



**THE HUMAN GENE MAP NEAR THE FRAGILE X**

by Graeme Kemble Suthers.

**Second of Two Volumes.**

**APPENDICES.**

## TABLE OF CONTENTS

### Volume One

	<i>Page</i>
DECLARATION	iii
TABLE OF CONTENTS	iv
LIST OF TABLES	viii
LIST OF FIGURES	ix
SUMMARY	x
ACKNOWLEDGEMENTS	xiii
Chapter 1:	
<u>Literature Review</u>	1
Introduction	3
The Fragile X syndrome: Clinical features	8
Cytogenetics	12
Population & segregation studies	17
Human gene map: Introduction	23
Polymorphic loci	25
Two-point linkage analysis	27
Multipoint linkage analysis	35
The Fragile X syndrome: Linkage studies	44
Hypotheses	53
Conclusion	55

## Chapter 2:

<u>Material and Methods</u>	59
Human DNA samples and cell lines	61
DNA probes and radiolabelling	69
Identification of RFLPs	76
Linkage analysis - implementation	78
- Normal (CEPH) pedigrees	90
- Fragile X pedigrees	92
- Homogeneity tests	95
- Estimation of confidence intervals	99

## Chapter 3:

<u>The locus for a non-specific form of X-linked mental retardation (XLMR) is located at Xp11.</u>	105
Introduction	107
Methods	109
Results	112
Discussion	116

## Chapter 4:

<u>Physical mapping of new DNA probes near the fragile X (FRAXA) with a panel of cell lines.</u>	121
Introduction	123
Methods	125
Results	127
Discussion	134

**Chapter 5:**

<u>Linkage analysis of new RFLPs at Xq27-q28 in the CEPH pedigrees.</u>	138
Introduction	140
Methods	142
Results	145
Discussion	150

**Chapter 6:**

<u>Genetic mapping of new RFLPs in fragile X families defines a strategy for DNA studies in the fragile X syndrome.</u>	156
Introduction	158
Methods	162
Results	164
Discussion	166

**Chapter 7:**

<u>Linkage homogeneity near the fragile X locus in normal and fragile X families.</u>	172
Introduction	174
Methods	176
Results	177
Discussion	181

**Chapter 8:**

<u>Conclusion.</u>	189
--------------------	-----

<b>References:</b>	193
--------------------	-----

## Volume Two

### **Appendix A:**

**Listing of computer programmes.**

DNASIZE	A-3
TEMPLATE	A-19
XPHASE	A-36
PLOT	A-47
MORGAN	A-60
BOOTMAP	A-64

### **Appendix B:**

**Pedigrees and genotypes of the normal (CEPH) pedigrees.**

### **Appendix C:**

**Pedigrees and genotypes of the fragile X pedigrees.**

### **Appendix D:**

**Published papers and manuscripts.**

## LIST OF TABLES

Table	<i>following page</i>
2-A Cell lines used for mapping studies.	63
2-B Loci detected with the VK series of probes.	70
2-C Details of established probes used in the study.	70
3-A Two-point LOD scores from <u>MRX1</u> analysis.	112
3-B Genotypes at loci near <u>MRX1</u> .	112
3-C Carrier risks for women of unknown carrier status.	114
4-A Location of breakpoints at Xq26-q28 in cell lines.	125
5-A Polymorphisms used in the CEPH linkage study.	143
5-B Two-point linkage analysis in the CEPH pedigrees.	147
5-C Phase-known recombinants in the CEPH pedigrees.	148
6-A Polymorphisms used in the fragile X linkage study.	163
6-B Two-point linkage analysis in the fragile X families.	164
6-C Strategy for DNA studies of fragile X families.	169
7-A Comparison of two-point LOD scores in fragile X and CEPH families.	177
7-B Two-point homogeneity tests in fragile X families.	179

## LIST OF FIGURES

Figure	<i>following page</i>
3-1 Pedigree of the MRX1 family.	109
3-2 Locations of polymorphic loci used in the <u>MRX1</u> study.	110
3-3 Multipoint linkage analysis of <u>MRX1</u> .	112
4-1 pc2S15 spans the breakpoint in cell line CY34.	127
4-2 <i>In situ</i> hybridisation studies of pc2S15.	127
4-3 Locations of cell line breakpoints and probes at Xq26-q28.	129
4-4 Physical mapping of VK clones using cell lines.	129
5-1 Mendelian inheritance of the RFLPs detected by VK21.	145
5-2 Mendelian inheritance of the RFLPs detected by VK23.	146
5-3 Mendelian inheritance of the RFLPs detected by pc2S15.	147
5-4 Multipoint linkage analysis in the CEPH pedigrees.	148
6-1 Locations of polymorphic loci used in the fragile X linkage study.	163
6-2 Multipoint linkage analysis of the fragile X families.	164



**Appendix A**

**LISTING OF COMPUTER PROGRAMS.**



*Appendix A. Computer programs. p. A-2*

Six computer programs were written in the course of this project. They were all written in Turbo-Pascal Version 3.02A (Borland International) for an IBM-compatible personal computer. Five programs were utility programs that simply aided the handling of data or the interpretation of output from other programs. Linkage analysis usually involves generating and manipulating a large volume of data, and any automated procedure that could reduce the risk of introducing "clerical" errors in the analysis seemed worth implementing. As a bonus, these programs were also faster than doing these procedures by hand. A second reason for writing these programs was that the output from some of the linkage or database programs did not encourage easy interpretation. The sixth program (BOOTMAP) was written to estimate approximate confidence intervals for gene location using multipoint linkage data from many pedigrees. The application of this program is detailed in Chapter 2.

The source codes of the six programs are listed in this Appendix:

<b>DNASIZE</b>	p. A-3
<b>TEMPLATE</b>	p. A-19
<b>XPHASE</b>	p. A-36
<b>PLOT</b>	p. A-47
<b>MORGAN</b>	p. A-60
<b>BOOTMAP</b>	p. A-64

**program DNASIZE**

This program estimated the size of a DNA fragment of unknown size by comparing its mobility in an agarose gel with that of fragments of known size. The mobility of DNA within a gel does not have a fixed relationship with the size of the fragment (Southern 1979; Elder & Southern 1983). It is usually necessary to determine the relationship for each gel studied. DNASIZE took the mobility of a number of fragments of known size, performed four transformations on that data ( $\ln[\text{fragment size}]$ ,  $1/[\text{fragment size}]$ ,  $1/\ln[\text{fragment size}]$ , and  $\ln[\text{mobility}]$  vs  $\ln[\text{fragment size}]$ ), and calculated the linear correlation coefficient of mobility versus fragment size for each transformation. The transformation which gave the best fit to the known data was then used to estimate the size of unknown fragments. Appropriate warnings were given regarding the number of standards that should be used and ensuring that the standards were of a size similar to that of the fragment being estimated.

```
program DNASIZE (input, output);
```

```
{sourcefile: DNASIZE.pas;  
sourcecode: Turbo-Pascal Version 3.02A;  
author : Graeme Suthers, 17/9/87.}
```

```
{ DNASIZE estimates the size of a DNA fragment by comparing its mobility  
with that of known standards. Details will be found in the procedure  
INFORMATION.
```

This is a simple program and I do not profess to be very  
proficient in Pascal. I hope that it will be easy to modify as errors  
or improved algorithms come to light.

Graeme Suthers 17/09/87}

```
const
```

```
{output format}  
fld=15;  
shortfld=6;  
prec=3;  
lowprec=1;
```

```
{input limits}  
maxfrag=20;  
maxkb=30;  
gellength=200;
```

```
var
```

```
{input var}
numfrag, fragment:integer;
stdkb, stddist : array [1..50] of real;
distunknown:real;
wronginput,useprinter:boolean;

{calc var}
tempsize, tempdist:array [1..50] of real;
intercept, slope, goodfit:real;
equationtype, bestequation: integer;
bestintercept, bestslope, bestgoodfit:real;
goodfit1, goodfit2, goodfit3, goodfit4:real;
kbunknown:real;

{menu var}
choice:integer;
decision:char;
```

label

```
start;
```

```
{PROCEDURES}
```

```
procedure inputerror;
begin
textcolor (red);
writeln;
writeln ('The value you have entered is outside the range expected.');
```

```
writeln ('Please try again. ');
writeln;
textcolor (yellow);
end;
```

```
procedure information;
begin
```

```
clrscr;
textcolor (cyan);
writeln ('Program DNASIZE');
writeln ('[source code: Turbo-Pascal.]');
writeln ('[source file: c:\bin\utils\dnasize.pas.]');
writeln;
writeln ('          INFORMATION');
writeln;
writeln ('This program estimates the size (in kb) of DNA fragments from');
writeln ('their mobility in a gel. The relationship between DNA size and ');
writeln ('mobility is not straightforward. On empirical grounds the best');
writeln ('description of the relationship is usually');
writeln;
writeln ('          mobility = slope * log (kb) + intercept ');
writeln;
writeln ('(where slope and intercept are constants.)');
writeln ('However this does not hold true for large DNA fragments i.e. ');
writeln ('> 15kb, and may vary according to the particular running ');
writeln ('conditions. Therefore any mathematical summary of the size');
writeln ('mobility relationship must be determined separately for each');
writeln ('gel.');
```

```

writeln;
writeln ('[press any key to continue]');
repeat until Keypressed;

clrscr;
writeln ;
writeln ('This program takes standard DNA fragments of known size and ');
writeln ('measured mobility and works out which equation best describes ');
writeln ('the relationship. The equations tested are of the form:');
writeln ;
writeln ('      mobility = slope / (kb) + intercept');
writeln ('      mobility = slope * log (kb) + intercept');
writeln ('      mobility = slope / log (kb) + intercept');
writeln ('      log (mobility)= slope * log (kb) + intercept');
writeln ;
writeln ('Linear equations have been chosen because the equations for ');
writeln ('regression analysis are very straightforward. Other forms of');
writeln ('linear equations or parabolic equations could also be tried. ');
writeln ('The program calculates the degree of scatter associated with');
writeln ('each formula and choses the one with the least scatter. The ');
writeln ('size of "unknown" DNA fragments is then estimated from that');
writeln ('formula. ');
writeln ;
writeln ('[press any key to continue]');
repeat until Keypressed;

clrscr;
writeln;
writeln ('Some tips:');
writeln ;

```

```

writeln ('a) calculate a new formula for each gel. ');
writeln ;
writeln ('b) you may get misleading estimates of DNA size if you use ');
writeln (' the one formula to estimate very small and large fragment ');
writeln (' sizes. If you have both SPP1 and lambda markers, use the ');
writeln (' SPP1 markers to estimate the size of small fragments. Then ');
writeln (' use the program to generate a new equation using the lambda ');
writeln (' markers to estimate the size of larger fragments. ');
writeln ;
writeln ('c) there are some arbitrary limits on what data you can enter. ');
writeln (' e.g. no more than 20 standard fragments of known size, gels ');
writeln (' cannot be more than 200mm, standard fragments must be less than ');
writeln (' 30 kb. These limits can be modified very easily. ');
writeln ;
writeln ('d) do check your data entry. Garbage in, garbage out. ');
writeln ;
writeln ('e) during data entry Genie will remember the values you entered ');
writeln (' before and show them in [square brackets]. If these values are ');
writeln (' correct just press ENTER and move on. ');
writeln ;
writeln ('[press any key to return to menu] ');
repeat until Keypressed;
clrscr;
textcolor (yellow);

end;

procedure dataentry;
begin

```

```

clrscr;
repeat
  decision:= 'n';
  write ('How many fragments of known size do you want to enter ? ');
  read (numfrag);
  writeln;
  wronginput:=(numfrag<2) or (numfrag>maxfrag);
  if wronginput
    then inputerror
    else begin
      for fragment:=1 to numfrag do
        begin
          writeln;
          write ('Type in the size of known fragment (in kb): ');
          if stdkb[fragment]>0
            then write ('['',stdkb[fragment]:shortf1d:prec,'] ');
          read (stdkb[fragment]);
          writeln;
          write ('Type in the mobility of known fragment (in mm): ');
          if stddist[fragment]>0
            then write ('['',stddist[fragment]:shortf1d:lowprec,'] ');
          read (stddist[fragment]);
          writeln;
        end;
      writeln;
      writeln;
      textcolor (lightgray);
      writeln ('  Fragment size      Fragment mobility');
      writeln ('      (kb)                (mm)');
      writeln;
    end;
end repeat;

```



```

for fragment:=1 to numfrag do
begin
write (stdkb[fragment]:fld:prec, stddist[fragment]:fld:lowprec);
wronginput:=not(((stdkb[fragment]) < maxkb) and (stddist[fragment] < gellength));
if wronginput
then write (' .... INCORRECT VALUE/S');
writeln;
end;
textcolor (yellow);
writeln;
write ('Is this listing correct ? (y/n) ');
read (decision);
clrscr;
end;

until decision='y';
if useprinter
then begin
writeln(1st, ' STANDARD FRAGMENTS AND MOBILITY:');
writeln(1st);
writeln(1st, ' Fragment size      Fragment mobility');
writeln(1st, '      (kb)                (mm)');
writeln(1st);
for fragment:=1 to numfrag do
begin
write (1st, stdkb[fragment]:fld:prec, stddist[fragment]:fld:lowprec);
wronginput:=not(((stdkb[fragment]) < maxkb) and (stddist[fragment] < gellength));
if wronginput
then write(1st, ' .... INCORRECT VALUE/S');
writeln(1st);

```

```

                end;
            writeln(1st);
        end;
end;

procedure regression;
var
    meansize, meandist:real;
    sumsize, sumsizeqr, sumdist:real;

begin
    meansize:=0;
    meandist:=0;
    sumsize:=0;
    sumsizeqr:=0;
    sumdist:=0;

    for fragment:=1 to numfrag do
        begin
            meansize:=meansize + tempsize[fragment];
            meandist:=meandist + tempdist[fragment];
        end;
    meansize:=meansize/numfrag;
    meandist:=meandist/numfrag;

    for fragment:=1 to numfrag do
        begin
            sumsize:=sumsize + ((tempsize[fragment]-meansize)*(tempdist[fragment]-meandist));
            sumsizeqr:=sumsizeqr + sqr(tempsize[fragment]-meansize);
        end;
    end;
end;

