q Max -12.7964 -1.8504 -0.1459 2.3969 10.6031

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.4151 1.0220 0.406 0.688
deltaE 182.1310 38.6050 4.718 5.55e-05 *** — Signif. codes: 0 ‘’ ***0.001*** ’’ *0.01* ’’ 0.05 ‘.’ 0.1 ’ ’ 1

Residual standard error: 4.995 on 29 degrees of freedom Multiple R-squared: 0.4342, Adjusted R-squared: 0.4147 F-statistic: 22.26 on 1 and 29 DF, p-value: 5.555e-05

```
#### Figure A ####
plot3A <- ggplot(dat3AB, aes(x=deltaE, y=r, colour = vaccine)) +
  theme_classic() + xlab("Observed Prevalence Change") +
  ylab("Standardized Fitness") + xlim(-0.03, 0.065) + ylim(-12,20) +
  geom_hline(yintercept=0, lty="dotted", alpha=.6) +
  geom_vline(xintercept=0, lty="dotted", alpha=.6) +
```

```
  annotate("rect", xmin=-0.03,xmax=0,ymin=-12,ymax=0,
           fill="darkslategray4", alpha= 0.15) +
  annotate("rect", xmin=0,xmax=0.065,ymin=0,ymax=20,
           fill="darkslategray4", alpha= 0.15) +
  geom_smooth(aes(group = 1), color="grey70", method=lm,
              formula = y~x,  show.legend=FALSE, se=FALSE) +
  geom_point(size = 3) + theme(legend.position = "none",
                               axis.text = element_text(colour = "black"),) +
  scale_colour_manual(values = c("#143c77","mediumorchid4")) +
  geom_text_repel(aes(label = paste("SC", BAPS2, sep="-")), data = highlight, size = 3.5)


#### Figure B ####
df_postV <- SCE2 %>% left_join(SC_freq_df)
SC_COG_postV <- as.matrix(t(df_postV %>% select(HMPREF0837_12128:HMPREF0837_10616)))

SC_freq_postV_obs <- SCE2 %>% mutate(Epoch1 = "E3") %>%
  left_join(SC_freq_df) %>%
  mutate(SC_freq=SC_n/sum(SC_n, na.rm = T)) %>%
  select(BAPS2, PCV7.actual, SC_freq) %>%
  mutate(SC_freq = replace_na(SC_freq, 0))

## Predict postV frequencies
SC_freq_postV_pred <- QP(SC_COG_postV, el) #Matrix: rows = COGs, columns = (SCs - VT)
SC_freq_postV_obs <- SC_freq_postV_obs %>%
  mutate(SC_pred = SC_freq_postV_pred) %>%
  left_join(vaccineT)

outlier3B <- SC_freq_postV_obs %>% mutate(diff = abs(SC_freq - SC_pred))
outlier3B <- outlier3B %>% filter(diff %in% boxplot(outlier3B$diff, plot = FALSE)$out)

W_model3B <- lm(data=SC_freq_postV_obs,SC_freq~SC_pred);
summary(W_model3B)
```

Call: lm(formula = SC_freq ~ SC_pred, data = SC_freq_postV_obs)

Residuals: Min 1Q Median 3Q Max -0.059467 -0.014478 -0.002457 0.015118 0.053442

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.012366 0.008616 1.435 0.16364
SC_pred 0.666107 0.198686 3.353 0.00255 ** — Signif. codes: 0 ‘’ ***0.001*** ’’ *0.01* ’’ 0.05 ‘.’ 0.1 ’ ’ 1

Residual standard error: 0.02329 on 25 degrees of freedom Multiple R-squared: 0.3101, Adjusted R-squared: 0.2826 F-statistic: 11.24 on 1 and 25 DF, p-value: 0.002551

```
confint(W_model3B)
```

```
                2.5 %     97.5 %
```

(Intercept) -0.005380061 0.03011185 SC_pred 0.256905970 1.07530899

```
linearHypothesis(W_model3B, c("(Intercept) = 0", "SC_pred = 1"), test = "Chisq")
```

Linear hypothesis test

Hypothesis: (Intercept) = 0 SC_pred = 1

Model 1: restricted model Model 2: SC_freq ~ SC_pred

Res.Df RSS Df Sum of Sq Chisq Pr(>Chisq) 1 27 0.015093
2 25 0.013561 2 0.0015319 2.8241 0.2436

```
linearHypothesis(W_model3B, c("(Intercept) = 0", "SC_pred = 1"))
```

Linear hypothesis test

Hypothesis: (Intercept) = 0 SC_pred = 1

Model 1: restricted model Model 2: SC_freq ~ SC_pred

Res.Df RSS Df Sum of Sq F Pr(>F) 1 27 0.015093
2 25 0.013561 2 0.0015319 1.412 0.2624

```
plot3B <- ggplot(SC_freq_postV_obs,
                 aes(x = SC_pred, y = SC_freq, colour = vaccine)) +
  geom_segment(aes(x=0,xend=0.12,y=0,yend=0.12),
               color="black",alpha=.7,lwd=0.5,lty=3) +
  theme(legend.position = "none") + theme_classic() +
  geom_smooth(method='lm', color="#899DA4" ,
              formula=y~x, alpha=0.3, lwd=.6,
              fullrange=T, linetype="blank", show.legend=F) +
  annotate(geom = "text", x=0.11, y =0.114,
           label = "1:1 line", angle = 45, size = 3) +
  geom_point(size=3) +  ##
  scale_x_continuous("Predicted Prevalence (NFDS)")+
  scale_y_continuous("Observed Prevalence") +
  coord_fixed(ratio = 1, xlim=c(0,0.12), ylim=c(0,0.12)) +
  scale_colour_manual(values = c("#143c77","mediumorchid4")) +
  theme(legend.position = "none",
        axis.text = element_text(colour = "black"),) +
  geom_text_repel(aes(label = paste("SC", BAPS2, sep = "-")), data = outlier3B, size = 3.5)


#### Figure C ####

SC_freq_E1 <- df_preV %>%
  mutate(SC_freq_E1=SC_n/sum(SC_n)) %>%
  select(BAPS2,PCV7.actual,SC_freq_E1)

SC_freq_postV_diff <- SC_freq_postV_obs %>%
  left_join(SC_freq_E1) %>%
  mutate(diff_pred = SC_freq_E1 - SC_pred,
         diff_obs = SC_freq_E1 - SC_freq) %>%
  left_join(vaccineT)

SC_freq_postV_diff <- SC_freq_postV_diff %>%
  mutate(diff = abs(diff_pred - diff_obs)) %>%
  arrange(diff)

stats <- summary(lm(SC_freq_postV_diff$diff_pred~SC_freq_postV_diff$diff_obs))
ars <- round(stats$adj.r.squared, digits = 3)
sseE <- round(sse(SC_freq_postV_diff$diff_obs, SC_freq_postV_diff$diff_pred), digits = 3)
maeE <- round(mae(SC_freq_postV_diff$diff_obs, SC_freq_postV_diff$diff_pred),digits = 3)##Mean Absolute
rmseE <- round(rmse(SC_freq_postV_diff$diff_obs, SC_freq_postV_diff$diff_pred),digits = 3) ##Root Mean

accNFDS <- data.frame(Model = "Accesory genome (NFDS)", nloci = length(el),
```

```
                      adj.r.squared = ars, SSE = sseE, RMSE = rmseE)


W_model3C <- lm(data=SC_freq_postV_diff,diff_obs~diff_pred);
summary(W_model3C)
```

Call: lm(formula = diff_obs ~ diff_pred, data = SC_freq_postV_diff)

Residuals: Min 1Q Median 3Q Max -0.057236 -0.011661 0.002054 0.011921 0.067643

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.003684 0.006783 -0.543 0.5919
diff_pred 0.795065 0.273307 2.909 0.0075 ** — Signif. codes: 0 ‘’ ***0.001*** ’’ *0.01* ’’ 0.05 ‘.’ 0.1 ’ ’ 1

Residual standard error: 0.0243 on 25 degrees of freedom Multiple R-squared: 0.2529, Adjusted R-squared: 0.223 F-statistic: 8.463 on 1 and 25 DF, p-value: 0.007504

```
confint(W_model3C)
```

```
               2.5 %     97.5 %
```

(Intercept) -0.01765401 0.01028661 diff_pred 0.23217925 1.35795061

```
linearHypothesis(W_model3C, c("(Intercept) = 0", "diff_pred = 1"), test = "Chisq")
```

Linear hypothesis test

Hypothesis: (Intercept) = 0 diff_pred = 1

Model 1: restricted model Model 2: diff_obs ~ diff_pred

Res.Df RSS Df Sum of Sq Chisq Pr(>Chisq) 1 27 0.015093
2 25 0.014761 2 0.00033197 0.5623 0.7549

```
outlier3C <- SC_freq_postV_diff %>%
  filter(diff %in% boxplot(SC_freq_postV_diff$diff, plot = FALSE)$out)

#### Figure C ####
plot3C <- ggplot(SC_freq_postV_diff, aes(x = diff_pred,
                                         y = diff_obs, colour = vaccine)) +
  geom_segment(aes(x=-0.1,xend=0.05,y=-0.1,yend=0.05),
               color="black",alpha=.7,lwd=0.5,lty=3) +
  geom_smooth(method='lm', color="gray80" ,formula=y~x,
              alpha=0.3, lwd=.6, fullrange=T,
              linetype="blank", show.legend=F) +
  annotate(geom = "text", x=0.0375, y =0.0425,
           label = "1:1 line", angle = 45, size = 3) +
  geom_point(size=3) + theme_classic() +
  scale_colour_manual(values = c("#143c77","mediumorchid4"),
                      labels = c("Non-vaccine type", "Mixed"),
                      name = "SC Composition") +
  theme(legend.position = "none",
        axis.text = element_text(colour = "black"),) +
  xlab("Predicted Prevalence Change (NFDS)")+
  ylab("Observed Prevalence Change") +
  coord_fixed(ratio = 1, xlim=c(-0.1,0.05), ylim=c(-0.1,0.05)) +
  annotate("text",x=-0.097,y=0.04, size=2.5,hjust = 0,
           label=paste("SSE = ", sseE, "\nRMSE = ",
                       rmseE, "\nAdj. R2 = ", ars)) +
```

```
  geom_text_repel(aes(label = paste("SC", BAPS2, sep = "-")), data = outlier3C, size = 3.5)

#### Figure D ####
SC_freq_postV_diff <- SC_freq_postV_diff %>%
  mutate(prorata = SC_freq_E1/sum(SC_freq_E1)) %>%
  mutate(diff_predPro = SC_freq_E1 - prorata) %>%
  mutate(diffP = abs(diff_predPro - diff_obs))

stats <- summary(lm(SC_freq_postV_diff$diff_predPro~SC_freq_postV_diff$diff_obs))
ars <- round(stats$adj.r.squared, digits = 3)
sseE <- round(sse(SC_freq_postV_diff$diff_obs, SC_freq_postV_diff$diff_predPro), digits = 3)
maeE <- round(mae(SC_freq_postV_diff$diff_obs, SC_freq_postV_diff$diff_predPro),digits = 3)##Mean Absol
rmseE <- round(rmse(SC_freq_postV_diff$diff_obs, SC_freq_postV_diff$diff_predPro),digits = 3) ##Root Me

accProrata <- data.frame(Model = "Accesory genome (Pro rata)", nloci = length(el),
                         adj.r.squared = ars, SSE = sseE, RMSE = rmseE)

W_model3D <- lm(data=SC_freq_postV_diff,diff_obs~diff_predPro);
summary(W_model3D)
```

Call: lm(formula = diff_obs ~ diff_predPro, data = SC_freq_postV_diff)

Residuals: Min 1Q Median 3Q Max -0.053397 -0.014551 -0.001313 0.021763 0.038569

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.027719 0.009368 -2.959 0.00666 ** diff_predPro -0.541862 0.431660 -1.255 0.22098
— Signif. codes: 0 ‘’ ***0.001*** ’’ *0.01* ’’ 0.05 ‘.’ 0.1 ’ ’ 1

Residual standard error: 0.02727 on 25 degrees of freedom Multiple R-squared: 0.05929, Adjusted R-squared: 0.02167 F-statistic: 1.576 on 1 and 25 DF, p-value: 0.221

```
confint(W_model3D)
```

```
                2.5 %       97.5 %
```

(Intercept) -0.04701286 -0.008425814 diff_predPro -1.43088194 0.347158089

```
linearHypothesis(W_model3D, c("(Intercept) = 0", "diff_predPro = 1"), test = "Chisq")
```

Linear hypothesis test

Hypothesis: (Intercept) = 0 diff_predPro = 1

Model 1: restricted model Model 2: diff_obs ~ diff_predPro

Res.Df RSS Df Sum of Sq Chisq Pr(>Chisq)
1 27 0.028071
2 25 0.018586 2 0.0094854 12.759 0.001696 ** — Signif. codes: 0 ‘’ ***0.001*** ’’ *0.01* ’’ 0.05 ‘.’ 0.1 ’ ’ 1

```
plot3D <- ggplot(SC_freq_postV_diff, aes(x = diff_predPro,
                                         y = diff_obs, colour = vaccine)) +
  geom_segment(aes(x=-0.1,xend=0.05,y=-0.1,yend=0.05),
               color="black",alpha=.7,lwd=0.5,lty=3) +
  geom_smooth(method='lm', color="gray80" ,formula=y~x,
              alpha=0.3, lwd=.6, fullrange=T,
              linetype="blank", show.legend=F) +
  annotate(geom = "text", x=0.0375, y =0.0425,
           label = "1:1 line", angle = 45, size = 3) +
  geom_point(size=3) + theme_classic() +
```

```
  scale_colour_manual(values = c("#143c77","mediumorchid4"),
                      labels = c("Non-vaccine type", "Mixed"),
                      name = "SC Composition") +
  theme(legend.position = "none",
        axis.text = element_text(colour = "black"),) +
  xlab("Predicted Prevalence Change (Pro rata)")+
  coord_fixed(ratio = 1, ylim=c(-0.1,0.05), xlim=c(-0.1,0.05)) +
  ylab("Observed Prevalence Change") +
  annotate("text",x=-0.097,y=0.04, size=2.5,hjust = 0,
           label=paste("SSE = ", sseE, "\nRMSE = ",
                       rmseE, "\nAdj. R2 = ", ars))


#### Combine figure 3 ####
ptitle1 <- ggplot() + theme_void() +
  annotate("rect", fill = "darkslategray4", alpha = 0.3,
           xmin = 0, xmax = 1, ymin = 0, ymax = 1)
ptitle1 <- ggdraw(ptitle1) + draw_label("Predicted Fitness",
                                        fontface='bold', colour = "darkslategrey") ## ,
pA <- plot_grid(plot3A, labels = c("A"))
pA <- plot_grid(ptitle1, pA, ncol = 1, scale = c(1,0.9), rel_heights=c(0.075, 1))


ptitle2 <- ggplot() + theme_void() +
  annotate("rect", fill = "gray93", xmin = 0, xmax = 1, ymin = 0, ymax = 1)
ptitle2 <- ggdraw(ptitle2) + draw_label("Post-vaccine Equilibrium Frequencies",
                                        fontface='bold', colour = "gray30") ## ,
pBCD <- plot_grid(plot3B, plot3C, plot3D, ncol = 3, labels = c("B","C","D"))
pBCD <- plot_grid(ptitle2, pBCD, ncol = 1, scale = c(1,0.9), rel_heights=c(0.075, 1))
plot3 <- plot_grid(pA, pBCD, rel_widths = c(1,3), rel_heights = c(1,1))

ggsave("figure3.png", plot3, width = 16, height = 4)
```

Fig. Suppl 1: Tree

```
## The tree was modified after being produced (style purposes)

readMatrix<-function(heatmapData){
  if (is.matrix(heatmapData)) {
    x = data.frame(heatmapData)
  }
  else if (is.data.frame(heatmapData)) {
    x = heatmapData
  }
  else {
    x<-read.csv(heatmapData,row.names=1)
  }
  x
}
getLayout<-function(infoFile,infoCols,heatmapData,barData,doBlocks,treeWidth=10,infoWidth=10,dataWidth=

  # m = layout matrix
  # w = layout widths vector
```

```
# h = layout height vector

# tree
w = c(edgeWidth,treeWidth)
m<-cbind(c(0,0,0),c(0,1,0)) # first two columns, edge + tree
x = 1

# info
if (!is.null(infoFile)) { # info is provided

  printCols = TRUE
  if (!is.null(infoCols)) {
    if (is.na(infoCols)) {
      printCols = FALSE
    }}

  if (printCols) {
    x = x + 1
    m<-cbind(m,c(0,x,0))
    w = c(w,infoWidth)
  }
}

# heatmap
if (!is.null(heatmapData)) {
  x = x + 1
  m<-cbind(m,c(x+1,x,0)) # add heatmap & labels
  x = x + 2
  m[1,2] = x # add heatmap scale above tree
  w = c(w,dataWidth)
}

# barplot
if (!is.null(barData)) {
  x = x + 1
  m<-cbind(m,c(0,x,x+1)) # barplot and scale bar
  x = x + 1
  w = c(w,barDataWidth)
}

if (doBlocks) {
  x = x + 1
  m<-cbind(m,c(0,x,0)) # recomb blocks
  w = c(w,blockPlotWidth)
}

# empty edge column
m<-cbind(m,c(0,0,0))
w = c(w,edgeWidth)

if (!is.null(heatmapData) | !is.null(barData)) { h = c(labelHeight,mainHeight,labelHeight) }
else { h = c(edgeWidth,mainHeight,edgeWidth) }
```

```
  return(list(m=as.matrix(m),w=w,h=h))
}
plotTree<-function(tree,ladderise=NULL,heatmapData=NULL,barData=NULL,infoFile=NULL,blockFile=NULL,snpFi

  require(ape)

  # PREPARE TREE, CHOOSE LADDERISATION OR NOT, AND GET TIP ORDER
  if (is.character(tree)){
    t<-read.tree(tree)
  }
  else t<-tree
  if (is.null(ladderise))
  {
    tl<-t
  }
  else if (ladderise=="descending")
  {
    tl<-ladderize(t, T)
  }
  else if (ladderise=="ascending")
  {
    tl<-ladderize(t, F)
  }
  else if (!is.null(ladderise))
  {
    print("Ladderise option should be exactly 'ascending' or 'descending'.  Any other command will raise
  }
  tips<-tl$edge[,2]
  tip.order<-tips[tips<=length(tl$tip.label)]
  tip.label.order<-tl$tip.label[tip.order] # for ordering data. note that for tiplabel(), the order is


  # PREPARE HEATMAP DATA
  if (!is.null(heatmapData)) {

    # read heatmap data and convert to data frame
    x<-readMatrix(heatmapData)

    # order rows of heatmap matrix to match tree
    y.ordered<-x[tip.label.order,]

    # reorder columns?
    if (!is.null(cluster)) {
      if (!(cluster==FALSE)) {

        if (cluster=="square" & ncol(y.ordered)==nrow(y.ordered)) {
          # order columns to match row order
          original_order<-1:nrow(x)
          names(original_order)<-rownames(x)
          reordered<-original_order[tip.label.order]
          y.ordered<-y.ordered[,rev(as.numeric(reordered))]
        }
```

