Pie charts for protein abundances

Jeffrey M. Dick

November 30, 2009

Abstract

This is an attempt to predict some characteristics of the relative abundances of organisms that are depicted in the pie charts shown by Spear et al., 2005[1]. For each type of organism present, we use a single model protein. We take the reported relative abundances of the four most abundant organisms to generate an assemblage of proteins that buffers the activies of CO_2, H_2O and NH_3. Then we calculate relative abundances of all proteins at different values of oxygen fugacity and H_2S activity and display them on pie charts that can be compared with organismal abundances. The results show that it is possible to predict conditions that foster high (many slices in the pie diagrams) or low (few slices) diversity, even if the make up of the community is not perfectly replicated.

1 Setup

Bacterial species and abundances from Spear et al., 2005 (Fig. 4):

```
> O1 <- c("Aquificales", "Thermotogales", "Bacillus", "Thermodesulfobacteria")
> O2 <- c("Thermus/Deinococcus", "Hydrogenobacter", "Hydrogenobaculum",
+     "Bacteroidetes")
> O3 <- c("a-proteobacterium", "d-proteobacterium", "g-proteobacterium",
+     "OD-1")
> ORGANISMS <- c(O1, O2, O3)
```

Which species are most abundant in low and high sulfur environments; the first three of each are aquificales, thermotogales, thermodesulfobacteria:

```
> lowS.which <- c(1, 2, 4, 3, 5)
> lowS.values <- c(75, 12, 4, 7, 2)
> highS.which <- c(1, 2, 4, 11, 6, 9, 7, 8, 10, 13)
> highS.values <- c(63, 2, 10, 2, 5, 2, 9, 2, 3, 1)
```

What chemical species we use to represent each organism:

```
> P1 <- c("ACCA_AQUAE", "A8F7V4_THELT", "RBL1_THIDA")
> P2 <- c("Q93EV7_9BACT", "ACCA_DEIRA", "Q05KD0_HYDTH", "A7WGI1_9AQUI")
> P3 <- c("Q64VW6_BACFR", "ACCA_CAUCR", "A1VC70_DESVV", "ACCA_PSEAE")
> PROTEINS <- c(P1, P2, P3)
```

2 Function definitions

To make pie charts for calculated abundances:

---

[1]Spear, J. R., Walker, J. J., McCollom, T. M. and Pace, N. R. Hydrogen and bioenergetics in the Yellowstone geothermal ecosystem. *Proc. Natl. Acad. Sci. U. S. A.*, 102, 2555-2560, 2005. `http://dx.doi.org/10.1073/pnas.0409574102`

```
> plot.pie.calc <- function(which = "high", T = 25, main = "") {
+     thermo$buffer <<- thermo$buffer[thermo$buffer$name != "PROTEINS",
+         ]
+     myprot <- PROTEINS[get(paste(which, "S.which", sep = ""))][1:4]
+     mypercent <- get(paste(which, "S.values", sep = ""))[1:4]
+     mybufprot <- paste(myprot, "RESIDUE", sep = ".")
+     mod.buffer("PROTEINS", mybufprot, "aq", log10(mypercent/100))
+     species(delete = TRUE)
+     add.protein(protein.residue(PROTEINS))
+     species(paste(PROTEINS, "RESIDUE", sep = "."))
+     basis(c("CO2", "H2O", "NH3"), "PROTEINS")
+     a <- affinity(return.buffer = TRUE, balance = 1, T = T)
+     sub <- c2s(paste(names(a)[1:3], round(as.numeric(a)[1:3])))
+     main <- paste("\n", main, "\n", sub)
+     species(1:nrow(species()), -99)
+     species(mybufprot, log10(mypercent/100))
+     basis(names(a)[1:3], as.numeric(a)[1:3])
+     thermo$basis$logact <<- as.numeric(thermo$basis$logact)
+     col <- rep("white", 99)
+     col[match(myprot, PROTEINS)] <- heat.colors(4)
+     mylogaH2O <- thermo$basis$logact[rownames(thermo$basis) ==
+         "H2O"]
+     a <- affinity(H2O = c(mylogaH2O, mylogaH2O - 1, 2), T = T)
+     logacts <- diagram(a, do.plot = FALSE, residue = TRUE, balance = 1)$logact
+     names <- character()
+     values <- numeric()
+     cols <- character()
+     logactmin <- -2
+     for (i in 1:length(logacts)) {
+         myvalue <- logacts[[i]][1]
+         if (myvalue > logactmin) {
+             names <- c(names, ORGANISMS[i])
+             values <- c(values, myvalue)
+             cols <- c(cols, col[i])
+         }
+     }
+     values <- 10^values
+     isort <- sort(values, index.return = TRUE, decreasing = TRUE)$ix
+     names <- names[isort]
+     values <- values[isort]
+     cols <- cols[isort]
+     pie(values, names, clockwise = FALSE, main = main, col = cols,
+         radius = 0.7)
+ }
```

