Sequence-Tagged Sites (STSS) Spanning 4p16.3 and the Huntington Disease Candidate Region

JAMES F. GUSELLA,* MICHAEL R. ALTHERR,† ANDREA I. MCCLATCHLEY,* LYNN A. DOUCETTE-STAMM,† DAN TAGE,‡ SARAH PLUMMER,† NICOLET GROOT,* GLENN BARNES,* HOLGER HUMMERICH,|| FRANCIS S. COLLINS,§ DAVID E. HOUSMAN,† HANS LEHRACH,‖ MARCY E. MACDONALD,* GILLIAN BATES,‖ AND JOHN J. WASMUTH†

*Neurogenetics Laboratory, Massachusetts General Hospital and Department of Genetics, Harvard Medical School, Charlestown, Massachusetts 02129; †Department of Biological Chemistry, University of California, Irvine, California 92717; ‡Department of Biology and Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; §Howard Hughes Medical Institute, University of Michigan, Ann Arbor, Michigan 48109; and ‖Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, United Kingdom

Received September 19, 1991; revised December 12, 1991

The generation of sequence-tagged sites (STSS) has been proposed as a unifying approach to correlating the disparate results generated by genetic and various physical techniques being used to map the human genome. We have developed an STS map to complement the existing physical and genetic maps of 4p16.3, the region containing the Huntington disease gene. A total of 18 STSSs span over 4 Mb of 4p16.3, with an average spacing of about 250 kb. Eleven of the STSSs are located within the primary candidate HD region of 2.5 Mb between D4S126 and D4S168. The availability of STSSs makes the corresponding loci accessible to the general community without the need for distribution of cloned DNA. These STSSs should also provide the means to isolate yeast artificial chromosome clones spanning the HD candidate region. © 1992 Academic Press, Inc.

INTRODUCTION

The maturing of technologies for manipulating, cloning, and sequencing DNA has made the mapping and sequencing of the entire human genome a feasible goal. Several different approaches are being employed to generate physical maps of human chromosomes, including long-range restriction mapping by pulsed-field gel electrophoresis, ordering of isolated clones by use of irradiation-reduced hybrid lines or fluorescence detection-based in situ hybridization, and isolation of overlapping clone sets (or “contigs”), particularly using cosmids or yeast artificial chromosomes. The completion of the genome map requires that information from each of the different mapping procedures can be readily compared and related to the genetic linkage map being constructed in parallel. It has been proposed that the use of sequence-tagged sites (STSSs), for which specific primer pairs yield a characteristic polymerase chain reaction product as common signposts on the maps, would fulfill the requirement of a “common language” for comparison of the different maps (Olson et al., 1989).

The p16.3 cytogenetic band of chromosome 4 is one of the more extensively mapped regions of the human genome as a result of the search for the defect causing Huntington’s disease (Gusella, 1991). The HD gene has been confined to a segment of 4p16.3 between D4S10 and the telomere (Gilliam et al., 1987a). A long-range restriction map of this region includes only one gap and spans 4.5 Mb from D4S10 to the telomere (Bučan et al., 1990; Bates et al., 1991). Irradiation-reduced hybrids have also been applied to mapping 4p16.3 (Cox et al., 1989; Doucette-Stamm et al., 1991). Numerous phage and cosmid clones have been mapped to the region and several have yielded DNA polymorphisms, permitting construction of a detailed genetic map (MacDonald et al., 1989a; Youngman et al., 1989; Allitto et al., 1991a).

To assist in comparing the evolving physical maps being generated with different cell lines and techniques and to lay the foundation for isolating overlapping YAC clones for the HD candidate region, we have generated a set of 18 STSSs spanning 4p16.3 at an average spacing of 250 kb.

MATERIALS AND METHODS

Origin of cell lines and clones. HHW416, a human–hamster hybrid containing only human chromosome 4, and HHW693, containing only human 4pter–4p15.1 and 5p15.1–5cen as part of a naturally occurring t(4;5) chromosome with 5q deleted, have been reported previously (Wasmuth et al., 1986). Clones used for generation of primer sets have been used successfully in Southern blot experiments, and their positions on genetic and physical maps have been described previously and are listed in Table 1. STSSs have been named by adding * followed by a sequential identification number to the locus symbol.