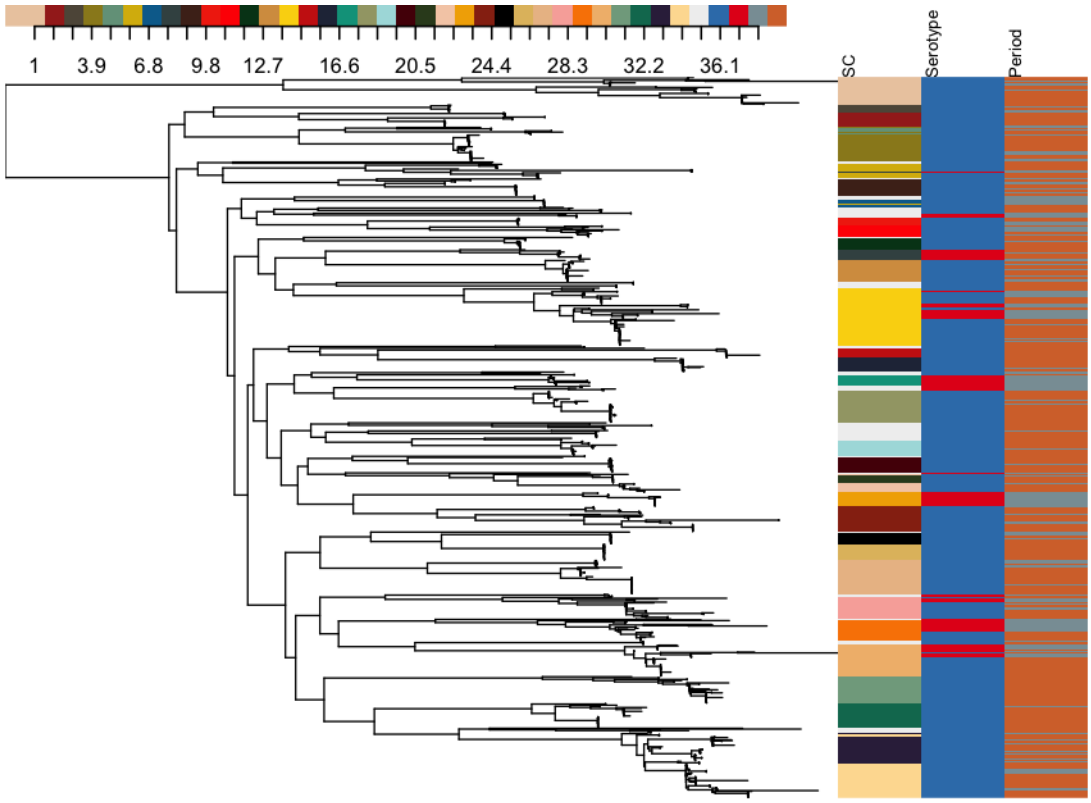
```
      else {
        # cluster columns
        if (cluster==TRUE) {cluster="ward"} # set default clustering algorithm
        h<-hclust(dist(t(na.omit(y.ordered))),cluster)
        y.ordered<-y.ordered[,h$order]
      }

    }} # finished reordering columns

} # finished setting up heatmap data


# PREPARE BAR PLOT
if (!is.null(barData)) {
  b<-readMatrix(barData)
  barData<-b[,1]
  names(barData)<-rownames(b)
}

# PREPARE INFO TO PRINT
if (!is.null(infoFile)) {
  info<-readMatrix(infoFile)
  info.ordered<-info[rev(tip.label.order),]
}
else {info.ordered=NULL}


# PREPARE DISCRETE TRAIT FOR COLOURING NODES AND INFERRING ANCESTRAL STATES
ancestral=NULL
nodeColourSuccess=NULL
if (!is.null(colourNodesBy) & !is.null(infoFile)) {

  if (colourNodesBy %in% colnames(info.ordered)) {
    nodeColourSuccess = TRUE
    loc1<-info.ordered[,which(colnames(info.ordered)==colourNodesBy)]

    # assign values
    tipLabelSet <- character(length(loc1))
    names(tipLabelSet) <- rownames(info.ordered)
    groups<-table(loc1,exclude="")
    n<-length(groups)
    groupNames<-names(groups)

    # set colours
    if (is.null(tipColours)){ colours<-rainbow(n) }
    else{ colours<-tipColours }

    # assign colours based on values
    for (i in 1:n) {
      g<-groupNames[i]
      tipLabelSet[loc1==g]<-colours[i]
    }
    tipLabelSet <- tipLabelSet[tl$tip]
```

```
    # ancestral reconstruction
    if (ancestral.reconstruction) { ancestral<-ace(loc1,tl,type="discrete") }

  }}
# finished with trait labels and ancestral reconstruction


# OPEN EXTERNAL DEVICE FOR DRAWING
# open PDF for drawing
if (!is.null(outputPDF)) {
  pdf(width=w,height=h,file=outputPDF)
}
# open PNG for drawing
if (!is.null(outputPNG)) {
  png(width=w,height=h,file=outputPNG)
}


# SET UP LAYOUT FOR PLOTTING
doBlocks <- (!is.null(blockFile) | !is.null(snpFile))
l <- getLayout(infoFile,infoCols,heatmapData,barData,doBlocks,treeWidth=treeWidth,infoWidth=infoWidth
layout(l$m, widths=l$w, heights=l$h)


# PLOT TREE
par(mar=rep(0,4))
tlp<-plot.phylo(tl,no.margin=T,show.tip.label=tip.labels,label.offset=offset,edge.width=lwd,edge.colo

# colour by trait
if (!is.null(nodeColourSuccess)) {
  tiplabels(col= tipLabelSet,pch=16,cex=tip.colour.cex)
  if (ancestral.reconstruction) { nodelabels(pie=ancestral$lik.anc, cex=pie.cex, piecol=colours) }
  if (legend) { legend(legend.pos,legend=groupNames,fill=colours) }
}

if (axis) { axisPhylo(axisPos) }

# PLOT INFO
if (!is.null(infoFile)) { # info is provided

  printCols = TRUE
  if (!is.null(infoCols)) {
    if (is.na(infoCols)) {
      printCols = FALSE
    }}

  if (printCols) {

    par(mar=rep(0,4))

    if (!is.null(infoCols)) {infoColNumbers = which(colnames(info.ordered) %in% infoCols)}
    else { infoColNumbers = 1:ncol(info.ordered)}
```

```
    plot(NA,axes=F,pch="",xlim=c(0,length(infoColNumbers)+1.5),ylim=c(0.5,length(tl$tip)+0.5),xaxs="i"

    # plot all info columns
    for (i in 1:length(infoColNumbers)) {
      j<-infoColNumbers[i]
      text(x=rep(i+1,nrow(info.ordered)+1),y=c((nrow(info.ordered)):1),info.ordered[,j],cex=infoCex)
    }

  }
}


# PLOT HEATMAP
if (!is.null(heatmapData)) {

  if (is.null(heatmapBreaks)) { heatmapBreaks = seq(min(y.ordered,na.rm=T),max(y.ordered,na.rm=T),leng

  # plot heatmap
  par(mar=rep(0,4), xpd=TRUE)
  image((1:ncol(y.ordered))-0.5,(1:nrow(y.ordered))-0.5, as.matrix(t(y.ordered)),col=heatmap.colours,l

  # draw vertical lines over heatmap
  if (!is.null(vlines.heatmap)) {
    for (v in vlines.heatmap) {abline(v=v, col=vlines.heatmap.col)}
  }

  # overlay blocks on heatmap
  if (!is.null(heatmap.blocks)) {
    for (coords in heatmap.blocks) {rect(xleft=coords[1], 0, coords[2], ncol(y.ordered), col=vlines.he
  }


  # data labels for heatmap
  par(mar=rep(0,4))
  plot(NA, axes=F, xaxs="i", yaxs="i", ylim=c(0,2), xlim=c(0.5,ncol(y.ordered)+0.5))
  text(1:ncol(y.ordered)-0.5,rep(0,ncol(x)),colnames(y.ordered), srt=90, cex=colLabelCex, pos=4)

  # scale for heatmap
  par(mar=c(2,0,0,2))
  #image(as.matrix(seq(min(y.ordered,na.rm=T),max(y.ordered,na.rm=T),length.out=length(heatmap.colours
  image(as.matrix(seq(min(y.ordered,na.rm=T),max(y.ordered,na.rm=T),length.out=length(heatmap.colours)
  axis(1,at=heatmapBreaks[-length(heatmapBreaks)]/max(y.ordered,na.rm=T),labels=round(heatmapBreaks[-1
}

# BARPLOT
if (!is.null(barData)) {
  par(mar=rep(0,4))
  barplot(barData[tip.label.order], horiz=T, axes=F, xaxs="i", yaxs="i", xlab="", ylab="", ylim=c(0.25

  # scale for barData plot
  par(mar=c(2,0,0,0))
  plot(NA, yaxt="n", xaxs="i", yaxs="i", xlab="", ylab="", ylim=c(0,2), xlim=c((-1)*max(barData,na.rm=
}
```

```
  # SNPS AND RECOMBINATION BLOCKS
  if (doBlocks) {
    par(mar=rep(0,4))
    plot(NA,axes=F,pch="",xlim=c(genome_offset,genome_offset+genome_size+1.5),ylim=c(0.5,length(tl$tip)+

    # plot snps
    if (!is.null(snpFile)) {
      snps<-read.csv(snpFile,header=F,row.names=1) # in case colnames start with numbers or contain das
      snps_strainCols <- snps[1,] # column names = strain names
      snps<-snps[-1,] # drop strain names

      for (strain in tip.label.order){
        # print SNPs compared to ancestral alleles in column 1
        s<-rownames(snps)[(as.character(snps[,1]) != as.character(snps[,which(snps_strainCols==strain)])]
        y <- which(tip.label.order==strain)
        if (length(s)>0) {
          for (x in s) {
            points(x,y,pch="|",col=snp_colour,cex=0.25)
          }
        }
      }
    }

    # plot blocks
    if (!is.null(blockFile)){
      blocks<-read.delim(blockFile,header=F)
      for (i in 1:nrow(blocks)) {
        if (as.character(blocks[i,1]) %in% tip.label.order) {
          y <- which(tip.label.order==as.character(blocks[i,1]))
          x1 <- blocks[i,2]
          x2 <- blocks[i,3]
          lines(c(x1,x2),c(y,y),lwd=blwd,lend=2,col=block_colour)
        }
      }
    }

  } # finished with SNPs and recomb blocks

  # CLOSE EXTERNAL DRAWING DEVICE
  if (!is.null(outputPDF) | !is.null(outputPNG)) {
    dev.off()
  }

  # RETURN ordered info and ancestral reconstruction object
  if (!is.null(heatmapData)){mat=as.matrix(t(y.ordered))}
  else {mat=NULL}
  return(list(info=info.ordered,anc=ancestral,mat=mat,strain_order=tip.label.order))
}


tree<-read.tree("RAxML_bestTree.All.Core.tre") #Core genome tree
#tree<-read.tree("RAxML_bestTree.All.Binary.tre") #Accessory genome tree
#tree <- ladderize(midpoint.root(tree), right = FALSE) #Ladderized the tree and midpoint root
```

```
heatmap_colors <- c("#ECCBAE", "#A42820", "#5F5647", "#9A8822", "#74A089", "#D8B70A",
                    "#046C9A", "#3F5151", "#4E2A1E", "#F2300F", "#FF0000", "#02401B",
                    "#D69C4E", "#FAD510", "#CB2314", "#273046", "#00A08A", "#A2A475",
                    "#ABDDDE", "#550307", "#354823", "#F5CDB4", "#F2AD00", "#972D15",
                    "#000000", "#E1BD6D", "#EABE94", "#F8AFA8", "#F98400", "#F1BB7B",
                    "#81A88D", "#0B775E", "#35274A", "#FDDDA0", "#f0f0f0", "#377eb8",
                    "#e41a1c","#899DA4", "#D67236")

plotTree(tree,
         heatmapData="NWMA_metadata_numerated_revised.csv",
         heatmap.colours=heatmap_colors,
         legend=T,
         tip.labels = FALSE,
         #tipColours=, tip.colour.cex=.8, legend.pos="bottomright",
         lwd=.85, treeWidth=10, dataWidth=3
         #infoCols="baps", infoCex=0
)
```



$info NULL

$anc NULL

$mat ERR907139_G5_19A_2010 ERR907140_G5_19A_2010 ERR907745_G2_19A_2007 SC 34 34 34 Serotype 36 36 36 Period 39 39 39 ERR907336_G5_19A_2010 ERR907314_G5_19A_2010 ERR907639_G2_19A_2007 SC 34 34 34 Serotype 36 36 36 Period 39 39 39 ERR907014_G5_19A_2010

ERR907723_G4_19A_2007 ERR907788_G3_19A_2007 SC 34 34 34 Serotype 36 36 36 Period 39 39 39 ERR907149_G5_19A_2010 ERR907101_G1_19A_1998 ERR907222_G1_19A_1998 SC 34 34 34 Serotype 36 36 36 Period 39 38 38 ERR907102_G1_19A_1998 ERR907704_G2_19A_2007 ERR907695_G2_19A_2007 SC 34 34 34 Serotype 36 36 36 Period 38 39 39 ERR907593_G4_19A_2006 ERR907340_G5_19A_2010 ERR907845_G2_19A_2007 SC 34 34 34 Serotype 36 36 36 Period 39 39 39 ERR907839_G2_19A_2007 ERR907869_G2_19A_2007 ERR907840_G3_19A_2007 SC 34 34 34 Serotype 36 36 36 Period 39 39 39 ERR907837_G3_19A_2007 ERR907773_G4_19A_2007 ERR906953_G5_19A_2010 SC 34 34 34 Serotype 36 36 36 Period 39 39 39 ERR907271_G1_19A_1999 ERR906955_G5_19A_2010 ERR907713_G3_19A_2007 SC 34 34 34 Serotype 36 36 36 Period 38 39 39 ERR907813_G3_19A_2007 ERR907860_G3_19A_2008 ERR907877_G3_19A_2008 SC 34 34 34 Serotype 36 36 36 Period 39 39 39 ERR907026_G5_19A_2011 ERR907399_G1_19A_1999 ERR907395_G1_19A_1999 SC 34 34 34 Serotype 36 36 36 Period 39 38 38 ERR907396_G1_19A_1999 ERR907398_G1_19A_1999 ERR907550_G1_19A_1999 SC 34 34 34 Serotype 36 36 36 Period 38 38 38 ERR907321_G5_19A_2010 ERR907233_G1_19A_1998 ERR907772_G2_19A_2007 SC 34 34 34 Serotype 36 36 36 Period 39 38 39 ERR907888_G3_19A_2007 ERR907824_G3_19A_2007 ERR907542_G3_19A_2007 SC 34 34 34 Serotype 36 36 36 Period 39 39 39 ERR907848_G2_19A_2007 ERR907672_G2_19A_2006 ERR907372_G5_15C_2011 SC 34 34 33 Serotype 36 36 36 Period 39 39 39 ERR907324_G5_15C_2010 ERR907373_G5_15C_2011 ERR907553_G1_15C_2000 SC 33 33 33 Serotype 36 36 36 Period 39 39 38 ERR907775_G2_7B_2007 ERR907835_G2_7B_2007 ERR907285_G1_7B_1999 SC 33 33 33 Serotype 36 36 36 Period 39 39 38 ERR907838_G3_7B_2007 ERR906998_G5_7B_2010 ERR907859_G2_7B_2007 SC 33 33 33 Serotype 36 36 36 Period 39 39 39 ERR907635_G2_7B_2007 ERR907634_G3_7B_2007 ERR907231_G1_7B_1998 SC 33 33 33 Serotype 36 36 36 Period 39 39 38 ERR907417_G1_15C_2000 ERR907875_G3_15C_2007 ERR907387_G1_15C_1999 SC 33 33 33 Serotype 36 36 36 Period 38 39 38 ERR907434_G1_15C_2001 ERR907103_G1_15C_1998 ERR907343_G5_15C_2010 SC 33 33 33 Serotype 36 36 36 Period 38 38 39 ERR907330_G5_15C_2010 ERR907318_G5_15C_2010 ERR907879_G3_15C_2007 SC 33 33 33 Serotype 36 36 36 Period 39 39 39 ERR907206_G1_15C_1999 ERR907755_G2_15B_2007 ERR907145_G5_15C_2010 SC 33 33 33 Serotype 36 36 36 Period 38 39 39 ERR907651_G4_15C_2007 ERR907236_G1_15C_1998 ERR907778_G4_15C_2007 SC 33 33 33 Serotype 36 36 36 Period 39 38 39 ERR907477_G2_15C_2007 ERR907472_G4_15C_2007 ERR907158_G5_15C_2011 SC 33 33 33 Serotype 36 36 36 Period 39 39 39 ERR907080_G1_15C_1998 ERR907389_G1_15C_1999 ERR907469_G4_15C_2007 SC 33 33 33 Serotype 36 36 36 Period 38 38 39 ERR907168_G5_15C_2011 ERR907725_G2_15C_2007 ERR907802_G4_19A_2007 SC 33 33 34 Serotype 36 36 36 Period 39 39 39 ERR907789_G2_19A_2007 ERR907697_G3_19A_2007 ERR907569_G1_15C_1999 SC 34 34 33 Serotype 36 36 36 Period 39 39 38 ERR907587_G4_9I_2006 ERR907774_G2_9I_2007 ERR907184_G5_9I_2012 SC 35 35 35 Serotype 36 36 36 Period 39 39 39 ERR907366_G5_9I_2011 ERR907291_G5_9I_2011 ERR907560_G1_23F_1999 SC 35 35 35 Serotype 36 36 37 Period 39 39 38 ERR907002_G5_19A_2010 ERR907489_G3_21_2006 ERR907490_G3_21_2006 SC 35 32 32 Serotype 36 36 36 Period 39 39 39 ERR907024_G5_21_2011 ERR907685_G3_21_2007 ERR907150_G5_21_2010 SC 32 32 32 Serotype 36 36 36 Period 39 39 39 ERR907640_G4_21_2006 ERR907645_G3_21_2006 ERR907696_G3_21_2007 SC 32 32 32 Serotype 36 36 36 Period 39 39 39 ERR907375_G5_21_2011 ERR907589_G3_21_2006 ERR907582_G4_21_2006 SC 32 32 32 Serotype 36 36 36 Period 39 39 39 ERR907687_G4_21_2007 ERR907579_G4_21_2006 ERR907598_G3_21_2007 SC 32 32 32 Serotype 36 36 36 Period 39 39 39 ERR907654_G2_21_2007 ERR907043_G5_15C_2012 ERR907198_G5_15C_2011 SC 32 32 32 Serotype 36 36 36 Period 39 39 39 ERR907193_G5_15C_2010 ERR907190_G5_15C_2010 ERR907817_G2_15C_2007 SC 32 32 32 Serotype 36 36 36 Period 39 39 39 ERR907522_G4_15C_2006 ERR907171_G5_15C_2011 ERR907305_G5_15C_2012 SC 32 32 32 Serotype 36 36 36 Period 39 39 39 ERR907306_G5_15C_2012 ERR907009_G5_15C_2010 ERR906986_G5_15C_2011 SC 32 32 32 Serotype 36 36 36 Period 39 39 39 ERR907039_G5_15C_2011 ERR907244_G1_21_1999 ERR907461_G1_21_1998 SC 32 32 32 Serotype 36 36 36 Period 39 38 38 ERR907131_G5_21_2012 ERR907304_G5_21_2012 ERR907155_G5_21_2010 SC 32 32 32 Serotype 36 36 36 Period 39 39 39 ERR907849_G2_23A_2007 ERR907829_G4_23A_2007 ERR907501_G2_23A_2007 SC 31 31 31 Serotype 36 36 36 Period 39 39 39 ERR907602_G3_23A_2006 ERR907702_G3_23A_2007 ERR907871_G3_23A_2007 SC 31 31 31 Serotype 36 36 36 Period 39 39 39 ERR907873_G3_23A_2007 ERR906993_G5_15A_2010 ERR907187_G5_23A_2010 SC 31

31 31 Serotype 36 36 36 Period 39 39 39 ERR907865_G4_23A_2006 ERR907809_G4_23A_2007 ERR907836_G3_23A_2007 SC 31 31 31 Serotype 36 36 36 Period 39 39 39 ERR907313_G5_23A_2010 ERR907010_G5_23A_2010 ERR907015_G5_23A_2010 SC 31 31 31 Serotype 36 36 36 Period 39 39 39 ERR907170_G5_23A_2011 ERR907731_G4_23A_2007 ERR906985_G5_23A_2011 SC 31 31 31 Serotype 36 36 36 Period 39 39 39 ERR907029_G5_23A_2011 ERR907027_G5_23A_2011 ERR907157_G5_23A_2010 SC 31 31 31 Serotype 36 36 36 Period 39 39 39 ERR907710_G4_23A_2007 ERR907709_G4_23A_2007 ERR907717_G4_23A_2007 SC 31 31 31 Serotype 36 36 36 Period 39 39 39 ERR906990_G5_23A_2010 ERR906958_G5_23A_2010 ERR907146_G5_23A_2010 SC 31 31 31 Serotype 36 36 36 Period 39 39 39 ERR907693_G3_23A_2006 ERR906978_G5_23A_2011 ERR907202_G5_23A_2011 SC 31 31 31 Serotype 36 36 36 Period 39 39 39 ERR907686_G2_23A_2007 ERR907156_G5_23A_2010 ERR907790_G3_23B_2007 SC 31 31 31 Serotype 36 36 36 Period 39 39 39 ERR907797_G2_23B_2007 ERR907311_G5_23B_2012 ERR907821_G2_35B_2007 SC 31 31 30 Serotype 36 36 36 Period 39 39 39 ERR907820_G3_35B_2007 ERR906984_G5_35B_2011 ERR907822_G2_35B_2007 SC 30 30 30 Serotype 36 36 36 Period 39 39 39 ERR907169_G5_35B_2011 ERR907715_G3_35B_2007 ERR907309_G5_35B_2012 SC 30 30 30 Serotype 36 36 36 Period 39 39 39 ERR907505_G4_35B_2007 ERR907338_G5_35B_2010 ERR907861_G3_35B_2008 SC 30 30 30 Serotype 36 36 36 Period 39 39 39 ERR907506_G4_35B_2007 ERR907516_G3_35B_2007 ERR906987_G5_35B_2012 SC 30 30 30 Serotype 36 36 36 Period 39 39 39 ERR907025_G5_35B_2011 ERR907864_G3_35B_2008 ERR907012_G5_35B_2010 SC 30 30 30 Serotype 36 36 36 Period 39 39 39 ERR906972_G5_15C_2011 ERR907125_G5_15C_2011 ERR907344_G5_15C_2010 SC 30 30 30 Serotype 36 36 36 Period 39 39 39 ERR907166_G5_15C_2011 ERR907137_G5_15C_2010 ERR907165_G5_15C_2011 SC 30 30 30 Serotype 36 36 36 Period 39 39 39 ERR906973_G5_15C_2011 ERR907367_G5_15C_2011 ERR907047_G1_9A_1998 SC 30 30 30 Serotype 36 36 37 Period 39 39 38 ERR907237_G1_9A_1998 ERR907046_G1_9A_1998 ERR907246_G1_9A_1999 SC 30 30 30 Serotype 37 37 37 Period 38 38 38 ERR907418_G1_9A_2000 ERR907223_G1_9A_1998 ERR907462_G4_19A_2007 SC 30 30 30 Serotype 37 37 36 Period 38 38 39 ERR907045_G1_9A_1998 ERR907436_G1_9A_2001 ERR907637_G3_9A_2007 SC 30 30 30 Serotype 37 37 37 Period 38 38 39 ERR907050_G1_9A_1998 ERR907051_G1_9A_1998 ERR907218_G1_9A_1999 SC 30 30 30 Serotype 37 37 37 Period 38 38 38 ERR907111_G1_9A_1999 ERR907269_G1_9A_1999 ERR907413_G1_9A_2000 SC 30 30 30 Serotype 37 37 37 Period 38 38 38 ERR907574_G1_9A_1999 ERR907806_G3_1_2007 ERR907659_G2_1_2006 SC 30 35 35 Serotype 37 36 36 Period 38 39 39 ERR907677_G2_1_2007 ERR907212_G1_1_1999 ERR907112_G1_1_1999 SC 35 35 35 Serotype 36 36 36 Period 39 38 38 ERR907334_G5_6A_2010 ERR907006_G5_6A_2010 ERR907144_G5_6A_2010 SC 29 29 29 Serotype 36 36 36 Period 39 39 39 ERR906967_G5_6A_2010 ERR907005_G5_6A_2010 ERR907690_G3_6A_2006 SC 29 29 29 Serotype 36 36 36 Period 39 39 39 ERR907129_G5_6A_2012 ERR907322_G5_6A_2010 ERR907300_G5_6A_2011 SC 29 29 29 Serotype 36 36 36 Period 39 39 39 ERR907310_G5_6A_2012 ERR906956_G5_6A_2010 ERR906960_G5_6A_2010 SC 29 29 29 Serotype 36 36 36 Period 39 39 39 ERR907061_G1_6A_1998 ERR907070_G1_6A_1998 ERR907557_G1_6A_1999 SC 29 29 29 Serotype 37 37 37 Period 38 38 38 ERR907069_G1_6A_1998 ERR907433_G1_6A_2001 ERR907552_G1_6A_2000 SC 29 29 29 Serotype 37 37 37 Period 38 38 38 ERR907209_G1_6A_1999 ERR907277_G1_6A_1999 ERR907119_G5_6A_2011 SC 29 29 29 Serotype 37 37 36 Period 38 38 39 ERR907422_G1_6A_2000 ERR907458_G1_6A_1998 ERR907225_G1_6A_1998 SC 29 29 29 Serotype 37 37 37 Period 38 38 38 ERR907248_G1_6A_1999 ERR907084_G1_6A_1998 ERR907459_G1_6A_1998 SC 29 29 29 Serotype 37 37 37 Period 38 38 38 ERR907270_G1_6A_1999 ERR907213_G1_6A_1999 ERR907743_G3_6A_2007 SC 35 35 28 Serotype 37 37 36 Period 38 38 39 ERR907647_G3_6A_2007 ERR907332_G5_5_2010 ERR906996_G5_5_2010 SC 28 28 28 Serotype 36 36 36 Period 39 39 39 ERR907018_G5_6A_2010 ERR907810_G4_5_2007 ERR907044_G5_6A_2012 SC 28 28 28 Serotype 36 36 36 Period 39 39 39 ERR907195_G5_6A_2010 ERR907719_G2_6A_2007 ERR907479_G3_6A_2007 SC 28 28 28 Serotype 36 36 36 Period 39 39 39 ERR907320_G5_6A_2010 ERR907460_G1_6A_1998 ERR907561_G1_6A_1999 SC 28 28 28 Serotype 36 36 36 Period 39 38 38 ERR907260_G1_6A_1999 ERR907173_G5_6A_2011 ERR907538_G4_6A_2006 SC 28 28 28 Serotype 36 36 36 Period 38 39 39 ERR907537_G4_6A_2006 ERR907541_G2_6A_2006 ERR907400_G1_6A_1999 SC 28 28 28 Serotype 36 36 36 Period 39 39 38 ERR907701_G2_6A_2007 ERR907128_G5_6A_2012 ERR907712_G2_6B_2007 SC 28 28 28 Serotype 36 36 37 Period 39 39 39 ERR907215_G1_6B_1999 ERR907216_G1_6B_1999 ERR907076_G1_6A_1998 SC 28 28 28 Serotype 37

