J Mol Evol (1988) 27:212–216

Journal of Molecular Evolution
© Springer-Verlag New York Inc. 1988

Calibration of the Change in Thermal Stability of DNA Duplexes and Degree of Base Pair Mismatch

Adalgisa Caccone,* Rob DeSalle, and Jeffrey R. Powell

Department of Biology, Yale University, PO Box 6666, New Haven, Connecticut 06511, USA

Summary. One method of determining the degree of base pair divergence between two sources of DNA (different strains, species, etc.) is to determine the decrease in thermal stability of hybrid duplex DNA due to mismatching of base pairs. Attempts to calibrate the change in median melting temperature (ΔT_m) to base pair mismatch have led to conflicting results. We have studied the ΔT_m between DNAs of known sequence over a range of from 0.55% to 7.2% base pair mismatch. The relationship of ΔT_m and percent base pair mismatch is remarkably linear over this range with a correlation coefficient >0.98. A ΔT_m of 1°C corresponds to 1.7% base pair mismatch. This conversion is higher than that usually assumed and, therefore, rates of DNA evolution estimated by DNA–DNA hybridization studies are likely faster than previously thought.

Key words: DNA evolution — DNA thermal stability — DNA–DNA hybridization

Introduction

Duplex DNA becomes single-stranded under certain conditions such as high temperature. When returned to conditions allowing stable duplex formation, strands with complementary sequences reanneal. The thermal stability of the reformed duplexes depends upon the fidelity of base pair matching, A with T and C with G. Perfectly matched duplexes have a higher median melting temperature (T_m) than do duplexes with mismatched bases. The relationship between change in T_m (ΔT_m) and percent base pair mismatch is quantitative and thought to be linear (Kohne 1970; Britten et al. 1974). This property of DNA has been exploited for a variety of studies concerning evolution, especially so-called DNA–DNA hybridization experiments. One can use ΔT_m to infer rates of DNA evolution (e.g., Britten 1986) as well as to reconstruct phylogenetic relationships (e.g., Sibley and Ahlquist 1984; O'Brien et al. 1985; Caccone and Powell 1987; Catzeflis et al. 1987; Sheldon 1987; Caccone et al. 1988). However, one aspect of change in thermal stability of DNA duplexes due to base pair mismatch has remained unclear: What is the conversion between ΔT_m and percent mismatch? With the plethora of DNA sequence data now available, it is possible to approach this problem directly.

Earlier studies on this problem were based either on synthetic oligonucleotides of known composition or on in vitro chemical modification of base pairing (Bautz and Bautz 1964; Uhlenbeck et al. 1968; Laird et al. 1969; Hutton and Wetmur 1973). Furthermore, some of these studies involved RNA–RNA or RNA–DNA duplexes. The results varied: a ΔT_m of 1°C corresponded to between 0.7% and 1.7% base pair mismatch. Despite this variation and lack of empirical data on natural DNA–DNA duplexes, most workers (e.g., Britten et al. 1974; Britten 1986; Koop et al. 1986; Catzeflis et al. 1987) have assumed a 1:1 conversion for DNA duplexes, i.e., 1°C ΔT_m corresponds to 1% mismatch. In a previous study on mitochondrial DNAs (mtDNA) of known sequences, we found that 1°C ΔT_m corresponds to between 1.5% and 2.0% mismatch (Powell et al. 1986). Here, we confirm this higher conversion over a range of percent mismatch.

Offprint requests to: J.R. Powell

* On leave from II Università di Roma 'Tor Vergata,' Rome, Italy

Materials and Methods

Most of this study was done on the rRNA-coding region of *Drosophila* mtDNA. The region is 906 bp in length and has been sequenced in several species of *Drosophila* (DeSalle et al. 1987; DeSalle, unpublished). For one comparison we used another region of *Drosophila* mtDNA, the *Hin*dIII-B fragment of *D. yakuba*, which is about 5 kb in length and contains the cytochrome oxidase genes (Wolstenholme and Clary 1985). This was hybridized to the corresponding region of *D. melanogaster* mtDNA. There were no insertions/deletions among the rRNA genes. There were 14 insertions/deletions (out of 5 kb) between the *D. yakuba* and *D. melanogaster Hin*dIII-B sequences; these include one seven-base, one three-base, one two-base, and two single-base insertions/deletions (Wolstenholme and Clary 1985). These occur in the intergenic regions and are not clustered. Therefore, we do not expect these insertions/deletions to significantly affect thermal stability.