To plot pie charts for observed abundances of organisms:

```
> plot.pie.obs <- function(which = "low") {
+     names <- ORGANISMS[get(paste(which, "S.which", sep = ""))]
+     values <- get(paste(which, "S.values", sep = ""))
+     main <- paste("observed at", which, "H2S")
+     mycol <- heat.colors(4)
+     mycol <- c(mycol, rep("white", length(names) - length(mycol)))
+     isort <- sort(values, index.return = TRUE, decreasing = TRUE)$ix
+     values <- values[isort]
```

```
+       names <- names[isort]
+       mycol <- mycol[isort]
+       pie(values, names, clockwise = FALSE, main = main, col = mycol,
+           radius = 0.7)
+ }
```

To plot both types of pie diagrams (showing calculated protein activities and observed abundances of organisms) at a given temperature:

```
> plot.pie <- function(T = 80) {
+       layout(matrix(1:4, byrow = TRUE, nrow = 2))
+       par(mar = c(1, 1, 1, 1))
+       basis(c("CO2", "H2O", "NH3", "H2S", "hydrogen", "H+"))
+       basis("pH", 7)
+       val <- function(text) round(as.numeric(thermo$basis$logact[rownames(thermo$basis) ==
+           text]))
+       basis(c("H2", "H2S"), c(-9, -9))
+       plot.pie.calc("low", T = T, main = paste("calculated for H2",
+           val("H2"), "H2S", val("H2S")))
+       label.plot("a")
+       plot.pie.obs("low")
+       label.plot("b")
+       basis(c("H2", "H2S"), c(-5, -3))
+       plot.pie.calc("high", T = T, main = paste("calculated for H2",
+           val("H2"), "H2S", val("H2S")))
+       label.plot("c")
+       plot.pie.obs("high")
+       label.plot("d")
+ }
```