37 37 Period 38 38 38 ERR907224_G1_6B_1998 ERR907376_G1_6B_1999 ERR907779_G2_15A_2008 SC 28 28 28 Serotype 37 37 36 Period 38 38 39 ERR907540_G4_19F_2007 ERR907545_G3_19F_2007 ERR907088_G1_19F_1998 SC 35 35 35 Serotype 37 37 37 Period 39 39 38 ERR907293_G5_19F_2011 ERR907507_G2_34_2007 ERR907643_G3_34_2006 SC 35 27 27 Serotype 37 36 36 Period 39 39 39 ERR907793_G2_34_2007 ERR907742_G3_34_2007 ERR907358_G5_34_2011 SC 27 27 27 Serotype 36 36 36 Period 39 39 39 ERR907126_G5_34_2011 ERR907868_G4_34_2007 ERR907867_G3_34_2007 SC 27 27 27 Serotype 36 36 36 Period 39 39 39 ERR907756_G4_34_2007 ERR907473_G4_34_2007 ERR907470_G4_34_2007 SC 27 27 27 Serotype 36 36 36 Period 39 39 39 ERR907405_G1_34_1999 ERR907349_G5_34_2011 ERR907348_G5_34_2011 SC 27 27 27 Serotype 36 36 36 Period 38 39 39 ERR907681_G3_34_2007 ERR907886_G3_34_2006 ERR907498_G3_34_2006 SC 27 27 27 Serotype 36 36 36 Period 39 39 39 ERR907368_G5_34_2011 ERR907736_G2_34_2007 ERR907514_G2_34_2007 SC 27 27 27 Serotype 36 36 36 Period 39 39 39 ERR907153_G5_34_2010 ERR907172_G5_34_2011 ERR907186_G5_34_2012 SC 27 27 27 Serotype 36 36 36 Period 39 39 39 ERR907533_G2_34_2007 ERR907194_G5_34_2010 ERR907192_G5_34_2010 SC 27 27 27 Serotype 36 36 36 Period 39 39 39 ERR907191_G5_34_2010 ERR907544_G2_34_2007 ERR907798_G4_34_2007 SC 27 27 27 Serotype 36 36 36 Period 39 39 39 ERR907801_G3_34_2007 ERR907536_G3_34_2007 ERR907345_G5_34_2010 SC 27 27 27 Serotype 36 36 36 Period 39 39 39 ERR907265_G1_34_1999 ERR907846_G2_34_2007 ERR907439_G1_6B_2000 SC 27 27 27 Serotype 36 36 36 Period 38 39 38 ERR907110_G1_22F_1999 ERR907214_G1_22F_1999 ERR907048_G1_22F_1998 SC 27 27 27 Serotype 36 36 36 Period 38 38 38 ERR907733_G3_22F_2007 ERR907699_G4_22F_2007 ERR907429_G1_22F_2000 SC 27 27 27 Serotype 36 36 36 Period 39 39 38 ERR907430_G1_22F_2000 ERR907287_G1_22F_1999 ERR907441_G1_22F_1998 SC 27 27 27 Serotype 36 36 36 Period 38 38 38 ERR907308_G5_17F_2012 ERR906991_G5_17F_2010 ERR907003_G5_17F_2010 SC 26 26 26 Serotype 36 36 36 Period 39 39 39 ERR907535_G4_17F_2006 ERR907197_G5_17F_2011 ERR907783_G3_17F_2007 SC 26 26 26 Serotype 36 36 36 Period 39 39 39 ERR907592_G2_17F_2007 ERR907657_G2_17F_2007 ERR907870_G2_17F_2008 SC 26 26 26 Serotype 36 36 36 Period 39 39 39 ERR907510_G4_17F_2007 ERR907509_G4_17F_2007 ERR907312_G5_17F_2010 SC 26 26 26 Serotype 36 36 36 Period 39 39 39 ERR907316_G5_17F_2010 ERR907632_G4_17F_2007 ERR907474_G3_17F_2007 SC 26 26 26 Serotype 36 36 36 Period 39 39 39 ERR906975_G5_17F_2011 ERR907707_G2_17F_2007 ERR907718_G2_17F_2007 SC 26 26 26 Serotype 36 36 36 Period 39 39 39 ERR907456_G1_17F_1998 ERR907619_G4_17F_2007 ERR907030_G5_17F_2011 SC 26 26 26 Serotype 36 36 36 Period 38 39 39 ERR907586_G2_15A_2007 ERR907585_G3_15A_2007 ERR907475_G4_15A_2007 SC 25 25 25 Serotype 36 36 36 Period 39 39 39 ERR907878_G3_15A_2007 ERR907179_G5_15A_2012 ERR907159_G5_15A_2011 SC 25 25 25 Serotype 36 36 36 Period 39 39 39 ERR907264_G1_15A_1999 ERR907880_G3_15A_2007 ERR907480_G4_15A_2007 SC 25 25 25 Serotype 36 36 36 Period 38 39 39 ERR907481_G4_15A_2007 ERR907671_G3_15A_2006 ERR907251_G1_15A_1999 SC 25 25 25 Serotype 36 36 36 Period 39 39 38 ERR907238_G1_15A_1998 ERR907227_G1_15A_1998 ERR907106_G1_15A_1999 SC 25 25 25 Serotype 36 36 36 Period 38 38 38 ERR907383_G1_20_1999 ERR907705_G4_6B_2007 ERR907791_G4_6B_2007 SC 35 24 24 Serotype 36 36 36 Period 38 39 39 ERR907792_G4_6B_2007 ERR907800_G3_6B_2007 ERR907528_G4_6B_2006 SC 24 24 24 Serotype 36 36 36 Period 39 39 39 ERR907529_G4_6B_2006 ERR907782_G2_6B_2007 ERR907777_G4_6B_2007 SC 24 24 24 Serotype 36 36 36 Period 39 39 39 ERR907780_G3_6B_2007 ERR907885_G3_6B_2006 ERR907449_G1_6B_1998 SC 24 24 24 Serotype 36 36 36 Period 39 39 38 ERR907249_G1_6B_1999 ERR907075_G1_6B_1998 ERR907483_G2_6B_2006 SC 24 24 24 Serotype 36 36 36 Period 38 38 39 ERR907853_G3_6B_2007 ERR907833_G3_6B_2007 ERR907496_G2_6B_2007 SC 24 24 24 Serotype 36 36 36 Period 39 39 39 ERR907497_G2_6B_2007 ERR907388_G1_6B_1999 ERR907757_G4_6B_2007 SC 24 24 24 Serotype 36 36 36 Period 39 38 39 ERR907515_G4_6B_2007 ERR907098_G1_6B_1998 ERR907289_G1_6B_1999 SC 24 24 24 Serotype 36 36 36 Period 39 38 38 ERR907513_G2_6B_2006 ERR907508_G3_6B_2006 ERR907463_G2_6B_2006 SC 24 24 24 Serotype 36 36 36 Period 39 39 39 ERR907416_G1_6B_2000 ERR907132_G5_6B_2012 ERR907676_G4_6B_2007 SC 24 24 24 Serotype 36 36 36 Period 38 39 39 ERR907729_G2_6B_2007 ERR907727_G3_6B_2007 ERR907065_G1_6B_1998 SC 24 24 24 Serotype 36 36 36 Period 39 39 38 ERR907276_G1_6B_1999 ERR907421_G1_6A_2000 ERR907420_G1_6A_2000 SC 24 23 23 Serotype 36 37 37 Period 38 38 38 ERR907406_G1_6A_1999 ERR907424_G1_6A_2000 ERR907563_G1_6A_2000 SC 23 23 23 Serotype 37 37 37 Period 38 38 38 ERR907382_G1_6A_1999 ERR907408_G1_6A_1999

ERR907390_G1_6A_1999 SC 23 23 23 Serotype 37 37 37 Period 38 38 38 ERR907211_G1_6A_1999 ERR907210_G1_6A_1999 ERR907059_G1_6A_1998 SC 23 23 23 Serotype 37 37 37 Period 38 38 38 ERR907254_G1_6A_1999 ERR907559_G1_6A_1999 ERR907097_G1_6A_1998 SC 23 23 23 Serotype 37 37 37 Period 38 38 38 ERR907082_G1_6A_1998 ERR907086_G1_6A_1998 ERR907564_G1_6A_1999 SC 23 23 23 Serotype 37 37 37 Period 38 38 38 ERR907243_G1_6A_1998 ERR907874_G3_10B_2007 ERR907814_G3_10A_2007 SC 23 22 22 Serotype 37 36 36 Period 38 39 39 ERR907823_G3_10A_2007 ERR907665_G3_10B_2007 ERR907035_G5_10B_2011 SC 22 22 22 Serotype 36 36 36 Period 39 39 39 ERR907294_G5_10B_2011 ERR907620_G2_10B_2007 ERR907295_G5_10B_2011 SC 22 22 22 Serotype 36 36 36 Period 39 39 39 ERR906994_G5_10A_2010 ERR907714_G2_10A_2007 ERR907089_G1_10A_1998 SC 22 22 22 Serotype 36 36 36 Period 39 39 38 ERR907549_G1_10A_1999 ERR907622_G4_35F_2007 ERR907624_G4_35F_2007 SC 22 21 21 Serotype 36 36 36 Period 38 39 39 ERR907547_G4_35F_2007 ERR907616_G4_35F_2007 ERR907013_G5_35F_2010 SC 21 21 21 Serotype 36 36 36 Period 39 39 39 ERR907021_G5_35F_2011 ERR907785_G2_35F_2007 ERR907543_G2_35F_2007 SC 21 21 21 Serotype 36 36 36 Period 39 39 39 ERR907530_G3_35F_2006 ERR907085_G1_35F_1998 ERR907099_G1_35F_1998 SC 21 21 21 Serotype 36 36 36 Period 39 38 38 ERR907371_G5_10A_2011 ERR907275_G1_6A_1999 ERR907573_G1_10A_1999 SC 22 35 35 Serotype 36 37 36 Period 39 38 38 ERR907649_G3_15A_2006 ERR907648_G3_15A_2006 ERR907546_G4_15A_2006 SC 20 20 20 Serotype 36 36 36 Period 39 39 39 ERR907627_G2_15A_2006 ERR906952_G5_15A_2010 ERR907328_G5_15A_2010 SC 20 20 20 Serotype 36 36 36 Period 39 39 39 ERR907323_G5_15A_2010 ERR907728_G3_15A_2007 ERR907020_G5_15A_2011 SC 20 20 20 Serotype 36 36 36 Period 39 39 39 ERR907882_G3_15A_2007 ERR907567_G1_15A_2000 ERR907748_G3_15A_2007 SC 20 20 20 Serotype 36 36 36 Period 39 38 39 ERR907749_G3_15A_2007 ERR907827_G4_15A_2007 ERR907663_G3_15A_2007 SC 20 20 20 Serotype 36 36 36 Period 39 39 39 ERR907662_G3_15A_2007 ERR907752_G2_15A_2007 ERR907884_G3_15A_2007 SC 20 20 20 Serotype 36 36 36 Period 39 39 39 ERR907735_G3_15A_2007 ERR907226_G1_6B_1998 ERR907667_G4_10F_2007 SC 20 35 35 Serotype 36 37 36 Period 39 38 39 ERR907781_G2_10F_2008 ERR907776_G4_16F_2007 ERR907670_G2_16F_2007 SC 35 19 19 Serotype 36 36 36 Period 39 39 39 ERR907196_G5_16F_2010 ERR907511_G2_16F_2007 ERR907354_G5_16F_2011 SC 19 19 19 Serotype 36 36 36 Period 39 39 39 ERR907167_G5_16F_2011 ERR907826_G4_16F_2007 ERR907730_G4_16F_2007 SC 19 19 19 Serotype 36 36 36 Period 39 39 39 ERR907452_G1_16F_1998 ERR907794_G3_16F_2007 ERR907764_G3_16F_2007 SC 19 19 19 Serotype 36 36 36 Period 38 39 39 ERR907493_G4_16F_2006 ERR907521_G2_16F_2007 ERR907327_G5_16F_2010 SC 19 19 19 Serotype 36 36 36 Period 39 39 39 ERR907572_G1_16F_1999 ERR907016_G5_16F_2010 ERR907333_G5_16F_2010 SC 19 19 19 Serotype 36 36 36 Period 38 39 39 ERR907669_G2_16F_2006 ERR907123_G5_16F_2011 ERR907362_G5_16F_2011 SC 19 19 35 Serotype 36 36 36 Period 39 39 39 ERR907374_G5_16F_2011 ERR907594_G4_16F_2006 ERR907711_G2_16F_2007 SC 35 35 35 Serotype 36 36 36 Period 39 39 39 ERR907683_G4_38_2007 ERR907872_G3_38_2007 ERR907819_G4_38_2007 SC 35 35 35 Serotype 36 36 36 Period 39 39 39 ERR907807_G4_38_2007 ERR906968_G5_38_2011 ERR906969_G5_38_2011 SC 35 35 35 Serotype 36 36 36 Period 39 39 39 ERR907037_G5_38_2011 ERR907467_G2_38_2006 ERR907380_G1_16F_1999 SC 35 35 35 Serotype 36 36 36 Period 39 39 38 ERR907500_G2_16F_2006 ERR907658_G2_12A_2006 ERR907724_G3_12A_2007 SC 35 35 35 Serotype 36 36 36 Period 39 39 39 ERR907703_G2_12B_2007 ERR907038_G5_12A_2011 ERR907100_G1_12A_1998 SC 35 35 35 Serotype 36 36 36 Period 39 39 38 ERR907638_G3_12B_2007 ERR907825_G4_12B_2007 ERR907164_G5_7F_2011 SC 35 35 35 Serotype 36 36 36 Period 39 39 39 ERR907163_G5_7F_2011 ERR907486_G2_22F_2006 ERR907787_G2_22A_2007 SC 35 35 18 Serotype 36 36 36 Period 39 39 39 ERR907138_G5_22A_2010 ERR907630_G2_22A_2007 ERR907518_G4_22A_2006 SC 18 18 18 Serotype 36 36 36 Period 39 39 39 ERR907519_G4_22A_2006 ERR906988_G5_22A_2012 ERR907722_G3_22A_2007 SC 18 18 18 Serotype 36 36 36 Period 39 39 39 ERR907628_G4_22A_2007 ERR907464_G4_22A_2006 ERR907326_G5_22A_2010 SC 18 18 18 Serotype 36 36 36 Period 39 39 39 ERR907636_G2_22A_2007 ERR907476_G3_22A_2007 ERR907600_G3_22A_2007 SC 18 18 18 Serotype 36 36 36 Period 39 39 39 ERR907603_G2_22A_2007 ERR907795_G3_22A_2008 ERR907655_G2_22A_2006 SC 18 18 18 Serotype 36 36 36 Period 39 39 39 ERR907642_G3_22A_2007 ERR907147_G5_22A_2010 ERR907876_G3_22A_2008 SC 18 18 18 Serotype 36 36 36 Period 39 39 39 ERR907617_G3_22A_2007 ERR907618_G2_22A_2007 ERR907698_G4_22A_2007 SC 18 18 18 Serotype 36 36 36 Period 39 39 39 ERR907520_G3_22A_2007 ERR907282_G1_22A_1999

ERR907181_G5_22A_2012 SC 18 18 18 Serotype 36 36 36 Period 39 38 39 ERR907359_G5_22A_2011 ERR907335_G5_22A_2010 ERR907495_G3_22A_2006 SC 18 18 18 Serotype 36 36 36 Period 39 39 39 ERR907432_G1_22A_2000 ERR907284_G1_22A_1999 ERR906961_G5_22A_2010 SC 18 18 18 Serotype 36 36 36 Period 38 38 39 ERR907808_G3_22A_2007 ERR907127_G5_22A_2011 ERR907122_G5_22A_2011 SC 18 18 18 Serotype 36 36 36 Period 39 39 39 ERR907124_G5_22A_2011 ERR907771_G3_22A_2007 ERR907353_G5_22A_2011 SC 18 18 18 Serotype 36 36 36 Period 39 39 39 ERR907661_G3_22A_2007 ERR907675_G3_22A_2006 ERR907682_G2_22A_2007 SC 18 18 18 Serotype 36 36 36 Period 39 39 39 ERR907716_G3_22A_2007 ERR907272_G1_4_1999 ERR907262_G1_4_1999 SC 18 35 35 Serotype 36 37 37 Period 39 38 38 ERR907577_G1_4_2000 ERR907083_G1_4_1998 ERR907409_G1_4_2000 SC 35 35 35 Serotype 37 37 37 Period 38 38 38 ERR907379_G1_4_1999 ERR907104_G1_4_1999 ERR907444_G1_14_1998 SC 35 35 17 Serotype 37 37 37 Period 38 38 38 ERR907447_G1_14_1998 ERR907081_G1_14_1998 ERR907093_G1_14_1998 SC 17 17 17 Serotype 37 37 37 Period 38 38 38 ERR907445_G1_14_1998 ERR907431_G1_14_1999 ERR907288_G1_14_1999 SC 17 17 17 Serotype 37 37 37 Period 38 38 38 ERR907247_G1_14_1999 ERR907446_G1_14_1998 ERR907450_G1_14_1998 SC 17 17 17 Serotype 37 37 37 Period 38 38 38 ERR907079_G1_14_1998 ERR907235_G1_14_1998 ERR907393_G1_14_1999 SC 17 17 17 Serotype 37 37 37 Period 38 38 38 ERR907525_G3_23B_2007 ERR907526_G2_5_2007 ERR907066_G1_18B_1998 SC 35 35 35 Serotype 36 36 36 Period 39 39 38 ERR907063_G1_18B_1998 ERR907451_G1_18B_1998 ERR907759_G3_35B_2007 SC 35 35 16 Serotype 36 36 36 Period 38 38 39 ERR907761_G3_35B_2007 ERR907680_G4_35B_2007 ERR907769_G3_35B_2007 SC 16 16 16 Serotype 36 36 36 Period 39 39 39 ERR907423_G1_35B_2000 ERR906989_G5_35B_2010 ERR907369_G5_35B_2011 SC 16 16 16 Serotype 36 36 36 Period 38 39 39 ERR907001_G5_35B_2010 ERR907033_G5_35B_2011 ERR907296_G5_35B_2011 SC 16 16 16 Serotype 36 36 36 Period 39 39 39 ERR907850_G3_35B_2007 ERR907851_G2_35B_2007 ERR907863_G2_35B_2008 SC 16 16 16 Serotype 36 36 36 Period 39 39 39 ERR907883_G3_35B_2007 ERR906999_G5_35B_2010 ERR907357_G5_35B_2011 SC 16 16 16 Serotype 36 36 36 Period 39 39 39 ERR906997_G5_35B_2010 ERR907183_G5_35B_2012 ERR907178_G5_19A_2012 SC 16 16 15 Serotype 36 36 36 Period 39 39 39 ERR907152_G5_19A_2010 ERR907471_G4_19A_2006 ERR907488_G2_19A_2007 SC 15 15 15 Serotype 36 36 36 Period 39 39 39 ERR907760_G3_19A_2007 ERR907762_G3_19A_2007 ERR907004_G5_19A_2010 SC 15 15 15 Serotype 36 36 36 Period 39 39 39 ERR907008_G5_19A_2010 ERR907644_G3_19A_2006 ERR906950_G5_19A_2010 SC 15 15 15 Serotype 36 36 36 Period 39 39 39 ERR907189_G5_19A_2010 ERR907034_G5_19A_2011 ERR907415_G1_6A_2000 SC 15 15 35 Serotype 36 36 37 Period 39 39 38 ERR907370_G5_15A_2011 ERR906992_G5_15A_2010 ERR907394_G1_14_1999 SC 35 35 17 Serotype 36 36 37 Period 39 39 38 ERR907758_G3_23B_2007 ERR907855_G2_23B_2007 ERR907737_G2_23B_2007 SC 14 14 14 Serotype 36 36 36 Period 39 39 39 ERR907626_G4_23B_2007 ERR907438_G1_23B_1999 ERR907425_G1_23B_1999 SC 14 14 14 Serotype 36 36 36 Period 39 38 38 ERR907361_G5_23B_2011 ERR907203_G5_23B_2011 ERR907092_G1_23B_1998 SC 14 14 14 Serotype 36 36 36 Period 39 39 38 ERR907763_G3_23B_2007 ERR907740_G2_23B_2007 ERR907011_G5_23B_2010 SC 14 14 14 Serotype 36 36 36 Period 39 39 39 ERR907641_G4_23B_2006 ERR907818_G2_23B_2007 ERR907517_G4_23B_2007 SC 14 14 14 Serotype 36 36 36 Period 39 39 39 ERR907601_G2_23B_2006 ERR907121_G5_23B_2011 ERR907019_G5_23B_2010 SC 14 14 14 Serotype 36 36 36 Period 39 39 39 ERR907666_G2_23B_2006 ERR907465_G3_23B_2007 ERR907141_G5_23B_2010 SC 14 14 14 Serotype 36 36 36 Period 39 39 39 ERR907267_G1_23B_1999 ERR907307_G5_23B_2012 ERR907847_G3_23B_2007 SC 14 14 14 Serotype 36 36 36 Period 38 39 39 ERR907352_G5_23B_2011 ERR906995_G5_23B_2010 ERR907154_G5_23B_2010 SC 14 14 14 Serotype 36 36 36 Period 39 39 39 ERR907562_G1_23B_2000 ERR907738_G3_23B_2007 ERR907468_G4_23B_2007 SC 14 14 14 Serotype 36 36 36 Period 38 39 39 ERR907621_G3_23B_2007 ERR907811_G4_23B_2007 ERR907142_G5_23B_2010 SC 14 14 14 Serotype 36 36 36 Period 39 39 39 ERR907844_G3_23B_2007 ERR907852_G3_23B_2007 ERR907072_G1_23F_1998 SC 14 14 14 Serotype 36 36 37 Period 39 39 38 ERR907053_G1_23F_1998 ERR907554_G1_23F_2000 ERR907252_G1_23F_1999 SC 14 14 14 Serotype 37 37 37 Period 38 38 38 ERR907234_G1_23F_1998 ERR907558_G1_23F_1999 ERR907457_G1_23F_1998 SC 14 14 14 Serotype 37 37 37 Period 38 38 38 ERR907239_G1_23F_1998 ERR907281_G1_23F_1999 ERR907058_G1_23F_1998 SC 14 14 14 Serotype 37 37 37 Period 38 38 38 ERR906946_G1_23F_1999 ERR907245_G1_23F_1999 ERR907527_G3_23A_2007 SC 14 14 14 Serotype 37 37 36 Period 38