All DNA was derived from recombinant plasmids. Tracer DNAs (radioactively labeled DNA) were restriction endonuclease fragments containing only the *Drosophila* DNA purified from agarose gels. Driver DNAs (unlabeled DNA) were the entire plasmid including the vector sequences. Controls showed that the *Drosophila* mtDNA did not cross-hybridize with vector sequences. The ratio of tracer to driver was 1:>2000.

In addition to base pair mismatch, the stability of DNA duplexes is affected by other factors including base composition (G–C pairs bond more strongly than A–T pairs) and the length of the duplex (shorter molecules melt at a lower temperature than longer molecules). Our methods allowed us to control these two factors. First, the DNA melting was done in a solvent of the chaotropic compound, tetraethylammonium chloride (abbreviated TEACL), which reduces the strength of G–C bonds more than that of A–T bonds. In 2.4 M TEACL, melting is independent of base composition (Melchior and von Hippel 1973; Chang et al. 1974; Orosz and Wetmur 1977). To correct for the effect of length of the duplexes, samples of DNA to be melted were run on an alkaline agarose gel and the tracer lengths determined by counting the radioactivity in slices along the lanes (McDonell et al. 1977). We used the formula derived by Hall et al. (1980) as presented in Hunt et al. (1981) to normalize the T_m to that expected for a duplex of 500 bp. In these experiments the sizes of the duplexes ranged from 116 to 560 bp with a mean of 206 bp.

Details of the TEACL method of DNA–DNA hybridization have been published and will not be repeated here (Hunt et al. 1981; Powell et al. 1986; Caccone et al. 1987; Caccone and Powell 1987). The main differences from the more familiar hydroxylapatite (HAP) method are that melting is done in 2.4 M TEACL and the amount of single-stranded DNA is determined by S1 nuclease digestion. The above references can be consulted for details.

Table 1. Results of DNA–DNA hybridizations between cloned DNAs of known sequence and base pair mismatch

Tracer	Driver	*n*	ΔT_m	SE	% base pair mismatch
1) SIL1	SIL2	4	0.28	0.09	0.55
2) SIL1	PIL	8	1.18	0.08	1.88
3) PIL	PIC	4	1.77	0.14	2.65
4) PIL	PSE	4	2.60	0.24	4.31
5) PIL	MIM	4	2.77	0.17	3.53
6) SIL1	SCA	8	3.44	0.13	5.76
7) PIL	SCA	8	3.56	0.11	5.74
8) SCA	FUN	4	4.17	0.19	6.95
9) YAK	MEL	5	4.11	0.31	7.20

The clones were derived from several drosophilids: *D. silvestris* (SIL1 and SIL2), *D. pilimana* (PIL), *D. picticornis* (PIC), *D. pseudoobscura* (PSE), *D. mimica* (MIM), *D. funebris* (FUN), *D. yakuba* (YAK), *D. melanogaster* (MEL), and *Scaptomyza adusta* (SCA). *n* is the number of replicates. The three comparisons with eight replicates were done reciprocally, i.e., each species used as the tracer and driver. Standard errors were calculated as in Caccone et al. (1987). The sequence data for lines 1–8 are from DeSalle et al. (1987) and for line 9 from Wolstenholme and Clary (1985)

Table 2. Analysis of data in Table 1

	Y-intercept	Slope	*r*	Conversion
No constraints				
All data	0.20	0.57	0.985	1.74
Remove PIL-MIM	0.08	0.58	0.997	1.71
Constrained through origin				
All data	0	0.61	0.998	1.64
Remove PIL-MIM	0	0.60	0.999	1.67
Mean of each point considered separately				
All data				1.63 (0.06)
Remove PIL-MIM				1.67 (0.04)