```

```

        sumdistsqr:=sumdistsqr + sqr(tempdist[fragment]-meandist);
    end;

if (sumsizedist=0) or (sumsizesqr=0) or (sumdistsqr=0)
then begin
    writeln;
    writeln ('There is an error in your data such that the sum of the squared');
    writeln ('residuals is 0. ');
    writeln ('Please check your figures and try again. ');
    writeln;
    halt;
end;

slope:=sumsizedist/sumsizesqr;
intercept:=meandist-(slope*meansize);
goodfit:=sqr(sumsumsizedist)/(sumsumsizesqr*sumsumdistsqr);

if goodfit>bestgoodfit
then begin
    bestgoodfit:=goodfit;
    bestslope:=slope;
    bestintercept:=intercept;
    bestequation:=equationtype;
end;

end;

{ BEGIN PROGRAM DNASIZE }

```

```

begin
  clrscr;

  {initialise arrays}
  for fragment:=1 to maxfrag do
    begin
      stdkb[fragment]:=0;
      stddist[fragment]:=0;
    end;

  start:
  choice:=0;
  decision:='n';
  useprinter:=false;
  writeln;
  writeln ('Program DNASIZE. ');
  writeln;
  writeln ('Estimates the size of a DNA fragment from its mobility. ');
  writeln ('Required input is a) length & mobility of known standards. ');
  writeln ('                and b) mobility of unknown DNA fragment. ');
  writeln;
  writeln ('Choose one of the following options: ');
  writeln ('      1      More information about the program. ');
  writeln ('      2      Data entry. ');
  writeln;
  writeln ('      0      Exit program. ');
  writeln;
  read (choice);
  case (choice) of

```

```

1:  begin
      information;
      goto start;
end;
2:  begin
      writeln;
      write('Do you want to have the results printed ? (y/n) : ');
      read (decision);
      writeln;
      if decision='y'
      then begin
            useprinter:=true;
            writeln('Please turn the printer on. ');
            writeln('[Press any key when ready]');
            repeat until Keypressed;
            end;
            dataentry;
            end;
0:  halt;
else
      begin
            inputerror;
            goto start;
            end;
end;

bestgoodfit:=0;

{analyse EQUATION 1:  linear(dist):inv.linear(kb)}

```

```
equationtype:=1;
for fragment:=1 to numfrag do
  begin
    tempsize[fragment]:=1/(stdkb[fragment]);
    tempdist[fragment]:=stddist[fragment];
  end;
```

```
regression;
goodfit1:=goodfit;
```

```
{analyse EQUATION 2:  linear(dist)/ln(kb)}
```

```
equationtype:=2;
for fragment:=1 to numfrag do
  begin
    tempsize[fragment]:=ln(stdkb[fragment]);
    tempdist[fragment]:=stddist[fragment];
  end;
```

```
regression;
goodfit2:=goodfit;
```

```
{analyse EQUATION 3:  linear(dist)/inverseln(kb)}
```

```
equationtype:=3;
for fragment:=1 to numfrag do
  begin
    if stdkb[fragment] = 1
```

```

        then stdkb[fragment]:=1.001;
        tempsize[fragment]:=1/ln(stdkb[fragment]);
        tempdist[fragment]:=stddist[fragment];
    end;

regression;
goodfit3:=goodfit;

{analyse EQUATION 4:  ln(dist)/ln(kb)}

equationtype:=4;
for fragment:=1 to numfrag do
    begin
        tempsize[fragment]:=ln(stdkb[fragment]);
        tempdist[fragment]:=ln(stddist[fragment]);
    end;

regression;
goodfit4:=goodfit;

{decide transformation of best fit; calculate unknown kb}

writeln;
writeln;
writeln ('The "goodness-of-fit" statistic (=r**2) for each transformation ');
writeln ('is shown below:');
writeln;
writeln ('1.',goodfit1:fld:prec,'          for linear(dist):inverse.linear(kb)');

```

```

writeln ('2.',goodfit2:fld:prec,'          for linear(dist):log(kb)');
writeln ('3.',goodfit3:fld:prec,'          for linear(dist):inverse.log(kb)');
writeln ('4.',goodfit4:fld:prec,'          for log(dist):log(kb)');
writeln;
writeln ('The following equation provides the highest goodness-of-fit value:');

case bestequation of
  1:writeln ('distance = ',bestslope:shortfld:prec,' /(kb) + ',bestintercept:shortfld:prec);
  2:writeln ('distance = ',bestslope:shortfld:prec,' log(kb) + ',bestintercept:shortfld:prec);
  3:writeln ('distance = ',bestslope:shortfld:prec,' /log(kb) + ',bestintercept:shortfld:prec);
  4:writeln ('log(distance) = ',bestslope:shortfld:prec,' log(kb) + ',bestintercept:shortfld:prec);
end;

writeln;
writeln;

if useprinter
then begin
  writeln(1st);
  writeln(1st,'The following equation provides the highest goodness-of-fit value:');
  case bestequation of
    1:writeln (1st,'distance = ',bestslope:shortfld:prec,' /(kb) + ',bestintercept:shortfld:prec);
    2:writeln (1st,'distance = ',bestslope:shortfld:prec,' log(kb) + ',bestintercept:shortfld:prec);
    3:writeln (1st,'distance = ',bestslope:shortfld:prec,' /log(kb) + ',bestintercept:shortfld:prec);
    4:writeln (1st,'log(distance) = ',bestslope:shortfld:prec,' log(kb) + ',bestintercept:shortfld:prec);
  end;
  writeln(1st);
  writeln(1st);
  writeln(1st,'SIZE OF UNKNOWN FRAGMENTS :');
  writeln(1st);

```



```

        writeln(1st,'Distance travelled (mm)      Size (kb)');
        writeln(1st);
        end;

decision:='n';
repeat
    writeln;
    write ('Type in mobility of unknown DNA fragment (in mm): ');
    read (distunknown);
    writeln;
    if (distunknown <=0) or (distunknown >gellength)
        then inputerror
        else begin
            case bestequation of
                1:kbunknown:=bestslope/(distunknown-bestintercept);
                2:kbunknown:=exp((distunknown-bestintercept)/bestslope);
                3:kbunknown:=exp(bestslope/(distunknown-bestintercept));
                4:kbunknown:=exp((ln(distunknown)-bestintercept)/bestslope);
            end;
            writeln ('The size of this fragment is          ',kbunknown:shortfld:prec,' kb. ');
            writeln;
            if useprinter then writeln(1st,distunknown:fld:lowprec,kbunknown:20:prec);
            write ('Calculate another ? (y/n) ');
            read (decision);
            writeln;
            end;
        until decision ='n';
    clrscr;
    goto start;
end {DNASIZE}.

```

**program TEMPLATE**

The program TEMPLATE was written to reduce the risk of introducing 'clerical' errors when genotyping hundreds of individuals from the CEPH pedigrees. The program used the pedigree numbers, individual identification numbers, and DNA sample numbers in one of the main CEPH data files to create a standard listing for each CEPH pedigree. The information was read directly from the CEPH database, thereby reducing the risk of errors. The standard listing for each pedigree was used when making a Southern blot for any CEPH pedigree. The listing was stored in the computer so that data could be entered in the same order that it was read from the autoradiograph. This approach greatly reduced the chance of errors during data entry, particularly as a number of people were involved in entering new genotypes into the database for both chromosome 16 and the X chromosome.

```
program TEMPLATE (input,infile,outfile);
```

```
{sourcecode: Turbo-Pascal Version 3.02A  
sourcefile:  TEMPLATE.PAS  
Graeme Suthers, 11/04/88. Updated:14/04/88
```

TEMPLATE is a utility program for the CEPH data base and is designed to minimise clerical errors when making filters or entering data. It allows the user to create input template files for use with the CEPH program INPUT. The necessary pedigree data is read from a binary file and written in ASCII code to specified files. At the same time files are created which can act as filter templates when making the corresponding filters.

The necessary infiles are shown in the "constant" declaration. The output files are .tmp or .ftr files for use as input templates or filter templates respectively.}

```
const
```

```
maxrelatives=25;  
linelength=80;
```

```
pedinfile='ped.dat';           {datafile provided by CEPH}  
filterinfile='cfilter.tmp';   {ASCII file with the top few lines of the template file you want to use}  
templateoutfile='template.inp'; {used to create a prompting file for data entry }
```

```
margin='      ';
```

```
type
```

```
pedrecord=record  
    pednumber: integer;  
    tabID, ID, fa, mo, sex, phenoref: array[1..maxrelatives] of integer;  
end;
```

```
filename=string[12];  
  
var  
  peddata:pedrecord;  
  pedfile:FILE of pedrecord;  
  
  words:string[lineLength];  
  filtertemplate:TEXT;  
  templatefile:TEXT;  
  filterfile:TEXT;  
  templateinfile:TEXT;  
  
  outfilename,checkfile:filename;  
  
  inputcorrect:boolean;  
  
  decision:char;  
  
  pedrequest:integer;  
  lanecount,person:integer;  
  
{I PRINTFIL.pas}  
{I FILEXIST.pas}  
  
PROCEDURE pednumberinput;  
  
begin  
  repeat  
    inputcorrect:=false;  
    reset(pedfile);
```

```

write('Type in pedigree number (0 to quit) : ');
readln(pedrequest);
if pedrequest=0 then exit;

    repeat
        read(pedfile,peddata);
    until (peddata.pednumber=pedrequest) or (EoF(pedfile));
    if EoF(pedfile)
        then begin
            textcolor(red);
            writeln('This pedigree number not found in CEPH pedigree file. ');
            textcolor(blue);
            end
        else inputcorrect:=true;

    until inputcorrect;

end; {PROCEDURE pednumberinput}

PROCEDURE familytemplate;

label
    abortprocedure;

begin
    clrscr;
    decision:='y';
    assign(pedfile,pedinfile);
    repeat
        pednumberinput;

```

```

if pedrequest=0 then goto abortprocedure;

str(pedrequest,outfilename);
if fileexist ('U'+outfilename+'.tmp')
  then begin
    decision:='n';
    writeln;
    textcolor(red);
    writeln('Family',outfilename,' is one of the UTAH kindreds. Special files have been');
    writeln('made for these families which contain both CEPH and UTAH I.D. numbers. ');
    writeln;
    writeln('The fields for this family are called U',outfilename,'.tmp and U',outfilename,'.ftr. ');
    writeln;
    write('Do you still want to make a ',outfilename,' set of files ? : y/n ');
    textcolor(blue);
    readln(decision);
    if decision='n' then exit;
    end;

if fileexist(outfilename+'.tmp')
  then begin
    textcolor(red);
    writeln('Template and filter files for Family ',outfilename,' already exist. ');
    writeln('You cannot rewrite these files. ');
    textcolor(blue);
    end

  else begin
    assign(templatefile,outfilename+'.tmp');
    rewrite(templatefile);
  
```

```

assign(filtertemplate,filterinfile);
reset(filtertemplate);
assign(filterfile,outfilename+'.ftr');
rewrite(filterfile);
writeln(filterfile,margin,'THIS FILTER TEMPLATE IS CALLED : ',outfilename+'.ftr');
writeln(filterfile);

repeat
  readln(filtertemplate,words);
  writeln(filterfile,margin,words);
until Eof(filtertemplate);
close(filtertemplate);

person:=3;
while (peddata.labID[person]>0) do
  begin

writeln(templatefile,pedrequest:10,peddata.labID[person]:10) ;
  writeln(filterfile);
  write(filterfile,pedrequest:14,'-');
  if peddata.ID[person]<10 then write(filterfile,'0');

writeln(filterfile,peddata.ID[person],peddata.labID[person]: 13);
  person:=person+1;
  end;

close(templatefile);
close(filterfile);

```

```

writeln;
writeln('Template and filter files for family ',pedrequest,' have been created. ');
writeln;
write('Do you want to create more family template/filter files ? y/n : ');
read(decision);
writeln;
end; {else}

until decision='n';
abortprocedure:
close(pedfile);

end; {PROCEDURE familytemplate}

PROCEDURE parenttemplate;

label
    SkipUtahParents,endprocedure;

begin
clrscr;
decision:='y';
assign(pedfile,pedinfile);
repeat
    write('Type in name of PARENT files to be created (max. of 8 characters) : ');
    readln(outfilename);
    if (Filexist(outfilename+'.tmp')) or (Filexist(outfilename+'.ftr'))
        then begin
            writeln;

```



```

        textcolor(red);
        writeln('A file with this name already exists. ');
        writeln('You cannot rewrite this file. ');
        textcolor(blue);
        writeln;
        end

    else begin
        assign(templatefile,outfilename+'.tmp');
        rewrite(templatefile);

assign(filtertemplate,filterinfile);
reset(filtertemplate);
assign(filterfile,outfilename+'.ftr');
rewrite(filterfile);
writeln(filterfile,margin,'THIS FILTER TEMPLATE IS CALLED :',outfilename+'.ftr');
writeln(filterfile);

repeat
    readln(filtertemplate,words);
    writeln(filterfile,margin,words);
until Eof(filtertemplate);
close(filtertemplate);

lanecount:=0;

repeat
    SkipUtahParents:
    pednumberinput;
    if pedrequest=0 then goto endprocedure;