38 39 ERR907734_G4_23A_2007 ERR907410_G1_23F_2000 ERR907455_G1_23F_1998 SC 14 14 14 Serotype 36 37 37 Period 39 38 38 ERR907453_G1_23F_1998 ERR907386_G1_23F_1999 ERR907207_G1_23F_1999 SC 14 14 14 Serotype 37 37 37 Period 38 38 38 ERR906957_G5_23B_2010 ERR907721_G2_23B_2007 ERR907803_G3_23B_2007 SC 14 14 14 Serotype 36 36 36 Period 39 39 39 ERR907804_G3_23B_2007 ERR907363_G5_23B_2011 ERR906964_G5_23B_2010 SC 14 14 14 Serotype 36 36 36 Period 39 39 39 ERR907341_G5_23B_2010 ERR906951_G5_23B_2010 ERR907161_G5_23B_2011 SC 14 14 14 Serotype 36 36 36 Period 39 39 39 ERR907055_G1_23B_1998 ERR907036_G5_23B_2011 ERR907414_G1_23A_2000 SC 14 14 14 Serotype 36 36 36 Period 38 39 38 ERR907280_G1_23A_1999 ERR907551_G1_23A_2000 ERR907268_G1_23A_1999 SC 14 14 14 Serotype 36 36 36 Period 38 38 38 ERR907257_G1_23F_1999 ERR907258_G1_23F_1999 ERR906971_G5_23A_2011 SC 14 14 14 Serotype 37 37 36 Period 38 38 39 ERR907815_G2_23A_2007 ERR907623_G2_23A_2006 ERR907599_G4_23A_2007 SC 14 14 14 Serotype 36 36 36 Period 39 39 39 ERR906959_G5_15C_2010 ERR907200_G5_15C_2011 ERR907298_G5_15C_2011 SC 35 35 35 Serotype 36 36 36 Period 39 39 39 ERR907784_G3_15C_2007 ERR907660_G2_16F_2007 ERR907606_G4_16F_2007 SC 35 35 35 Serotype 36 36 36 Period 39 39 39 ERR907607_G4_16F_2007 ERR907605_G4_16F_2007 ERR907217_G1_35B_1999 SC 35 35 13 Serotype 36 36 36 Period 39 39 38 ERR907435_G1_35B_2001 ERR907385_G1_35B_1999 ERR907232_G1_35B_1998 SC 13 13 13 Serotype 36 36 36 Period 38 38 38 ERR907650_G2_35B_2006 ERR907120_G5_35B_2011 ERR907325_G5_35B_2010 SC 13 13 13 Serotype 36 36 36 Period 39 39 39 ERR907060_G1_35B_1998 ERR907279_G1_35B_1999 ERR907841_G2_35B_2007 SC 13 13 13 Serotype 36 36 36 Period 38 38 39 ERR907747_G3_35B_2007 ERR907765_G2_35B_2007 ERR907531_G2_35B_2007 SC 13 13 13 Serotype 36 36 36 Period 39 39 39 ERR907889_G3_35B_2007 ERR907842_G2_35B_2007 ERR907240_G1_35B_1998 SC 13 13 13 Serotype 36 36 36 Period 39 39 38 ERR907402_G1_35B_1999 ERR907487_G3_35B_2007 ERR907766_G3_35B_2007 SC 13 13 13 Serotype 36 36 36 Period 38 39 39 ERR907610_G2_35B_2007 ERR907656_G3_35B_2007 ERR907175_G5_35B_2011 SC 13 13 13 Serotype 36 36 36 Period 39 39 39 ERR907378_G1_35B_2000 ERR907365_G5_35B_2011 ERR907609_G4_35B_2007 SC 13 13 13 Serotype 36 36 36 Period 38 39 39 ERR907342_G5_35B_2010 ERR907524_G2_35B_2006 ERR907078_G1_35B_1998 SC 13 13 13 Serotype 36 36 36 Period 39 39 38 ERR907556_G1_19F_1999 ERR907255_G1_19F_1999 ERR907664_G2_19F_2007 SC 8 8 8 Serotype 37 37 37 Period 38 38 39 ERR907443_G1_19F_1998 ERR907130_G5_19F_2012 ERR907364_G5_19F_2011 SC 8 8 8 Serotype 37 37 37 Period 38 39 39 ERR907339_G5_19F_2010 ERR907746_G3_19F_2007 ERR907056_G1_19F_1998 SC 8 8 8 Serotype 37 37 37 Period 39 39 38 ERR907068_G1_19F_1998 ERR907377_G1_19F_1999 ERR907629_G4_19F_2007 SC 8 8 8 Serotype 37 37 37 Period 38 38 39 ERR907631_G4_19F_2007 ERR907064_G1_19F_1998 ERR907858_G3_35A_2007 SC 8 8 12 Serotype 37 37 36 Period 39 38 39 ERR907856_G2_35A_2007 ERR907484_G4_35A_2007 ERR907678_G4_35A_2007 SC 12 12 12 Serotype 36 36 36 Period 39 39 39 ERR907504_G4_35A_2007 ERR907503_G4_35A_2007 ERR907384_G1_35A_1999 SC 12 12 12 Serotype 36 36 36 Period 39 39 38 ERR907031_G5_35A_2011 ERR906977_G5_35A_2011 ERR907625_G3_35A_2006 SC 12 12 12 Serotype 36 36 36 Period 39 39 39 ERR907381_G1_35A_1999 ERR907590_G4_35A_2006 ERR907887_G3_35A_2006 SC 12 12 12 Serotype 36 36 36 Period 38 39 39 ERR907532_G4_35A_2006 ERR907263_G1_35A_1999 ERR907816_G2_5_2007 SC 12 12 35 Serotype 36 36 36 Period 39 38 39 ERR907583_G4_17F_2006 ERR907584_G2_17F_2006 ERR907094_G1_17F_1998 SC 11 11 11 Serotype 36 36 36 Period 39 39 38 ERR906974_G5_17F_2011 ERR906965_G5_17F_2010 ERR907437_G1_17F_1999 SC 11 11 11 Serotype 36 36 36 Period 39 39 38 ERR907499_G2_17F_2007 ERR907115_G1_17F_1999 ERR907831_G4_17F_2007 SC 11 11 11 Serotype 36 36 36 Period 39 38 39 ERR907555_G1_17F_2000 ERR907017_G5_17F_2010 ERR907054_G1_13_1998 SC 11 11 11 Serotype 36 36 36 Period 38 39 38 ERR907095_G1_13_1998 ERR906966_G5_13_2010 ERR906976_G5_13_2011 SC 11 11 11 Serotype 36 36 36 Period 38 39 39 ERR907442_G1_35F_1998 ERR907062_G1_35F_1998 ERR907057_G1_35F_1998 SC 10 10 10 Serotype 36 36 36 Period 38 38 38 ERR907407_G1_35F_1999 ERR907116_G1_35F_1999 ERR907220_G1_35F_1998 SC 10 10 10 Serotype 36 36 36 Period 38 38 38 ERR907744_G3_35F_2007 ERR907290_G5_35F_2011 ERR907768_G3_35F_2007 SC 10 10 10 Serotype 36 36 36 Period 39 39 39 ERR907741_G3_35F_2007 ERR907242_G1_23F_1998 ERR907448_G1_23F_1998 SC 10 35 35 Serotype 36 37 37 Period 39 38 38 ERR907403_G1_23F_1999 ERR907397_G1_23F_1999 ERR907568_G1_23F_2000 SC 35 35 35 Serotype 37 37 37 Period 38 38 38 ERR907419_G1_NT_2000 ERR907266_G1_NT_1999

ERR907862_G2_37_2008 SC 35 35 35 Serotype 36 36 36 Period 38 38 39 ERR907796_G2_37_2007 ERR907478_G2_37_2007 ERR907219_G1_33A_1998 SC 35 35 35 Serotype 36 36 37 Period 39 39 38 ERR907751_G3_7A_2007 ERR907834_G4_7F_2007 ERR907404_G1_7A_1999 SC 35 35 35 Serotype 36 36 36 Period 39 39 38 ERR907595_G3_19A_2006 ERR907633_G4_19A_2006 ERR906947_G3_19A_2007 SC 7 7 7 Serotype 36 36 36 Period 39 39 39 ERR907105_G1_5_1999 ERR907107_G1_19A_1999 ERR907108_G1_19A_1999 SC 6 7 7 Serotype 36 36 36 Period 38 38 38 ERR907278_G1_19A_1999 ERR907440_G1_19A_2000 ERR907113_G1_19A_1999 SC 7 7 7 Serotype 36 36 36 Period 38 38 38 ERR907077_G1_19A_1998 ERR907096_G1_9I_1998 ERR907250_G1_9I_1999 SC 7 35 35 Serotype 36 36 36 Period 38 38 38 ERR907090_G1_9I_1998 ERR907071_G1_9I_1998 ERR907454_G1_9I_1998 SC 35 35 35 Serotype 36 36 36 Period 38 38 38 ERR907615_G2_31_2006 ERR907494_G3_31_2007 ERR907114_G1_31_1999 SC 9 9 9 Serotype 36 36 36 Period 39 39 38 ERR907117_G1_31_1999 ERR906980_G5_31_2011 ERR907032_G5_31_2011 SC 9 9 9 Serotype 36 36 36 Period 38 39 39 ERR907604_G2_31_2006 ERR907854_G3_31_2007 ERR907426_G1_31_1999 SC 9 9 9 Serotype 36 36 36 Period 39 39 38 ERR907028_G5_31_2011 ERR907302_G5_31_2012 ERR907299_G5_31_2011 SC 9 9 9 Serotype 36 36 36 Period 39 39 39 ERR907185_G5_31_2012 ERR907612_G3_31_2007 ERR907109_G1_31_1999 SC 9 9 9 Serotype 36 36 36 Period 39 39 38 ERR907204_G1_31_1999 ERR907087_G1_31_1998 ERR907767_G3_31_2007 SC 9 9 9 Serotype 36 36 36 Period 38 38 39 ERR907739_G2_31_2007 ERR907684_G4_31_2007 ERR907700_G3_31_2007 SC 9 9 9 Serotype 36 36 36 Period 39 39 39 ERR907049_G1_31_1998 ERR907812_G2_7C_2007 ERR907843_G3_7C_2007 SC 35 6 6 Serotype 36 36 36 Period 38 39 39 ERR907832_G2_7C_2007 ERR906954_G5_7C_2010 ERR907041_G5_7C_2011 SC 6 6 6 Serotype 36 36 36 Period 39 39 39 ERR907040_G5_7C_2011 ERR907182_G5_7C_2012 ERR907355_G5_7C_2011 SC 6 6 6 Serotype 36 36 36 Period 39 39 39 ERR907205_G1_25F_1999 ERR907391_G1_5_1999 ERR907199_G5_5_2011 SC 8 6 6 Serotype 37 36 36 Period 38 38 39 ERR907201_G5_5_2011 ERR907575_G1_19F_1999 ERR907356_G5_7C_2011 SC 6 8 6 Serotype 36 37 36 Period 39 38 39 ERR907674_G4_7C_2007 ERR907162_G5_7C_2011 ERR907160_G5_7C_2011 SC 6 6 6 Serotype 36 36 36 Period 39 39 39 ERR907317_G5_7C_2010 ERR907118_G5_7C_2011 ERR907588_G3_8_2006 SC 6 6 35 Serotype 36 36 36 Period 39 39 39 ERR907286_G1_8_1999 ERR907151_G5_11A_2010 ERR907392_G1_11D_1999 SC 35 35 4 Serotype 36 36 36 Period 38 39 38 ERR907570_G1_11D_1999 ERR907228_G1_11D_1998 ERR907866_G2_11D_2007 SC 4 4 4 Serotype 36 36 36 Period 38 38 39 ERR907350_G5_11D_2011 ERR906963_G5_11D_2010 ERR906962_G5_11D_2010 SC 4 4 4 Serotype 36 36 36 Period 39 39 39 ERR907613_G3_11D_2007 ERR907259_G1_11D_1999 ERR907256_G1_11D_1999 SC 4 4 4 Serotype 36 36 36 Period 39 38 38 ERR907253_G1_11D_1999 ERR907229_G1_11D_1998 ERR907273_G1_11D_1999 SC 4 4 4 Serotype 36 36 36 Period 38 38 38 ERR907337_G5_11D_2010 ERR907136_G5_11D_2010 ERR907135_G5_11D_2010 SC 4 4 4 Serotype 36 36 36 Period 39 39 39 ERR907652_G4_11D_2006 ERR907539_G3_11D_2007 ERR907679_G4_11D_2007 SC 4 4 4 Serotype 36 36 36 Period 39 39 39 ERR907608_G4_11D_2006 ERR907694_G2_11D_2007 ERR907668_G3_11D_2007 SC 4 4 4 Serotype 36 36 36 Period 39 39 39 ERR907708_G3_11D_2007 ERR907578_G2_11D_2006 ERR907581_G4_11D_2006 SC 4 4 4 Serotype 36 36 36 Period 39 39 39 ERR907042_G5_11D_2012 ERR907534_G3_11D_2006 ERR907188_G5_11D_2010 SC 4 4 4 Serotype 36 36 36 Period 39 39 39 ERR907770_G4_11D_2007 ERR907732_G3_11D_2007 ERR907726_G2_11D_2007 SC 4 4 4 Serotype 36 36 36 Period 39 39 39 ERR907706_G2_11D_2007 ERR907720_G3_11D_2007 ERR907133_G5_11D_2010 SC 4 4 4 Serotype 36 36 36 Period 39 39 39 ERR907571_G1_11D_1999 ERR907523_G2_37_2007 ERR907653_G2_37_2007 SC 4 5 5 Serotype 36 36 36 Period 38 39 39 ERR907673_G4_33A_2007 ERR907174_G5_37_2011 ERR907799_G3_37_2007 SC 4 5 5 Serotype 36 36 36 Period 39 39 39 ERR907750_G3_37_2007 ERR907274_G1_37_1999 ERR907828_G4_37_2007 SC 5 5 5 Serotype 36 36 36 Period 39 38 39 ERR907576_G1_37_2000 ERR907208_G1_33F_1999 ERR907591_G4_3_2006 SC 5 4 2 Serotype 36 36 36 Period 38 38 39 ERR907261_G1_3_1999 ERR907331_G5_3_2010 ERR907805_G3_3_2007 SC 2 2 2 Serotype 36 36 36 Period 38 39 39 ERR907007_G5_3_2010 ERR907022_G5_3_2011 ERR907512_G2_3_2006 SC 2 2 2 Serotype 36 36 36 Period 39 39 39 ERR907315_G5_3_2010 ERR907830_G4_3_2007 ERR907689_G4_3_2006 SC 2 2 2 Serotype 36 36 36 Period 39 39 39 ERR907688_G4_3_2006 ERR907466_G2_3_2007 ERR907297_G5_3_2011 SC 2 2 2 Serotype 36 36 36 Period 39 39 39 ERR907351_G5_3_2011 ERR907134_G5_3_2010 ERR907786_G2_3_2007 SC 2 2 2 Serotype 36 36 36 Period 39 39 39 ERR907492_G4_3_2006 ERR907857_G3_3_2007 ERR907482_G4_3_2007

SC 2 2 2 Serotype 36 36 36 Period 39 39 39 ERR907074_G1_10A_1998 ERR907073_G1_10A_1998 ERR907230_G1_10A_1998 SC 3 3 3 Serotype 36 36 36 Period 38 38 38 ERR906979_G5_10A_2011 ERR907646_G2_10A_2006 ERR907881_G3_10A_2007 SC 3 3 3 Serotype 36 36 36 Period 39 39 39 ERR907241_G1_10A_1998 ERR907565_G1_10A_1999 ERR907491_G3_10A_2007 SC 3 3 3 Serotype 36 36 36 Period 38 38 39 ERR907067_G1_10A_1998 ERR907000_G5_NT_2010 ERR907329_G5_NT_2010 SC 3 1 1 Serotype 36 36 36 Period 38 39 39 ERR906981_G5_NT_2011 ERR906949_G5_NT_2010 ERR907346_G5_NT_2011 SC 1 1 1 Serotype 36 36 36 Period 39 39 39 ERR907301_G5_NT_2011 ERR907611_G4_NT_2007 ERR907597_G3_NT_2007 SC 1 1 1 Serotype 36 36 36 Period 39 39 39 ERR907023_G5_NT_2011 ERR907428_G1_NT_2000 ERR907580_G4_NT_2007 SC 1 1 1 Serotype 36 36 36 Period 39 38 39 ERR907143_G5_NT_2010 ERR907502_G2_NT_2007 ERR907692_G4_NT_2007 SC 1 1 1 Serotype 36 36 36 Period 39 39 39 ERR907691_G4_NT_2007 ERR906983_G5_NT_2011 ERR907596_G2_NT_2007 SC 1 1 1 Serotype 36 36 36 Period 39 39 39 ERR907614_G3_NT_2007 ERR907427_G1_NT_1999 ERR907221_G1_NT_1998 SC 1 1 1 Serotype 36 36 36 Period 39 38 38 ERR907176_G5_NT_2011 ERR907177_G5_NT_2011 ERR906982_G5_NT_2011 SC 1 1 1 Serotype 36 36 36 Period 39 39 39 ERR907347_G5_NT_2011 ERR907412_G1_NT_2000 ERR907411_G1_NT_2000 SC 1 1 1 Serotype 36 36 36 Period 39 38 38 ERR907283_G1_NT_1999 ERR907052_G1_NT_1998 ERR907303_G5_NT_2012 SC 1 1 1 Serotype 36 36 36 Period 38 38 39 ERR907319_G5_NT_2010 ERR907401_G1_NT_1999 ERR907091_G1_NT_1998 SC 1 1 1 Serotype 36 36 36 Period 39 38 38 ERR906970_G5_NT_2011 ERR907753_G3_NT_2007 ERR907754_G2_NT_2007 SC 1 1 1 Serotype 36 36 36 Period 39 39 39 ERR907148_G5_NT_2010 SC 1 Serotype 36 Period 39

$strain_order [1] “ERR907139_G5_19A_2010” “ERR907140_G5_19A_2010” “ERR907745_G2_19A_2007”
[4] “ERR907336_G5_19A_2010” “ERR907314_G5_19A_2010” “ERR907639_G2_19A_2007”
[7] “ERR907014_G5_19A_2010” “ERR907723_G4_19A_2007” “ERR907788_G3_19A_2007”
[10] “ERR907149_G5_19A_2010” “ERR907101_G1_19A_1998” “ERR907222_G1_19A_1998”
[13] “ERR907102_G1_19A_1998” “ERR907704_G2_19A_2007” “ERR907695_G2_19A_2007”
[16] “ERR907593_G4_19A_2006” “ERR907340_G5_19A_2010” “ERR907845_G2_19A_2007”
[19] “ERR907839_G2_19A_2007” “ERR907869_G2_19A_2007” “ERR907840_G3_19A_2007”
[22] “ERR907837_G3_19A_2007” “ERR907773_G4_19A_2007” “ERR906953_G5_19A_2010”
[25] “ERR907271_G1_19A_1999” “ERR906955_G5_19A_2010” “ERR907713_G3_19A_2007”
[28] “ERR907813_G3_19A_2007” “ERR907860_G3_19A_2008” “ERR907877_G3_19A_2008”
[31] “ERR907026_G5_19A_2011” “ERR907399_G1_19A_1999” “ERR907395_G1_19A_1999”
[34] “ERR907396_G1_19A_1999” “ERR907398_G1_19A_1999” “ERR907550_G1_19A_1999”
[37] “ERR907321_G5_19A_2010” “ERR907233_G1_19A_1998” “ERR907772_G2_19A_2007”
[40] “ERR907888_G3_19A_2007” “ERR907824_G3_19A_2007” “ERR907542_G3_19A_2007”
[43] “ERR907848_G2_19A_2007” “ERR907672_G2_19A_2006” “ERR907372_G5_15C_2011”
[46] “ERR907324_G5_15C_2010” “ERR907373_G5_15C_2011” “ERR907553_G1_15C_2000”
[49] “ERR907775_G2_7B_2007” “ERR907835_G2_7B_2007” “ERR907285_G1_7B_1999” [52] “ERR907838_G3_7B_2007” “ERR906998_G5_7B_2010” “ERR907859_G2_7B_2007” [55] “ERR907635_G2_7B_2007” “ERR907634_G3_7B_2007” “ERR907231_G1_7B_1998” [58] “ERR907417_G1_15C_2000” “ERR907875_G3_15C_2007” “ERR907387_G1_15C_1999” [61] “ERR907434_G1_15C_2001” “ERR907103_G1_15C_1998” “ERR907343_G5_15C_2010” [64] “ERR907330_G5_15C_2010” “ERR907318_G5_15C_2010” “ERR907879_G3_15C_2007” [67] “ERR907206_G1_15C_1999” “ERR907755_G2_15B_2007” “ERR907145_G5_15C_2010” [70] “ERR907651_G4_15C_2007” “ERR907236_G1_15C_1998” “ERR907778_G4_15C_2007” [73] “ERR907477_G2_15C_2007” “ERR907472_G4_15C_2007” “ERR907158_G5_15C_2011” [76] “ERR907080_G1_15C_1998” “ERR907389_G1_15C_1999” “ERR907469_G4_15C_2007” [79] “ERR907168_G5_15C_2011” “ERR907725_G2_15C_2007” “ERR907802_G4_19A_2007” [82] “ERR907789_G2_19A_2007” “ERR907697_G3_19A_2007” “ERR907569_G1_15C_1999” [85] “ERR907587_G4_9I_2006” “ERR907774_G2_9I_2007” “ERR907184_G5_9I_2012” [88] “ERR907366_G5_9I_2011” “ERR907291_G5_9I_2011” “ERR907560_G1_23F_1999” [91] “ERR907002_G5_19A_2010” “ERR907489_G3_21_2006” “ERR907490_G3_21_2006” [94] “ERR907024_G5_21_2011” “ERR907685_G3_21_2007” “ERR907150_G5_21_2010” [97] “ERR907640_G4_21_2006” “ERR907645_G3_21_2006” “ERR907696_G3_21_2007” [100] “ERR907375_G5_21_2011” “ERR907589_G3_21_2006” “ERR907582_G4_21_2006” [103] “ERR907687_G4_21_2007” “ERR907579_G4_21_2006” “ERR907598_G3_21_2007” [106]