Regression lines were calculated with and without the single datum (PIL-MIM) more than 2 standard errors off the regression line and with and without the constraint of passing through the origin. *r* is the correlation coefficient. "Conversion" is the percent base pair mismatch corresponding to 1°C ΔT_m and was calculated as the reciprocal of the slope. The lower part of the table presents the means of the conversion calculated separately for each line in Table 1; the standard error is in parentheses

Results and Discussion

Table 1 and Fig. 1 present the results. The range of base pair mismatch was from 0.55% to 7.2%. Over this range the relationship between ΔT_m and base pair mismatch is remarkably linear. Only one point (PIL-MIM) lies more than 2 standard errors off the regression line. Table 2 presents the regression analyses including and excluding this point. It makes little difference whether this point is included. In theory the regression line should pass through the origin and indeed the Y-intercept is very close to zero. Because of this expectation we have also calculated the regressions with the constraint of passing through the origin; again this makes little difference in the conversion factor. We present another way of analyzing the data in the lower part of Table 2. The conversion factor was calculated separately for each line in Table 1 and the means were determined. The six conversions in Table 2 range from 1.63 to 1.74. Because of experimental error (see SEs in Table 1), the second digit after the decimal in these

Table 3. Comparison of HAP and TEACL methods to determine ΔT_m

Species compared	ΔT_m	
	HAP	TEACL
Drosophila melanogaster–D. simulans	1.9	2.17 ± 0.09
D. melanogaster–D. orena	2.3	2.51 ± 0.10
Strongylocentrotus purpuratus–S. drobachiensis	2.5	2.5
S. purpuratus–S. franciscanus	15	13

Drosophila HAP data are from Zwiebel et al. (1982), *Drosophila* TEACL data from Caccone et al. (1988), and *Strongylocentrotus* data from Hall et al. (1980). Where available, standard errors are indicated

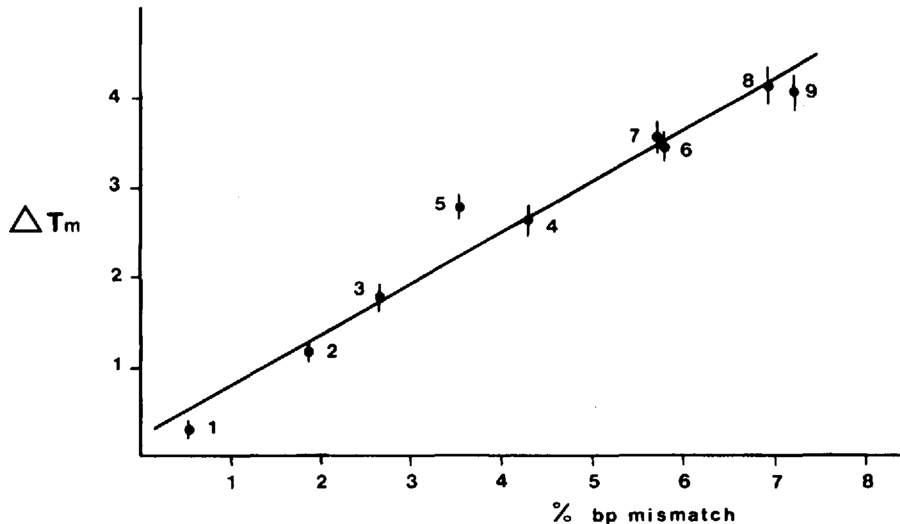

Fig. 1. Graph of data in Table 1. Numbers correspond to left-most column in Table 1. Bars indicate 1 standard error. Regression line is drawn using all points without constraint of passing through the origin. See Table 2 for alternative analyses.

conversion estimates is meaningless. Therefore, we conclude that the best estimate of the conversion is that a 1°C ΔT_m indicates 1.7% base pair mismatch.