```

```

str(pedrequest,checkfile);
if fileexist('U'+checkfile+'.tmp')
then begin
  textcolor(red);
  writeln('Family',checkfile,' is one of the UTAH kindreds. There are special');
  writeln('template and filter files for these parents which contain both CEPH');
  writeln('and UTAH I.D. numbers. These fields are called UParentA and UParentB.');
```

```

  writeln;
  decision:='n';
  writeln('Do you still want to include the parents of family',checkfile);
  write('in this file ? y/n : ');
  textcolor(blue);
  readln(decision);
  if decision='n' then goto SkipUtahParents;
  end;

for person:=1 to 2 do
  begin
    writeln(templatefile,pedrequest:10,peddata.labID[person]:10) ;
    writeln(filterfile);
    write(filterfile,pedrequest:14,'-');
    if peddata.ID[person]<10 then write(filterfile,'0');
    writeln(filterfile,peddata.ID[person],peddata.labID[person]: 13);
  end;

lanecount:=lanecount+2;
writeln;
writeln('There are now ',lanecount:3,' people in this template file.');
```

```

writeln('Do you want to add more parents to this file ? (y/n) : ');

```

```

        readln(decision);

until decision='n';
endprocedure;

close(templatefile);
close(filterfile);

writeln('Template and filter fields for PARENTS ',outfilename,' have been created. ');
write('Do you want to create more parent template/filter files ? y/n :');
readln(decision);
end; {else}

until decision='n';
close(pedfile);

end; {PROCEDURE parenttemplate}

PROCEDURE mergetemplate;

begin
  clrscr;
  decision:='y';
  assign(TEMPLATEFILE,templateoutfile);
  rewrite(TEMPLATEFILE);
  writeln('This procedure allows you to merge any number of template files');
  writeln('for data entry using the CEPH program INPUT. The merged fields are');
  writeln('placed in a file called TEMPLATE.INP which is then used by INPUT. ');
  writeln;
  writeln('You will be prompted to enter each pedigree or parent file that you');

```

```

writeln('want to include in the merged file. You need only enter the pedigree');
writeln('number or parent file name (e.g 1234 or CParentA ); do not enter the');
writeln('filename extension .tmp. ');
writeln;
repeat
  write('Type in the pedigree number or parent file name : ');
  readln(outfilename);
  if fileexist(outfilename+'.tmp')
    then begin
      assign(TEMPLATEINFILE,outfilename+'.tmp');
      reset(TEMPLATEINFILE);
      repeat
        readln(TEMPLATEINFILE,words);
        writeln(TEMPLATEFILE,words);
      until Eof(TEMPLATEINFILE);
      close(TEMPLATEINFILE);
      end
    else begin
      textcolor(red);
      writeln('File ',outfilename,'.tmp not found. ');
      textcolor(blue);
      end;

  writeln;
  write('Merge another file ? y/n : ');
  readln(decision);

until decision='n';
close(TEMPLATEFILE);

```

```

end;

begin {TEMPLATE}
  textbackground(lightgray);
  textcolor(blue);
  repeat
    clrscr;
    gotoXY(1,5);
    writeln('          Program TEMPLATE. ');
    writeln;
    writeln(' This program creates files which act as template files for CEPH data');
    writeln(' entry (file extension= .tmp) and files which can be used');
    writeln(' in making filters (file extension= .ftr). ');
    writeln;
    writeln(' Do you want to create template files for ');
    writeln('      P   Parents only');
    writeln('      F   Families excluding parents');
    writeln;
    writeln('      O   print OUT a file on printer');
    writeln('      M   Merge template files for data entry');
    writeln;
    writeln('      e   exit from program');
    readln(decision);

    case decision of
      'P':parenttemplate;
      'p':parenttemplate;
      'F':familytemplate;
      'f':familytemplate;

```

```

'o':begin
  clrscr;
  decision:='n';
  writeln('You may print either a template or a filter file. ');
  writeln;
  repeat
    writeln('Type in the full name of the file (e.g. 1234.tmp or 1234.ftr) : ');
    read(outfilename);
    printfile(outfilename);
    writeln;
    write('Do you want to print another ? y/n : ');
    readln(decision);
  until decision='n';
end;
'o':begin
  clrscr;
  decision:='n';
  writeln('You may print either a template or a filter file. ');
  writeln;
  repeat
    write('Type in the full name of the file (e.g. 1234.tmp or 1234.ftr) : ');
    printfile(outfilename);
    writeln;
    write('Do you want to print another ? y/n : ');
    readln(decision);
  until decision='n';
end;
'M':mergetemplate;
'm':mergetemplate;

```

```
        end;  
        until decision='e';  
        textcolor(yellow);  
        textbackground(black);  
        clrscr;  
end. {TEMPLATE}
```

{sourcecode:Turbo-Pascal. V3.02A  
sourcefile:FILEEXIST.pas.  
Graeme Suthers. 13/04/88.

This an INCLUDE file to be included in the compilation of another .PAS file.  
The program defines the function FILEEXIST which returns the Boolean expression  
TRUE if the file specified in the function call exists in the current directory.  
The parent program must contain the TYPE declaration:

```
                filename:string[12]
    }
FUNCTION filexist(queryfile:filename) : Boolean;
var
    CHECKFILE:file;
begin
    assign(CHECKFILE,queryfile);
    {$I-}
    reset(CHECKFILE);
    close(CHECKFILE);
    {$I+}
    filexist:=(IOresult=0);
end;
```



{sourcecode:Turbo-Pascal.V3.02A  
sourcefile:PRINTFIL.PAS  
Graeme Suthers. 13/04/88

PRINTFIL prints ASCII files to the default printer.

This program is designed as an INCLUDE file to be included with another .PAS file during compiling. This program defines a procedure PRINTFILE with the name of the file to be printed included in the procedure call. The filename must be specified in full and not exceed 12 characters. The parent program must contain the TYPE declaration:

```
Filename:string[12].
```

If the file to be printed does not exist an error message is printed.)

```
PROCEDURE printfile (printfilename:filename);
```

```
var
```

```
  TEXTFILE:text;  
  LineOfWords:string[80];
```

```
begin
```

```
  assign(TEXTFILE,printfilename);  
  {$I-}  
  reset(TEXTFILE);  
  close(TEXTFILE);  
  {$I+}  
  if (IOresult=0)  
    then begin
```

```
writeln;
writeln('Turn printer on and set page position, please. ');
writeln;
writeln('Press any key when ready... ');
repeat until KeyPressed;
reset(TEXTFILE);
repeat
    readln(TEXTFILE, LineOfWords);
    writeln(lst, LineOfWords);
until EOF(TEXTFILE);
close(TEXTFILE);
end

else writeln(printfilename, ' does not exist in this directory. ');

end; {PROCEDURE printfile}
```

**program XPHASE**

The software provided by CEPH included a program to check for inconsistent genotypes at autosomal loci. There was no provision for checking for inconsistent genotypes at X-linked loci. CEPH also distributed a pedigree plotting program that drew a three-generation pedigree with specified genotypes listed below each individual. These genotypes were not listed in phase, making it difficult to detect recombination events without redrawing the pedigree.

The program XPHASE was written to fulfill two functions. Firstly, it checked for inconsistent genotypes at X-linked loci. The genotype at each locus was compared with those of the parents, with appropriate allowance made for sex. Second, the program inferred the phase of X-linked loci from the genotypes of the father and of the maternal grandfather. The output from XPHASE was used by the CEPH plotting program to draw pedigrees with the genotypes shown in phase (as in Appendix B). XPHASE listed in a separate file (CHARTS.log) any genotype inconsistencies and any loci at which the phase could not be inferred.

```

program XPHASE (infile,outfile,output);

{sourcefile: XPHASE.pas
 sourcecode: Turbo-Pascal Version 3.02A
 author    : Graeme Suthers, 10/11/88 }

CONST
  MaxLineLength=132;
  MaxLocusNumber=10;
  NumberStr=' No. : ';
  Dashes='-----';
  Margin='          ';

  Debug=false;

TYPE
  FileName=string[12];
  AlleleArray=array[1..MaxLocusNumber] of integer;

VAR
  Infile,Outfile,Oldfile,Logfile:text;

  PedNumber,PedCount,i,locus,LocusTotal,allele1,allele2: integer;
  PedID,labID,ID,Fa,Mb,PGF,PGM,MGF,MGM,p1,p2,p3,Sex,Proband: integer;

  FirstChar:string[1];
  FirstWord:string[8];
  WholeOfLine:string[132];

  FaAllele,MGFAllele,MaAllele1,MaAllele2:alleleArray;

```

```

EndofPedigrees:boolean;

{$I Fileexist.pas}

PROCEDURE Genotypewarning;
begin
  writeln(logfile,margin,'Inconsistent genotype for person ',labID:5,' at locus ',locus:2);
  writeln(MARGIN,'INCONSISTENT GENOTYPE for person ',labID:5);
end;

PROCEDURE CheckGenotype;
begin
  if ID=MGF
  then begin
    if allele1<>allele2 then Genotypewarning
    end
  else begin
    case ID of
      1:if allele1<>allele2 then Genotypewarning;
      2:if MGFallele[locus]>0
        then begin
          if ((allele1=allele2)and(allele1>0)and(MGFallele[locus]<>allele1)) then Genotypewarning;
          if ((allele1<>allele2)and(allele1<MGFallele[locus])and(allele2<MGFallele[locus])) then Genotypewarning;
        end;
    else begin
      if Sex=1
      then begin
        if allele1<>allele2 then Genotypewarning;
        if ((allele1>0)and(allele1<MaAllele1[locus])and(allele1<MaAllele2[locus])) then Genotypewarning;
      end;
    end;
  end;
end;

```

```

        end
    else
        begin
            if ((allele1=allele2)and(allele1>0))
                then begin
                    if ((allele1<>FaAllele[locus])and(FaAllele[locus]>0)) then GenotypeWarning;
                    if ((allele2<>MaAllele1[locus])and(allele2<>MaAllele2[locus])) then GenotypeWarning;
                    end;
                    if ((allele1<>allele2)and(FaAllele[locus]>0))
                        then begin
                            if ((allele1<>FaAllele[locus])and(allele2<>FaAllele[locus]))
                                then GenotypeWarning
                            else begin
                                if ((allele1=FaAllele[locus])and(allele2<>MaAllele1[locus])and(allele2<>MaAllele2[locus]))
                                    then GenotypeWarning;
                                if ((allele2=FaAllele[locus])and(allele1<>MaAllele1[locus])and(allele1<>MaAllele2[locus]))
                                    then GenotypeWarning;
                                end;
                            end;
                        end;
                    end;
                end;
            end; {CASE else}
        end; {CASE}
    end; {else}
end;

PROCEDURE SearchforPed;
begin
    if Debug then writeLn('SearchforPed');
    reset(infile);
    Pedcount:=0;

```

```

EndofPedigrees:=false;
for i:=1 to 15 do readln(infile);
repeat
  readln(infile,FirstWord,WholeofLine);
  if FirstWord=' >>>>>>' then EndofPedigrees:=true
  else if FirstWord='Pedigree' then Pedcount:=Pedcount+1;
  if EOF(infile) then EndofPedigrees:=true;
until ((EndofPedigrees) or (PedCount=Pednumber));
end;

PROCEDURE GetData;
begin
  readln(infile);
  read(infile,FirstChar,PedID,labID, ID,PGF,PGM,p1,p2,p3,Sex,Proband);
  writeln('Analysing Pedigree No. : ',PedID:8);
  writeln(logfile,'Pedigree No. : ',PedID:8);
  if Debug then writeln(PedID,labID, ID,PGF,PGM,p1,p2,p3,Sex,Proband);
  for locus:=1 to LocusTotal do read(infile,FirstChar,FirstChar,FaAllele[locus],allele2);
  readln(infile);
  read(infile,FirstChar,PedID,labID, ID,MGF,MGM,p1,p2,p3,Sex,Proband);
  if Debug then writeln(PedID,labID, ID,MGF,MGM);
  for locus:=1 to LocusTotal do read(infile,FirstChar,FirstChar,MaAllele1[locus],MaAllele2[locus]);
  if MGF>0
  then begin
    repeat
      readln(infile);
      read(infile,FirstChar,PedID,labID, ID);
    until ID=MGF;
    read(infile,Fa,Mo,p1,p2,p3,Sex,Proband);
    for locus:=1 to LocusTotal do read(infile,FirstChar,FirstChar,MGFallele[locus],allele2);
  end;
end;

```

```

        end
        else for locus:=1 to LocusTotal do MGFallele[locus]:=0;
for locus:=1 to LocusTotal do
begin
    if (MGFallele[locus]=0) and (MaAllele1[locus]<>MaAllele2[locus])
    then writeln(logfile,margin,'Maternal phase not known at locus ',locus:2);
    if (FaAllele[locus]=0) and (MaAllele1[locus]<>MaAllele2[locus])
    then writeln(logfile,margin,'Daughters phases not known at locus ',locus:2);
    end;
writeln(logfile);

end;

PROCEDURE WriteData;
begin
    writeln(outfile,FirstWord,NumberStr,PedID:5);
    writeln(outfile);
    readln(infile);
    read(infile,FirstChar);
    while (FirstChar=' ') or (FirstChar='+') do
        begin
            read(infile,PedID,labID,ID,Fa,Mo,p1,p2,p3,Sex,Proband);
            write(outfile,PedID:6,labID:5,ID:3,Fa:3,Mo:3,p1:3,p2:3,p3:3,Sex:3,Proband:3);
            if (ID=PGF) or (ID=PGM) or (ID=MGM)
            then for locus:=1 to LocusTotal do write(outfile,'|':2,0:3,0:3)
            else
                for locus:=1 to LocusTotal do
                    begin
                        write(outfile,'|':2);
                        read(infile,FirstChar,FirstChar,allele1,allele2);
                    end;
                end;
            end;
        end;
    end;
end;

```



```

CheckGenotype;
if ID=2 then
  if MGFallele[locus]=0 then write(outfile,allele1:3,allele2:3)
  else
    if allele1=MGFallele[locus] then write(outfile,MGFallele[locus]:3,allele2:3)
    else write(outfile,MGFallele[locus]:3,allele1:3)
  else
    if MaAllele1[locus]=MaAllele2[locus] then write(outfile,0:3,0:3)
    else
      if Sex=1 then write(outfile,0:3,allele1:3)
      else
        if FaAllele[locus]=0 then write(outfile,allele1:3,allele2:3)
        else
          if allele1=FaAllele[locus] then write(outfile,FaAllele[locus]:3,allele2:3)
          else write(outfile,FaAllele[locus]:3,allele1:3);
        end;
      end;
    writeln(outfile);
    readln(infile);
    read(infile,FirstChar);
  end; {while loop}

  writeln(outfile,dashes);

end; {procedure WriteData}

{PROGRAM xphase}
begin {program}
  clrscr;
  writeln('program XPHASE :');
  writeln;

