Help file for R function SpatialClust(), which implements the spatial clustering algorithm in:

> Reich BJ, Bondell HD. A spatial Dirichlet process mixture model for clustering population genetic data. In press, Biometrics.

The inputs (default values) are:

z
An n x M x A array containing genetic data. z[i,m,a] is the number of copies (0,1,or 2) of allele a at marker m for individual i. A is the largest number of alleles for any marker.

n.all
An M vector with the number of possible alleles for each marker. For example, if n.all[2]=3, then there are three possible alleles for marker 2 and z[,2,a] should be zero for a=3,...,A.

x
The n x 2 matrix of spatial coordinates, assumed to be in $[0,1]^2$.

m1 (25)
The maximum number of clusters.

m2 (25)
The number of mixture components in each spatial distribution.

spatial (TRUE)
TRUE/FALSE indicator of whether to include the spatial information in the model.

genetic (TRUE)
TRUE/FALSE indicator of whether to include the genetic information in the model (genetic = F is useful for examining the prior number of clusters).

runs (5000)
Number of MCMC samples to generate.

burn (1000)
Number of MCMC samples to discard.

update (10)
Number of iterations between displaying the current iteration.

common.range (TRUE)
TRUE/FALSE indicator of whether each mixture component in the spatial distribution should have the same range.

p0 (0.5)
Prior inclusion probability for each marker.

b.a (1), b.b (1)
Hyperparameters that control the prior number of clusters, b_U ~ Gamma(b.a,b.b) (following the notation of Reich and Bondell).

The outputs are:

g
Posterior samples of the cluster labels.

rho.star
Posterior samples of rho.star (following notation of Reich and Bondell), the parameters that control the genetic variability across clusters.

clusters
Posterior samples of the number of clusters, defined as the number of unique cluster labels.

clusters2
Posterior samples of the number of clusters, defined as the number of unique cluster labels with at least two members.

pequal
n x n matrix of pair-wise cluster probabilities.

membership
An n x m1 matrix of cluster memberships for each number of clusters, estimated using the posterior cluster probabilities and the k-means clustering algorithm, as described in Reich and Bondell.

A Simulated example:

```
n=100
M=10
A=2
n.clust=3

#Allele frequencies for each cluster:

all.freq=array(1/A,c(M,A,n.clust))
all.freq[1,,1]=all.freq[2,,1]=all.freq[3,,1]=all.freq[4,,1]=c(.2,.8)
all.freq[1,,2]=all.freq[2,,2]=all.freq[3,,2]=all.freq[4,,2]=c(.8,.2)

#Spatial distribution of each cluster

s1=s2=c(-.5,.5,0)
bw=0.25

#Generate the genetic data:

n.all=rep(A,M)
z=array(0,c(n,M,A))
g=sample(1:n.clust,n,replace=T)
x=cbind(rnorm(n,s1[g],bw), rnorm(n,s2[g],bw))
for(i in 1:n){for(l in 1:M){for(rep in 1:2){
     samp<-sample(1:A,1, prob=all.freq[l,,g[i]])
     z[i,l,samp]= z[i,l,samp]+1
}}}
```

```
#Calculate the prior on the number of clusters:

fit=SpatialClust(z=z,n.all=n.all,x=x,spatial=F,genetic=F)
table(fit$clusters2)/length(fit$clusters2)


#Calculate the posterior of the number of clusters:

fit=SpatialClust(z=z,n.all=n.all,x=x)
table(fit$clusters2)/length(fit$clusters2)


#Plot the membership classification assuming three clusters:

plot(x[,1],x[,2],pch=19,col=fit$membership[,3])
```