In addition to the ΔT_m statistic, some authors use $\Delta T_{50}H$ (Sibley and Ahlquist 1984), ΔT_mR (Benveniste 1985), or ΔT median (Britten 1986). These statistics take into consideration the fraction of the DNA that did not hybridize under the conditions of the reassocation. We used conditions (1 M TEACL, 45°C) that require about 75% or greater base pair match to form stable duplexes. Because the most distant comparison was 7.2% divergent, the degree of hybridization should be high. The mean percent reassociation for heteroduplexes (duplexes formed with tracer and driver from different clones) normalized to the homoduplex (duplexes formed from tracer and driver from the same clone) was 97% with a standard deviation of 8%. Therefore our results can only be used to calibrate ΔT_m. However, this calibration is certainly relevant to interpreting the other statistics.

We have used the TEACL method of DNA–DNA hybridization to calibrate the conversion. There is evidence that the conversion would be similar for the more familiar hydroxylapatite procedure. Table 3 presents comparisons of the few cases where both methods have been used on the same species. The two procedures appear to yield very similar results. Thus, this conversion is probably fairly general for liquid hybridization. Whether it is valid for other types of hybridization such as Southern blotting (Southern 1975) remains to be determined.

Although the HAP and TEACL procedures appear to yield very similar results, there is another procedure for following DNA melting using spectrophotometry. The hyperchromicity of DNA solutions increases with increasing percent single-stranded DNA. The change in absorbency due to temperature change can be followed in a spectrophotometer fitted with a temperature controlled cuvette. In such a system, partially as well as fully denatured DNA molecules will contribute to the absorbency change. However, in both the HAP and TEACL procedure, partially denatured molecules are probably counted as double-stranded. This is because partially denatured molecules can bind to HAP. With the TEACL procedure, when the samples are removed from the temperature gradient for nuclease S1 digestion, partially denatured sequences have an opportunity to "zipper" again. This could be an explanation of the apparent discrepancy between our results and those of Hutton and Wetmur (1973) who studied the effect of chemical modification of base pairing using spectrophotometry. The

important point is that most evolutionary studies employing DNA–DNA hybridization have used either the HAP or TEACL procedure and thus our results would seem to be the pertinent ones for such studies.

We believe our results are not limited to the particular sequences we studied. Because we have used the TEACL procedure, the effect of base composition is negated. Therefore our result is applicable to any DNA regardless of A+T/G+C ratio. One could argue that rRNA genes are peculiar, but the fact that the protein-coding (primarily) sequence gives the same result (line 9, Table 1) as the rRNA genes argues against this. There is some indication of clustering of substitutions in the rRNA genes (DeSalle et al. 1987). However, the *Hin*dIII-B region appears to be like a "typical" coding sequence, i.e., most substitutions are synonymous and are fairly "randomly" spread over the region. Whether there is nonrandomness in the pattern of substitutions in the DNAs we studied is a moot point. We used "natural" DNAs that had diverged due to evolutionary processes. It is highly unlikely that evolution produces truly random patterns of substitutions. We studied *real* sequences, which are what are of interest in evolutionary studies.

These results provide an apparently quite accurate conversion of change in median melting temperature and percent base pair mismatch. The accuracy of the conversion is indicated by the high correlation coefficient, >0.98 (Table 2). Studies that employed the previously accepted 1:1 conversion must now be reevaluated in light of these new data. Rates of DNA evolution based on ΔT_m's must be faster than previously assumed. The difference is not trivial: 70% faster. Finally, this more accurate conversion will allow better comparison between results from DNA–DNA hybridization experiments and sequence data.

Acknowledgments. We thank Amanda Benson and George Amato for assistance, John Kochmer for help with the analysis, and Roy Britten, Donald Crothers, Dieter Maier, and Anette Preiss for helpful comments and advice. Financial support was provided by grants to J.R.P. from the National Science Foundation (BSR 8500207) and the National Institutes of Health (GM 36648).