```

```

writeln('XPHASE modifies the file PED.OUT so that genotypes are in phase. ');
writeln('The modified PED.OUT file can be used with the CEPH program SEEALL ');
writeln('to create pedigrees with genotypes in correct phase. ');
writeln;
writeln('XPHASE is designed for X-linked loci that have been genotyped. It will ');
writeln('NOT work with autosomal or phenotyped data. ');
writeln;
writeln('If XPHASE cannot determine the phase of a locus a warning message is ');
writeln('printed in the file CHARTS.LOG. ');
writeln;
writeln('For XPHASE to work the file PED.OUT must be present in this directory. ');
writeln('(the unmodified version of PED.OUT is stored in OLDPED.OUT. ');
writeln;
write('Press any key when ready ( ^C to abort ) : ..... ');
repeat until KeyPressed;
writeln;
if not Fileexist('PED.OUT')
  then begin
    writeln('PED.OUT is not in this directory. ');
    writeln('Program aborted. ');
    halt;
  end;
if Fileexist('OLDPED.OUT')
  then begin
    assign(oldfile, 'OLDPED.OUT');
    close(oldfile);
    erase(oldfile);
  end;
assign(infile, 'PED.OUT');
rename(infile, 'OLDPED.OUT');

```

```

reset(infile);
assign(outfile,'PED.OUT');
rewrite(outfile);
assign(logfile,'CHARTS.LOG');
rewrite(logfile);

for i:=1 to 10 do
  begin
    readln(infile,wholeOfLine);
    writeln(outfile,wholeOfLine);
    writeln(logfile,wholeOfLine);
  end;

  readln(infile,LocusTotal,wholeOfLine);
  writeln(outfile,'      ',LocusTotal,wholeOfLine);
  writeln(logfile,'      ',LocusTotal,wholeOfLine);
  readln(infile,Firstword,wholeOfLine);
  while not (Firstword='Pedigree') do
    begin
      writeln(outfile,Firstword,wholeOfLine);
      writeln(logfile,Firstword,wholeOfLine);
      readln(infile,Firstword,wholeOfLine);
    end;
  writeln(logfile,dashes);
  writeln(logfile);
  writeln(logfile,margin,'PHASE LOG FOR PED.OUT AND CHARTS');
  writeln(logfile);

  Pednumber:=1;

```

```
SearchForPed;
while not EndofPedigrees do
  begin
    GetData;
    SearchforPed;
    Writedata;
    Pednumber:=pednumber+1;
    SearchforPed;
  end;

  close(infile);
  writeln(outfile,FirstWord,WholeofLine);
  close(outfile);
  writeln(logfile);
  writeln(logfile,dashes);
  close(logfile);
  writeln('XPHASE completed. ');
  writeln('Unknown phases listed in CHARTS.LOG');
end. {program XPHASE}
```

{sourcecode:Turbo-Pascal.  
sourcefile:FILEXIST.pas.  
Graeme Suthers. 13/04/88.

This an INCLUDE file to be included in the compilation of another .PAS file.  
The program defines the function FILEXIST which returns the Boolean expression  
TRUE if the file specified in the function call exists in the current directory.  
The parent program must contain the TYPE declaration:

```
                                filename:string[12]
)

FUNCTION filexist(queryfile:filename) : Boolean;

var
    CHECKFILE:file;

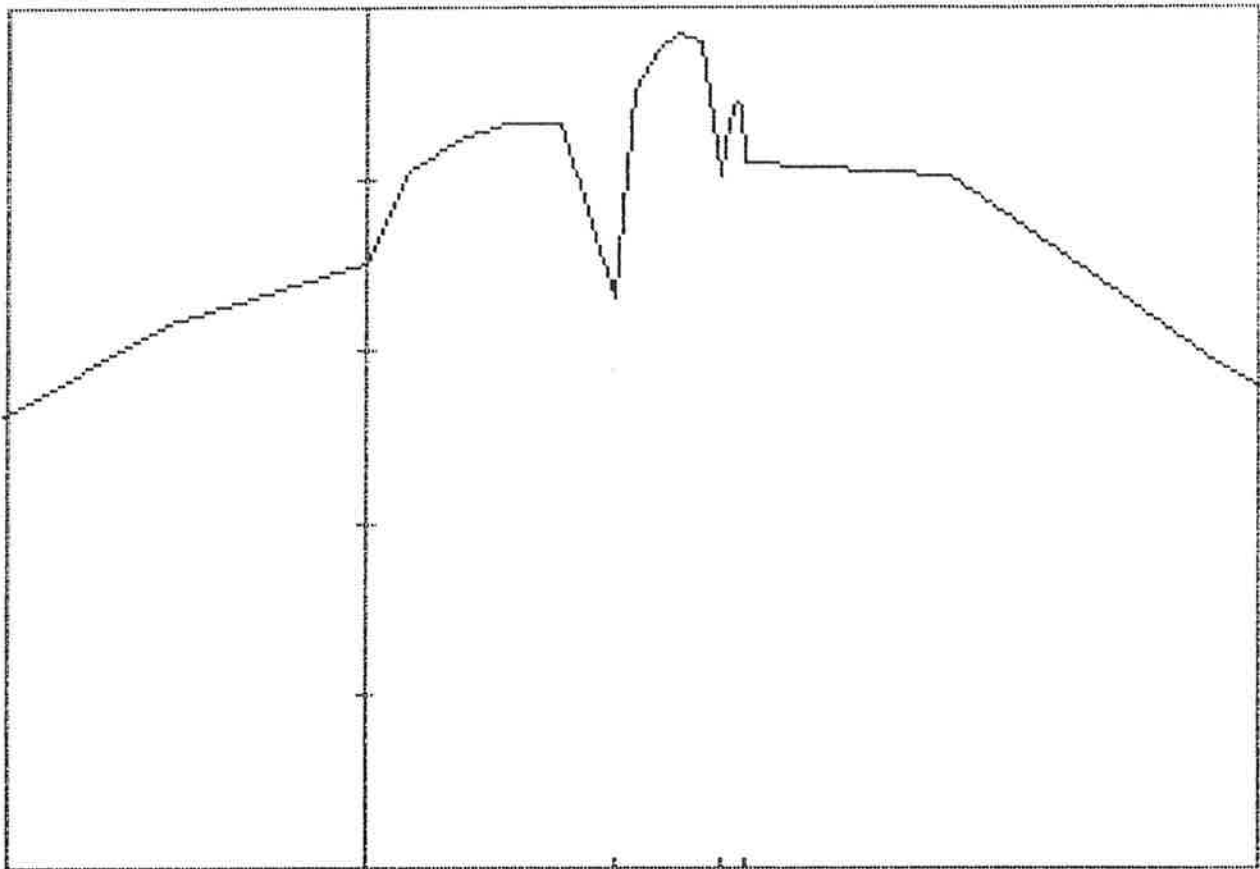
begin
    assign(CHECKFILE,queryfile);
    {$I-}
    reset(CHECKFILE);
    close(CHECKFILE);
    {$I+}
    filexist:=(IOresult=0);
end;
```

## program PLOT

This was a simple program that plotted points on the computer screen. The resulting plot did not look perfectly smooth on the screen for two reasons. First, straight lines were drawn between the points; there was no attempt to fit a curve. Second, the resolution of the screen itself limited how smoothly a diagonal appeared. The plot looked smoother as more points were plotted. The plot could be printed on a dot-matrix printer by running the DOS program GRAPHICS before running PLOT, and then using the 'PRINT SCREEN' key. PLOT drew axes and scale marks, but did not write scale values on the graph.

The stimulus for writing this program was to clarify the output from the linkage analysis program, LINKMAP. Early versions of LINKMAP generated a list of recombination fractions between loci and the location score for that set of recombination fractions. A program MAP was written to convert the set of recombination fractions to a genetic location, a parameter that was easier to plot and interpret. Subsequent versions of the LINKAGE programs rendered MAP redundant, and provided a list of paired genetic locations and location scores. The program PLOT allowed this output to be promptly plotted and checked. An example of such a plot is shown overleaf.

This figure is the PLOT version of the multipoint linkage map shown in Figure 6-2. The X-axis is the background genetic map on which FRAXA was localised. The ticks along the X-axis indicate the positions of the loci that constituted the map. The Y-axis indicates the multipoint LOD score at various points along the map. The ticks on the Y-axis are at intervals of 10 LOD score units, ranging from 0 to 50.



```
Program PLOT (input,infile,output);
{sourcefile:PLOT.pas
 sourcecode:Turbo-Pascal V3.02A
 author    :Graeme Suthers, 26/4/89.
```

```
PLOT.PAS is a shell program for running the procedure DRAWxy.PAS.
DRAWxy.PAS is designed to be readily incorporated into other programs, and
PLOT.PAS simply provides the environment to run DRAWxy.PAS.
}
```

```
type
  filename=string[12];
var
  data:text;
  datafile:filename;
  answer:char;
  scaledata:boolean;
```

```
{I DrawXY.PAS}
{I Fileexist.pas}
```

```
begin
  clrscr;
  writeln('program PLOT. ');
  writeln;
  writeln('(written by Graeme Suthers, 24/489). ');
  writeln;
  writeln('This program will plot a series of points on a graph, and join them ');
  writeln('with straight lines. To get a printed copy of the graph press the ');
  writeln('PRINT SCREEN key.');
```



```

writeln;
writeln('The data must be in a textfile in tabulated form ');
writeln('          ( X1 Y1');
writeln('          X2 Y2 etc.)');
writeln('with the X values in order of increasing size. ');
writeln;
writeln('You are asked to specify the range for both axes. ');
writeln('If the point 0,0 lies within these bounds the axes will be drawn through');
writeln('that point. If 0,0 is not in the range of the axes you will be asked');
writeln('to specify a new point through which the axes will pass.');
```

writeln;

```

writeln('Any data points that lie outside the range of the axes are ignored.');
```

writeln;

```

writeln('Press any key to continue...');
repeat until keypressed;
clrscr;
writeln('You can place up to 20 marks on each axis to indicate scale. The scale values ');
writeln('may be entered manually or be listed in the datafile. If the values are in the ');
writeln('datafile, values for the X-axis must be on the first line and the values for ');
writeln('the Y-axis on the second line. On each line the number of scale values must');
writeln('be indicated at the beginning of the line.');
```

writeln('	e.g. 3	-0.2	-0.1	0.1	(X-axis values)');
writeln('	4	-10	-5	5	10 (Y-axis values)');

```

writeln;
writeln('To exit this program after the graph is drawn, type Q (for quit).');
```

writeln;

```

writeln('Now strike any key to continue or press Ctrl-Break to exit.');
```

repeat until keypressed;

```

clrscr;
write('Enter the name of your datafile : ');
```

```
readln(datafile);
if not Fileexist(datafile) then
  begin
    writeln('File ',datafile,' not found in this directory. ');
    writeln('Program aborted. ');
    halt(1);
  end;
assign(data,datafile);
write('Does this file contain the X and Y scale values ? (y/n) : ');
readln(answer);
if (answer='y')or(answer='Y') then scaledata:=true
  else scaledata:=false;
DrawXY(data,1,1,1,1,0,0,scaledata);

end {program PLOT}.
```

{sourcecode:Turbo-Pascal.  
sourcefile:FILEXIST.pas.  
Graeme Suthers. 13/04/88.

This an INCLUDE file to be included in the compilation of another .PAS file.  
The program defines the function FILEXIST which returns the Boolean expression  
TRUE if the file specified in the function call exists in the current directory.  
The parent program must contain the TYPE declaration:

```
                filename:string[12]
    }
FUNCTION filexist(queryfile:filename) : Boolean;
var
    CHECKFILE:file;
begin
    assign(CHECKFILE,queryfile);
    {$I-}
    reset(CHECKFILE);
    close(CHECKFILE);
    {$I+}
    filexist:=(IOresult=0);
end {FILEXIST include file};
```

{sourcefile:DRAWxy.PAS  
sourcecode:Turbo-Pascal V3.02A  
author :Graeme Suthers, 26/4/89

DRAWxy.PAS is an INCLUDE file for drawing graphs on a 320x200 pixel graphics screen. The procedure call is

```
DRAWxy (datafile,xmin,xmax,ymin,ymax,xorigin,yorigin,readscale);
```

where datafile : is a previously assigned textfile with the data in tabulated form ( X1 Y1  
X2 Y2 etc.) and the X values are in order of increasing size. The datafile has the logical name XYDATA in the procedure DRAWxy.

xmin,xmax,ymin,ymax,xorigin,yorigin : are real. If xmin=xmax or ymin=ymax then you are prompted to enter these values. If xorigin or yorigin do not lie within the min-max intervals then you are prompted to enter these. If these 6 values are set at  
1,1,1,1,0,0  
then you will be prompted to enter all xmin,xmax,ymin, and ymax  
If the point 0,0 lies within these bounds the origin will be placed there.

readscale : is Boolean. If true the scale values for the X-axis are on the first line of Datafile, and the values for the Y-axis are on the second line. On each line the number of scale values must be indicated at the beginning of the line.

