Volume 307, number 1, 113–115 FEBS 10967 July 1992


*Minireview*

# Recent advances in minisatellite biology

John A.L. Armour and Alec J. Jeffreys

*Department of Genetics, University of Leicester, Leicester LE1 7RH, UK*

Received 17 February 1992

Highly polymorphic tandemly repeated 'minisatellite' loci are very abundant in the human genome, and of considerable utility in human genetic analysis. This review describes the use of an ordered-array Charomid library in the systematic and efficient cloning of these regions, and in the analysis of the relative overlap between the different probes used to screen for hypervariable loci. Recent work on the process of mutation leading to the generation of new-length alleles is also discussed, including the observation that at least some mutations may be due to unequal exchanges.

Minisatellite; VNTR; Charomid; Mutation

## 1. INTRODUCTION

Highly polymorphic 'minisatellite' regions of the genome include the most genetically informative human loci so far described [1–3]. These regions are composed of tandem repeats of a short (8–90 bp) repeated sequence; allelic variation at these loci derives from variation in the number of tandem repeats, and hence in the length of the region. Since the number of possible allelic states is in principle unlimited, it is possible for a large number of alleles to be generated, each of which is rare in the population. Indeed, the most variable loci so far characterized show such a distribution, with a large number of rare alleles, and consequently have a high level of genetic informativeness [4–7]. For this reason they have found uses in a wide variety of human genetic analysis [8,9]. In this brief review we will outline some recent work from our laboratory which addresses two major problems in the field: how may highly informative minisatellites be isolated efficiently by cloning, and what mechanisms are operating in the mutations by which minisatellite variation is generated?

## 2. MINISATELLITE ISOLATION BY CLONING

The first highly polymorphic minisatellites to be studied were those which had been isolated fortuitously; these include the random DNA clone, D14S1 [1], and minisatellites discovered near gene sequences [10–12]. The discovery of probes which hybridized at low stringency to large numbers of dispersed hypervariable regions not only allowed the development of the multilocus minisatellite profile or 'DNA fingerprint' [8,13,14], but also provided, in principle, a simple method for the systematic isolation of hypervariable loci, namely screening genomic libraries by hybridization with these probes.

This observation formed the basis of the work of Nakamura et al. [3,15] in the screening of human cosmid libraries with oligonucleotides similar in sequence to the DNA fingerprinting probes. While this resulted in the successful isolation of many useful genetic markers, the average informativeness was low. The work of Wong et al. [6,7] was more specifically addressed to the isolation of the most variable loci, and exploited an observed relation between size and variability among the loci detected by DNA fingerprinting probes: while many of the hybridizing fragments smaller than about 2 kb were frequently derived from monomorphic or only minimally variable loci, fragments larger than about 4 kb were nearly always derived from hypervariable loci [14]. Moreover, since the tandem repeat array composing a minisatellite allele often lacks restriction sites even for frequently cutting enzymes such as *Mbo*I, most of the fragments derived from highly polymorphic loci would be found on unusually large restriction fragments for these frequently cutting enzymes. Thus size selection before cloning of large restriction fragments not only enriches the selected DNA for tandemly repeated loci, but also enriches for fragments derived from highly polymorphic loci at the expense of the monomorphic or minimally variable loci.

This approach, in which a library of size-selected human DNA was screened by hybridization with DNA fingerprinting probes, was successfully applied by Wong et al. [6,7], who isolated six highly informative

*Correspondence address:* J.A.L. Armour, Department of Genetics, University of Leicester, Leicester LE1 7RH, UK. Fax: (44) (533) 52 33 78.

loci from a phage lambda genomic library. Despite the initial success of this approach, further screening of the same library led to the repeated isolation of the same small subset of loci, suggesting that only a few of the many minisatellite loci potentially available for cloning were actually represented in the library. Since lambda phage require to be packaged into viable particles at every generation of growth, and since that packaging process is size-dependent, any loss of tandem repeat units from the cloned insert during growth might lead to inviability of some recombinant phage, and consequent under-representation of some loci in the library.

The problem of clone inviability due to loss of minisatellite repeats can be obviated by the use of cosmid vectors; however, most cosmid vectors have been designed to clone large (35–45 kb) fragments. Charomids [16] are a series of vectors designed for the cosmid cloning of smaller fragments, and appear well-suited to the efficient cloning of the fragments (approximately 4–20 kb) richest in hypervariable minisatellites. A Charomid library enriched in hypervariable minisatellites was constructed by cloning 4–9 kb human *Mbo*I fragments into the *Bam*HI site of Charomid 9–36. The resulting library was screened in ordered array, and was used successfully in the isolation of 27 new polymorphic minisatellites [17–19]. The use of an ordered array allows not only efficient screening and isolation of clones, but also comparison of the sets of clones which hybridize positively with each of the multilocus DNA fingerprinting probes used for screening. In this connection we are currently collaborating (with G. Vergnaud, Vert le Petit, France) on a study of sequence motifs within the sets of polymorphic clones; specifically, to what extent the naturally occurring DNA 'fingerprinting' probes recognize sets of loci overlapping with those detected by (random sequence) synthetic tandem repeats (STRs) [20,21].

## 3. MUTATIONAL ANALYSIS

The high level of variability within human populations at some minisatellite loci is generated by a high level of mutation to new length alleles in the germline. Indeed, at the most variable loci so far studied, the spontaneous rate of germline mutation is high enough to be measured directly by pedigree analysis [22]. These surveys of mutation among the loci studied in our laboratory have shown that the mutation process appears to be symmetrical with respect to allele size; mutations which increase the length of an allele are as likely as mutations which shorten alleles.

