

# Interfacing With Common Resources: NetCDF

Nishant Gopalakrishnan  
Fred Hutchinson Cancer Research Center

17-18 February, 2011

Introduction

ncdf R package

Resources

# Handling Large Data

- ▶ text files
- ▶ binary files
- ▶ read and process in chunks
- ▶ NetCDF

# Network Common Data Form (NetCDF)

Set of data formats, programming interfaces and software libraries to read/write array oriented scientific data

- ▶ NetCDF software library
  - ▶ NetCDF version 3
  - ▶ NetCDF version 4
- ▶ R packages
  - ▶ ncdf
    - ▶ Warning: character array implementation is inefficient.
  - ▶ ncdf4
    - ▶ Multiple unlimited dimensions
    - ▶ Data compression
    - ▶ Not available on Windows

# NetCDF Model

- ▶ Variables : N-dimensional arrays of data: byte, short, integer, float, double
- ▶ Dimensions
  - ▶ Axes of data arrays
  - ▶ Name, length
  - ▶ Unlimited dimension
- ▶ Attributes : Annotate Variables with meta data

# Using ncdf R package

- ▶ Define dimensions

```
sampDim <- dim.def.ncdf(name = "sampleDim",  
  units = "id", vals = seq_len(NROWS))  
snpDim <- dim.def.ncdf(name = "snpDim",  
  units = "id", vals = seq_len(NCOLS))
```

- ▶ Define variable

```
snpDat <- var.def.ncdf(name = "snpData",  
  units = "0: missing, 1: AA, 2: AB, 3: BB",  
  dim = list(sampDim, snpDim),  
  missval = 0L, prec = "byte")
```

# Using ncdf R package

- ▶ Create file, write variable

```
nc <-open.ncdf("myFile.nc")  
put.var.ncdf(nc, varid = "snpData", vals = mat)
```

- ▶ Write slice

```
## samples 1:10, snps 1:20  
put.var.ncdf(nc, varid = "snpData", vals = slice,  
             start = c(1,1), count =c(10, 20))  
close(nc)
```

## Using ncdf R package

- ▶ Read variable from file

```
nc <-open.ncdf("myFile.nc")  
myVar <- get.var.ncdf(nc, varid = "snpData")  
close(nc)
```

- ▶ Read slices of data

```
nc <-open.ncdf("myFile.nc")  
## samples 30:40, snps 100:120  
slice1 <- get.var.ncdf(nc, "snpData",  
  start =c(30, 100), count =c(10, 20))  
## all samples, snps 1000:1100  
slice2 <- get.var.ncdf(nc, "snpData",  
  start =c(1, 1000), count =c(-1,100))  
close(nc)
```



# Resources

- ▶ <http://cran.r-project.org/web/packages/ncdf/ncdf.pdf>
- ▶ <http://www.unidata.ucar.edu/software/netcdf/docs/netcdf-tutorial.html>