e.g. 3 -0.2 -0.1 0.1  
4 -10 -5 5 10

If readscale is false, you are prompted to provide these values.  
The maximum number of scale marks on each axis is 20.

```
}  
  
PROCEDURE DrawXY (var xydata:text;  
                  xmin,xmax,ymin,ymax,xorigin,yorigin:real;  
                  readscale:boolean);  
  
const  
  windowwidth=319;  
  windowdepth=199;  
  debug=false;  
  
var  
  xvalue,yvalue,nextx,nexty:real;  
  i,j,xscalenum,yscalenum:integer;  
  quit:char;  
  xscale,yscale:array[1..20]of real;  
  fatalerror:boolean;  
  
PROCEDURE SetUp;  
  
begin  
  if (xmin>=xmax) then  
    repeat  
      write('Enter minimum X value on graph : ');  
      readln(xmin);
```

```

        write('Enter maximum X value on graph : ');
        readln(xmax);
        writeln;
    until xmin<xmax;

if (ymin>ymax) then
    repeat
        write('Enter minimum Y value on graph : ');
        readln(ymin);
        write('Enter maximum Y value on graph : ');
        readln(ymax);
        writeln;
    until ymin<ymax;

if (xorigin<xmin)or(xorigin>xmax)or(yorigin<ymin)or(yorigin>ymax) then
    repeat
        write('Enter X and Y coordinates of the origin : ');
        readln(xorigin,yorigin);
        writeln;
    until (xorigin>=xmin)and(xorigin<=xmax)and(yorigin>=ymin)and(yorigin<=ymax);

if readscale then
    begin
        read(xydata,xscalenum);
        for i:=1 to xscalenum do read(xydata,xscale[i]);
        readln(xydata);
        read(xydata,yscalenum);
        for i:=1 to yscalenum do read(xydata,yscale[i]);
        readln(xydata);
    end

```

```

end
else
  begin
    write('How many X scale values will you want ? ');
    readln(xscalenum);
    writeln;
    for j:=1 to xscalenum do
      begin write('Enter X scale value #',j,' : ');
        readln(xscale[j]);
      end;
    writeln;
    write('How many Y scale values will you want ? ');
    readln(yscalenum);
    writeln;
    for j:=1 to yscalenum do
      begin write('Enter Y scale value #',j,' : ');
        readln(yscale[j]);
      end;
    end;

  end;

readln(xydata,xvalue);
while not EOF(xydata) do
  begin
    readln(xydata,nextx);
    if nextx<xvalue then
      begin
        close(xydata);
        writeln('ERROR in DRAW procedure: X values incorrectly ordered. ');
        writeln('Procedure aborted. ');
        fatalerror:=true;
      end;
    end;
  end;
end;

```

```

                exit;
            end;
            xvalue:=nextx;
        end;
end; {PROCEDURE SetUp}

PROCEDURE DoPlot;

FUNCTION CalcX (X:real):integer; {scales X value for graph}
begin
    CalcX:=round((((x-xmin)/(xmax-xmin))*(windowwidth-1))+1);
end;

FUNCTION CalcY (Y:real):integer; {scales Y value for graph}
begin
    CalcY:=round(((1-((y-ymin)/(ymax-ymin)))*(windowdepth-1))+1);
end;

begin {PROCEDURE DoPlot}
    clrscr;
    Graphcolormode;
    Graphbackground(0);
    Palette(3);

    {draw box}
    draw(1,1,windowwidth,1,2);
    draw(1,1,1,windowdepth,2);

```



```

draw(1,windowdepth,windowwidth,windowdepth,2);
draw(windowwidth,1,windowwidth,windowdepth,2);

{draw axes}
draw(1,calcy(yorigin),windowwidth,calcy(yorigin),2);
draw(calcx(xorigin),1,calcx(xorigin),windowdepth,2);

{draw scales}
for i:=1 to xscalenum do
    draw(calcx(xscale[i]),(calcy(yorigin)-2),calcx(xscale[i]),(calcy(yorigin)+2),2);
for i:=1 to yscalenum do
    draw((calcx(xorigin)-2),calcy(yscale[i]),(calcx(xorigin)+2),calcy(yscale[i]),2);

{plot XY values}
readln(xydata,xvalue,yvalue);
while not EOF(xydata) do
    begin
        readln(xydata,nextx,nexty);
        draw(calcx(xvalue),calcy(yvalue),calcx(nextx),calcy(nexty),3);
        xvalue:=nextx;
        yvalue:=nexty;
    end; {while}

end; {PROCEDURE DoPlot}

begin {PROCEDURE DrawXY}
    fatalerror:=false;
    reset(xydata);
    SetUp;
    if fatalerror then exit;

```

```
reset(xydata);  
if readscale then for i:=1 to 2 do readln(xydata);  
DoPlot;  
repeat  
    read(kbd,quit);  
until ((quit='q') or (quit='Q'));  
close(xydata);  
textmode;  
clrscr;  
end {procedure DRAWXY};
```

**program MORGAN**

The LINKAGE programs frequently required interconversion of recombination fractions and genetic distances (in cM). MORGAN was written to do these calculations using the mapping functions described by Haldane and Kosambi (Ott 1985, p.8). Ott has subsequently distributed a similar program which also performs these calculations using the mapping functions of Rao and Carter-Falconer.

```

program MORGAN {input, output};

{sourcefile:MORGAN.pas
 sourcecode:Turbo-Pascal
 author   :Graeme Suthers, 16/12/88}

{calculates the recombination fraction (theta) corresponding to a given
 map distance (in M) and vica versa using Kosambi and Haldane functions. }

var
  mapdistance, theta: real;
  calculation: integer;
  choice: char;

procedure MORGANS;
begin
  repeat
    write ('Type in map distance (in Morgans): ');
    read (mapdistance);
    writeln;
    theta:=((exp(4*mapdistance))-1)/((exp(4*mapdistance)+1)*2);
    writeln ('>>> theta = ',theta:4:3, ' (Kosambi function)');
    theta:=0.5*(1-exp(-2*mapdistance));
    writeln ('>>> theta = ',theta:4:3, ' (Haldane function)');
    write('Do another ? (y/n) : ');
    readln(choice);
    writeln;
  until choice='n';
end;

```

```

procedure THETAS;
begin
  repeat
    write('Type in recombination fraction (as decimal fraction): ');
    read(theta);
    writeln;
    mapdistance:=0.25*(ln((1+2*theta)/(1-2*theta)));
    writeln('>>> Mapdistance = ',mapdistance:4:3,' Morgans (Kosambi function)');
    mapdistance:=-0.5*ln(1-2*theta);
    writeln('>>> Mapdistance = ',mapdistance:4:3,' Morgans (Haldane function)');
    write('Do another ? (y/n) : ');
    readln(choice);
    writeln;
  until choice='n';
end;

begin {PROGRAM}
  repeat
    clrscr;
    writeln('program MORGAN');
    writeln;
    writeln('Please indicate whether you want to calculate...');
    writeln('      1:  mapdistance (given rec fraction)');
    writeln('      2:  rec fraction (given mapdistance)');
    writeln;
    writeln('      0:  exit program');
    writeln;
    write('                                Your choice : ');
    readln(calculation);
  case calculation of

```

```
1:thetas;  
2:morgans;  
end;  
until (calculation<>1)and(calculation<>2);  
end {program MORGAN}.
```

program **BOOTMAP**

The program **BOOTMAP** was written to estimate an approximate confidence interval for FRAXA location from multipoint linkage data using the resampling or 'bootstrap' methodology (Chapter 2). The location of FRAXA was analyzed in a series of 101 pedigrees with program **LINKMAP** (Chapter 6). The multipoint LOD scores corresponding to each pedigree were stored in a series of datafiles, one for each pedigree. This was readily achieved with the aid of some small programs to manipulate the data (programs not shown). **BOOTMAP** randomly resampled (with replacement) the sets of LOD scores 101 times. Multipoint LOD scores from different pedigrees can be summed (in the same way as two-point LOD scores), and **BOOTMAP** simply calculated the most likely FRAXA location in the resample. This process was repeated 1000 times. The range of 95% of the FRAXA locations (centered at the median) indicated the approximate 95% confidence interval for FRAXA location. An example of the output using this conservative approach is shown after the program listing.

The program was then modified and the analysis repeated using a non-conservative resampling approach whereby a least one recombinant between FRAXA and each of the other loci was included in each resample. This ensured that no resample indicated a FRAXA location that was impossible using the original data set.

```
program BOOTMAP (datafile, outfile, output);
```

```
{sourcefile:BOOTMAP.pas  
sourcecode:Turbo-Pascal V3.02A  
author :Graeme Suthers, 12/5/89, 9/5/90.
```

This program calculates a confidence interval for genetic location of a locus against a background map using location scores from the LINKAGE program LINKMAP. The rationale behind this approach is described in the thesis, by Suthers & Wilson [Am J Hm Genet, July 1990], and in Chapters 2 and 6 of this thesis.

The necessary data files are described in the procedure READDATA below.  
}

```
const  
  title='FRAXA';           {of output file}  
  nped=101;                {number of pedigrees; best over 30}  
  nlocation=30;           {number of location points in study; best at 1 cM intervals}  
  niteration=1000;         {number of iterations; best at ~1000}  
  CI=0.95;                {desired confidence interval}  
  outfile='BOOTMAP.OUT';  {name of outfile}  
  minscore=-10E6;         {don't change}  
  debug=false;           {if true, you get a lot of screen feedback about resampling; best if false}  
  
type  
  pednumbers=1..nped;  
  
var
```



```

datafile,outfile:text;

pedsample,pedfreq:array [1..nped] of integer;
scoresum,location:array [1..nlocation] of real;
pedscore:array [pednumbers,1..nlocation] of real;
tally:array [1..nlocation] of integer;

iteration,tallypoint,tallycount,pedchoice,pedcount,mediancount,i,j:integer;
topscore,chisqr,totallocation:real;

procedure READDATA;
{reads location scores or multipoint LOD scores from a series of data files.
There is one datafile (in ASCII) per pedigree. The datafile has the pedigree number
(up to 4-digit integer) on the first line followed by nlocation location or LOD scores on
successive lines. The scores must be real, have no more than 11 significant digits,
have an exponent of no more than two digits, and be in the range 1E-38 to 1E+38.
The list of datafile names must be an ASCII file called LODFILE.lst with each datafile name
(no more than 14 characters) on successive lines. The scores are all stored in the array pedscore.
The list of location points at which the scores were calculated must be an ASCII file with
nlocation location points listed on successive lines (same constraints on these real values
as for the scores).
}

var
lodlist,lodfile,locationlist:text;
lodfilename:string[14];
lod,locationpoint:real;

```

```

pednum: integer;

begin
  writeln('Reading data from individual LOD files....');
  writeln;
  assign(lodlist, 'LODFILE.lst');
  reset(lodlist);

  for i:=1 to nped do
    begin
      readln(lodlist, lodfilename);
      assign(lodfile, lodfilename);
      reset(lodfile);
      readln(lodfile, pednum);
      writeln('      ...pedigree ', pednum:5);
      for j:=1 to nlocation do
        begin
          readln(lodfile, lod);
          pedscore[i,j]:=lod;
        end;
      close(lodfile);
    end;
  close(lodlist);

  assign(locationlist, 'LOCATION.lst');
  reset(locationlist);
  for i:=1 to nlocation do
    begin
      readln(locationlist, locationpoint);
      location[i]:=locationpoint;
    end;
  end;
end;

```

```

        end;
close(locationlist);

clrscr;
writeln('Finished reading data.');
```

end;

```

procedure SELECTSAMPLE;
{chooses a resample of nped pedigrees such that each sample includes pedigrees
with recombination between FRAXA and 4D8, FRAXA and RN1/VK23B, FRAXA and VK21/
IDS, and FRAXA and U6.2. This step is necessary to ensure that the resampled sets of
location scores don't place FRAXA at an impossible location e.g. at a locus where there
was phase-known recombination in the original data set. See Suthers & Wilson, 1990. }

var
  p4D8rec,RN1_VK23Brec,VK21_IDSrec,U6_2rec:boolean;

begin
  repeat
    p4D8rec:=false;
    RN1_VK23Brec:=false;
    VK21_IDSrec:=false;
    U6_2rec:=false;

    for i:=1 to nped do pedsample[i]:=random(nped)+1;
    i:=0;
    repeat
      i:=i+1;
```

```

        { The numbers in the [sets] below are the numbers of the pedigrees with recombinants as listed in LODFILE.lst}
        if pedsample[i] in [89] then p4D8rec:=true;
        if pedsample[i] in [42,49,54,66,70,71] then RN1_VK23Brec:=true;
        if pedsample[i] in [42,57] then VK21_IDSrec:=true;
        if pedsample[i] in [4] then U6_2rec:=true;
    until (p4D8rec and RN1_VK23Brec and VK21_IDSrec and U6_2rec) or (i=nped);
until (p4D8rec and RN1_VK23Brec and VK21_IDSrec and U6_2rec);
if debug then
begin
    writeln;
    for i:=1 to nped do writeln(pedsample[i]:4);
    writeln;
end;
end;

```

```
begin {program BOOTMAP}
```

```
{SET UP}
```

```
assign(outfile,outfilename);
rewrite(outfile);
```

```
clrscr;
writeln('Program BOOTMAP :');
writeln;
writeln('Location scores from ',nped:4,' pedigrees at ',nlocation:4,' locations.');
```

```
writeln;
writeln('Resampling scores by pedigrees ',niteration:6,' times.');
```

```
writeln;
```

```
writeln(outfile,outfilename,' : ',title);
writeln(outfile);
writeln(outfile,'Number of pedigrees : ',nped:6);
writeln(outfile,'Number of locations : ',nlocation:6);
writeln(outfile);
```

```
for i:=1 to nped do pedfreq[i]:=0;
for i:=1 to nlocation do tally[i]:=0;
```

```
readdata;
```

```
{RESAMPLING and determining location of maximum score in resample}
```

```
gotoXY(1,10);
```

```
for iteration:=1 to niteration do
```

```
begin
```

```
write('Iteration : ',iteration:8);
```

```
for i:=1 to nlocation do scoresum[i]:=0;
```

```
Selectsample;
```

```
for pedcount:=1 to nped do
```

```
begin
```

```
pedchoice:=pedsample[pedcount];
```

```
pedfreq[pedchoice]:=pedfreq[pedchoice]+1;
```

```
for i:=1 to nlocation do scoresum[i]:=scoresum[i]+pedscore[pedchoice,i];
```

```
end;
```

```
topscore:=minscore;
```

```
for i:=1 to nlocation do
```

```
if scoresum[i]>topscore then
```

```
begin
```

```

        topscore:=scoresum[i];
        tallypoint:=i;
    end;
tally[tallypoint]:=tally[tallypoint]+1;
if debug then
    begin
        for i:=tallypoint-3 to tallypoint+3 do write(lst,scoresum[i]:10:4);
        writeln(lst);
    end;
gotoXY(1,10);
end;

