

dnaDist, v. 0.4: Compute Distances Between DNA Sequences

Bernhard Haubold

Max-Planck-Institute for Evolutionary Biology, Plön, Germany

November 6, 2018

1 Introduction

Given a set of DNA sequences in FASTA format, `dnaDist` computes the Jukes-Cantor distance between all pairs.

2 Getting Started

`dnaDist` was written in C on a computer running Linux and should work on any standard UNIX system. However, please contact me at `haubold@evolbio.mpg.de` if you have any problems with the program.

- Unpack the program

```
tar -xvzf dnaDist_XXX.tgz
```

where XXX indicates the version.

- Change into the newly created directory

```
cd DnaDist_XXX
```

and list its contents

```
ls
```

- Generate `dnaDist`

```
make
```

- List its options

```
./dnaDist -h
```

3 Listing

The following listing documents the driver program for `dnaDist`.

```
1  /***** dnaDist.c *****/
   * Description:
   * Author: Bernhard Haubold, haubold@evolbio.mpg.de
   * Date: Mon Jun 1 16:43:51 2015
   *****/
6  #include <stdio.h>
   #include <stdlib.h>
```

```

#include <fcntl.h>
#include <unistd.h>
#include "gsl_rng.h"
11 #include "sequenceData.h"
#include "alignment.h"
#include "interface.h"
#include "eprintf.h"
#include "distance.h"

16
void runAnalysis(Args *args, int fd){
    gsl_rng *r;
    int i;
    Sequence *seq;
21    Alignment *al;

    seq = readFasta(fd);
    al = seq2al(seq);
    jukesCantor(al);
26    printDist(al);
    if(args->b || args->p){
        r = ini_gsl_rng(args);
        for(i=0;i<args->b-1;i++){
            jukesCantorBoot(al,r);
31            printDist(al);
        }
        for(i=0;i<args->p-1;i++){
            jukesCantorPairBoot(al,r);
            printDist(al);
36        }
        free_gsl_rng(r,args);
    }
    freeSequence(seq);
    freeAlignment(al);
41 }

int main(int argc, char *argv[]){
    int i, fd;
    char *version;
46    Args *args;

    version = "0.4";
    setprogname2("dnaDist");
    args = getArgs(argc, argv);
51    if(args->v)
        printSplash(version);
    if(args->h || args->e)
        printUsage(version);
    if(args->numInputFiles){
61        for(i=0;i<args->numInputFiles;i++){
            fd = open(args->inputFiles[i],O_RDONLY,0);
            runAnalysis(args,fd);
            close(fd);
        }
    }
    else{

```

```
        fd = 0;
        runAnalysis(args, fd);
    }
    free(args);
66 free(progname());
    return 0;
}
```

4 Change Log

- Version 0.1; June 2, 2015
 - Fixed a number of problems with `sequenceData.c` and `sequenceData.h`, which were due to my using `int` instead of `size_t`.
- Version 0.2; June 3, 2015
 - Implemented bootstrapping. Random number generation is based on the GSL library.
- Version 0.3; June 25, 2015
 - Implemented pairwise bootstrap.
- Version 0.4; November 6, 2018
 - Fixed bug in `interface.c`.