DNA sequencing and preparation of oligonucleotides. DNA sequencing was obtained by the procedure of Sanger et al. (1977). Primers were synthesized using a Biosearch Cyclone DNA synthesizer or an Applied Biosystems DNA synthesizer or were purchased from Genosys (Houston).
TABLE 1
Description of 4p16.3 Sequence-Tagged Sites

<table>
<thead>
<tr>
<th>STS</th>
<th>Clone</th>
<th>PCR primer sequences</th>
<th>PCR product size (bp)</th>
<th>PCR reaction mix</th>
<th>PCR program</th>
</tr>
</thead>
<tbody>
<tr>
<td>D4S10+1</td>
<td>G5</td>
<td>5' GACATGGGATCGTCGGGCTCCTGCTGG 3'</td>
<td>192</td>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>D4S10+2</td>
<td>I5</td>
<td>5' GAGAATGGGCTGGAGGCGAGCCCGGAG 3'</td>
<td>213</td>
<td>B</td>
<td>2</td>
</tr>
<tr>
<td>D4S126+1</td>
<td>p309</td>
<td>5' GCACTGCTCACTGGTCAAGCTGGAG 3'</td>
<td>184</td>
<td>B</td>
<td>3</td>
</tr>
<tr>
<td>D4S125+1</td>
<td>pYNZ32</td>
<td>5' GGTGGTCTCCGTCAAGCTGGAG 3'</td>
<td>159</td>
<td>B</td>
<td>4</td>
</tr>
<tr>
<td>D4S180+1</td>
<td>L19ps11</td>
<td>5' GCTGTTCTACAAGTTGCTCCTGCTG 3'</td>
<td>103</td>
<td>A</td>
<td>4</td>
</tr>
<tr>
<td>D4S90+1</td>
<td>pBS674-VR</td>
<td>5' GAGCTCTACCGGTCCCAAGAGGAG 3'</td>
<td>1066</td>
<td>A</td>
<td>3</td>
</tr>
<tr>
<td>D4S182+1</td>
<td>pV19Rco2.3</td>
<td>5' GCTGTTCTCCACAGTGGAGCTGCTCCTG 3'</td>
<td>213</td>
<td>B</td>
<td>3</td>
</tr>
<tr>
<td>D4S181+1</td>
<td>L14ps7</td>
<td>5' GAGGCCAGTTGAGCCCTCAAGGTTA 3'</td>
<td>141</td>
<td>B</td>
<td>3</td>
</tr>
<tr>
<td>D4S43+1</td>
<td>C39Kpn4.5</td>
<td>5' GCTCAGATCTGAGATTGCT 3'</td>
<td>243</td>
<td>B</td>
<td>3</td>
</tr>
<tr>
<td>D4S45+2</td>
<td>LCD450</td>
<td>5' GACTGTTCTGAGGAGCTG 3'</td>
<td>378</td>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>D4S166+1</td>
<td>L6-H10</td>
<td>5' UCACCGATGTTGAGGAGCTG 3'</td>
<td>899</td>
<td>A</td>
<td>3</td>
</tr>
<tr>
<td>FGFR3+1</td>
<td>pBS338H-A</td>
<td>5' AAAGTACCGTACAGGAG 3'</td>
<td>202</td>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>D4S168+1</td>
<td>E4ps2</td>
<td>5' TCCAGACGGCTGAGCCAGAGAG 3'</td>
<td>118</td>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>D4S115+1</td>
<td>p252.3</td>
<td>5' CACCGAGGAGCCTCAGAGCTG 3'</td>
<td>164</td>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>IDUA+1</td>
<td>p157.9</td>
<td>5' GATGACCTCTCGACATCTG 3'</td>
<td>230</td>
<td>B</td>
<td>3</td>
</tr>
<tr>
<td>D4S232+1</td>
<td>H4ps18</td>
<td>5' CACCGAGGAGCCTCAGAGCTG 3'</td>
<td>117</td>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>D4S60+1</td>
<td>D5</td>
<td>5' CACCGAGGAGCCTCAGAGCTG 3'</td>
<td>569</td>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>D4S142+1</td>
<td>S8-18</td>
<td>5' GATGACCTCTCGACATCTG 3'</td>
<td>267</td>
<td>B</td>
<td>3</td>
</tr>
</tbody>
</table>