“ERR907654_G2_21_2007” “ERR907043_G5_15C_2012” “ERR907198_G5_15C_2011” [109]
“ERR907193_G5_15C_2010” “ERR907190_G5_15C_2010” “ERR907817_G2_15C_2007” [112]
“ERR907522_G4_15C_2006” “ERR907171_G5_15C_2011” “ERR907305_G5_15C_2012” [115]
“ERR907306_G5_15C_2012” “ERR907009_G5_15C_2010” “ERR906986_G5_15C_2011” [118]
“ERR907039_G5_15C_2011” “ERR907244_G1_21_1999” “ERR907461_G1_21_1998” [121] “ERR907131_G5_21_2012”
“ERR907304_G5_21_2012” “ERR907155_G5_21_2010” [124] “ERR907849_G2_23A_2007” “ERR907829_G4_23A_2007”
“ERR907501_G2_23A_2007” [127] “ERR907602_G3_23A_2006” “ERR907702_G3_23A_2007”
“ERR907871_G3_23A_2007” [130] “ERR907873_G3_23A_2007” “ERR906993_G5_15A_2010”
“ERR907187_G5_23A_2010” [133] “ERR907865_G4_23A_2006” “ERR907809_G4_23A_2007”
“ERR907836_G3_23A_2007” [136] “ERR907313_G5_23A_2010” “ERR907010_G5_23A_2010”
“ERR907015_G5_23A_2010” [139] “ERR907170_G5_23A_2011” “ERR907731_G4_23A_2007”
“ERR906985_G5_23A_2011” [142] “ERR907029_G5_23A_2011” “ERR907027_G5_23A_2011”
“ERR907157_G5_23A_2010” [145] “ERR907710_G4_23A_2007” “ERR907709_G4_23A_2007”
“ERR907717_G4_23A_2007” [148] “ERR906990_G5_23A_2010” “ERR906958_G5_23A_2010”
“ERR907146_G5_23A_2010” [151] “ERR907693_G3_23A_2006” “ERR906978_G5_23A_2011”
“ERR907202_G5_23A_2011” [154] “ERR907686_G2_23A_2007” “ERR907156_G5_23A_2010”
“ERR907790_G3_23B_2007” [157] “ERR907797_G2_23B_2007” “ERR907311_G5_23B_2012”
“ERR907821_G2_35B_2007” [160] “ERR907820_G3_35B_2007” “ERR906984_G5_35B_2011”
“ERR907822_G2_35B_2007” [163] “ERR907169_G5_35B_2011” “ERR907715_G3_35B_2007”
“ERR907309_G5_35B_2012” [166] “ERR907505_G4_35B_2007” “ERR907338_G5_35B_2010”
“ERR907861_G3_35B_2008” [169] “ERR907506_G4_35B_2007” “ERR907516_G3_35B_2007”
“ERR906987_G5_35B_2012” [172] “ERR907025_G5_35B_2011” “ERR907864_G3_35B_2008”
“ERR907012_G5_35B_2010” [175] “ERR906972_G5_15C_2011” “ERR907125_G5_15C_2011”
“ERR907344_G5_15C_2010” [178] “ERR907166_G5_15C_2011” “ERR907137_G5_15C_2010”
“ERR907165_G5_15C_2011” [181] “ERR906973_G5_15C_2011” “ERR907367_G5_15C_2011”
“ERR907047_G1_9A_1998” [184] “ERR907237_G1_9A_1998” “ERR907046_G1_9A_1998” “ERR907246_G1_9A_1999”
[187] “ERR907418_G1_9A_2000” “ERR907223_G1_9A_1998” “ERR907462_G4_19A_2007”
[190] “ERR907045_G1_9A_1998” “ERR907436_G1_9A_2001” “ERR907637_G3_9A_2007” [193]
“ERR907050_G1_9A_1998” “ERR907051_G1_9A_1998” “ERR907218_G1_9A_1999” [196] “ERR907111_G1_9A_1999”
“ERR907269_G1_9A_1999” “ERR907413_G1_9A_2000” [199] “ERR907574_G1_9A_1999” “ERR907806_G3_1_2007”
“ERR907659_G2_1_2006”
[202] “ERR907677_G2_1_2007” “ERR907212_G1_1_1999” “ERR907112_G1_1_1999”
[205] “ERR907334_G5_6A_2010” “ERR907006_G5_6A_2010” “ERR907144_G5_6A_2010” [208]
“ERR906967_G5_6A_2010” “ERR907005_G5_6A_2010” “ERR907690_G3_6A_2006” [211] “ERR907129_G5_6A_2012”
“ERR907322_G5_6A_2010” “ERR907300_G5_6A_2011” [214] “ERR907310_G5_6A_2012” “ERR906956_G5_6A_2010”
“ERR906960_G5_6A_2010” [217] “ERR907061_G1_6A_1998” “ERR907070_G1_6A_1998” “ERR907557_G1_6A_1999”
[220] “ERR907069_G1_6A_1998” “ERR907433_G1_6A_2001” “ERR907552_G1_6A_2000” [223]
“ERR907209_G1_6A_1999” “ERR907277_G1_6A_1999” “ERR907119_G5_6A_2011” [226] “ERR907422_G1_6A_2000”
“ERR907458_G1_6A_1998” “ERR907225_G1_6A_1998” [229] “ERR907248_G1_6A_1999” “ERR907084_G1_6A_1998”
“ERR907459_G1_6A_1998” [232] “ERR907270_G1_6A_1999” “ERR907213_G1_6A_1999” “ERR907743_G3_6A_2007”
[235] “ERR907647_G3_6A_2007” “ERR907332_G5_5_2010” “ERR906996_G5_5_2010”
[238] “ERR907018_G5_6A_2010” “ERR907810_G4_5_2007” “ERR907044_G5_6A_2012” [241]
“ERR907195_G5_6A_2010” “ERR907719_G2_6A_2007” “ERR907479_G3_6A_2007” [244] “ERR907320_G5_6A_2010”
“ERR907460_G1_6A_1998” “ERR907561_G1_6A_1999” [247] “ERR907260_G1_6A_1999” “ERR907173_G5_6A_2011”
“ERR907538_G4_6A_2006” [250] “ERR907537_G4_6A_2006” “ERR907541_G2_6A_2006” “ERR907400_G1_6A_1999”
[253] “ERR907701_G2_6A_2007” “ERR907128_G5_6A_2012” “ERR907712_G2_6B_2007” [256]
“ERR907215_G1_6B_1999” “ERR907216_G1_6B_1999” “ERR907076_G1_6A_1998” [259] “ERR907224_G1_6B_1998”
“ERR907376_G1_6B_1999” “ERR907779_G2_15A_2008” [262] “ERR907540_G4_19F_2007”
“ERR907545_G3_19F_2007” “ERR907088_G1_19F_1998” [265] “ERR907293_G5_19F_2011”
“ERR907507_G2_34_2007” “ERR907643_G3_34_2006” [268] “ERR907793_G2_34_2007” “ERR907742_G3_34_2007”
“ERR907358_G5_34_2011” [271] “ERR907126_G5_34_2011” “ERR907868_G4_34_2007” “ERR907867_G3_34_2007”
[274] “ERR907756_G4_34_2007” “ERR907473_G4_34_2007” “ERR907470_G4_34_2007” [277]
“ERR907405_G1_34_1999” “ERR907349_G5_34_2011” “ERR907348_G5_34_2011” [280] “ERR907681_G3_34_2007”
“ERR907886_G3_34_2006” “ERR907498_G3_34_2006” [283] “ERR907368_G5_34_2011” “ERR907736_G2_34_2007”

“ERR907514_G2_34_2007” [286] “ERR907153_G5_34_2010” “ERR907172_G5_34_2011” “ERR907186_G5_34_2012” [289] “ERR907533_G2_34_2007” “ERR907194_G5_34_2010” “ERR907192_G5_34_2010” [292] “ERR907191_G5_34_2010” “ERR907544_G2_34_2007” “ERR907798_G4_34_2007” [295] “ERR907801_G3_34_2007” “ERR907536_G3_34_2007” “ERR907345_G5_34_2010” [298] “ERR907265_G1_34_1999” “ERR907846_G2_34_2007” “ERR907439_G1_6B_2000” [301] “ERR907110_G1_22F_1999” “ERR907214_G1_22F_1999” “ERR907048_G1_22F_1998” [304] “ERR907733_G3_22F_2007” “ERR907699_G4_22F_2007” “ERR907429_G1_22F_2000” [307] “ERR907430_G1_22F_2000” “ERR907287_G1_22F_1999” “ERR907441_G1_22F_1998” [310] “ERR907308_G5_17F_2012” “ERR906991_G5_17F_2010” “ERR907003_G5_17F_2010” [313] “ERR907535_G4_17F_2006” “ERR907197_G5_17F_2011” “ERR907783_G3_17F_2007” [316] “ERR907592_G2_17F_2007” “ERR907657_G2_17F_2007” “ERR907870_G2_17F_2008” [319] “ERR907510_G4_17F_2007” “ERR907509_G4_17F_2007” “ERR907312_G5_17F_2010” [322] “ERR907316_G5_17F_2010” “ERR907632_G4_17F_2007” “ERR907474_G3_17F_2007” [325] “ERR906975_G5_17F_2011” “ERR907707_G2_17F_2007” “ERR907718_G2_17F_2007” [328] “ERR907456_G1_17F_1998” “ERR907619_G4_17F_2007” “ERR907030_G5_17F_2011” [331] “ERR907586_G2_15A_2007” “ERR907585_G3_15A_2007” “ERR907475_G4_15A_2007” [334] “ERR907878_G3_15A_2007” “ERR907179_G5_15A_2012” “ERR907159_G5_15A_2011” [337] “ERR907264_G1_15A_1999” “ERR907880_G3_15A_2007” “ERR907480_G4_15A_2007” [340] “ERR907481_G4_15A_2007” “ERR907671_G3_15A_2006” “ERR907251_G1_15A_1999” [343] “ERR907238_G1_15A_1998” “ERR907227_G1_15A_1998” “ERR907106_G1_15A_1999” [346] “ERR907383_G1_20_1999” “ERR907705_G4_6B_2007” “ERR907791_G4_6B_2007” [349] “ERR907792_G4_6B_2007” “ERR907800_G3_6B_2007” “ERR907528_G4_6B_2006” [352] “ERR907529_G4_6B_2006” “ERR907782_G2_6B_2007” “ERR907777_G4_6B_2007” [355] “ERR907780_G3_6B_2007” “ERR907885_G3_6B_2006” “ERR907449_G1_6B_1998” [358] “ERR907249_G1_6B_1999” “ERR907075_G1_6B_1998” “ERR907483_G2_6B_2006” [361] “ERR907853_G3_6B_2007” “ERR907833_G3_6B_2007” “ERR907496_G2_6B_2007” [364] “ERR907497_G2_6B_2007” “ERR907388_G1_6B_1999” “ERR907757_G4_6B_2007” [367] “ERR907515_G4_6B_2007” “ERR907098_G1_6B_1998” “ERR907289_G1_6B_1999” [370] “ERR907513_G2_6B_2006” “ERR907508_G3_6B_2006” “ERR907463_G2_6B_2006” [373] “ERR907416_G1_6B_2000” “ERR907132_G5_6B_2012” “ERR907676_G4_6B_2007” [376] “ERR907729_G2_6B_2007” “ERR907727_G3_6B_2007” “ERR907065_G1_6B_1998” [379] “ERR907276_G1_6B_1999” “ERR907421_G1_6A_2000” “ERR907420_G1_6A_2000” [382] “ERR907406_G1_6A_1999” “ERR907424_G1_6A_2000” “ERR907563_G1_6A_2000” [385] “ERR907382_G1_6A_1999” “ERR907408_G1_6A_1999” “ERR907390_G1_6A_1999” [388] “ERR907211_G1_6A_1999” “ERR907210_G1_6A_1999” “ERR907059_G1_6A_1998” [391] “ERR907254_G1_6A_1999” “ERR907559_G1_6A_1999” “ERR907097_G1_6A_1998” [394] “ERR907082_G1_6A_1998” “ERR907086_G1_6A_1998” “ERR907564_G1_6A_1999” [397] “ERR907243_G1_6A_1998” “ERR907874_G3_10B_2007” “ERR907814_G3_10A_2007” [400] “ERR907823_G3_10A_2007” “ERR907665_G3_10B_2007” “ERR907035_G5_10B_2011” [403] “ERR907294_G5_10B_2011” “ERR907620_G2_10B_2007” “ERR907295_G5_10B_2011” [406] “ERR906994_G5_10A_2010” “ERR907714_G2_10A_2007” “ERR907089_G1_10A_1998” [409] “ERR907549_G1_10A_1999” “ERR907622_G4_35F_2007” “ERR907624_G4_35F_2007” [412] “ERR907547_G4_35F_2007” “ERR907616_G4_35F_2007” “ERR907013_G5_35F_2010” [415] “ERR907021_G5_35F_2011” “ERR907785_G2_35F_2007” “ERR907543_G2_35F_2007” [418] “ERR907530_G3_35F_2006” “ERR907085_G1_35F_1998” “ERR907099_G1_35F_1998” [421] “ERR907371_G5_10A_2011” “ERR907275_G1_6A_1999” “ERR907573_G1_10A_1999” [424] “ERR907649_G3_15A_2006” “ERR907648_G3_15A_2006” “ERR907546_G4_15A_2006” [427] “ERR907627_G2_15A_2006” “ERR906952_G5_15A_2010” “ERR907328_G5_15A_2010” [430] “ERR907323_G5_15A_2010” “ERR907728_G3_15A_2007” “ERR907020_G5_15A_2011” [433] “ERR907882_G3_15A_2007” “ERR907567_G1_15A_2000” “ERR907748_G3_15A_2007” [436] “ERR907749_G3_15A_2007” “ERR907827_G4_15A_2007” “ERR907663_G3_15A_2007” [439] “ERR907662_G3_15A_2007” “ERR907752_G2_15A_2007” “ERR907884_G3_15A_2007” [442] “ERR907735_G3_15A_2007” “ERR907226_G1_6B_1998” “ERR907667_G4_10F_2007” [445] “ERR907781_G2_10F_2008” “ERR907776_G4_16F_2007” “ERR907670_G2_16F_2007” [448] “ERR907196_G5_16F_2010” “ERR907511_G2_16F_2007” “ERR907354_G5_16F_2011” [451] “ERR907167_G5_16F_2011” “ERR907826_G4_16F_2007” “ERR907730_G4_16F_2007” [454] “ERR907452_G1_16F_1998” “ERR907794_G3_16F_2007” “ERR907764_G3_16F_2007” [457] “ERR907493_G4_16F_2006” “ERR907521_G2_16F_2007” “ERR907327_G5_16F_2010” [460]

“ERR907572_G1_16F_1999” “ERR907016_G5_16F_2010” “ERR907333_G5_16F_2010” [463]
“ERR907669_G2_16F_2006” “ERR907123_G5_16F_2011” “ERR907362_G5_16F_2011” [466]
“ERR907374_G5_16F_2011” “ERR907594_G4_16F_2006” “ERR907711_G2_16F_2007” [469]
“ERR907683_G4_38_2007” “ERR907872_G3_38_2007” “ERR907819_G4_38_2007” [472] “ERR907807_G4_38_2007”
“ERR906968_G5_38_2011” “ERR906969_G5_38_2011” [475] “ERR907037_G5_38_2011” “ERR907467_G2_38_2006”
“ERR907380_G1_16F_1999” [478] “ERR907500_G2_16F_2006” “ERR907658_G2_12A_2006”
“ERR907724_G3_12A_2007” [481] “ERR907703_G2_12B_2007” “ERR907038_G5_12A_2011”
“ERR907100_G1_12A_1998” [484] “ERR907638_G3_12B_2007” “ERR907825_G4_12B_2007”
“ERR907164_G5_7F_2011” [487] “ERR907163_G5_7F_2011” “ERR907486_G2_22F_2006” “ERR907787_G2_22A_2007”
[490] “ERR907138_G5_22A_2010” “ERR907630_G2_22A_2007” “ERR907518_G4_22A_2006” [493]
“ERR907519_G4_22A_2006” “ERR906988_G5_22A_2012” “ERR907722_G3_22A_2007” [496]
“ERR907628_G4_22A_2007” “ERR907464_G4_22A_2006” “ERR907326_G5_22A_2010” [499]
“ERR907636_G2_22A_2007” “ERR907476_G3_22A_2007” “ERR907600_G3_22A_2007” [502]
“ERR907603_G2_22A_2007” “ERR907795_G3_22A_2008” “ERR907655_G2_22A_2006” [505]
“ERR907642_G3_22A_2007” “ERR907147_G5_22A_2010” “ERR907876_G3_22A_2008” [508]
“ERR907617_G3_22A_2007” “ERR907618_G2_22A_2007” “ERR907698_G4_22A_2007” [511]
“ERR907520_G3_22A_2007” “ERR907282_G1_22A_1999” “ERR907181_G5_22A_2012” [514]
“ERR907359_G5_22A_2011” “ERR907335_G5_22A_2010” “ERR907495_G3_22A_2006” [517]
“ERR907432_G1_22A_2000” “ERR907284_G1_22A_1999” “ERR906961_G5_22A_2010” [520]
“ERR907808_G3_22A_2007” “ERR907127_G5_22A_2011” “ERR907122_G5_22A_2011” [523]
“ERR907124_G5_22A_2011” “ERR907771_G3_22A_2007” “ERR907353_G5_22A_2011” [526]
“ERR907661_G3_22A_2007” “ERR907675_G3_22A_2006” “ERR907682_G2_22A_2007” [529]
“ERR907716_G3_22A_2007” “ERR907272_G1_4_1999” “ERR907262_G1_4_1999”
[532] “ERR907577_G1_4_2000” “ERR907083_G1_4_1998” “ERR907409_G1_4_2000”
[535] “ERR907379_G1_4_1999” “ERR907104_G1_4_1999” “ERR907444_G1_14_1998” [538]
“ERR907447_G1_14_1998” “ERR907081_G1_14_1998” “ERR907093_G1_14_1998” [541] “ERR907445_G1_14_1998”
“ERR907431_G1_14_1999” “ERR907288_G1_14_1999” [544] “ERR907247_G1_14_1999” “ERR907446_G1_14_1998”
“ERR907450_G1_14_1998” [547] “ERR907079_G1_14_1998” “ERR907235_G1_14_1998” “ERR907393_G1_14_1999”
[550] “ERR907525_G3_23B_2007” “ERR907526_G2_5_2007” “ERR907066_G1_18B_1998” [553]
“ERR907063_G1_18B_1998” “ERR907451_G1_18B_1998” “ERR907759_G3_35B_2007” [556]
“ERR907761_G3_35B_2007” “ERR907680_G4_35B_2007” “ERR907769_G3_35B_2007” [559]
“ERR907423_G1_35B_2000” “ERR906989_G5_35B_2010” “ERR907369_G5_35B_2011” [562]
“ERR907001_G5_35B_2010” “ERR907033_G5_35B_2011” “ERR907296_G5_35B_2011” [565]
“ERR907850_G3_35B_2007” “ERR907851_G2_35B_2007” “ERR907863_G2_35B_2008” [568]
“ERR907883_G3_35B_2007” “ERR906999_G5_35B_2010” “ERR907357_G5_35B_2011” [571]
“ERR906997_G5_35B_2010” “ERR907183_G5_35B_2012” “ERR907178_G5_19A_2012” [574]
“ERR907152_G5_19A_2010” “ERR907471_G4_19A_2006” “ERR907488_G2_19A_2007” [577]
“ERR907760_G3_19A_2007” “ERR907762_G3_19A_2007” “ERR907004_G5_19A_2010” [580]
“ERR907008_G5_19A_2010” “ERR907644_G3_19A_2006” “ERR906950_G5_19A_2010” [583]
“ERR907189_G5_19A_2010” “ERR907034_G5_19A_2011” “ERR907415_G1_6A_2000” [586]
“ERR907370_G5_15A_2011” “ERR906992_G5_15A_2010” “ERR907394_G1_14_1999” [589]
“ERR907758_G3_23B_2007” “ERR907855_G2_23B_2007” “ERR907737_G2_23B_2007” [592]
“ERR907626_G4_23B_2007” “ERR907438_G1_23B_1999” “ERR907425_G1_23B_1999” [595]
“ERR907361_G5_23B_2011” “ERR907203_G5_23B_2011” “ERR907092_G1_23B_1998” [598]
“ERR907763_G3_23B_2007” “ERR907740_G2_23B_2007” “ERR907011_G5_23B_2010” [601]
“ERR907641_G4_23B_2006” “ERR907818_G2_23B_2007” “ERR907517_G4_23B_2007” [604]
“ERR907601_G2_23B_2006” “ERR907121_G5_23B_2011” “ERR907019_G5_23B_2010” [607]
“ERR907666_G2_23B_2006” “ERR907465_G3_23B_2007” “ERR907141_G5_23B_2010” [610]
“ERR907267_G1_23B_1999” “ERR907307_G5_23B_2012” “ERR907847_G3_23B_2007” [613]
“ERR907352_G5_23B_2011” “ERR906995_G5_23B_2010” “ERR907154_G5_23B_2010” [616]
“ERR907562_G1_23B_2000” “ERR907738_G3_23B_2007” “ERR907468_G4_23B_2007” [619]
“ERR907621_G3_23B_2007” “ERR907811_G4_23B_2007” “ERR907142_G5_23B_2010” [622]
“ERR907844_G3_23B_2007” “ERR907852_G3_23B_2007” “ERR907072_G1_23F_1998” [625]
“ERR907053_G1_23F_1998” “ERR907554_G1_23F_2000” “ERR907252_G1_23F_1999” [628]