The mechanism of these length (repeat copy number) change mutations has been the subject of much recent work, and has been approached by studies on genetic markers immediately flanking the minisatellite array [23,24], as well as studies of the *internal* structure of minisatellite alleles [25]. The internal structure of minisatellite alleles can be investigated by taking advantage of a second level of variation at these loci: in addition to variation in the number of tandem repeated units, some loci have repeat units which differ slightly in sequence between repeats, such that any allele at such a locus is composed of a complex interspersion pattern of two or more repeat unit types. These repeat unit types can be mapped within the tandem array as a whole by restriction mapping [25], or, more recently, with repeat type-specific PCR primers [26].

One unexpected but informative finding from internal mapping studies has been the discovery of a gradient or polarity of variation within the tandem array. At the D1S8 locus [25] and another highly variable minisatellite (unpublished work), there appears to be a relatively invariant end of the array, at which there is relatively little variation in internal structure, and an 'ultravariable' end, at which most variation and mutation seems to occur. Thus most, if not all, turnover of repeat units during the evolution of these loci can be attributed to events occurring in the last few repeats at one end of the array. No evidence for the involvement of unequal meiotic recombination in the mutation process was forthcoming from initial studies [23–25], and indeed this mechanism could be ruled out as the major mechanism of allele length change at one locus [24], at which entirely intra-allelic processes such as replication slippage or unequal sister chromatid exchange events account for most length change mutations.

However, more recent work [26] has shown that a significant minority of mutation events at some loci can only be satisfactorily explained by postulating an unequal exchange or gene conversion event between alleles. Thus true unequal recombination events may after all contribute to the generation of variation at human minisatellite loci. If so, then the high rate of germline mutation at these loci, coupled with a significant fraction of mutations attributable to recombination, may ultimately confirm early speculation that at least some human minisatellite loci may act as hot-spots for homologous recombination in human meiosis [27].

*Acknowledgements:* J.A.L.A. is a Wellcome Trust Postdoctoral Fellow. A.J.J. is Wolfson Professor of the Royal Society. This work was supported by grants to A.J.J. from the MRC and the Wolfson Foundation.

## REFERENCES

[1] Wyman, A.R. and White, R. (1980) Proc. Natl. Acad. Sci. USA 77, 6754–6758.
[2] Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985) Nature 314, 67–73.
[3] Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. (1987) Science 235, 1616–1622.
[4] Balazs, I., Baird, M., Clyne, M. and Maede, E. (1989) Am. J. Hum. Genet. 44, 182–190.
[5] Flint, J., Boyce, A.J., Martinson, J.J. and Clegg, J.B. (1989) Hum. Genet. 83, 257–263.

[6] Wong, Z., Wilson, V., Jeffreys, A.J. and Thein, S.L. (1986) Nucleic Acids Res. 14, 4605–4616.
[7] Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A.J. (1987) Ann. Hum. Genet. 51, 269–288.
[8] Gill, P., Jeffreys, A.J. and Werrett, D.J. (1985) Nature 318, 577–579.
[9] Knowlton, R.G., Brown, V.A., Braman, J.C., Barker, D., Schumm, J.W., Murray, C., Takvorian, T., Ritz, J. and Donis-Keller, H. (1986) Blood 68, 378–385.
[10] Higgs, D.R., Goodbourn, S.E.Y., Wainscoat, J.S., Clegg, J.B. and Weatherall, D.J. (1981) Nucleic Acids Res. 9, 4213–4224.
[11] Bell, G.I., Selby, M.J. and Rutter, W.J. (1982) Nature 295, 31–35.
[12] Capon, D.J., Chen, E.Y., Levinson, A.D., Seeberg, P.H. and Goeddel, D.V. (1983) Nature 302, 33–37.
[13] Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985) Nature 316, 76–79.
[14] Jeffreys, A.J., Wilson, V., Thein, S.L., Weatherall, D.J. and Ponder, B.A.J. (1986) Am. J. Hum. Genet. 39, 11–24.
[15] Nakamura, Y., Carlson, M., Krapcho, K., Kanamori, M. and White, R. (1988) Am. J. Hum. Genet. 3, 854–859.
[16] Saito, I. and Stark, G.R. (1986) Proc. Natl. Acad. Sci. USA 83, 8664–8668.
[17] Armour, J.A.L., Povey, S., Jeremiah, S. and Jeffreys, A.J. (1990) Genomics 8, 501–512.
[18] Armour, J.A.L., Crosier, M. and Jeffreys, A.J. (1991) Nucleic Acids Res. 19, 5447.
[19] Crosier, M., Armour, J.A.L. and Jeffreys, A.J. (1991) Nucleic Acids Res. 19, 5446.
[20] Vergnaud, G. (1989) Nucleic Acids Res. 17, 7623–7630.
[21] Vergnaud, G., Mariat, D., Apiou, F., Aurias, A., Lathrop, M. and Lauthier, V. (1991) Genomics 11, 135–144.
[22] Jeffreys, A.J., Royle, N.J., Wilson, V. and Wong, Z. (1988) Nature 332, 278–281.
[23] Wolff, R.K., Nakamura, Y. and White, R. (1988) Genomics 3, 347–351.
[24] Wolff, R.K., Plaetke, R., Jeffreys, A.J. and White, R. (1989) Genomics 5, 382–394.
[25] Jeffreys, A.J., Neumann, R. and Wilson, V. (1990) Cell 60, 473–485.
[26] Jeffreys, A.J., MacLeod, A., Tamaki, K., Neil, D.L. and Monckton, D.G. (1991) Nature 354, 204–209.
[27] Jarman, A.P. and Wells, R.A. (1989) Trends Genet. 5, 367–371.