3   Now run it!

At 80 °C:

```
> plot.pie(80)
```

```
basis: changed basis to CO2 H2O NH3 H2S H2 H+.
mod.buffer: added PROTEINS.
add.protein: added 11 proteins
protein: found ACCA_AQUAE.RESIDUE (C5.06H8.1N1.36O1.41S0.04, 1 residues)
protein: found A8F7V4_THELT.RESIDUE (C5.02H7.88N1.32O1.45S0.03, 1 residues)
protein: found RBL1_THIDA.RESIDUE (C4.99H7.67N1.36O1.45S0.05, 1 residues)
protein: found Q93EV7_9BACT.RESIDUE (C5.15H7.88N1.41O1.45S0.06, 1 residues)
protein: found ACCA_DEIRA.RESIDUE (C4.73H7.66N1.43O1.44S0.03, 1 residues)
protein: found Q05KD0_HYDTH.RESIDUE (C5H8.04N1.38O1.44S0.04, 1 residues)
protein: found A7WGI1_9AQUI.RESIDUE (C5H7.91N1.32O1.49S0.03, 1 residues)
protein: found Q64VW6_BACFR.RESIDUE (C4.97H7.92N1.4O1.48S0.05, 1 residues)
protein: found ACCA_CAUCR.RESIDUE (C4.74H7.62N1.41O1.45S0.03, 1 residues)
protein: found A1VC70_DESVV.RESIDUE (C5.01H7.82N1.44O1.44S0.04, 1 residues)
protein: found ACCA_PSEAE.RESIDUE (C4.87H7.79N1.37O1.46S0.05, 1 residues)
affinity: temperature is 80 C
energy.args: pressure is Psat
affinity: loading buffer species
affinity: loading ionizable protein groups
subcrt: 34 species at 353.15 K and 1 bar (wet)
```
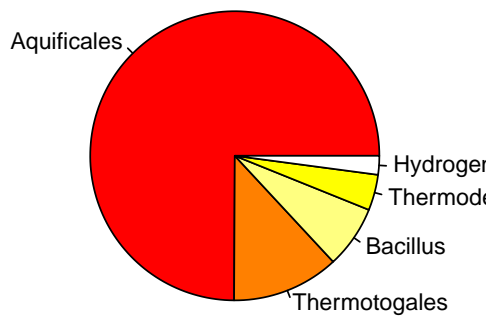
```
buffer: ( PROTEINS ) for activity of CO2 H2O NH3 (active), product (conserved).
affinity: temperature is 80 C
energy.args: pressure is Psat
energy.args: variable 1 is H2O at 2 increments from -37.1300833012037 to -38.1300833012037
affinity: loading ionizable protein groups
subcrt: 34 species at 353.15 K and 1 bar (wet)
diagram: conservation coefficients are 1 1 1 1 1 1 1 1 1 1 1
diagram: using residue equivalents
diagram: log total activity of species (from species) is -0.008773924
mod.buffer: added PROTEINS.
add.protein: added 0 proteins
affinity: temperature is 80 C
energy.args: pressure is Psat
affinity: loading buffer species
affinity: loading ionizable protein groups
subcrt: 34 species at 353.15 K and 1 bar (wet)
buffer: ( PROTEINS ) for activity of CO2 H2O NH3 (active), product (conserved).
affinity: temperature is 80 C
energy.args: pressure is Psat
energy.args: variable 1 is H2O at 2 increments from -25.3801871576367 to -26.3801871576367
affinity: loading ionizable protein groups
subcrt: 34 species at 353.15 K and 1 bar (wet)
diagram: conservation coefficients are 1 1 1 1 1 1 1 1 1 1 1
diagram: using residue equivalents
diagram: log total activity of species (from species) is -0.1135093
```
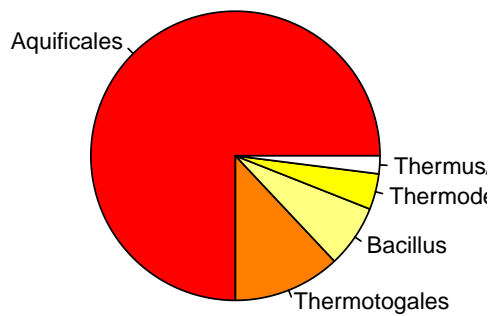
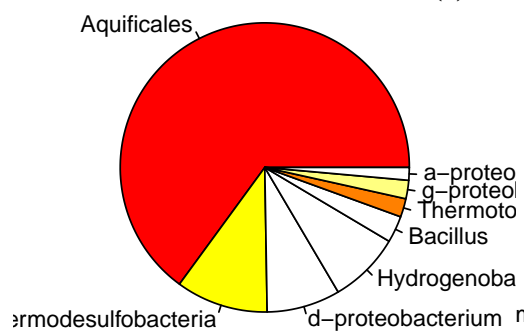**calculated for H2 −9 H2S −9 CO2 −55 H2O −37 NH3 −20**
(*a*)

Aquificales
Hydrogen
Thermode
Bacillus
Thermotogales

observed at low H2S
(*b*)

Aquificales
Thermus
Thermode
Bacillus
Thermotogales

calculated for H2 −5 H2S −3 CO2 −43 H2O −25 NH3 3
(*c*)

Aquificales
a−proteo
g−proteo
Thermoto
Bacillus
Hydrogenoba
ermodesulfobacteria
d−proteobacterium

observed at high H2S
(*d*)

Aquificales
Bacteroic
a−proteob
g−proteob
Thermotog
d−proteoba
Hydrogenobact
modesulfobacteria
Hydrogenobaculum