Polymerase chain reaction. The polymerase chain reaction and cycling conditions for each primer set are listed in Table 1 as buffer mixture A or B and programs 1-4. Mixture A consists of 67 mM Tris-HCl, pH 8.8; 6.7 mM MgCl₂; 16.6 mM ammonium sulfate; 10 mM β-mercaptoethanol; 10% DMSO; and 1.25 mM dNTPs. Mixture B is 10 mM Tris-HCl, pH 8.3; 1.5 mM MgCl₂; 50 mM KCl; 0.01% gelatin; and 200 μM dNTPs. Both sets of reactions included 10 ng/μl of each primer, 0.02 unit/μl Taq DNA polymerase (Perkin-Elmer/Cetus), and 10 ng/μl genomic DNA and were carried out using one of the following programs designated in Table 1:

1. 94°C,30s
2. 94°C,30s
3. 94°C,30s
4. 94°C,30s

RESULTS

Generation of STSs

Individual clones from a variety of sources have been positioned previously on the long-range restriction map of 4p16.3 (Bucan et al., 1990; Whaley et al., 1991; Bates et al., 1991; Liu et al., 1991). The list of those selected for sequence analysis to provide a source of STSs is presented in Table 1, along with the primer pair, amplification conditions, and predicted size of the amplification product for each resulting STS.

Each of the STS reactions was initially tested using DNA from HHW416, a human × hamster hybrid line containing only human chromosome 4, or from HHW693, a human × hamster hybrid line containing 4p15.1→4pter as part of a translocation chromosome and DNA from tsH1, the hamster parent (Wasmuth et al., 1986). Since 5p15.1→cen is the only other human DNA present in HHW693, the chromosome 4 origin of each PCR product could be confirmed using hybrids containing either chromosome 4 (HHW416) or chromosome 5 (HHW599) as the only human DNA. Subsequently, the STS reactions were demonstrated also to yield the expected fragment as the primary product in amplifying human genomic DNA from a variety of sources, such as lymphoblastoid cell lines, placenta, or other tissues, and as the only product from the clone used to generate the corresponding STS.

Figure 1 shows typical examples of PCR amplification products from genomic DNA for several of the STSs.
reported here. In each case, the major PCR product closely matches the predicted size. Occasionally, a fragment was also amplified from hamster DNA but was readily distinguishable from the human product based on its size. For example, the ~450-bp product from the D4S126*1 STS reaction seen in both the HHW416 and hamster lanes is due to amplification of hamster DNA (Fig. 1B, lanes 2 and 3), as are the ~450-bp and ~700-bp products in the D4S180*1 STS amplification (Fig. 1A, lane 7) and the numerous light bands in the D4S125*1 amplification (Fig. 1B, lanes 5 and 6).

Location of STSs

The location and corresponding locus symbol for the 18 STSs and for eight polymorphisms for which PCR assays are available are given in Fig. 2. Two STSs are given for the extended locus D4S43, which spans 170 kb. The more centromeric D4S43*1 was generated from the sequence adjacent to the VNTR polymorphism that has yet to be sequenced. D4S181*1 maps within the 250-kb NotI fragment distal to D4S182, but has not yet been localized more precisely. This will require the more detailed restriction map of this segment that will be generated from YAC cloning of the region.

Two STSs are reported for the extended locus D4S43, which spans 170 kb. The more centromeric D4S43*1 was generated from the sequence adjacent to the VNTR polymorphism that has yet to be sequenced. D4S181*1 maps within the 250-kb NotI fragment distal to D4S182, but has not yet been localized more precisely. This will require the more detailed restriction map of this segment that will be generated from YAC cloning of the region.

D4S182*1 and D4S181*1 were derived from subclones of cosmids Y12 and L14, respectively (Lin et al., 1991). D4S182*1 is adjacent to a VNTR polymorphism that has yet to be sequenced. D4S181*1 maps within the 250-kb NotI fragment distal to D4S182, but has not yet been localized more precisely. This will require the more detailed restriction map of this segment that will be generated from YAC cloning of the region.
FIG. 2. Physical map locations of 4p16.3 STS and polymorphism PCR assays. The locations of 18 new STSs (listed in Table 1), two site polymorphisms, two dinucleotide repeat polymorphisms, and four VNTR polymorphisms for which PCR assays have been established are shown below the long-range restriction map of 4p16.3 with a black line to indicate the smallest interval to which each has been localized. The corresponding locus numbers are given above the map. The proximal and distal portions of the map are separated by a gap of unknown size, where fragments have not yet been visualized. The map of the proximal region is drawn from Bates et al. (1991), Lin et al. (1991), and Whaley et al. (1991), and the map of the distal region is drawn from Whaley et al. (1991). Sites (N, NotI; M, MluI; R, NruI) prone to partial digestion are shown in lighter typeface. The Huntington disease gene candidate region (Bates et al., 1991) is shown as the cross-hatched portion of the map. Square brackets around the NotI sites at FGFR3 denote the presence of additional sites that have not all been mapped precisely. The PCR assays shown in the figure are: 1, D4S10*1; 2, D4S10*2; 3, D4S10 HindIII polymorphism assay (McIntosh et al., 1989); 4, D4S126*1; 5, D4S125*1; 6, D4S125 VNTR assay (Richards et al., 1991); 7, D4S180*1; 8, D4S127 dinucleotide repeat repeat assay (Taylor et al., 1991); 9, D4S96*1; 10, D4S95 VNTR assay (Allitto et al., in preparation). Together, D4S168*1 and D4S115*1 may permit the isolation of YAC clones that will span the gap in the physical map, which has not been bridged by other techniques.

