J Mol Evol (1999) 48:245–247

JOURNAL OF MOLECULAR EVOLUTION

Letters to the Editor

Microsatellite Allele Size Homoplasy in the Guppy (*Poecilia reticulata*)

John S. Taylor, Justina S.P. Sanny, Felix Breden

Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

Received: 31 March 1998 / Accepted: 6 August 1998

Homoplasy refers to trait similarity due to convergent or parallel evolution (Brooks and McLennan 1991). Mistaking homoplasy for homology can lead to inaccurate estimation of paternity, relatedness, within-population genetic diversity, among-population genetic divergence, and interspecific phylogenies. Microsatellites have become the marker of choice for intraspecific genetic studies and have also been used in interspecific investigations. However, microsatellites have two features which make them especially prone to homoplasy. First, theoretical considerations (Levinson and Gutman 1987; Shriver et al. 1993; Valdes et al. 1993) and empirical data (Weber and Wong 1993; Crawford and Cuthbertson 1996; Primmer et al. 1996; Schug et al. 1997) suggest that microsatellites mutate by gaining or losing repeats one unit at a time, a pattern referred to as the stepwise mutation model (Ohta and Kimura 1973; Shriver et al. 1993). The stepwise mutation model restricts possible mutational outcomes which increases the chance that taxa sharing alleles have converged on those alleles. Second, the high mutation rate at many microsatellite loci (Weber and Wong 1993; Primmer et al. 1996; Crawford and Cuthbertson 1996) increases the likelihood of the same mutation occurring independently (i.e., in parallel) in different lineages. Given this predisposition for convergent and parallel evolution, it is important to estimate homoplasy in real microsatellite data sets in order to evaluate its impact.

Homoplasious alleles can be identified by mapping allele size data onto a phylogeny and by sequencing alleles. Mapping allele sizes onto a phylogeny reveals instances where different lineages have evolved same-sized alleles independently. Authors who have uncovered homoplasious alleles using this phylogenetic approach include Zardoya et al. (1996), Jin et al. (1996), Ho et al. (1996), Orti et al. (1997), and Doyle et al. (1998). Sequence data may also provide evidence that same-sized alleles evolved independently. Grimaldi and Crouau-Roy (1997) discovered flanking sequence variation in the human MIB microsatellite locus which resulted in different, and presumably independently derived, 350-bp alleles. Variation in interruptions within repeats suggests that same-sized alleles evolved independently at the TNFa microsatellite locus in humans and gorillas (Blanquer-Maumont and Crouau-Roy 1995) and at the A113 locus in the bee genus *Apis* (Estoup et al. 1995).

Despite the success of the sequencing studies mentioned above in detecting homoplasious alleles, the pattern of microsatellite mutation (i.e., the addition and deletion of repeat units) predicts that for most microsatellite loci same-sized alleles will have the same sequence whether they are homoplasious or homologous. Compound microsatellites are an exception because they contain at least two different repeated regions. The advantage of this class of repeats for detecting homoplasy

Correspondence to: J.S. Taylor

Table 1. Allele size and sequence data at a compound microsatellite locus in 10 guppy (*Poecilia reticulata*) populations[a]