```

{CALC RELATIVE LOCATION SCORES}

```

writeln;
for i:=1 to nlocation do scoresum[i]:=0;
for i:=1 to nlocation do
    for j:=1 to nped do scoresum[i]:= scoresum[i]+pedscore[j,i];
topscore:=minscore;
for i:=1 to nlocation do
    if scoresum[i]>topscore
        then begin
            topscore:=scoresum[i];
            tallypoint:=i;
        end;
writeln(outfile);
writeln(outfile,'Maximum location score is ',scoresum[tallypoint]:10:4,' at ',location[tallypoint]:10:3);
writeln(outfile);

```

{CHECK RANDOMNESS OF SAMPLING}

```

writeln(outfile,'Number of resamplings : ',niteration:6, 'samples of ',nped:3,' pedigrees each.');
```

```

writeln(outfile);
{for i:=1 to nped do writeln(outfile,'Pedigree ',i:2,' sampled ',pedfreq[i]:4,' times.')}
writeln(outfile);}
chisqr:=0;
for i:=1 to nped do chisqr:=chisqr+((sqr(pedfreq[i]-niteration))/(niteration));
writeln(outfile,'Chisquared value for ped sampling was ',chisqr:10:4);
writeln(outfile,'          with ',(nped-1):3,' degrees of freedom.')}
writeln(outfile);

{PRINT OUT LOCATION SCORES AND TALLY COUNTS}
writeln(outfile,'Location':20,'LOD score':20,'Tally':20);
for i:=1 to nlocation do
begin
write(outfile,location[i]:20:3,scoresum[i]:20:4,tally[i]:20);
writeln(outfile);
end;

{CALCULATE MEDIAN, MEAN, AND CI}
i:=0;
mediancount:=0;
repeat
i:=i+1;
mediancount:=mediancount+tally[i];
until mediancount>=trunc(niteration/2);
writeln('Median location : ',location[i]:10:3);
writeln(outfile,'Median location : ',location[i]:10:3);
totallocation:=0;
for i:=1 to nlocation do totallocation:=totallocation+(location[i]*tally[i]);

```

```

writeln('Mean location : ',totallocation/niteration:10:3);
writeln(outfile,'Mean location : ',totallocation/niteration:10:3);
i:=0;
tallycount:=0;
repeat
    i:=i+1;
    tallycount:=tallycount+tally[i];
until tallycount>=trunc(niteration*(0.5-(CI/2)));
write(CI*100:3:0,'% confidence interval is ',location[i]:10:3,' to ');
write(outfile,CI*100:3:0,'% confidence interval is ',location[i]:10:3,' to ');
i:=0;
tallycount:=0;
repeat
    i:=i+1;
    tallycount:=tallycount+tally[i];
until tallycount>=trunc(niteration*(0.5+(CI/2)));
writeln(location[i]:10:3);
writeln(outfile,location[i]:10:3);

close(outfile);

end.

```



Example of output from BOOTMAP:

BOOTMAP.OUT : FRAXA

Number of pedigrees : 101  
Number of locations : 45

Maximum location score is 48.4926 at 0.180

Number of resamplings : 1000 samples of 101 pedigrees each.  
Chisquared value for ped sampling was 122.1960  
with 100 degrees of freedom.

Location (M)	LOD score	Tally
-10.000	0.0000	0
-0.805	5.4134	0
-0.458	13.6583	0
-0.255	22.8721	0
-0.112	31.5807	0
-0.000	35.0835	0
0.000	35.0835	0

0.025	40.3874	0
0.052	42.1401	2
0.080	43.2613	14
0.110	43.2542	1
0.141	33.0846	0
0.141	33.0846	0
0.144	40.5351	0
0.147	42.6681	0
0.150	44.0563	0
0.153	45.0841	0
0.156	45.8880	0
0.159	46.5346	1
0.162	47.0616	4
0.165	47.4916	16
0.168	47.8385	39
0.171	48.1103	67
0.174	48.3108	130
0.177	48.4397	181
0.180	48.4926	204
0.183	48.4599	168
0.186	48.3241	102
0.189	48.0543	50
0.192	47.5922	15
0.195	46.8129	5
0.198	45.3617	0
0.202	40.2769	0
0.202	40.2769	0
0.204	43.0751	0
0.207	44.0071	0
0.209	44.3906	0

*Appendix A. Computer programs. p. A-75*

0.211	44.3161	0
0.214	40.9718	0
0.214	40.9718	0
0.325	40.1844	1
0.469	29.8838	0
0.672	18.0090	0
1.019	6.9552	0
10.214	-0.0000	0

Median gene location : 0.180 Morgan  
Mean gene location : 0.177 Morgan  
95% confidence interval for gene location is 0.165 to 0.189 Morgan



**Appendix B**

**GENOTYPES IN THE CEPH PEDIGREES.**

This Appendix lists the genotypes at nine polymorphic loci at Xq26-q28 in the normal Centre d'Etude du Polymorphisme Humain (CEPH) pedigrees. The analysis of these data is presented in Chapter 5.

The nine loci (and probes) listed are:

<u>F9</u>	pVIII
<u>DXS105</u>	cX55.7
<u>DXS98</u>	4D-8
<u>DXS369</u>	RN1
<u>DXS297</u>	VK23B
<u>DXS296</u>	VK21A and VK21C
<u>IDS</u>	pc2S15
<u>DXS304</u>	U6.2
<u>DXS52</u>	St14-1

Details of which RFLP was used at each locus are given in Chapter 5 (Table 5-A). Most of the data regarding RFLPs detected by the probes F9, cX55.7, 4D-8, RN1, U6.2, and St14 were provided by Dr I. Oberle. Permission to reproduce or utilize the data which were generated with these probes should be sought from Dr Oberle<sup>1</sup>. The data regarding RFLPs detected by VK23B, VK21A, VK21C, and IDS were obtained by the candidate. They have been lodged with CEPH, and are in the public domain.

---

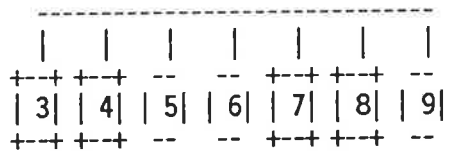
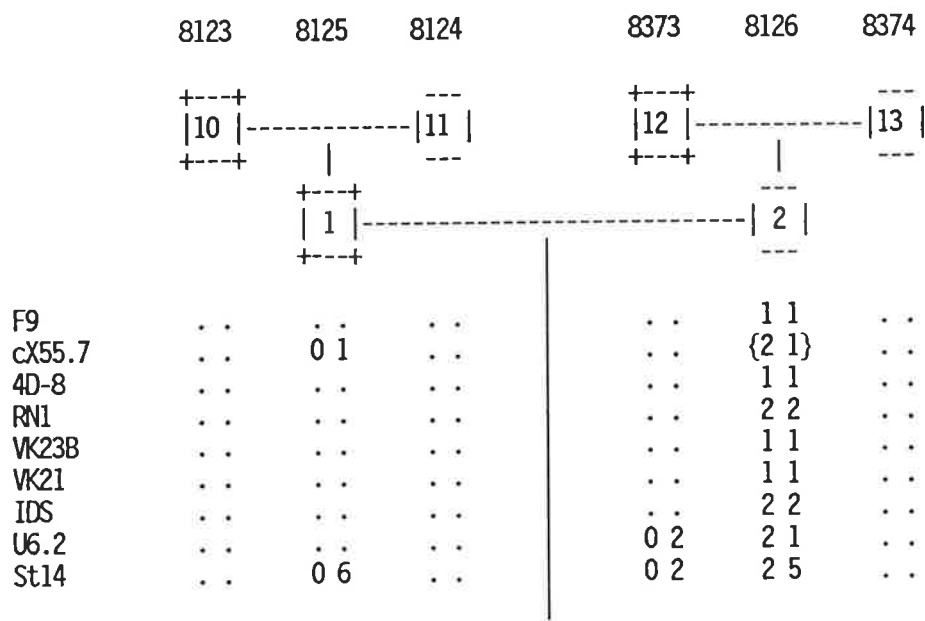
<sup>1</sup> Dr I Oberle, L.G.M.E./C.N.R.S., INSERM U184, Faculte de Medecine, 11 Rue Humann, 67085 Strasbourg Cedex, France.

Where a number of RFLPs have been identified at a single locus, the haplotypes were reduced to a pair of alleles (Chapter 2). The program XPHASE (Appendix A) was then used to check the pedigree and genotype data and to infer the phase of the alleles. The plotting of the pedigrees with the genotypes was performed by the CEPH program SEEALL. The output from SEEALL was then modified to remove redundant genotypes and to indicate recombination events. The paternal grandparents and maternal grandmother do not provide information in the linkage analysis of X-linked loci, and these genotypes are not listed below.

The loci are listed in order down the X chromosome. Alleles are given as pairs of numbers, but because some haplotypes have been reduced, these numbers do not correspond to the allele numbers listed by Kidd et al. (1989). The two columns of figures under each female correspond to the two X chromosomes with the alleles in phase. In males there is one column of "0"s followed by a column of the alleles on the single X chromosome. Where the phase at a locus has been inferred the alleles in the mother have been {bracketed}. Where there is no data at a locus (usually because the mother was homozygous at that locus) the alleles have been shown as '..'. Where recombination has occurred the flanking loci are indicated by arrows (<). The numbers within, below, or above

male/female symbols are identification numbers used by the CEPH programs.

Pedigree No. : 13291

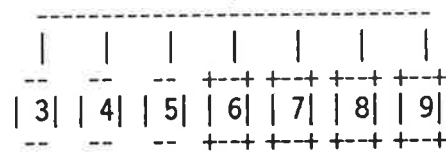
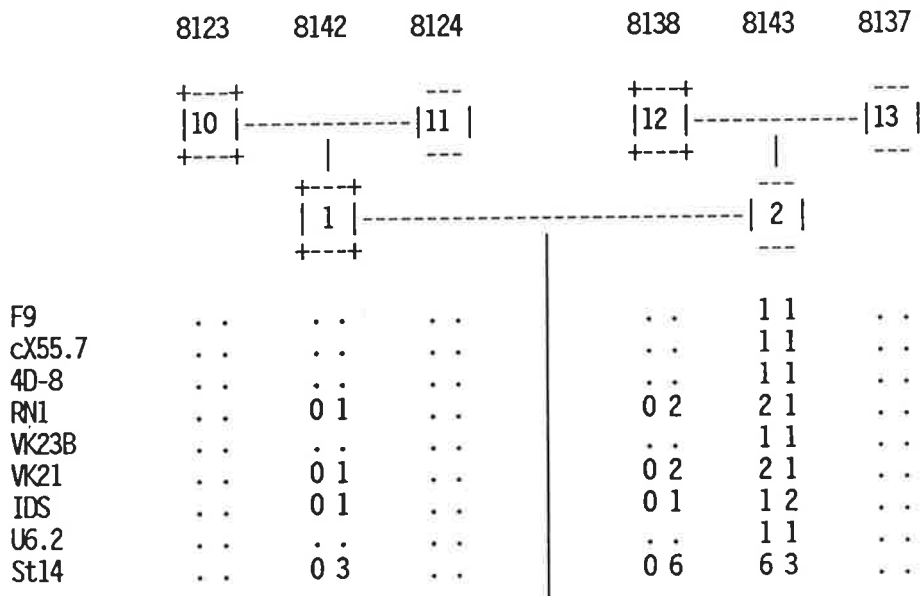


8128 8129 8145 8144 8127 8428 8427

F9	..	..	..	..	..	..	..
cX55.7	..	0 1	1 1	1 2	0 2<	0 2	1 2
4D-8	..	..	..	..	..	..	..
RN1	..	..	..	..	..	..	..
VK23B	..	..	..	..	..	..	..
VK21	..	..	..	..	..	..	..
IDS	..	..	..	..	..	..	..
U6.2	..	0 1	1 1	1 2	..	0 2<	1 2
St14	0 5	0 5	6 5	6 2	0 5<	0 5<	6 2



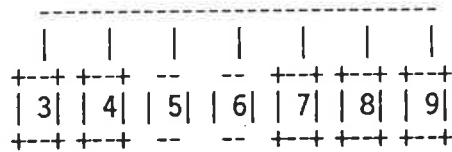
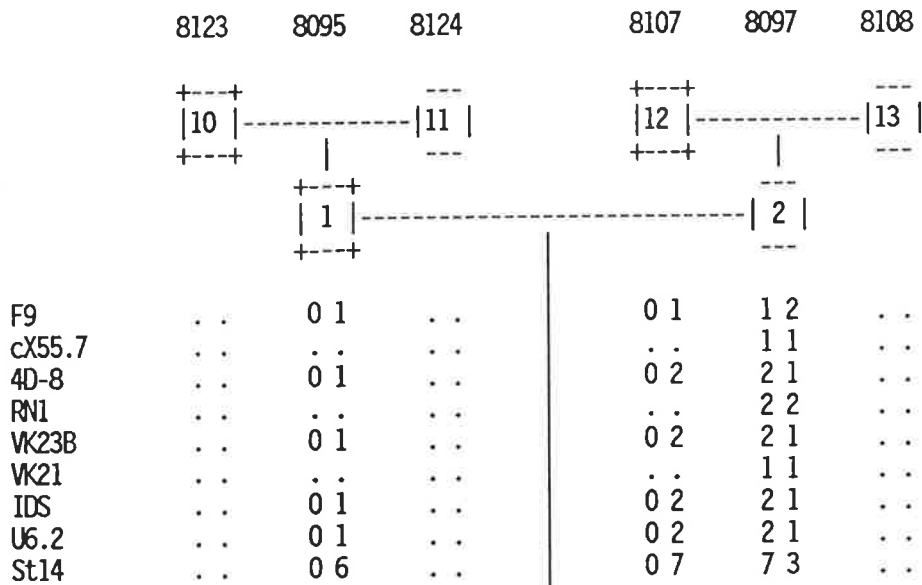
Pedigree No. : 13292