IDUA*1 lies within the D4S111 locus, which has recently been renamed IDUA, encoding a L-iduronidase, the disease gene in Hurler syndrome (MacDonald et al., 1991b). Two other segments can be amplified from this locus, one crossing a polymorphic KpnI site (Scott et al., 1991a) and a second spanning a VNTR polymorphism (Scott et al., 1991b). D4S232*1 is from cosmID14, which lies within a 400-kb segment of the telomeric NotI fragment of 4p (Whaley et al., 1991), but has not been assigned more precisely due to differences in methylation patterns affecting the NotI sites in pulsed-field gel mapping of this region. It is probably located close to PDEB, encoding the B polypeptide of cGMP phosphodiesterase, the human homologue of the mouse rd gene (Weber et al., 1991; Bowes et al., 1990).

D4S90*1 was generated from the single-copy probe D5, which detects RFLP with HincII and Stul (Youngman et al., 1989). This locus is currently the most telomeric marker that has been formally placed on the linkage map of the chromosome. The PCR product crosses both a HincII and a Stul site within D5, but we have not determined whether either of these represents a polymorphic site. D4S142*1 was derived from 88-18, a single-copy probe detecting an MspI RFLP and located at the proximal end of a YAC clone spanning the 100 kb that includes the telomere of 4p (Bates et al., 1990). It is located proximal to the recently reported D4S169 locus, for which a simple-sequence repeat can be assayed by PCR (Pritchard et al., 1991).

DISCUSSION

The average spacing between our new STSs from D4S10 to the telomere is about 250 kb, with the density being highest in the HD candidate region. The most notable segments without STSs are the 450 kb between D4S180 and D4S95 and the gaps of indeterminate size between IDUA, D4S232, and D4S90. The first of these gaps contains a dinucleotide repeat polymorphism at D4S127 for which a PCR reaction is available (Taylor et al., 1992). Filling the gaps between IDUA, D4S232, and D4S90 will then require a more specific positioning of D4S232. Also unrepresented is the region between D4S168 and D4S115 for which pulsed-field fragments have yet to be visualized. It is unlikely that this unmapped gap is large, given the number of random clones already isolated from 4p16.3, and it may be possible to cross it by isolating YAC clones using the D4S168 and D4S115 primer sets.

Taken together with the single-site, dinucleotide repeat, and VNTR polymorphisms for which PCR reactions have been worked out, the 18 STSs reported here bring to 26 the number of PCR assays that could be used to screen for YACs in 4p16.3. The 11 new STSs between and including D4S126 and D4S168 should prove sufficient to isolate YAC clones spanning most or all of the HD candidate region (Bates et al., 1991). Moreover, the density of STSs in the segment can be increased if...
needed by developing additional primer sets from the ends of YACs as they are isolated. We have already isolated YAC clones using the D4S95 and D4S166 STS assays and are pursuing the goal of cloning the HD candidate region and the rest of 4p16.3 in its entirety. Of more immediate impact, the development of these assays and are pursuing the goal of cloning the candidate region and the rest of 4p16.3 makes the corresponding loci and their surrounding regions readily accessible to the entire research community without the cost or difficulty of distributing cloned DNAs.

ACKNOWLEDGMENTS

The authors thank Dr. Phyllis McAlpine for helpful suggestions concerning the naming of STSs. This work was supported by PHS Grants HG00169 and NS16387 (J.F.G.) and HG00322 and NS25631 (J.J.W.). Additional funding was received from the Hereditary Disease Foundation, the Joan and William Schreyer Research Fund to Cure Huntington’s Disease, and the Huntington’s Disease Society of America.

REFERENCES


