BIOINFORMATICS APPLICATIONS NOTE Vol. 15 no. 2 1999
Pages 174–175

DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis

J. Rozas and R. Rozas

Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Diagonal 645, 08071 Barcelona, Spain

Received on July 17, 1998; revised on September 2, 1998; accepted on September 4, 1998

Abstract

Summary: *DnaSP is a Windows integrated software package for the analysis of the DNA polymorphism from nucleotide sequence data. DnaSP version 3 incorporates several methods for estimating the amount and pattern of DNA polymorphism and divergence, and for conducting neutrality tests.*
Availability: *For academic uses, DnaSP is available free of charge from: http://www.bio.ub.es/~julio/DnaSP.html*
Contact: *julio@porthos.bio.ub.es*

DnaSP (from **DNA S**equence **P**olymorphism) (Rozas and Rozas, 1995, 1997) is an integrated software package for the analysis of the DNA polymorphism from nucleotide sequence data. The major new features of DnaSP version 3 include: (i) analysis of the DNA polymorphism and divergence in synonymous, non-synonymous and silent (both synonymous sites in the coding region and non-coding positions) sites separately for different functional regions (Nei and Gojobori, 1986); (ii) analysis of the number of segregating sites and of the pairwise differences distribution (mismatch distribution) in constant size and in growing populations (Rogers and Harpending, 1992; Slatkin and Hudson, 1991; Tajima, 1989b); (iii) algorithms for conducting the McDonald and Kreitman (1991) and the McDonald (1996, 1998) tests; and (iv) a module for estimating the confidence intervals of some test-statistics by the coalescent model (Hudson, 1990). DnaSP runs on IBM-compatible personal computers under 32-bit Microsoft Windows.

Data files

DnaSP can read four different nucleotide sequence file formats: MEGA, NBRF/PIR, NEXUS, and PHYLIP. DnaSP can also convert (export) sequences from one file format to another. Additionally, DnaSP also allows conversion of the data file to the file format recognized by the Hudson *et al.* (1992) program for detecting population subdivision.

Analytical tools

DnaSP estimates several measures of the DNA polymorphism within and between populations, linkage disequilibrium, recombination, gene flow and gene conversion (Nei, 1987; Tajima, 1983; Rozas and Rozas, 1997). DnaSP can also carry out several tests of neutrality and estimate the confidence intervals of some test-statistics by the coalescent (see below).

DnaSP allows the analysis in a subset of sites, or in a subset of sequences of the data file. The software also allows analyses in synonymous and non-synonymous sites; for the analyses DnaSP can use four different genetic codes (nuclear universal; and the mitochondrial of *Drosophila*, mammalian and yeast). Additionally, DnaSP can perform several analyses by the sliding window method.

Neutrality tests

DnaSP can conduct the neutrality tests of Hudson *et al.* (1987), Tajima (1989a), McDonald and Kreitman (1991), and Fu and Li (1993). The Hudson *et al.* (1987) test (HKA test) is based on the neutral theory of molecular evolution prediction (Kimura, 1983) that regions of the genome that evolve at high rates will also present high levels of polymorphism within species. DnaSP can perform the HKA test comparing: (i) two regions of the data file of arbitrary size; (ii) autosomal and sex-linked regions; and (iii) regions differing in the number of sequences (intraspecific data) or differing in the number of sites (between the intraspecific and the interspecific data). The Tajima (1989a) and Fu and Li (1993) tests contrast different estimates of θ ($\theta = 4N\mu$, where N is the effective population size, and μ is the mutation rate per sequence and per generation). For the latter, DnaSP can conduct tests with or without outgroup. The McDonald and Kreitman (1991) test compares the synonymous and non-synonymous variation within and between species. Under neutrality, the ratio of non-synonymous to synonymous fixed substitutions between species should be the same as the ratio of non-synonymous to synonymous polymorphisms within species. Additionally, DnaSP can also generate the input data file for performing the tests proposed by McDo-

nald (1996, 1998) to detect heterogeneity in the polymorphism to divergence ratio across a region of DNA. These tests are based on the distribution, across a region of DNA, of polymorphic sites and fixed differences.

Coalescent simulations

DnaSP can perform computer simulations based on the coalescent process for a neutral infinite-sites model without recombination and assuming a large constant population size (Hudson, 1990). DnaSP performs computer simulations, (i) fixing the value of θ (i.e. assuming a value of θ), or (ii) fixing *S*, the number of segregating sites (mutations) on the genealogy. From the simulations DnaSP generates the distribution of the Tajima's *D* (Tajima, 1989a) and of the raggedness *r* (Harpending, 1994) test statistics. Therefore, DnaSP can estimate both the confidence limits for a given confidence interval and the probability of obtaining values of the statistic lower (or higher) than the observed. Thus, both one-sided and two-sided tests can be conducted.

Acknowledgements

We thank M. Aguadé and C. Segarra for critical comments on the manuscript. We also thank the numerous people who tested the program with their data, especially members of the Molecular Evolutionary Genetics group in the Departament de Genètica, Universitat de Barcelona. This work was supported by grant PB94-0923 from the Dirección General de Investigación Científica y Técnica, Spain, and by grant 1997SGR00059 from the Comissió Interdepartamental de Recerca i Tecnologia, Catalonia, Spain, conferred on M. Aguadé.

References

Fu,Y.-X. and Li,W.-H. (1993) Statistical tests of neutrality of mutations. *Genetics*, **133,** 693–709.

Harpending,H. (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum. Biol*, **66,** 591–600.

Hudson,R.R. (1990). Gene genealogies and the coalescent process. *Ox. Surv. Evol. Biol*., **7,** 1–44.

Hudson,R.R., Kreitman,M. and Aguadé,M. (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116,** 153–159.

Hudson,R.R., Boos,D.D. and Kaplan,N.L. (1992) A statistical test for detecting population subdivision. *Mol. Biol. Evol*., **9,** 138–151.

Kimura,M. (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, MA.

McDonald,J.H. (1996) Detecting non-neutral heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. *Mol. Biol. Evol*., **13,** 253–260.

McDonald,J.H. (1998) Improved tests for heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. *Mol. Biol. Evol*., **15,** 377–384.

McDonald,J.H. and Kreitman,M. (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*, **351,** 652–654.

Nei,M. (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei,M. and Gojobori,T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol*., **3,** 418–426.

Rogers,A.R. and Harpending,H. (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol*., **9,** 552–569.

Rozas,J. and Rozas,R. (1995) DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. *Comput. Applic. Biosci*., **11,** 621–625.

Rozas,J. and Rozas,R. (1997) DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. *Comput. Applic. Biosci*., **13,** 307–311.

Slatkin,M. and Hudson,R.R. (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129,** 555–562.

Tajima,F. (1989a) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123,** 585–595.

Tajima,F. (1989b) The effect of change in population size on DNA polymorphism. *Genetics*, **123,** 597–601.

Tajima,F. (1993) Measurement of DNA polymorphism. In Takahata,N. and Clark,A.G. (eds), *Mechanisms of Molecular Evolution*. Sinauer Associates Inc., Sunderland, MA, pp. 37–59.