ComP^{3}D

A SAS Program for Compression and P3D of Mixed Linear Model for GWAS

Version: 2.10
Last update: March 4, 2010

Author: Zhiwu Zhang

Institute for Genomic Diversity
Cornell University

THE USER MANUAL OF COMP3D

ComP3D is a SAS program to implement the compression (Com) and P3D of mixed linear model for genome-wide association study.

What are included in the package?

SAS source code
- ComP3D_Demo.sas (main program)
- getAvgKin.sas (macro)
- LORG.SAS (macro)
- Matrixconvert.SAS (macro)
- SetParms.sas (macro)

Demonstration data
- kin_SNP553_taca277_from_SAS_All.txt (kinship)
- popStructure_taxa286_rn_rename.txt (population structure)
- SNP553_taca277_from_SAS_major_chr.txt (genotype)
- three_traits_rn_rename.txt (phenotype)

Result files
The ModelFit folder contains the result of optimization for compression level with UPGMA (AVE) clustering algorithm. The NoCompression folder contains the association study results with no compression (equivalent to compression level one). The Compression3.5 folder contains the association study results with compression level of 3.5. The files in the Results folder are the association study results with compression level of 1.07.

Getting start

Stages I (Grouping optimization)

The first stage is to do grouping optimization. The optimization is to evaluate cluster algorithms and compression levels to achieve optimum model fit indicated by maximum likelihood etc. This can be done by testing clustering algorithm one at a time. For each clustering algorithm, the computing time for optimizing group numbers can be reduced by screening only part of grouping numbers: such as n, 0.9n, 0.9*0.9n, …, 1.

The parameters can be set as follows:

```
*ClusterMeth=AVE;*Try each of: AVE, CEN, COM, FLE, MCQ, MED, SIN, WAR;
*grpStartAt=1;
*grpEndA=51
*snpStartAt=1;
*snpEndAt=1;
*OPTAppraoch=0;
```

Stages II (Screening markers)
The parameters can be set as follows:

```
*ClusterMeth=AVE;*Use the best of: AVE, CEN, COM, FLE, MCQ, MED, SIN, WAR;
*grpStartAt=0;
```

*grpEndA=0
 *snpStartAt=1;
 *snpEndAt=71378;*the total number of markers;
 *OPTAppraoch=2;*use 1-5, 2 is recommended:
 *CF=5;*set it to the compression level with the best model fit;

The value of CF has to be changed to the value derived from stage I.

How to cite ComP3D

ComP3D is part of supporting documents for the paper entitled "Mixed Linear Model Adapted for Genome-Wide Association Study" on Nature Genetics:

Zhiwu Zhang, Elhan Ersoz, Chao-Qiang Lai, Rory J Todhunter, Hemant K Tiwari, Michael A Gore, Peter J Bradbury, Jianming Yu, Donna K Arnett, Jose M Ordovas and Edward S Buckler. Nature Genetics (7 March 2010) doi:10.1038/ng.546.