“ERR907234_G1_23F_1998” “ERR907558_G1_23F_1999” “ERR907457_G1_23F_1998” [631]
“ERR907239_G1_23F_1998” “ERR907281_G1_23F_1999” “ERR907058_G1_23F_1998” [634]
“ERR906946_G1_23F_1999” “ERR907245_G1_23F_1999” “ERR907527_G3_23A_2007” [637]
“ERR907734_G4_23A_2007” “ERR907410_G1_23F_2000” “ERR907455_G1_23F_1998” [640]
“ERR907453_G1_23F_1998” “ERR907386_G1_23F_1999” “ERR907207_G1_23F_1999” [643]
“ERR906957_G5_23B_2010” “ERR907721_G2_23B_2007” “ERR907803_G3_23B_2007” [646]
“ERR907804_G3_23B_2007” “ERR907363_G5_23B_2011” “ERR906964_G5_23B_2010” [649]
“ERR907341_G5_23B_2010” “ERR906951_G5_23B_2010” “ERR907161_G5_23B_2011” [652]
“ERR907055_G1_23B_1998” “ERR907036_G5_23B_2011” “ERR907414_G1_23A_2000” [655]
“ERR907280_G1_23A_1999” “ERR907551_G1_23A_2000” “ERR907268_G1_23A_1999” [658]
“ERR907257_G1_23F_1999” “ERR907258_G1_23F_1999” “ERR906971_G5_23A_2011” [661]
“ERR907815_G2_23A_2007” “ERR907623_G2_23A_2006” “ERR907599_G4_23A_2007” [664]
“ERR906959_G5_15C_2010” “ERR907200_G5_15C_2011” “ERR907298_G5_15C_2011” [667]
“ERR907784_G3_15C_2007” “ERR907660_G2_16F_2007” “ERR907606_G4_16F_2007” [670]
“ERR907607_G4_16F_2007” “ERR907605_G4_16F_2007” “ERR907217_G1_35B_1999” [673]
“ERR907435_G1_35B_2001” “ERR907385_G1_35B_1999” “ERR907232_G1_35B_1998” [676]
“ERR907650_G2_35B_2006” “ERR907120_G5_35B_2011” “ERR907325_G5_35B_2010” [679]
“ERR907060_G1_35B_1998” “ERR907279_G1_35B_1999” “ERR907841_G2_35B_2007” [682]
“ERR907747_G3_35B_2007” “ERR907765_G2_35B_2007” “ERR907531_G2_35B_2007” [685]
“ERR907889_G3_35B_2007” “ERR907842_G2_35B_2007” “ERR907240_G1_35B_1998” [688]
“ERR907402_G1_35B_1999” “ERR907487_G3_35B_2007” “ERR907766_G3_35B_2007” [691]
“ERR907610_G2_35B_2007” “ERR907656_G3_35B_2007” “ERR907175_G5_35B_2011” [694]
“ERR907378_G1_35B_2000” “ERR907365_G5_35B_2011” “ERR907609_G4_35B_2007” [697]
“ERR907342_G5_35B_2010” “ERR907524_G2_35B_2006” “ERR907078_G1_35B_1998” [700]
“ERR907556_G1_19F_1999” “ERR907255_G1_19F_1999” “ERR907664_G2_19F_2007” [703]
“ERR907443_G1_19F_1998” “ERR907130_G5_19F_2012” “ERR907364_G5_19F_2011” [706]
“ERR907339_G5_19F_2010” “ERR907746_G3_19F_2007” “ERR907056_G1_19F_1998” [709]
“ERR907068_G1_19F_1998” “ERR907377_G1_19F_1999” “ERR907629_G4_19F_2007” [712]
“ERR907631_G4_19F_2007” “ERR907064_G1_19F_1998” “ERR907858_G3_35A_2007” [715]
“ERR907856_G2_35A_2007” “ERR907484_G4_35A_2007” “ERR907678_G4_35A_2007” [718]
“ERR907504_G4_35A_2007” “ERR907503_G4_35A_2007” “ERR907384_G1_35A_1999” [721]
“ERR907031_G5_35A_2011” “ERR906977_G5_35A_2011” “ERR907625_G3_35A_2006” [724]
“ERR907381_G1_35A_1999” “ERR907590_G4_35A_2006” “ERR907887_G3_35A_2006” [727]
“ERR907532_G4_35A_2006” “ERR907263_G1_35A_1999” “ERR907816_G2_5_2007”
[730] “ERR907583_G4_17F_2006” “ERR907584_G2_17F_2006” “ERR907094_G1_17F_1998”
[733] “ERR906974_G5_17F_2011” “ERR906965_G5_17F_2010” “ERR907437_G1_17F_1999”
[736] “ERR907499_G2_17F_2007” “ERR907115_G1_17F_1999” “ERR907831_G4_17F_2007”
[739] “ERR907555_G1_17F_2000” “ERR907017_G5_17F_2010” “ERR907054_G1_13_1998”
[742] “ERR907095_G1_13_1998” “ERR906966_G5_13_2010” “ERR906976_G5_13_2011” [745]
“ERR907442_G1_35F_1998” “ERR907062_G1_35F_1998” “ERR907057_G1_35F_1998” [748]
“ERR907407_G1_35F_1999” “ERR907116_G1_35F_1999” “ERR907220_G1_35F_1998” [751]
“ERR907744_G3_35F_2007” “ERR907290_G5_35F_2011” “ERR907768_G3_35F_2007” [754]
“ERR907741_G3_35F_2007” “ERR907242_G1_23F_1998” “ERR907448_G1_23F_1998” [757]
“ERR907403_G1_23F_1999” “ERR907397_G1_23F_1999” “ERR907568_G1_23F_2000” [760]
“ERR907419_G1_NT_2000” “ERR907266_G1_NT_1999” “ERR907862_G2_37_2008” [763] “ERR907796_G2_37_2007”
“ERR907478_G2_37_2007” “ERR907219_G1_33A_1998” [766] “ERR907751_G3_7A_2007” “ERR907834_G4_7F_2007”
“ERR907404_G1_7A_1999” [769] “ERR907595_G3_19A_2006” “ERR907633_G4_19A_2006”
“ERR906947_G3_19A_2007” [772] “ERR907105_G1_5_1999” “ERR907107_G1_19A_1999” “ERR907108_G1_19A_1999”
[775] “ERR907278_G1_19A_1999” “ERR907440_G1_19A_2000” “ERR907113_G1_19A_1999”
[778] “ERR907077_G1_19A_1998” “ERR907096_G1_9I_1998” “ERR907250_G1_9I_1999” [781]
“ERR907090_G1_9I_1998” “ERR907071_G1_9I_1998” “ERR907454_G1_9I_1998” [784] “ERR907615_G2_31_2006”
“ERR907494_G3_31_2007” “ERR907114_G1_31_1999” [787] “ERR907117_G1_31_1999” “ERR906980_G5_31_2011”
“ERR907032_G5_31_2011” [790] “ERR907604_G2_31_2006” “ERR907854_G3_31_2007” “ERR907426_G1_31_1999”
[793] “ERR907028_G5_31_2011” “ERR907302_G5_31_2012” “ERR907299_G5_31_2011” [796]

"ERR907185_G5_31_2012" "ERR907612_G3_31_2007" "ERR907109_G1_31_1999" [799] "ERR907204_G1_31_1999"
"ERR907087_G1_31_1998" "ERR907767_G3_31_2007" [802] "ERR907739_G2_31_2007" "ERR907684_G4_31_2007"
"ERR907700_G3_31_2007" [805] "ERR907049_G1_31_1998" "ERR907812_G2_7C_2007" "ERR907843_G3_7C_2007"
[808] "ERR907832_G2_7C_2007" "ERR906954_G5_7C_2010" "ERR907041_G5_7C_2011" [811]
"ERR907040_G5_7C_2011" "ERR907182_G5_7C_2012" "ERR907355_G5_7C_2011" [814] "ERR907205_G1_25F_1999"
"ERR907391_G1_5_1999" "ERR907199_G5_5_2011"
[817] "ERR907201_G5_5_2011" "ERR907575_G1_19F_1999" "ERR907356_G5_7C_2011" [820]
"ERR907674_G4_7C_2007" "ERR907162_G5_7C_2011" "ERR907160_G5_7C_2011" [823] "ERR907317_G5_7C_2010"
"ERR907118_G5_7C_2011" "ERR907588_G3_8_2006"
[826] "ERR907286_G1_8_1999" "ERR907151_G5_11A_2010" "ERR907392_G1_11D_1999" [829]
"ERR907570_G1_11D_1999" "ERR907228_G1_11D_1998" "ERR907866_G2_11D_2007" [832]
"ERR907350_G5_11D_2011" "ERR906963_G5_11D_2010" "ERR906962_G5_11D_2010" [835]
"ERR907613_G3_11D_2007" "ERR907259_G1_11D_1999" "ERR907256_G1_11D_1999" [838]
"ERR907253_G1_11D_1999" "ERR907229_G1_11D_1998" "ERR907273_G1_11D_1999" [841]
"ERR907337_G5_11D_2010" "ERR907136_G5_11D_2010" "ERR907135_G5_11D_2010" [844]
"ERR907652_G4_11D_2006" "ERR907539_G3_11D_2007" "ERR907679_G4_11D_2007" [847]
"ERR907608_G4_11D_2006" "ERR907694_G2_11D_2007" "ERR907668_G3_11D_2007" [850]
"ERR907708_G3_11D_2007" "ERR907578_G2_11D_2006" "ERR907581_G4_11D_2006" [853]
"ERR907042_G5_11D_2012" "ERR907534_G3_11D_2006" "ERR907188_G5_11D_2010" [856]
"ERR907770_G4_11D_2007" "ERR907732_G3_11D_2007" "ERR907726_G2_11D_2007" [859]
"ERR907706_G2_11D_2007" "ERR907720_G3_11D_2007" "ERR907133_G5_11D_2010" [862]
"ERR907571_G1_11D_1999" "ERR907523_G2_37_2007" "ERR907653_G2_37_2007" [865] "ERR907673_G4_33A_2007"
"ERR907174_G5_37_2011" "ERR907799_G3_37_2007" [868] "ERR907750_G3_37_2007" "ERR907274_G1_37_1999"
"ERR907828_G4_37_2007" [871] "ERR907576_G1_37_2000" "ERR907208_G1_33F_1999" "ERR907591_G4_3_2006"
[874] "ERR907261_G1_3_1999" "ERR907331_G5_3_2010" "ERR907805_G3_3_2007"
[877] "ERR907007_G5_3_2010" "ERR907022_G5_3_2011" "ERR907512_G2_3_2006"
[880] "ERR907315_G5_3_2010" "ERR907830_G4_3_2007" "ERR907689_G4_3_2006"
[883] "ERR907688_G4_3_2006" "ERR907466_G2_3_2007" "ERR907297_G5_3_2011"
[886] "ERR907351_G5_3_2011" "ERR907134_G5_3_2010" "ERR907786_G2_3_2007"
[889] "ERR907492_G4_3_2006" "ERR907857_G3_3_2007" "ERR907482_G4_3_2007"
[892] "ERR907074_G1_10A_1998" "ERR907073_G1_10A_1998" "ERR907230_G1_10A_1998"
[895] "ERR906979_G5_10A_2011" "ERR907646_G2_10A_2006" "ERR907881_G3_10A_2007"
[898] "ERR907241_G1_10A_1998" "ERR907565_G1_10A_1999" "ERR907491_G3_10A_2007"
[901] "ERR907067_G1_10A_1998" "ERR907000_G5_NT_2010" "ERR907329_G5_NT_2010"
[904] "ERR906981_G5_NT_2011" "ERR906949_G5_NT_2010" "ERR907346_G5_NT_2011"
[907] "ERR907301_G5_NT_2011" "ERR907611_G4_NT_2007" "ERR907597_G3_NT_2007"
[910] "ERR907023_G5_NT_2011" "ERR907428_G1_NT_2000" "ERR907580_G4_NT_2007"
[913] "ERR907143_G5_NT_2010" "ERR907502_G2_NT_2007" "ERR907692_G4_NT_2007"
[916] "ERR907691_G4_NT_2007" "ERR906983_G5_NT_2011" "ERR907596_G2_NT_2007"
[919] "ERR907614_G3_NT_2007" "ERR907427_G1_NT_1999" "ERR907221_G1_NT_1998"
[922] "ERR907176_G5_NT_2011" "ERR907177_G5_NT_2011" "ERR906982_G5_NT_2011"
[925] "ERR907347_G5_NT_2011" "ERR907412_G1_NT_2000" "ERR907411_G1_NT_2000"
[928] "ERR907283_G1_NT_1999" "ERR907052_G1_NT_1998" "ERR907303_G5_NT_2012"
[931] "ERR907319_G5_NT_2010" "ERR907401_G1_NT_1999" "ERR907091_G1_NT_1998"
[934] "ERR906970_G5_NT_2011" "ERR907753_G3_NT_2007" "ERR907754_G2_NT_2007" [937]
"ERR907148_G5_NT_2010"

Fig. Suppl 2: Mass US Data

```
dfMA <- read_csv("data_MassUS.csv")

vaccineMA <- dfMA %>% distinct(SC, PCV7.actual)
vaccineMA <- vaccineMA %>% subset(PCV7.actual == "VT") %>%
```

```
  dplyr::rename(W = PCV7.actual) %>%
  full_join(subset(vaccineMA,PCV7.actual == "NVT")) %>%
  unite("vaccine", W:PCV7.actual, na.rm = TRUE)

dfE1MA <- dfMA %>% subset(Epoch1 == "E1") ## E1 prevaccine

dfFMA <- dfMA %>% select(SC, Epoch1) %>% group_by(Epoch1) %>%
  count(SC) %>% mutate(freq = prop.table(n)) %>% ungroup() %>%
  select(Epoch1, SC, freq) %>%
  spread(Epoch1, freq, fill = 0)


zero_E1MA <- dfE1MA %>% distinct(SC) %>%
  mutate(n = 0, freq = 0)

#### Replicates  - null expectation Pro rata ####
dfE1_NVT_allMA <- data.frame(NULL)
replicates <- 10000

for(i in 1:replicates){
  #sub-sampling from each epoch independently with replacemnt - best subsampling strategy
  dfE1_NVTMA_i <- dfE1MA %>% sample_frac(1, replace = TRUE) %>%
    subset(PCV7.actual == "NVT") %>% count(SC) %>%
    mutate(freq = prop.table(n)) %>% bind_rows(zero_E1MA) %>%
    group_by(SC) %>% summarise(n = sum(n), freq = sum(freq)) %>%
    arrange(SC) %>% ungroup %>% mutate(iter = i)

  dfE1_NVT_allMA <- bind_rows(dfE1_NVT_allMA, dfE1_NVTMA_i)
}

dfE1_NVT_allMA <- dfE1_NVT_allMA %>% group_by(SC) %>%
  summarise(expected_E3 = quantile(freq, 0.5), cil = quantile(freq, 0.025),
            ciu = quantile(freq, 0.975)) %>% ungroup

dfFMA <- left_join(dfFMA,dfE1_NVT_allMA) %>%
  replace(., is.na(.), 0) %>%
  mutate(delta = E3 - E1, deltaExp = expected_E3-E1,
         CI_low = cil-E1, CI_up = ciu-E1) %>%
  mutate(expectation = "Null exp.",
         signif = ifelse(delta > CI_up, "pos",NA)) %>%
  mutate(signif = ifelse(delta < CI_low, "neg", signif)) %>%
  mutate(signif = ifelse(E1 == 0 | E3 == 0, NA, signif)) %>%
  as_tibble(dfFMA) %>% mutate(SC = as.double(SC))

dfFMA <- left_join(as_tibble(dfFMA),vaccineMA)

#### Plot A: Prevalence by sequence cluster ####
dat1MA <- mutate(dfFMA, signif = ifelse(E3 > CI_up | E3 < CI_low, -0.005,NA), Epoch = "E1")
dat1MA <- subset(dat1MA, select = c(SC, Epoch, signif))

datPlotAMA <- dfFMA %>% select(SC, E1, E3) %>%
  pivot_longer(-SC, names_to =  "Epoch", values_to = "Prevalence") %>%
  left_join(dat1MA) %>% left_join(select(dfFMA, SC, vaccine)) %>%
```

```
  mutate(SC = as.character(SC)) %>% arrange(Epoch,-Prevalence)

rank <- as.character(unique(datPlotAMA$SC))

datPlotAMA <- datPlotAMA %>% mutate(SC = factor(SC, levels = rank))
dfFMA <- dfFMA %>% mutate(SC = factor(SC, levels = rank))

figureSA <- ggplot(datPlotAMA, aes(x=SC, y=Prevalence, alpha=Epoch, fill=vaccine)) +
  geom_bar(stat='identity', position='dodge') +
  scale_alpha_manual(values = c(1,0.4),
                     labels = c("Pre-vaccine", "Post-vaccine")) +
  scale_fill_manual(values = c("NVT"="#143c77","VT"="darkred","VT_NVT"="mediumpurple4"),
                    labels = c("NVT"="Non-vaccine type", "VT"="Vaccine type", "VT_NVT"="Mixed"),
                    name = "Composition") +
  xlab("Strain (SC)") + theme_classic() +
  scale_y_continuous(expand = c(0, 0), limits = c(0,0.3)) +
  theme(legend.title = element_text(face="bold", size = 10),
        legend.justification = c(1, 1), legend.box = "horizontal",
        legend.position = c(1, 1), # legend.position = c(0.725, 0.85),
        legend.spacing.y = unit(0.2, "cm"),
        legend.text=element_text(size=9),
        legend.background = element_blank(),
        legend.box.background = element_rect(fill = gray(0.96), color = NA))


#### Plot B: Change in prevalence ####
figureSB <- ggplot() + ylim(c(-0.2,0.35)) +
  geom_hline(yintercept=0, lty="dashed",size=0.65) +
  scale_colour_manual(values = c("NVT"="#143c77","VT"="darkred","VT_NVT"="mediumpurple4"),
                      labels = c("NVT"="Non-vaccine type", "VT"="Vaccine type", "VT_NVT"="Mixed"),
                      name = "SC Composition") +
  scale_linetype_manual(values = c(1), name = NULL) +
  scale_shape_manual(values = c("-", "+")) + theme_classic() +
  geom_point(data=dfFMA, aes(SC, delta, col = vaccine),
             size =4, alpha  = 0.85, show.legend = F) +
  geom_pointrange(data=dfFMA, aes(x=SC, y=deltaExp,
                                  ymin=CI_low, ymax=CI_up, lty=expectation),
                  size=.15, fatten = 6, show.legend = F) +
  geom_point(data=dfFMA, aes(x=SC, y=-0.2, shape = signif), size=6,
             col = "lightsteelblue4", show.legend = F) +
  labs(x="Strain (SC)") + labs(y="Change in prevalence") +
  theme(legend.title = element_text(face="bold", size = 10),
        legend.text=element_text(size=9),
        legend.position = c(0.01, 1),
        legend.justification = c(0.01, 1),
        legend.box = "horizontal",
        legend.background = element_blank(),
        legend.box.background = element_rect(fill = gray(0.95), color = NA))

######################

dfFVTMA <- dfMA %>% select(SC,PCV7.actual, Epoch1) %>%
  group_by(Epoch1) %>%  count(SC,PCV7.actual) %>%
```

```
  mutate(freq = round(prop.table(n), digits = 3)) %>%
  ungroup() %>% select(Epoch1, SC, PCV7.actual, freq) %>%
  spread(Epoch1, freq, fill = 0) %>% arrange(SC,PCV7.actual)

#### Present at E1 (17 SCs)####
SCE1MA <- dfFVTMA %>% subset(E1 > 0) %>% ##
  select(SC, PCV7.actual) %>%
  mutate(Epoch1 = "E1")

#### NVT present at E1 (9 SCs)####
SCE2MA <- SCE1MA %>% subset(PCV7.actual == "NVT")


SC_freq_dfMA <- dfMA %>% select(SC,PCV7.actual,Epoch1,pbp1a:CLS343169) %>%
  arrange(SC) %>% group_by(SC,PCV7.actual,Epoch1) %>%
  mutate(SC_n = n()) %>% ungroup() %>%
  group_by(SC,PCV7.actual,Epoch1,SC_n) %>%
  summarise_at(vars(pbp1a:CLS343169),mean) %>%
  ungroup()

### Get the matrix and the SC for the pre-vaccine epch "E1"
df_preVMA <- SCE1MA %>% left_join(SC_freq_dfMA)
SC_freq_preVMA <- as.matrix(df_preVMA %>%
                              mutate(SC_freq=SC_n/sum(SC_n)) %>% select(SC_freq))
SC_COG_preVMA <- as.matrix(t(df_preVMA %>% select(pbp1a:CLS343169)))

## Get e_l for the Mass data (1056 COGs)
el_MA <- SC_COG_preVMA %*% SC_freq_preVMA

#### Imputed data ####
dfImputedMA <- dfFVTMA %>%
  subset(PCV7.actual == "NVT" & E1 == 0 & E2 > 0) %>%
  select(SC,PCV7.actual) %>% mutate(Epoch1 = "E1", n=1) %>%
  select(Epoch1,SC,PCV7.actual,n)

dfFVTImputedMA <- dfMA %>% select(SC,PCV7.actual, Epoch1) %>%
  group_by(Epoch1) %>%  count(SC,PCV7.actual) %>%
  ungroup() %>% bind_rows(dfImputedMA) %>% group_by(Epoch1) %>%
  mutate(freq = round(prop.table(n), digits = 3)) %>% ungroup() %>%
  select(Epoch1, SC, PCV7.actual, freq) %>%
  spread(Epoch1, freq, fill = 0) %>% arrange(SC,PCV7.actual)

### SC13 is not present at E2
dfFNVTImputedMA <- dfFVTImputedMA %>%
  subset(PCV7.actual == "NVT" & SC != "13") %>%
  mutate(deltaE = E3 - E1) %>% arrange(SC)

#### E2 - frequencies just after vaccine intro ####
x_imputedMA <- dfFNVTImputedMA$E1
x_imputedMA <- as.matrix(round(x_imputedMA/sum(x_imputedMA), digits = 5))

dat2_imputedMA <- dfImputedMA %>% mutate(Epoch1 = "E2") %>%
  select(SC,PCV7.actual,Epoch1) %>% bind_rows(SCE2MA) %>%
```

```
  arrange(SC) %>% left_join(SC_freq_dfMA) %>%
  select(pbp1a:CLS343169)
dat2_imputedMA <- as.matrix(dat2_imputedMA)

fl_impMA <- t(dat2_imputedMA) %*% x_imputedMA

#### fitness function "omega" just after vaccine intro ####
wholeMA <- as.numeric(el_MA - fl_impMA)
omegaMA <- as.vector(dat2_imputedMA %*% wholeMA) ## length SCs, similar to FFS

phiMA <- sum(x_imputedMA * omegaMA) ## average fitness
rateOfChangeMA <- omegaMA - phiMA ## "rate of change": omega_g - phi

dat3ABMA <- dfFNVTImputedMA %>% select(SC, deltaE) %>%
  mutate(r = rateOfChangeMA) %>% subset(deltaE != 0) %>%
  mutate(change = ifelse(deltaE < 0, "Decreased", "Increased")) %>%
  mutate(col = ifelse(sign(deltaE) == sign(r), "same", "diff")) %>%
  left_join(vaccineMA)


lmS2 <- lm(r~deltaE,dat3ABMA)
summary(lmS2)
```