8175 8598 8599 8141 8146 8425 8426

F9	..	..	..	..	..	..	..
cX55.7	..	..	..	..	..	..	..
4D-8	..	..	..	..	..	..	..
RN1	1 2	1 1	1 1	0 2	0 1	0 1	0 2
VK23B	..	..	..	..	..	..	..
VK21	1 2	1 1	1 1	0 2	0 1	0 2	0 2
IDS	1 1	1 2	1 2	0 1	0 2	0 1	0 1
U6.2	..	..	..	..	..	..	..
St14	3 6	3 6	3 3	0 6	..	0 6	0 6

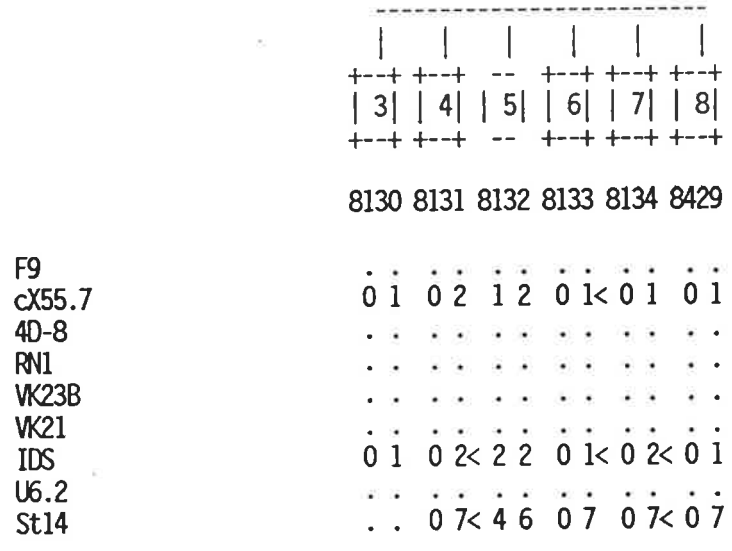
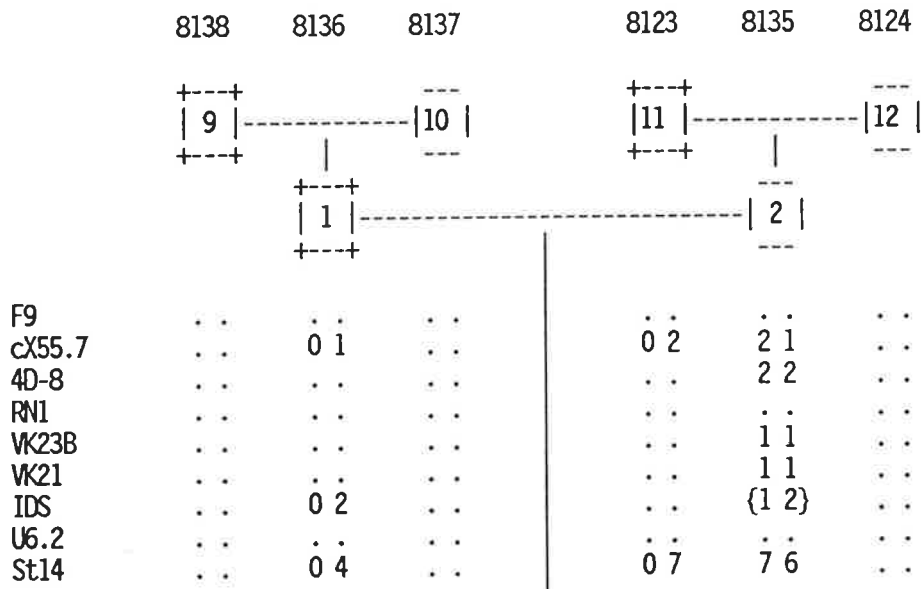
Pedigree No. : 13293



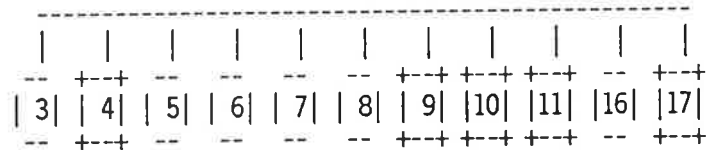
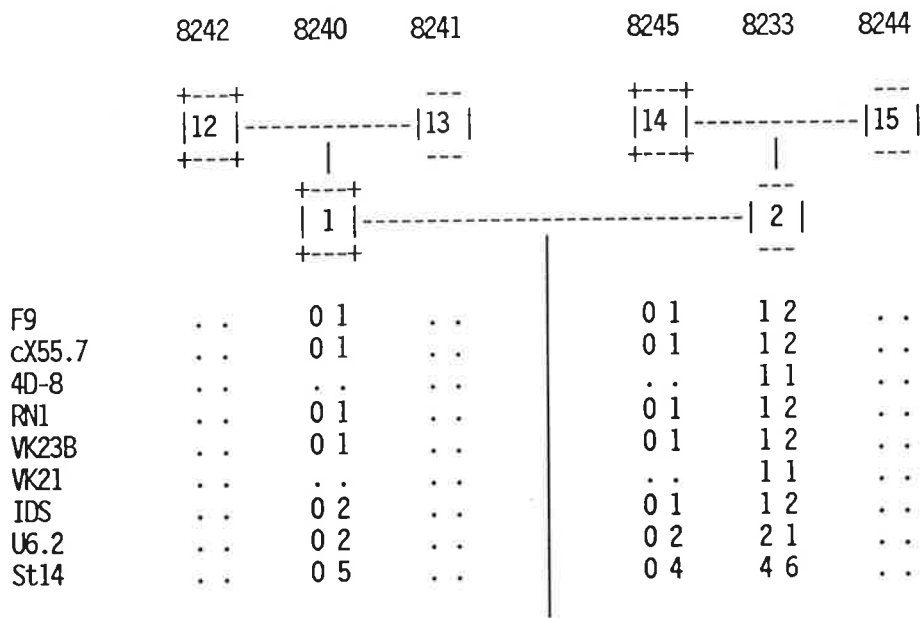
8090 8091 8092 8093 8096 8432 8433

F9	0 2<	0 2	1 2	1 2	0 1	0 2	0 1
cX55.7	0 .	0 .	1 .	1 .	0 .	0 .	0 .
4D-8	0 2<	0 1	1 1	1 1	0 2	0 1	0 2
RN1	0 .	0 .	1 .	1 .	0 .	0 .	0 .
VK23B	0 2	0 1	1 1	1 1	0 2	0 1	0 2
VK21	0 .	0 .	1 .	1 .	0 .	0 .	0 .
IDS	0 2	0 1	1 1	1 1	0 2	0 1	0 2
U6.2	0 2	0 1	1 1<	1 1	0 2<	0 1	0 2
St14	0 7	0 3	6 7<	6 3	0 3<	0 3	0 7

Pedigree No. : 13294



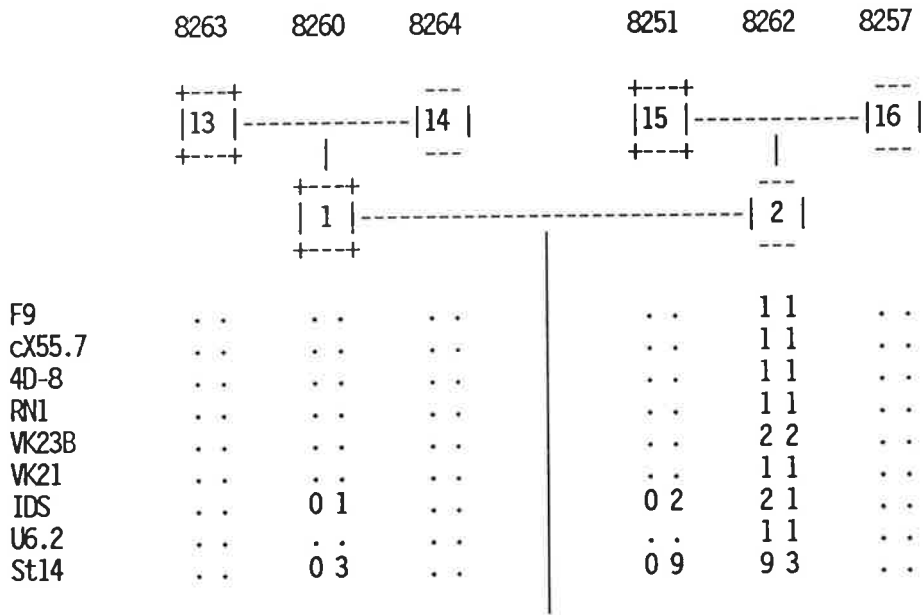
Pedigree No. : 1331



8310 8549 8237 8236 8235 8234 8238 8243 8239 2426 2427

F9	1 1	0 2	1 2	1 1	1 1	< 1 2	0 1	0 1	0 2	. .	0 1
cX55.7	1 1	0 2	1 2	1 1	1 2	< 1 2	0 1	< 0 1	0 2	< . .	0 1
4D-8	. .	. .	. .	. .	. .	. .	. .	. .	. .	. .	. .
RN1	1 1	0 2	1 2	1 1	1 2	1 2	0 2	< 0 1	0 1	< . .	0 1
VK23B	1 1	0 2	1 2	1 1	1 2	1 2	0 2	0 1	0 1	1 1	0 1
VK21	. .	. .	. .	. .	. .	. .	. .	. .	. .	. .	. .
IDS	2 1	0 2	2 2	2 1	2 2	2 2	0 2	0 1	0 1	2 1	0 1
U6.2	2 2	0 1	2 1	2 2	2 1	2 1	0 1	0 2	0 2	2 2	0 2
St14	5 4	0 6	5 6	5 4	5 6	5 6	0 6	0 4	0 4	. .	0 4

Pedigree No. : 1332

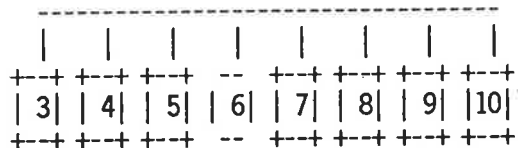
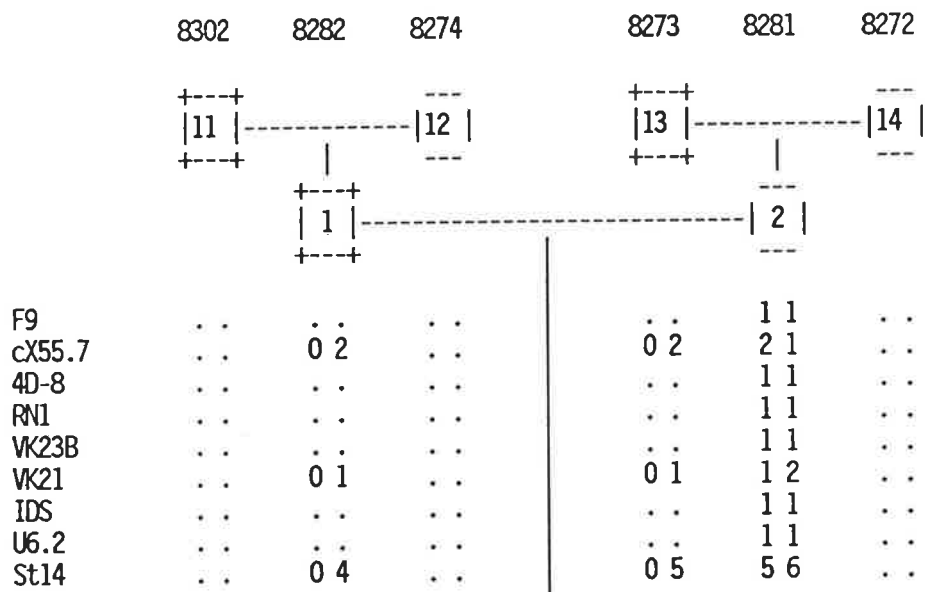


	8263	8260	8264	8251	8262	8257
F9	..	..	..	..	1 1	..
cX55.7	..	..	..	..	1 1	..
4D-8	..	..	..	..	1 1	..
RN1	..	..	..	..	1 1	..
VK23B	..	..	..	..	2 2	..
VK21	..	..	..	..	1 1	..
IDS	..	0 1	..	0 2	2 1	..
U6.2	..	..	..	..	1 1	..
St14	..	0 3	..	0 9	9 3	..

