```
===================== DARC program README file =========================

Extract the executables by first typing "gunzip darc1.0.6_i86linux64.tar.gz" and
then "tar xvf darc1.0.6_i86linux64.tar" This will create the directory
'darc1.0.6_i86linux64', which contains the darc and sbsmpl_dca programs.  To
create a cma-formatted MSA as input go to the MAPGAPS website
   http://www.igs.umaryland.edu/labs/neuwald/software/mapgaps/
and download mapgaps.  Type 'darc' or 'sbsmpl_dca' on the command line shows
usage and input and output file descriptions.


These programs have been tested under the following Linux operating systems:
RedHat version 7.6, and Ubuntu version X.X.  No non-standard hardware
is required and the program needs no installation requirements other than placing
the executable on a user's path. This program and the source code are freely
available under the MIT License (see LICENSE file).

Extract the source code by first typing "gunzip darc_src_1.0.6.tar.gz" and
then "tar xvf darc_src_1.0.6.tar" This will create the directory 'darc_src_1.0.6',
which contains several subdirectories.


-------------------- Demo in example subdirectory ----------------------
To perform DARC and sbsmpl_dca analyses of C-type lectin domains go into the 'example/'
subdirectory and type './run_example'.  This example, which includes a minimal
subsampling option, should take about 10 minutes to finish on a typical Linux
workstation (or less time without subsampling). From the input multiple
sequence alignment (MSA) file (C_Lectin.mma), DARC produces the following
output files:

C_Lectin.chk:  checkpoint file for additional analysis using our BPPS program.
C_Lectin.hsw:  sequence weights (to reduce redundancy).
C_Lectin.paths: paths to pdb coordinate files for PyMOL scripts.
C_Lectin.sprs: SIPRIS input file (see reference 4 in usage).
C_Lectin_1.pml: pymol script for proteins of known structure included in foreground MSA.
   :    :      :     :     :     :     :     :     :     :       :    :        :
C_Lectin_8.pml:  :    :     :     :     :     :     :     :     :       :    :        :
C_Lectin_aln.rtf: rich text format 'contrast alignment' showing BPPS-defined patterns.
C_Lectin_fg:   query family (foreground) MSA (in cma format) as defined by BPPS
C_Lectin.darc: summary of the results.


Temporary files:
C_Lectin_fg.in:       query MSA in PSICOV format for DCA analysis:
C_Lectin_fg_X.dca: CCMpred output file (input for STARC).
C_Lectin_fg_X.mst: STARC input file.
C_Lectin_fg_X: EVcouplings formated DCA file (required by STARC).


-------------------- Reproducing the published analysis ----------------------
To reproduce the analysis in the DARC paper, with slight variation due to DARC's stochastic
nature, download darc_data.tar.gz, then, within the darc1.0.6 directory, uncompress and
extract from darc_data.tar.gz the directory 'data'.  Go into the data subdirectory and
type 'run_gdp' (for the gamma + delta prime versus other AAA+ analysis) or 'run_gmm'
(for the gamma versus delta prime analysis).  Run the script 'run_gdp_sbsmpl' script
to reproduce the subsampling analysis.  Note that this analysis requires quite a bit of RAM.
```