Call: lm(formula = r ~ deltaE, data = dat3ABMA)

Residuals: Min 1Q Median 3Q Max -14.1551 -1.9226 0.3457 3.0135 8.5055

Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) 4.413 2.481 1.779 0.101 deltaE 20.799 58.154 0.358 0.727

Residual standard error: 6.156 on 12 degrees of freedom Multiple R-squared: 0.01055, Adjusted R-squared: -0.07191 F-statistic: 0.1279 on 1 and 12 DF, p-value: 0.7268

```
highlightS <- dat3ABMA %>% filter(col == "diff")

#### Figure C ####
plotSC <- ggplot(dat3ABMA, aes(x=deltaE, y=r, colour = vaccine)) +
  theme_classic() +
  xlab("Observed Prevalence Change") +
  ylab("Standardized Fitness") + #xlim(-0.03, 0.065) + ylim(-12,20) +
  geom_hline(yintercept=0, lty="dotted", alpha=.6) +
  geom_vline(xintercept=0, lty="dotted", alpha=.6) +
  annotate("rect", xmin=-0.025,xmax=0,ymin=-10,ymax=0,
           fill="darkslategray4", alpha= 0.15) +
  annotate("rect", xmin=0,xmax=0.09,ymin=0,ymax=14,
           fill="darkslategray4", alpha= 0.15) +
  geom_smooth(aes(group = 1), color="grey70", method=lm,
              formula = y~x,  show.legend=FALSE, se=FALSE) +
  geom_point(size = 3) +
  theme(legend.position = "none",
        axis.text = element_text(colour = "black"),) +
  scale_colour_manual(values = c("#143c77","mediumorchid4")) +
  geom_text_repel(aes(label = paste("SC", SC, sep="-")),
                  data = highlightS, size = 3.5)

##################################
```

```
df_postVMA <- SCE2MA %>% left_join(SC_freq_dfMA)
SC_COG_postVMA <- as.matrix(t(df_postVMA %>% select(pbp1a:CLS343169)))

#### Predict postV frequencies ####
SC_freq_postV_predMA <- QP(SC_COG_postVMA, el_MA) #Matrix: rows = COGs, columns = (SCs - VT)

SC_freq_postV_obsMA <- SCE2MA %>% mutate(Epoch1 = "E3") %>%
  left_join(SC_freq_dfMA) %>%
  mutate(SC_freq=SC_n/sum(SC_n, na.rm = T)) %>%
  select(SC, PCV7.actual, SC_freq) %>%
  mutate(SC_freq = replace_na(SC_freq, 0))

SC_freq_postV_obsMA <- SC_freq_postV_obsMA %>%
  mutate(SC_pred = SC_freq_postV_predMA) %>%
  left_join(vaccineMA)

W_modelSB <- lm(data=SC_freq_postV_obsMA,SC_freq~SC_pred);
summary(W_modelSB)
```

Call: lm(formula = SC_freq ~ SC_pred, data = SC_freq_postV_obsMA)

Residuals: Min 1Q Median 3Q Max -0.073407 -0.018085 -0.008067 0.027381 0.068476

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.02328 0.02959 0.787 0.4572
SC_pred 0.79044 0.22626 3.494 0.0101 * — Signif. codes: 0 ‘’ ***0.001*** ’’ *0.01* ’’ 0.05 ‘.’ 0.1 ’ ’ 1

Residual standard error: 0.04684 on 7 degrees of freedom Multiple R-squared: 0.6355, Adjusted R-squared: 0.5834 F-statistic: 12.2 on 1 and 7 DF, p-value: 0.01008

```
confint(W_modelSB)
```

```
               2.5 %     97.5 %
```

(Intercept) -0.04669562 0.09326407 SC_pred 0.25542617 1.32545776

```
linearHypothesis(W_modelSB, c("(Intercept) = 0", "SC_pred = 1"), test = "Chisq")
```

Linear hypothesis test

Hypothesis: (Intercept) = 0 SC_pred = 1

Model 1: restricted model Model 2: SC_freq ~ SC_pred

Res.Df RSS Df Sum of Sq Chisq Pr(>Chisq) 1 9 0.017243
2 7 0.015361 2 0.0018824 0.8578 0.6512

```
## no outliers
outlierSB <- SC_freq_postV_obsMA %>% mutate(diff = abs(SC_freq - SC_pred))
outlierSB <- outlierSB %>%
  filter(diff %in% boxplot(outlierSB$diff, plot = FALSE)$out)

#### Figure D ####
plotSD <- ggplot(SC_freq_postV_obsMA,
                 aes(x = SC_pred, y = SC_freq, colour = vaccine)) +
  theme_classic() +
  geom_segment(aes(x=0,xend=0.3,y=0,yend=0.3),
               color="black",alpha=.7,lwd=0.5,lty=3) +
  geom_smooth(method='lm',color="#899DA4",
```

```
              formula=y~x, alpha=0.3, lwd=.6,
              fullrange=T, linetype="blank", show.legend=F) +
  annotate(geom = "text", x=0.28, y =0.29,
           label = "1:1 line", angle = 45, size = 3) +
  geom_point(size=3) +

  scale_x_continuous("Predicted Prevalence (NFDS)")+
  scale_y_continuous("Observed Prevalence") +
  coord_fixed(ratio = 1, ylim=c(0,0.3), xlim=c(0,0.3)) +
  scale_colour_manual(values = c("#143c77","mediumorchid4")) +
  theme(legend.position = "none",
        axis.text = element_text(colour = "black"),)

###################################

SC_freq_E1MA <- df_preVMA %>%
  mutate(SC_freq_E1=SC_n/sum(SC_n)) %>%
  select(SC,PCV7.actual,SC_freq_E1)

SC_freq_postV_diffMA <- SC_freq_postV_obsMA %>%
  left_join(SC_freq_E1MA) %>%
  mutate(diff_pred = SC_freq_E1 - SC_pred,
         diff_obs = SC_freq_E1 - SC_freq) %>%
  left_join(vaccineMA)

SC_freq_postV_diffMA <- SC_freq_postV_diffMA %>%
  mutate(diff = abs(diff_pred - diff_obs)) %>%
  arrange(diff)

stats <- summary(lm(SC_freq_postV_diffMA$diff_pred~SC_freq_postV_diffMA$diff_obs))
ars <- round(stats$adj.r.squared, digits = 3)
sseE <- round(sse(SC_freq_postV_diffMA$diff_obs, SC_freq_postV_diffMA$diff_pred), digits = 3)
maeE <- round(mae(SC_freq_postV_diffMA$diff_obs, SC_freq_postV_diffMA$diff_pred),digits = 3)##Mean Abso
rmseE <- round(rmse(SC_freq_postV_diffMA$diff_obs, SC_freq_postV_diffMA$diff_pred),digits = 3) ##Root M

W_modelSC <- lm(data=SC_freq_postV_diffMA,diff_obs~diff_pred);
summary(W_modelSC)
```

Call: lm(formula = diff_obs ~ diff_pred, data = SC_freq_postV_diffMA)

Residuals: Min 1Q Median 3Q Max -0.065605 -0.017260 -0.003470 0.006301 0.071991

Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) -0.02685 0.02422 -1.108 0.304 diff_pred 0.54087 0.33000 1.639 0.145

Residual standard error: 0.04393 on 7 degrees of freedom Multiple R-squared: 0.2773, Adjusted R-squared: 0.1741 F-statistic: 2.686 on 1 and 7 DF, p-value: 0.1452

```
confint(W_modelSC)
```

```
               2.5 %     97.5 %
```

(Intercept) -0.08413255 0.03043303 diff_pred -0.23946801 1.32120619

```
linearHypothesis(W_modelSC, c("(Intercept) = 0", "diff_pred = 1"), test = "Chisq")
```

Linear hypothesis test

Hypothesis: (Intercept) = 0 diff_pred = 1

Model 1: restricted model Model 2: diff_obs ~ diff_pred

Res.Df RSS Df Sum of Sq Chisq Pr(>Chisq) 1 9 0.017243
2 7 0.013508 2 0.0037353 1.9357 0.3799

```
#### No outlier ####
outlier3C <- SC_freq_postV_diffMA %>%
  filter(diff %in% boxplot(SC_freq_postV_diffMA$diff, plot = FALSE)$out)

#### Figure E ####
plotSE <- ggplot(SC_freq_postV_diffMA,
                 aes(x = diff_pred, y = diff_obs, colour = vaccine)) +
  geom_segment(aes(x=-0.2,xend=0.025,y=-0.2,yend=0.025),
               color="black",alpha=.7,lwd=0.5,lty=3) +
  geom_smooth(method='lm', color="gray80" ,formula=y~x,
              alpha=0.3, lwd=.6, fullrange=T,
              linetype="blank", show.legend=F) +
  annotate(geom = "text", x=0.01, y =0.0175,
           label = "1:1 line", angle = 45, size = 3) +
  geom_point(size=3) + theme_classic() +
  scale_colour_manual(values = c("#143c77","mediumorchid4"),
                      labels = c("Non-vaccine type", "Mixed"),
                      name = "SC Composition") +
  theme(legend.position = "none",
        axis.text = element_text(colour = "black"),) +
  xlab("Predicted Prevalence Change (NFDS)")+
  ylab("Observed Prevalence Change") +
  coord_fixed(ratio = 1, xlim=c(-0.2,0.025), ylim=c(-0.2,0.025)) +
  annotate("text",x=-0.19,y=0.01, size=2.5,hjust = 0,
           label=paste("SSE = ", sseE, "\nRMSE = ",
                       rmseE, "\nAdj. R2 = ", ars))

#### Figure F ####
SC_freq_postV_diffMA <- SC_freq_postV_diffMA %>%
  mutate(prorata = SC_freq_E1/sum(SC_freq_E1)) %>%
  mutate(diff_predPro = SC_freq_E1 - prorata) %>%
  mutate(diffP = abs(diff_predPro - diff_obs))

stats <- summary(lm(SC_freq_postV_diffMA$diff_predPro~SC_freq_postV_diffMA$diff_obs))
ars <- round(stats$adj.r.squared, digits = 3)
sseE <- round(sse(SC_freq_postV_diffMA$diff_obs, SC_freq_postV_diffMA$diff_predPro), digits = 3)
maeE <- round(mae(SC_freq_postV_diffMA$diff_obs, SC_freq_postV_diffMA$diff_predPro),digits = 3)##Mean A
rmseE <- round(rmse(SC_freq_postV_diffMA$diff_obs, SC_freq_postV_diffMA$diff_predPro),digits = 3) ##Roo

W_modelSD <- lm(data=SC_freq_postV_diffMA,diff_obs~diff_predPro);
summary(W_modelSD)
```

Call: lm(formula = diff_obs ~ diff_predPro, data = SC_freq_postV_diffMA)

Residuals: Min 1Q Median 3Q Max -0.05935 -0.04736 0.00284 0.01786 0.07442

Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) -0.04741 0.02675 -1.772 0.12 diff_predPro 0.18930 0.35483 0.533 0.61

Residual standard error: 0.05065 on 7 degrees of freedom Multiple R-squared: 0.03907, Adjusted R-squared:

-0.0982 F-statistic: 0.2846 on 1 and 7 DF, p-value: 0.6102

```
confint(W_modelSD)
```

```
               2.5 %     97.5 %
```

(Intercept) -0.1106683 0.01584975 diff_predPro -0.6497446 1.02834770

```
linearHypothesis(W_modelSD, c("(Intercept) = 0", "diff_predPro = 1"), test = "Chisq")
```

Linear hypothesis test

Hypothesis: (Intercept) = 0 diff_predPro = 1

Model 1: restricted model Model 2: diff_obs ~ diff_predPro

Res.Df RSS Df Sum of Sq Chisq Pr(>Chisq)
1 9 0.031356
2 7 0.017961 2 0.013394 5.22 0.07353 . — Signif. codes: 0 ‘’ ***0.001*** ’’ *0.01* ’’ 0.05 ‘.’ 0.1 ’ ’ 1

```
plotSF <- ggplot(SC_freq_postV_diffMA, aes(x = diff_predPro,
                                           y = diff_obs, colour = vaccine)) +
  geom_segment(aes(x=-0.2,xend=0.025,y=-0.2,yend=0.025),
               color="black",alpha=.7,lwd=0.5,lty=3) +
  geom_smooth(method='lm', color="gray80" ,formula=y~x,
              alpha=0.3, lwd=.6, fullrange=T,
              linetype="blank", show.legend=F) +
  annotate(geom = "text", x=0.01, y =0.0175,
           label = "1:1 line", angle = 45, size = 3) +
  geom_point(size=3) + theme_classic() +
  scale_colour_manual(values = c("#143c77","mediumorchid4"),
                      labels = c("Non-vaccine type", "Mixed"),
                      name = "SC Composition") +
  theme(legend.position = "none",
        axis.text = element_text(colour = "black"),) +
  xlab("Predicted Prevalence Change (Pro rata)")+
  ylab("Observed Prevalence Change") +
  coord_fixed(ratio = 1, xlim=c(-0.2,0.025), ylim=c(-0.2,0.025)) +
  annotate("text",x=-0.19,y=0.01, size=2.5,hjust = 0,
           label=paste("SSE = ", sseE, "\nRMSE = ",
                       rmseE, "\nAdj. R2 = ", ars))



figureSAB <- plot_grid(figureSA, figureSB, labels = c("A","B"), nrow = 1)
figureSCF <- plot_grid(plotSC, plotSD, plotSE, plotSF,
                       nrow = 1, labels=c("C","D","E","F"))

plotSPARC <- plot_grid(figureSAB,figureSCF, ncol=1)
ggsave("figureS_MASS.png", plotSPARC, width = 14, height = 7)
```

Fig. Suppl 3: Similarity of accessory genes among strains

```
#Accessory genome phylogeny
AG.tree <- read.tree("RAxML_bestTree.All.Binary.tre") #All

#Creating distance matrix from Tee
```

```
PatristicDistMatrix <- cophenetic(AG.tree) # patristic distances
PatristicDist <- as.dist(PatristicDistMatrix, diag = TRUE, upper = TRUE)

#Setting up group/SC assignments for between SC patristic distance
seq.labels <- as.data.frame(rownames(PatristicDistMatrix)) #Obtaining ordered taxa
colnames(seq.labels) <- "taxa"
clades <- as.data.frame(cbind(df$FinalName,df$BAPS2)) #All data set
colnames(clades) <- c("taxa","clade")
seq.labels$id  <- 1:nrow(clades) #adding row number to maintain order for sorting after merge
labels.clades <- merge(seq.labels, clades, by="taxa") #merging
labels.clades <- labels.clades[order(labels.clades$id), ] #ordering

#Creating Final Matrix
md <- meandist(PatristicDist, labels.clades$clade) #calculating mean distance between clades
md.matrix <- as.dist(md,diag = FALSE, upper = TRUE)
md.matrix <- as.matrix(md.matrix)
diag(md.matrix) <- NA

#Saving Accessory geneome distances in long format - needed for Sup Figure 4
PatristicDist.long <- as.data.frame(as.table(md.matrix))
PatristicDist.long <- PatristicDist.long[! (PatristicDist.long$Var1 == PatristicDist.long$Var2),]
PatristicDist.long <- PatristicDist.long[!is.na(PatristicDist.long$Freq),]
AccGenomeDistances <- PatristicDist.long #Sup figure 4

#Heatmap of between strain accessory genome distances
my_palette <- colorRampPalette(c("#4d004b", "#8c96c6", "#e0ecf4"))(n = 100)

pdf('figureS_Heatmap.pdf', width = 10, height = 10) #Change as needed
gplots::heatmap.2(md.matrix,
                  col = my_palette,
                  scale = "none",
                  na.rm = TRUE,
                  dendrogram = "row",
                  trace="none",
                  cexCol = .7, cexRow=.7,
                  labRow = rownames(md),
                  srtCol = 70, #Changes angle of X-axis
                  key=TRUE,
                  key.title=NA,
                  key.xlab = "Patristic \nDistance"
)
dev.off()
```

```
## quartz_off_screen
##                 2
```

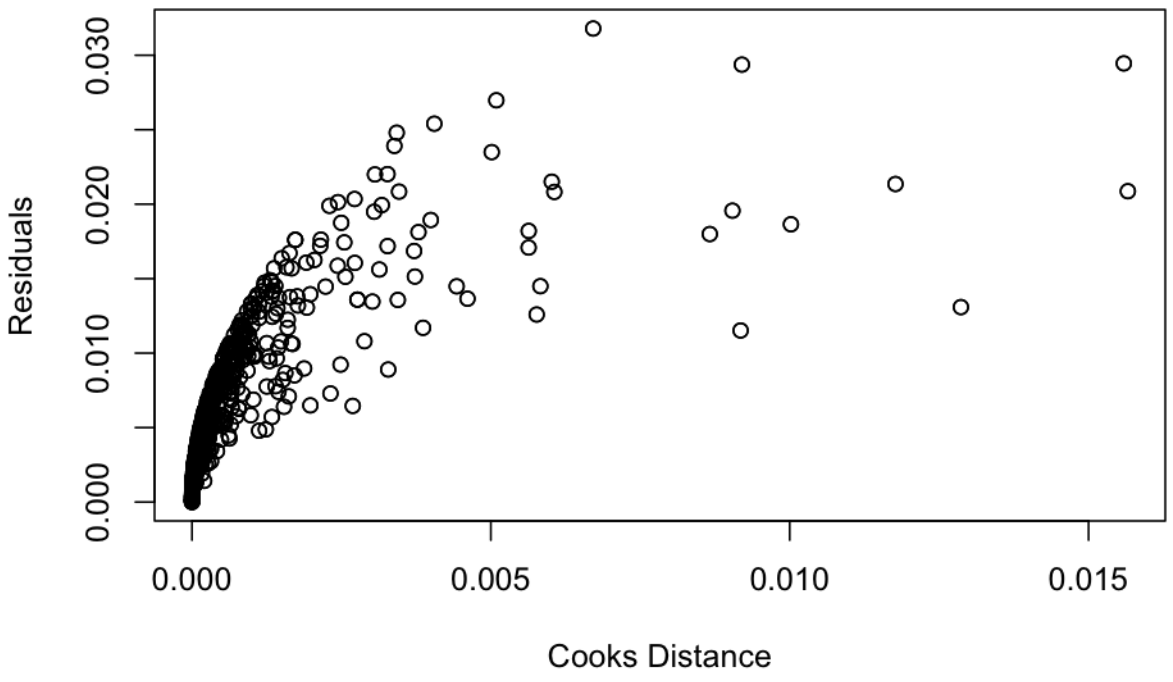
Fig. Suppl 4: Core versus Accessory Genome distances

```
##Scatterplot comparing core genome patristic
## distance and accessory genome distance

#### Figure A ####
#Core genome distances
```

```
CG.tree<-read.tree("RAxML_bestTree.All.Core.tre") #Core genome tree

PatristicDistMatrix <- cophenetic(CG.tree) #patristic distances
PatristicDist <- as.dist(PatristicDistMatrix,diag = TRUE, upper = TRUE)
#Setting up group/SC assignments for between SC patristic distance
seq.labels <- as.data.frame(rownames(PatristicDistMatrix)) #Obtaining ordered taxa
colnames(seq.labels) <- "taxa"
clades <- as.data.frame(cbind(df$FinalName,df$BAPS2)) #All data set
colnames(clades) <- c("taxa","clade")
seq.labels$id  <- 1:nrow(clades) #adding row number to maintain order for sorting after merge
labels.clades <- merge(seq.labels, clades, by="taxa") #merging
labels.clades <- labels.clades[order(labels.clades$id), ] #ordering

md <- meandist(PatristicDist, labels.clades$clade) #calculating mean distance between clades
md.matrix <- as.dist(md,diag = FALSE, upper = TRUE)
md.matrix <- as.matrix(md.matrix)
diag(md.matrix) <- NA

#Accessory geneome distances
PatristicDist.long <- as.data.frame(as.table(md.matrix))
PatristicDist.long <- PatristicDist.long[! (PatristicDist.long$Var1 == PatristicDist.long$Var2),]
PatristicDist.long <- PatristicDist.long[!is.na(PatristicDist.long$Freq),]
CoreGenomeDistances <- PatristicDist.long

MergedDistances <- as.data.frame(cbind(CoreGenomeDistances,AccGenomeDistances))
colnames(MergedDistances) <- c("Var1a", "Var2b", "PCore", "Var1c", "Var2d","PAcc")

MergedDistances$MSE  <- (MergedDistances$PCore-MergedDistances$PAcc)^2
MergedDistancesMedian <- MergedDistances[(MergedDistances$PCore > 0.06 & MergedDistances$PCore < 0.15),]
MergedDistancesMedian <- within(MergedDistancesMedian, A.quartile <- as.integer(cut(MergedDistancesMedia
MergedDistancesMedian <- within(MergedDistancesMedian, P.quartile <- as.integer(cut(MergedDistancesMedia

model <- lm(MergedDistancesMedian$PCore~MergedDistancesMedian$PAcc)
car::outlierTest(model)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##      rstudent unadjusted p-value Bonferroni p
## 207 -3.554855         0.00039412      0.43826
```

```
MergedDistancesMedian$residuals <- abs(resid(model)) #Risiduals
MergedDistancesMedian$cooks <- cooks.distance(model)
plot(MergedDistancesMedian$cooks, MergedDistancesMedian$residuals,
     xlab="Cooks Distance", ylab="Residuals")
```




```
#Make name varianble
MergedDistancesMedian$comp <- paste(MergedDistancesMedian$Var1a, MergedDistancesMedian$Var2b, sep = "-")

R2 <- round(cor(MergedDistancesMedian$PCore,MergedDistancesMedian$PAcc,method = "pearson"),2)

plotS4A <- ggplot(MergedDistancesMedian, aes(PCore,PAcc)) +
  geom_point(color="black",size=1) +
  geom_density_2d(color="#08519c", alpha=.7, show.legend =FALSE) +
  geom_smooth(color="#636363", method=lm, alpha=.3, linetype="dashed",
              size=.6, formula = y~x,  show.legend=FALSE, se=TRUE) +
  labs(x="Core Genome Divergence (Patristic Distance)") +
  labs(y="Accessory Genome Divergence \n(Patristic Distance)") +
  annotate("text", x = .075, y = 2.3,
           label = paste("R^2 == ", R2), parse=TRUE)

#### Figure B ####

RelativeFitness.E1 <- dat3AB %>% filter(BAPS2!="01")

RF.E1.distmat <- as.matrix(dist(RelativeFitness.E1$r))
RF.E1.distmat[upper.tri(RF.E1.distmat)] <- NA; diag(RF.E1.distmat) <- NA
rownames(RF.E1.distmat) <- RelativeFitness.E1$BAPS2; colnames(RF.E1.distmat) <- RelativeFitness.E1$BAPS
RF.E1.distlong <- na.omit(as.data.frame.table(RF.E1.distmat))
colnames(RF.E1.distlong) <- c("Var1a","Var2b","RelFit")

CorDist_RelFit <- merge(RF.E1.distlong, MergedDistances, by=c("Var1a", "Var2b"))
```

```
CorDist_RelFit <- CorDist_RelFit[,c("Var1a", "Var2b","RelFit","PCore","PAcc")]

####Supplemental figures - Core and accessory divergence including fitness
plotS4B <- ggplot(CorDist_RelFit, aes(PCore,RelFit)) +
  geom_point(color="black",size=1) +
  geom_density_2d(color="#08519c", alpha=.7, show.legend =FALSE) +
  labs(x="Core Genome Divergence (Patristic Distance)") +
  labs(y="Absolute fitness difference")

plotS4C <- ggplot(CorDist_RelFit, aes(PAcc,RelFit)) +
  geom_point(color="black",size=1) +
  geom_density_2d(color="#08519c", alpha=.7, show.legend =FALSE) +
  labs(x="Accessory Genome Divergence (Patristic Distance)") +
  labs(y="Absolute fitness difference")

#theme_set(theme_cowplot(font_size=13))
figureS4 <- plot_grid(ncol=3, nrow=1, plotS4A, plotS4B,
                      plotS4C, labels = "AUTO", align = 'h')
ggsave("figureS_distance.png", figureS4, width = 15, height = 5)
```