References

Bautz EK, Bautz FA (1964) The influence of noncomplementary bases on the stability of ordered polynucleotides. Proc Natl Acad Sci USA 52:1476–1481

Benvenista RE (1985) The contribution of retroviruses to the study of mammalian evolution. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum, New York, pp 359–417

Britten RJ (1986) Rates of DNA sequence evolution differ between taxonomic groups. Science 231:1393–1398

Britten RJ, Graham DE, Neufeld BR (1974) Analysis of repeating DNA sequences by reassociation. In: Grossman L, Moldave K (eds) Methods in enzymology, vol 29e. Academic Press, New York, pp 363–418

Caccone A, Powell JR (1987) Molecular evolutionary divergence among North American cave crickets. II. DNA–DNA hybridization. Evolution 41:1215–1238

Caccone A, Amato G, Powell JR (1987) Intraspecific DNA divergence in *Drosophila:* a study on parthenogenetic *D. mercatorum.* Mol Biol Evol 4:343–350

Caccone A, Amato G, Powell JR (1988) Rates and pattern of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup. Genetics (in press)

Catzeflis FM, Sheldon FH, Ahlquist JE, Sibley CG (1987) DNA–DNA hybridization evidence of the rapid rate of muroid rodent DNA evolution. Mol Biol Evol 4:242–253

Chang CT, Hain TC, Hutton JR, Wetmur JG (1974) Effects of microscopic and macroscopic viscosity on the rate of renaturation of DNA. Biopolymers 13:1847–1858

DeSalle R, Freedman T, Prager E, Wilson AC (1987) Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila.* J Mol Evol 26:157–164

Hall TJ, Grula JW, Davidson EH, Britten RJ (1980) Evolution of sea urchin non-repetitive DNA. J Mol Evol 16:95–110

Hunt JA, Hall TJ, Britten RJ (1981) Evolutionary distance in Hawaiian *Drosophila* measured by DNA reassociation. J Mol Evol 17:361–367

Hutton JR, Wetmur JG (1973) Effect of chemical modification on the rate of renaturation of deoxyribonucleic acid. Deaminated and glyoxalated deoxyribonucleic acid. Biochemistry 12:558–563

Kohne DE (1970) Evolution of higher organism DNA. Q Rev Biophys 33:327–375

Koop BF, Goodman M, Zu P, Chan JL, Slighton JL (1986) Primate η-globin DNA sequences and man's place among the great apes. Nature 319:234–238

Laird CD, McConaughy BL, McCarthy BJ (1969) Rate of fixation of nucleotide substitutions in evolution. Nature 224:149–154

McDonell MW, Simon MN, Studier FW (1977) Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. J Mol Biol 110:119–131

Melchior WB, von Hippel PH (1973) Alteration of the relative stability of dA–dT and dG–dC base pairs in DNA. Proc Natl Acad Sci USA 70:298–302

O'Brien SJ, Nash WG, Wildt DE, Bush ME, Benveniste RE (1985) A molecular solution to the riddle of the giant panda's phylogeny. Nature 317:140–144

Orosz JM, Wetmur JG (1977) DNA melting temperature and renaturation rates in concentrated alkylammonium salt solutions. Biopolymers 16:1183–1199

Powell JR, Caccone A, Amato GD, Yoon C (1986) Rates of nucleotide substitutions in *Drosophila* mitochondrial DNA and nuclear DNA are similar. Proc Natl Acad Sci USA 83:9090–9093

Sheldon FH (1987) Rates of single-copy DNA evolution in herons. Mol Biol Evol 4:56–69

Sibley CG, Ahlquist JE (1984) The phylogeny of the hominoid primates as indicated by DNA–DNA hybridization. J Mol Evol 20:2–15

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517

Uhlenbeck O, Harrison R, Doty P (1968) Some effects of noncomplementary bases on the stability of helical complexes of

polyribonucleotides. In: Pullman B (ed) Molecular associations in biology. Academic Press, New York, pp 107–114
Wolstenholme DR, Clary DO (1985) Sequence evolution of *Drosophila* mitochondrial DNA. Genetics 109:725–744
Zwiebel LJ, Cohn VH, Wright DR, Moore GP (1982) Evolution of single-copy DNA and the ADH gene in seven drosophilids. J Mol Evol 19:62–71

Received October 15, 1987/Revised and accepted January 25, 1988