Population	Allele length (bp)	N1	Aligned allele sequence	N2
Isla de Margarita (Venezuela)	22	8	ggga$(tc)_2$gg$(TC)_6$ TT$(TG)_4$. . ttta	1
Springlands (Guyana)	22	18	ggga$(tc)_2$gg$(TC)_6$ TT$(TG)_4$. . ttta	1
Lelydorp (Surinam)	22	20	ggga$(tc)_2$gg$(TC)_6$ TT$(TG)_4$. . ttta	2
New Amsterdam (Guyana)	22	20	ggga$(tc)_2$gg$(TC)_6$ . . $(TG)_5$. . ttta	1
Paria River (Trinidad)	22	19	ggga$(tc)_2$gg$(TC)_6$ . . $(TG)_5$. . ttta	2
Arima River (Trinidad)	22	19	ggga$(tc)_2$gg$(TC)_6$ . . $(TG)_5$. . ttta	2
Aripo River (Trinidad)	22	14	ggga$(tc)_2$gg$(TC)_7$ . . $(TG)_4$. . ttta	2
Quare River (Trinidad)	22	36	ggga$(tc)_2$gg$(TC)_7$ . . $(TG)_4$. . ttta	1
Oropuche River (Trinidad)	22	4	ggga$(tc)_2$gg$(TC)_7$ . . $(TG)_4$. . ttta	4
Guanare River (Venezuela)	24	24	ggga$(tc)_2$gg$(TC)_7$ TT$(TG)_3$TAttta	1
Isla de Margarita (Venezuela)	24	12	ggga$(tc)_2$gg$(TC)_7$ . . $(TG)_5$. . ttta	1
Arima River (Trinidad)	24	1	ggga$(tc)_2$gg$(TC)_7$ . . $(TG)_5$. . ttta	1
Paria River (Trinidad)	26	1	ggga$(tc)_2$gg$(TC)_8$ . . $(TG)_5$. . ttta	1
Quare River (Trinidad)	30	4	ggga$(tc)_2$gg$(TC)_{11}$. . $(TG)_4$. . ttta	1
Oropuche River (Trinidad)	30	2	ggga$(tc)_2$gg$(TC)_{11}$. . $(TG)_4$. . ttta	2

[a] N1 refers to the number of alleles scored by size. N2 refers to the number of alleles sequenced. Flanking sequence data are in lowercase and the repeat sequence is in uppercase letters.

is that mutations in different repeat regions creating alleles of the same size can be recognized by different combinations of repeat number (Garza and Freimer 1996). Including our study, three studies have sequenced compound repeats specifically to uncover microsatellite homoplasy. Garza and Freimer (1996) sequenced a compound repeat called Mfd59 and uncovered 11 chimpanzee alleles where only 8 had been detected using size variation alone and 6 human alleles where only 4 were detected using size variation. Metzgar et al. (1998) sequenced alleles of ERK1, a compound repeat in *Candida albicans. Candida albicans* isolates from 15 human patients were all different with respect to ERK1 sequence, but there were only six allele sizes. Metzgar et al. (1998) discovered two 237-bp alleles, five 246-bp alleles, and five 249-bp alleles. We discovered a compound repeat with the sequence $(TC)_n(TG)_n$, in the guppy (*Poecilia reticulata*). We surveyed 10 guppy populations at this locus and uncovered four different-sized alleles. We sequenced 23 alleles and discovered three 22-bp alleles and two 24-bp alleles (Table 1). Altogether, our sample contained 7 allele sequences but only 4 allele sizes.

Assuming that different repeat combinations detected by sequencing reflect convergent evolution in allele size, and assuming that compound microsatellites have the same mutation rate and pattern as perfect microsatellites, compound microsatellites can provide us with an estimate of microsatellite homoplasy. We estimated microsatellite homoplasy using the compound repeat data in our study, in Garza and Freimer's (1996) Mfd59 survey of chimpanzees and humans, and in Metzgar and coworkers' (1998) *C. albicans* ERK1 survey. By dividing the number of allele sequences by the number of allele sizes [(7/4 + 11/8 + 6/4 + 12/6)/4], we estimate that for every microsatellite allele size there are approximately 1.66 alleles. In this calculation we included 12 rather than 15 *C. albicans* sequences. Three of the 246-bp alleles reported by Metzgar et al. (1998) differed only in single nucleotide substitutions in the flanking sequence, and we wanted to estimate microsatellite homoplasy without the confounding influence of flanking sequence variation.

The consequences of not recognizing homoplasy in a microsatellite data set are demonstrated by some of the studies discussed above. An individual in Grimaldi and Crouau-Roy's (1997) study who would have been scored as homozygous based upon allele size was heterozygous, possessing different 350-bp alleles. This heterozygous individual demonstrates how homoplasy may lead to underestimates of within-population and, indeed, within-individual genetic variation. Our study and the studies by Garza and Freimer (1996) and Estoup et al. (1995) demonstrate that homoplasy can lead to underestimates of population divergence. Almost all guppies surveyed from Trinidad, Guyana, and Surinam possessed 22-bp alleles. However, sequence data divide guppies from these three countries into three groups depending upon whether they possess alleles with the sequence $(TC)_6TT(TG)_4$, $(TC)_6(TG)_5$, or $(TC)_7(TG)_4$. Garza and Freimer (1996) show that microsatellites would underestimate genetic divergence within humans and between humans and chimpanzees. Estoup et al. (1995) illustrated the effect of microsatellite homoplasy by contrasting the high divergence estimated from mtDNA sequences with the low divergence estimated from the A113 microsatellite locus. By comparing ERK1 data from *C. albicans* and *C. dubliniensis*, Metzgar et al. (1998) demonstrated that homoplasy conceals interspecific divergence.