Fig. Suppl 5: COGs by SCs

```
SC_freq_E1 <- SC_freq_postV_obs %>% select(BAPS2, PCV7.actual) %>%
  left_join(df) %>% subset(Epoch1 == "E1") %>%
  select(BAPS2, HMPREF0837_12128:HMPREF0837_10616) %>%
  arrange(BAPS2) %>% group_by(BAPS2) %>%
  summarise_at(vars(HMPREF0837_12128:HMPREF0837_10616),mean) %>%
  ungroup() %>%
  pivot_longer(cols=HMPREF0837_12128:HMPREF0837_10616, names_to = "COG", values_to = "f") %>%
  mutate(Epoch = "Pre-vaccine")

SC_freq_E3 <- SC_freq_postV_obs %>% select(BAPS2, PCV7.actual) %>%
  left_join(df) %>% subset(Epoch1 == "E3") %>%
  select(BAPS2, HMPREF0837_12128:HMPREF0837_10616) %>%
  arrange(BAPS2) %>% group_by(BAPS2) %>%
  summarise_at(vars(HMPREF0837_12128:HMPREF0837_10616),mean) %>%
  ungroup() %>%
  pivot_longer(cols=HMPREF0837_12128:HMPREF0837_10616, names_to = "COG", values_to = "f") %>%
  mutate(Epoch = "Post-vaccine")

SC_freq_E13 <- bind_rows(SC_freq_E1, SC_freq_E3)

plotCOGs_SC <- ggplot(SC_freq_E13, aes(f, fill = Epoch)) +
  geom_histogram(aes(y = ..count..), position = "dodge", bins = 10) +
  facet_wrap(~BAPS2, nrow = 3) +
  xlab("COG frequency") + theme_minimal() +
  scale_fill_manual("", values = c("#D3723D", "#8A9DA4")) +
  annotate("segment", x=-Inf, xend=Inf, y=-Inf, yend=-Inf)+
  annotate("segment", x=-Inf, xend=-Inf, y=-Inf, yend=Inf) +
  theme(legend.position = "bottom")

ggsave("figureS_COGsSCs.png", plotCOGs_SC, width = 15, height = 5)
```

Table Statistics

```
#### Accessory genome SA ####
## Sensitivity analysis using a subsample of 119 isolates
## collected in 2010 prior to the initiation of PCV13

#### Data ####
dfSenst <- df %>% separate(FinalName, c("ID1", "ID2", "ID3", "Year"))
dfSenstE12 <- dfSenst %>% filter(Epoch1 != "E3")
dfSenstE3 <- dfSenst %>% filter(Epoch1 == "E3" & Year == "2010")
dfS <- bind_rows(dfSenstE12,dfSenstE3)

dfFVTS <- dfS %>% select(BAPS2,PCV7.actual, Epoch1) %>%
  group_by(Epoch1) %>%  count(BAPS2,PCV7.actual) %>%
  mutate(freq = round(prop.table(n), digits = 3)) %>% ungroup() %>%
  select(Epoch1, BAPS2, PCV7.actual, freq) %>%
  spread(Epoch1, freq, fill = 0) %>% arrange(BAPS2,PCV7.actual)

#### Present at E1 ####
SCE1S <- dfFVTS %>% subset(E1 > 0) %>% ## & BAPS2 !="27"
  select(BAPS2, PCV7.actual) %>%
  mutate(Epoch1 = "E1")

#### NVTs Present at E1 ####
SCE2S <- SCE1S %>% subset(PCV7.actual == "NVT"  & BAPS2 !="27")

SC_freq_dfS <- dfS %>% select(BAPS2, PCV7.actual, Epoch1,
                              HMPREF0837_12128:HMPREF0837_10616) %>%
  arrange(BAPS2) %>% group_by(BAPS2,PCV7.actual,Epoch1) %>%
  mutate(SC_n = n()) %>% ungroup() %>%
  group_by(BAPS2,PCV7.actual,Epoch1,SC_n) %>%
  summarise_at(vars(HMPREF0837_12128:HMPREF0837_10616),mean) %>%
  ungroup()

### Get the matrix and the SC for the pre-vaccine epch "E1"
df_preVS <- SCE1S %>% left_join(SC_freq_dfS)
SC_freq_preVS <- as.matrix(df_preVS %>% mutate(SC_freq=SC_n/sum(SC_n)) %>% select(SC_freq))
SC_COG_preVS <- as.matrix(t(df_preVS %>% select(HMPREF0837_12128:HMPREF0837_10616)))

#### Get e_l ####
el_S <- SC_COG_preVS %*% SC_freq_preVS

df_postVS <- SCE2S %>% left_join(SC_freq_dfS)
SC_COG_postVS <- as.matrix(t(df_postVS %>% select(HMPREF0837_12128:HMPREF0837_10616)))

SC_freq_postV_obsS <- SCE2S %>% mutate(Epoch1 = "E3") %>%
  left_join(SC_freq_dfS) %>%
  mutate(SC_freq=SC_n/sum(SC_n, na.rm = T)) %>%
  select(BAPS2, PCV7.actual, SC_freq) %>%
  mutate(SC_freq = replace_na(SC_freq, 0))

## Predict postV frequencies
SC_freq_postV_predS <- QP(SC_COG_postVS, el_S) #Matrix: rows = COGs, columns = (SCs - VT)
SC_freq_postV_obsS <- SC_freq_postV_obsS %>%
```

```
  mutate(SC_pred = SC_freq_postV_predS)

SC_freq_E1S <- df_preVS %>%
  mutate(SC_freq_E1=SC_n/sum(SC_n)) %>%
  select(BAPS2,PCV7.actual,SC_freq_E1)

SC_freq_postV_diffS <- SC_freq_postV_obsS %>%
  left_join(SC_freq_E1S) %>%
  mutate(diff_pred = SC_freq_E1 - SC_pred,
         diff_obs = SC_freq_E1 - SC_freq)

SC_freq_postV_diffS <- SC_freq_postV_diffS %>%
  mutate(diff = abs(diff_pred - diff_obs)) %>%
  arrange(diff)

stats <- summary(lm(SC_freq_postV_diffS$diff_pred~SC_freq_postV_diffS$diff_obs))
ars <- round(stats$adj.r.squared, digits = 3)
sseE <- round(sse(SC_freq_postV_diffS$diff_obs, SC_freq_postV_diffS$diff_pred), digits = 3)
rmseE <- round(rmse(SC_freq_postV_diffS$diff_obs, SC_freq_postV_diffS$diff_pred),digits = 3)

accNFDSSA <- data.frame(Model = "Accesory genome (NFDS) SA",
                        nloci = length(el_S), adj.r.squared = ars,
                        SSE = sseE, RMSE = rmseE)

SC_freq_postV_diffS <- SC_freq_postV_diffS %>%
  mutate(prorata = SC_freq_E1/sum(SC_freq_E1)) %>%
  mutate(diff_predPro = SC_freq_E1 - prorata) %>%
  mutate(diffP = abs(diff_predPro - diff_obs))

stats <- summary(lm(SC_freq_postV_diffS$diff_predPro~SC_freq_postV_diffS$diff_obs))
ars <- round(stats$adj.r.squared, digits = 3)
sseE <- round(sse(SC_freq_postV_diffS$diff_obs, SC_freq_postV_diffS$diff_predPro), digits = 3)
rmseE <- round(rmse(SC_freq_postV_diffS$diff_obs, SC_freq_postV_diffS$diff_predPro),digits = 3)

accProrataSA <- data.frame(Model = "Accesory genome (Prorata) SA",
                           nloci = length(el_S), adj.r.squared = ars,
                           SSE = sseE, RMSE = rmseE)

#####################

#### Core genome ####
dfcore <- read.csv("CoreSNPpresenceAbsence.txt", header = F)
dfcore <- dfcore %>% as_tibble() %>% rename(FinalName = V1)
dfcore <- df %>% select(FinalName, BAPS2, PCV7.actual, Epoch1) %>%
  left_join(dfcore)

dfFVTcore <- dfcore %>% select(BAPS2,PCV7.actual, Epoch1) %>%
  group_by(Epoch1) %>%  count(BAPS2,PCV7.actual) %>%
  mutate(freq = round(prop.table(n), digits = 3)) %>% ungroup() %>%
  select(Epoch1, BAPS2, PCV7.actual, freq) %>%
  spread(Epoch1, freq, fill = 0) %>% arrange(BAPS2,PCV7.actual)

#### Present at E1 ####
```

```
SCE1core <- dfFVTcore %>% subset(E1 > 0) %>% ## & BAPS2 !="27"
  select(BAPS2, PCV7.actual) %>%
  mutate(Epoch1 = "E1")

#### NVTs Present at E1 ####
SCE2core <- SCE1core %>% subset(PCV7.actual == "NVT"  & BAPS2 !="27")


SC_freq_dfcore <- dfcore %>% select(BAPS2, PCV7.actual, Epoch1,
                                    V2:V62654) %>%
  arrange(BAPS2) %>% group_by(BAPS2,PCV7.actual,Epoch1) %>%
  mutate(SC_n = n()) %>% ungroup() %>%
  group_by(BAPS2,PCV7.actual,Epoch1,SC_n) %>%
  summarise_at(vars(V2:V62654),mean) %>%
  ungroup()

### Get the matrix and the SC for the pre-vaccine epch "E1"
df_preVcore <- SCE1core %>% left_join(SC_freq_dfcore)
SC_freq_preVcore <- as.matrix(df_preVcore %>% mutate(SC_freq=SC_n/sum(SC_n)) %>% select(SC_freq))
SC_COG_preVcore <- as.matrix(t(df_preVcore %>% select(V2:V62654)))

#### Get e_l for the core (62653 loci) ####
el_core <- SC_COG_preVcore %*% SC_freq_preVcore

#### Observed versus prediced prevalence ####
df_postVcore <- SCE2core %>% left_join(SC_freq_dfcore)
SC_COG_postVcore <- as.matrix(t(df_postVcore %>% select(V2:V62654)))

SC_freq_postV_obsCore <- SCE2core %>% mutate(Epoch1 = "E3") %>%
  left_join(SC_freq_dfcore) %>%
  mutate(SC_freq=SC_n/sum(SC_n, na.rm = T)) %>%
  select(BAPS2, PCV7.actual, SC_freq) %>%
  mutate(SC_freq = replace_na(SC_freq, 0))

## Predict postV frequencies
SC_freq_postV_predCore <- QP(SC_COG_postVcore, el_core)
SC_freq_postV_obsCore <- SC_freq_postV_obsCore %>%
  mutate(SC_pred = SC_freq_postV_predCore)

SC_freq_E1core <- df_preVcore %>%
  mutate(SC_freq_E1=SC_n/sum(SC_n)) %>%
  select(BAPS2,PCV7.actual,SC_freq_E1)

SC_freq_postV_diffcore <- SC_freq_postV_obsCore %>%
  left_join(SC_freq_E1core) %>%
  mutate(diff_pred = SC_freq_E1 - SC_pred,
         diff_obs = SC_freq_E1 - SC_freq)

SC_freq_postV_diffcore <- SC_freq_postV_diffcore %>%
  mutate(diff = abs(diff_pred - diff_obs)) %>%
  arrange(diff)

#### Pro rata ####
```

```
SC_freq_postV_diffcore <- SC_freq_postV_diffcore %>%
  mutate(prorata = SC_freq_E1/sum(SC_freq_E1)) %>%
  mutate(diff_predPro = SC_freq_E1 - prorata) %>%
  mutate(diffP = abs(diff_predPro - diff_obs))

#### Stats core NFDS ####
stats <- summary(lm(SC_freq_postV_diffcore$diff_pred~SC_freq_postV_diffcore$diff_obs))
ars <- round(stats$adj.r.squared, digits = 3)
sseE <- round(sse(SC_freq_postV_diffcore$diff_obs, SC_freq_postV_diffcore$diff_pred), digits = 3)
rmseE <- round(rmse(SC_freq_postV_diffcore$diff_obs, SC_freq_postV_diffcore$diff_pred),digits = 3)

coreNFDS <- data.frame(Model = "Core genome (NFDS)", nloci = length(el_core),
                       adj.r.squared = ars, SSE = sseE, RMSE = rmseE)

#### Stats core Pro rata ####
stats <- summary(lm(SC_freq_postV_diffcore$diff_predPro~SC_freq_postV_diffcore$diff_obs))
ars <- round(stats$adj.r.squared, digits = 3)
sseE <- round(sse(SC_freq_postV_diffcore$diff_obs, SC_freq_postV_diffcore$diff_predPro), digits = 3)
rmseE <- round(rmse(SC_freq_postV_diffcore$diff_obs, SC_freq_postV_diffcore$diff_predPro),digits = 3)

coreProrata <- data.frame(Model = "Core genome (Pro rata)", nloci = length(el_core),
                          adj.r.squared = ars, SSE = sseE, RMSE = rmseE)

##############################

#### Metabolic loci ####
dfmeta <- read.csv("Core_Metabolic_SNPpresenceAbsence.txt", header = F)
dfmeta <- dfmeta %>% as_tibble() %>% rename(FinalName = V1)
dfmeta <- df %>% select(FinalName, BAPS2, PCV7.actual, Epoch1) %>%
  left_join(dfmeta)

dfFVTmeta <- dfmeta %>% select(BAPS2,PCV7.actual, Epoch1) %>%
  group_by(Epoch1) %>%  count(BAPS2,PCV7.actual) %>%
  mutate(freq = round(prop.table(n), digits = 3)) %>% ungroup() %>%
  select(Epoch1, BAPS2, PCV7.actual, freq) %>%
  spread(Epoch1, freq, fill = 0) %>% arrange(BAPS2,PCV7.actual)

#### Present at E1 ####
SCE1meta <- dfFVTmeta %>% subset(E1 > 0) %>% ## & BAPS2 !="27"
  select(BAPS2, PCV7.actual) %>%
  mutate(Epoch1 = "E1")

#### NVTs Present at E1 ####
SCE2meta <- SCE1meta %>% subset(PCV7.actual == "NVT"  & BAPS2 !="27")


SC_freq_dfmeta <- dfmeta %>% select(BAPS2, PCV7.actual, Epoch1,
                                    V2:V22434) %>%
  arrange(BAPS2) %>% group_by(BAPS2,PCV7.actual,Epoch1) %>%
  mutate(SC_n = n()) %>% ungroup() %>%
  group_by(BAPS2,PCV7.actual,Epoch1,SC_n) %>%
  summarise_at(vars(V2:V22434),mean) %>%
  ungroup()
```

```
### Get the matrix and the SC for the pre-vaccine epch "E1"
df_preVmeta <- SCE1meta %>% left_join(SC_freq_dfmeta)
SC_freq_preVmeta <- as.matrix(df_preVmeta %>% mutate(SC_freq=SC_n/sum(SC_n)) %>% select(SC_freq))
SC_COG_preVmeta <- as.matrix(t(df_preVmeta %>% select(V2:V22434)))

#### Get e_l for the meta (62653 loci) ####
el_meta <- SC_COG_preVmeta %*% SC_freq_preVmeta

#### Observed versus prediced prevalence ####
df_postVmeta <- SCE2meta %>% left_join(SC_freq_dfmeta)
SC_COG_postVmeta <- as.matrix(t(df_postVmeta %>% select(V2:V22434)))

SC_freq_postV_obsMeta <- SCE2meta %>% mutate(Epoch1 = "E3") %>%
  left_join(SC_freq_dfmeta) %>%
  mutate(SC_freq=SC_n/sum(SC_n, na.rm = T)) %>%
  select(BAPS2, PCV7.actual, SC_freq) %>%
  mutate(SC_freq = replace_na(SC_freq, 0))

## Predict postV frequencies
SC_freq_postV_predMeta <- QP(SC_COG_postVmeta, el_meta)
SC_freq_postV_obsMeta <- SC_freq_postV_obsMeta %>%
  mutate(SC_pred = SC_freq_postV_predMeta)

SC_freq_E1meta <- df_preVmeta %>%
  mutate(SC_freq_E1=SC_n/sum(SC_n)) %>%
  select(BAPS2,PCV7.actual,SC_freq_E1)

SC_freq_postV_diffmeta <- SC_freq_postV_obsMeta %>%
  left_join(SC_freq_E1meta) %>%
  mutate(diff_pred = SC_freq_E1 - SC_pred,
         diff_obs = SC_freq_E1 - SC_freq)

SC_freq_postV_diffmeta <- SC_freq_postV_diffmeta %>%
  mutate(diff = abs(diff_pred - diff_obs)) %>%
  arrange(diff)

#### Pro rata ####
SC_freq_postV_diffmeta <- SC_freq_postV_diffmeta %>%
  mutate(prorata = SC_freq_E1/sum(SC_freq_E1)) %>%
  mutate(diff_predPro = SC_freq_E1 - prorata) %>%
  mutate(diffP = abs(diff_predPro - diff_obs))

#### Stats metabolic loci NFDS ####
stats <- summary(lm(SC_freq_postV_diffmeta$diff_pred~SC_freq_postV_diffmeta$diff_obs))
ars <- round(stats$adj.r.squared, digits = 3)
sseE <- round(sse(SC_freq_postV_diffmeta$diff_obs, SC_freq_postV_diffmeta$diff_pred), digits = 3)
rmseE <- round(rmse(SC_freq_postV_diffmeta$diff_obs, SC_freq_postV_diffmeta$diff_pred),digits = 3)

metaNFDS <- data.frame(Model = "Metabolic loci (NFDS)", nloci = length(el_meta),
                       adj.r.squared = ars, SSE = sseE, RMSE = rmseE)

#### Stats metabolic loci Pro rata ####
stats <- summary(lm(SC_freq_postV_diffmeta$diff_predPro~SC_freq_postV_diffmeta$diff_obs))
```

```r
ars <- round(stats$adj.r.squared, digits = 3)
sseE <- round(sse(SC_freq_postV_diffmeta$diff_obs, SC_freq_postV_diffmeta$diff_predPro), digits = 3)
rmseE <- round(rmse(SC_freq_postV_diffmeta$diff_obs, SC_freq_postV_diffmeta$diff_predPro),digits = 3)

metaProrata <- data.frame(Model = "Metabolic loci (Pro rata)", nloci = length(el_meta),
                          adj.r.squared = ars, SSE = sseE, RMSE = rmseE)

##############################

#### create table ####
dfOut <- bind_rows(accNFDS, accProrata, accNFDSSA,
                   accProrataSA, coreNFDS, coreProrata,
                   metaNFDS, metaProrata)
write.csv(dfOut, "statsTable.csv", row.names = F)
```