More studies which map microsatellite allele sizes onto phylogenies or report microsatellite sequences will elucidate the ubiquity and impact of microsatellite homoplasy. Only then can researchers balance the expense and time required to reconstruct phylogenies or sequence alleles against the dangers of ignoring homoplasy.

Acknowledgments. We would like to thank Dr. Bernard Crespi and Dr. Andrew Beckenbach and two anonymous reviewers for their comments on the manuscript. This work was supported by a grant from NSERC to Felix Breden.

References

Blanquer-Maumont A, Crouau-Roy B (1995) Polymorphism, monomorphism, and sequences in conserved microsatellites in primate species. J Mol Evol 41:492–497

Brooks DR, McLennan DA (1991) Phylogeny, ecology, and behaviour. University of Chicago Press, Chicago and London, p 14

Crawford AM, Cuthbertson RP (1996) Mutations in sheep microsatellites. Genome Res 6:876–879

Doyle JJ, Morgante M, Tingey SV, Powell W (1998) Size homoplasy in chloroplast microsatellites of wild perennial relatives of soybean (*Glycine* subgenus *Glycine*). Mol Biol Evol 15:215–218

Estoup A, Tailliez C, Cornuet J-M, Solignac M (1995) Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). Mol Biol Evol 12:1074–1084

Garza JC, Freimer NB (1996) Homoplasy for size at microsatellite loci in humans and chimpanzees. Genome Res 6:211–217

Grimaldi M, Crouau-Roy B (1997) Microsatellite allelic homoplasy due to variable flanking sequences. J Mol Evol 44:336–340

Ho K-F, Craddock EM, Piano F, Kambysellis MP (1996) Phylogenetic analysis of DNA length mutations in a repetitive region of the Hawaiian *Drosophila* yolk protein gene *Yp2*. J Mol Evol 43:116–124

Jin L, Macuabas C, Hallmayer J, Kimura A, Mignot E (1996) Mutation rate varies among alleles at a microsatellite locus: Phylogenetic evidence. Proc Natl Acad Sci USA 93:15285–15288

Levinson G, Gutman GA (1987) Slipped-strand mispairing: A major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221

Metzgar D, Field D, Haubrich R, Wills C (1998) Sequence analysis of a compound coding-region microsatellite in *Candiada albicans* resolves homoplasies and provides a high-resolution tool for genotyping. FEMS Immunol Med Microbiol 20:103–109

Ohta T, Kimura M (1973) The model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a genetic population. Genet Res 22:201–204

Orti G, Pearse DE, Avise JC (1997) Phylogenetic assessment of length variation at a microsatellite locus. Proc Natl Acad Sci USA 94: 10745–10749

Primmer CR, Ellegren H, Saino N, Møller AP (1996) Directional evolution in germline microsatellite mutations. Nature Genet 13:391–393

Schug MD, Mackay TFC, Aquadro CF (1997) Low mutation rates of microsatellite loci in *Drosophila melanogaster*. Nature Genet 15: 99–102

Shriver MD, Li J, Chakraborty R, Boerwinkle E (1993) VNTR allele frequency distributions and the stepwise mutation model: A computer simulation. Genetics 134:983–993

Valdes AM, Slatkin M, Freimer N (1993) Allele frequencies at microsatellite loci: The stepwise mutation model revisited. Genetics 133:737–749

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2:1123–1128

Zardoya R, Vollmer DM, Craddock C, Streelman T, Karl S, Meyer A (1996) Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (Pisces: Perciformes). Proc R Soc Lond B 263:1589–1598