	8252	1133	8265	8261	8256	8255	8258	8259	8253	8254
F9	..	..	..	..	..	..	..	..	..	..
cX55.7	..	..	..	..	..	..	..	..	..	..
4D-8	..	..	..	..	..	..	..	..	..	..
RN1	..	..	..	..	..	..	..	..	..	..
VK23B	..	..	..	..	..	..	..	..	..	..
VK21	..	..	..	..	..	..	..	..	..	..
IDS	1 2	0 2	1 2	0 1	1 1	1 1	..	1 2	0 2	0 2
U6.2	..	..	..	..	..	..	..	..	..	..
St14	3 9	..	3 9	0 3	3 3	3 3	..	3 9	0 9	0 9

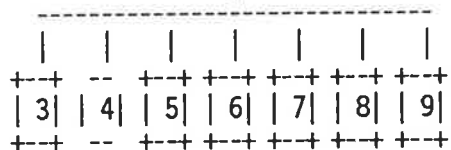
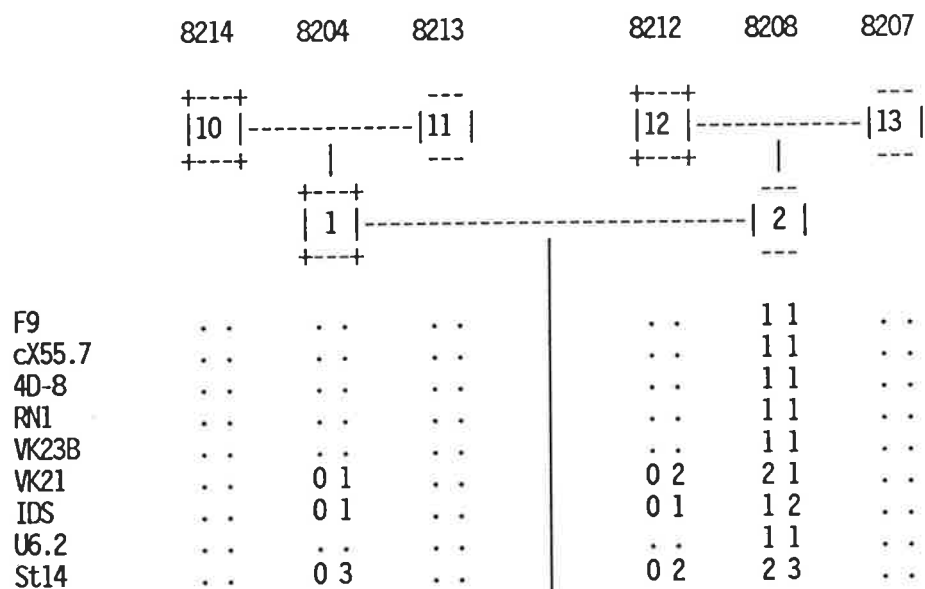
Pedigree No. : 1333



8280 8276 8303 8275 8278 8277 8279 8337

F9		0 1	0 1	0 2	2 2<	0 2	0 1	0 1	0 2
cX55.7		0 1	0 1	0 2	2 2<	0 2	0 1	0 1	0 2
4D-8		..	..	..	..	..	..	..	..
RN1		..	..	..	..	..	..	..	..
VK23B		0 2	0 2	0 1	1 2<	0 1	0 2	0 2	0 1
VK21		0 2	0 2	0 1	1 2<	0 1	0 2	0 2	0 1
IDS		..	..	..	..	..	..	..	..
U6.2		..	..	..	..	..	..	..	..
St14		0 6	0 6	0 5	4 6	0 5	0 6	0 6	..

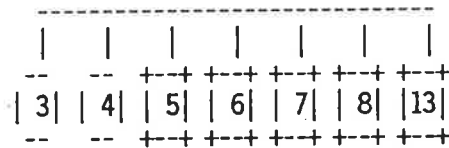
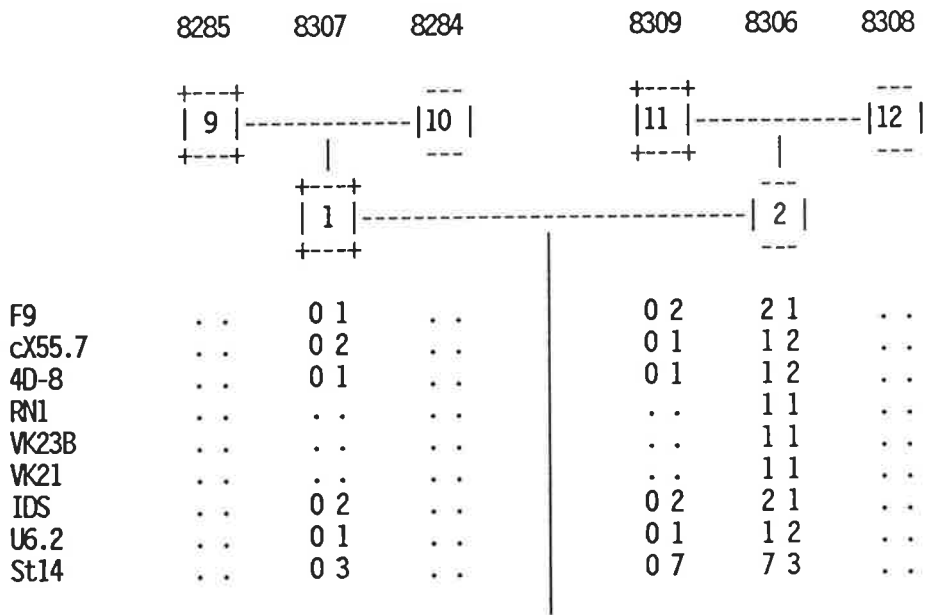
Pedigree No. : 1334



8201 8202 8210 8211 8203 8205 8206

F9	..	..	..	..	..	..	..
cX55.7	..	..	..	..	..	..	..
4D-8	..	..	..	..	..	..	..
RNI	..	..	..	..	..	..	..
VK23B	..	..	..	..	..	..	..
VK21	01	12	02	01	02	02	02
IDS	02<	11<	01<	02	01	01	01
U6.2	..	..	..	..	..	..	..
St14	02<	33<	03<	03	02	02	02

Pedigree No. : 1340

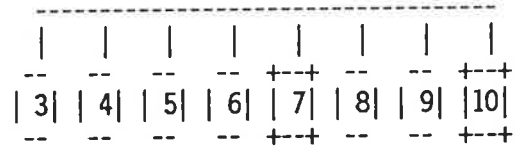
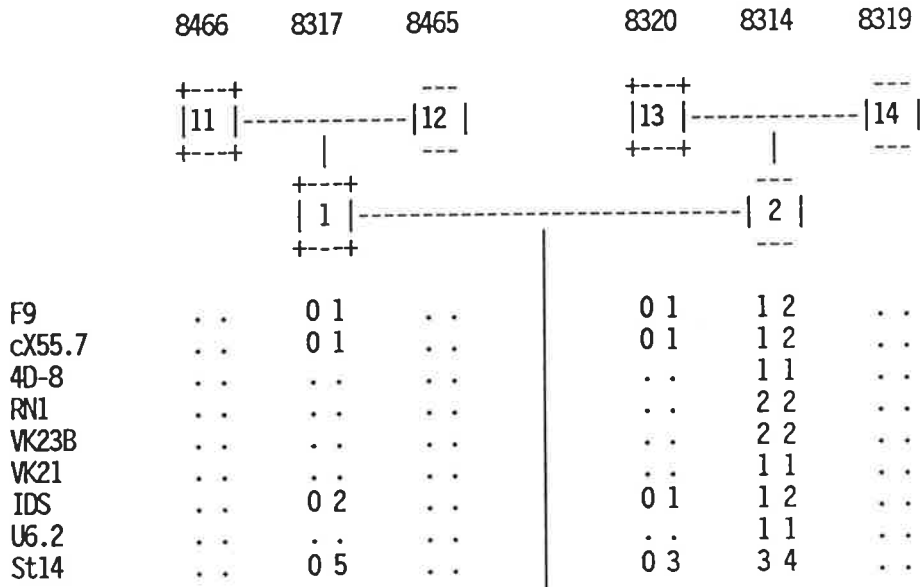


8283 8392 8209 8336 8305 8286 1371

F9	1 2	1 2	0 2	0 1	0 2	0 1	..
cX55.7	2 1	2 1	0 1	0 2	0 1	0 2	0 1
4D-8	1 1	1 1	0 1	0 2	0 1	0 2	0 1
RN1	..	..	..	..	..	..	..
VK23B	..	..	..	..	..	..	..
VK21	..	..	..	..	..	..	..
IDS	2 2	2 2	0 2	0 2	0 2	0 1	0 2
U6.2	1 1	1 1	0 1	0 1	0 1	0 2	0 1
St14	3 7	3 7	0 7	0 7	0 7	0 3	..



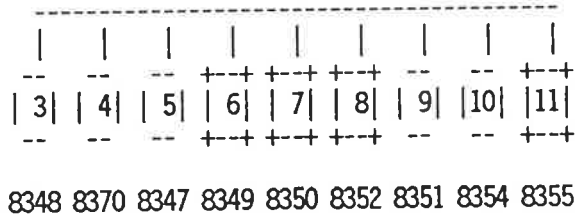
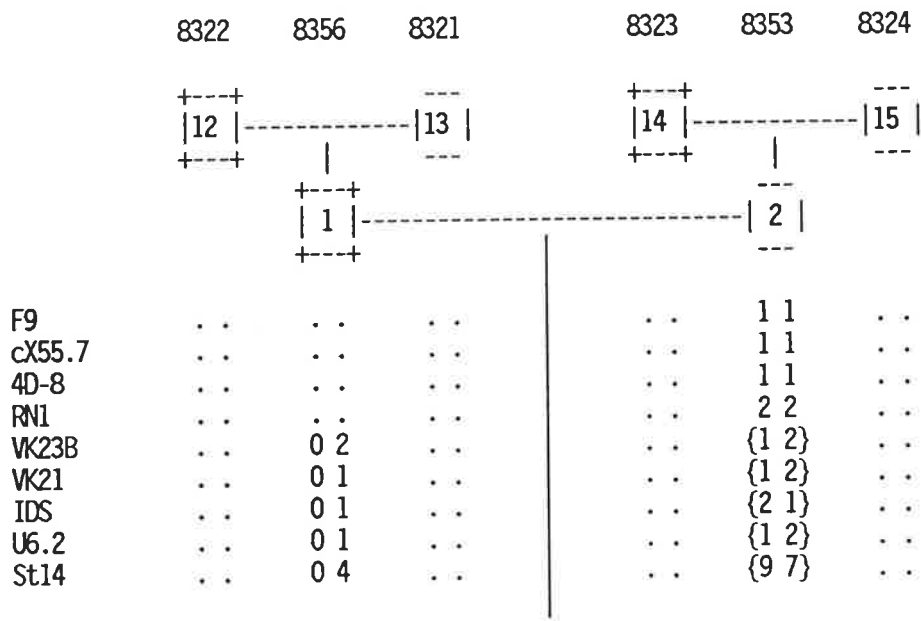
Pedigree No. : 1341



8346 8313 8311 8631 8318 8312 8315 8316

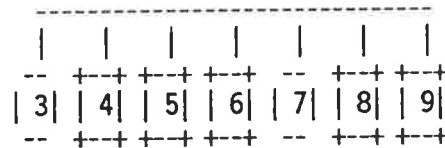
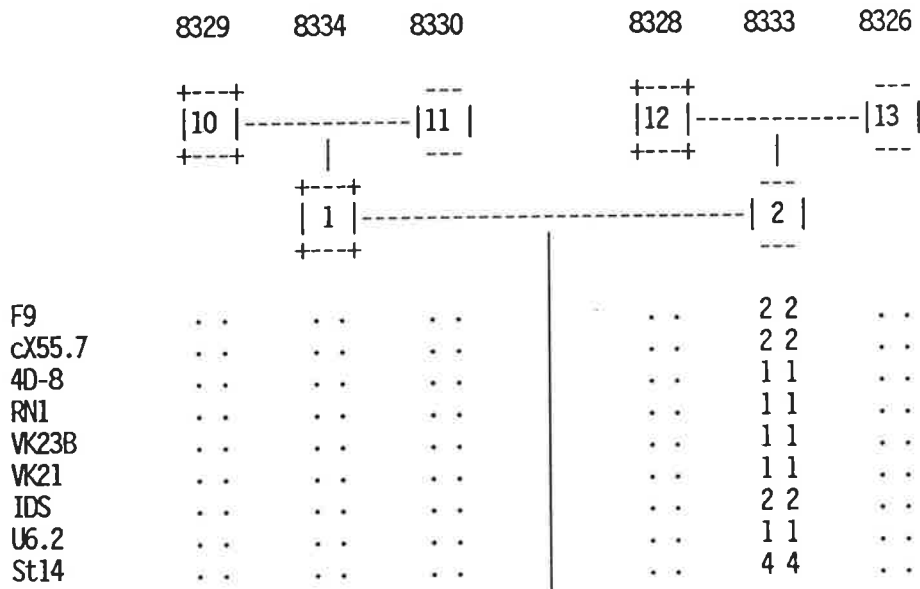
F9	1 2	1 2	1 1	1 0	0 2	1 1	1 2	0 1<
cX55.7	1 0	1 0	1 1	1 1<	0 2<	1 1	1 2	..
4D-8	..	..	..	..	..	..	..	..
RNI	..	..	..	..	..	..	..	..
VK23B	..	..	..	..	..	..	..	..
VK21	..	..	..	..	..	..	..	..
IDS	2 2	2 2	2 0	2 2<	0 1<	2 1<	2 2	0 2<
U6.2	..	..	..	..	..	..	..	..
St14	5 4	5 4	5 3	5 0	0 3	5 4<	5 4	0 4

Pedigree No. : 1344



F9	..	..	..	..	..	..	..	..	..	..
cX55.7	..	..	..	..	..	..	..	..	..	..
4D-8	..	..	..	..	..	..	..	..	..	..
RN1	..	..	..	..	..	..	..	..	..	..
VK23B	2 1	2 1	2 1<	0 2	..	..	2 1	2 2	0 2<	..
VK21	1 1	1 1	1 2<	0 2	..	0 1	1 1	1 2	0 1<	..
IDS	1 2	1 2	1 1	0 1	..	..	1 2	1 1	0 2	..
U6.2	1 1	1 1<	1 2	0 2	0 1	0 1	1 1	1 2<	..	..
St14	4 9	4 7<	4 7	..	0 7	0 9	4 9	4 9<	0 9	..

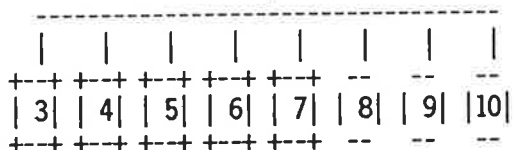
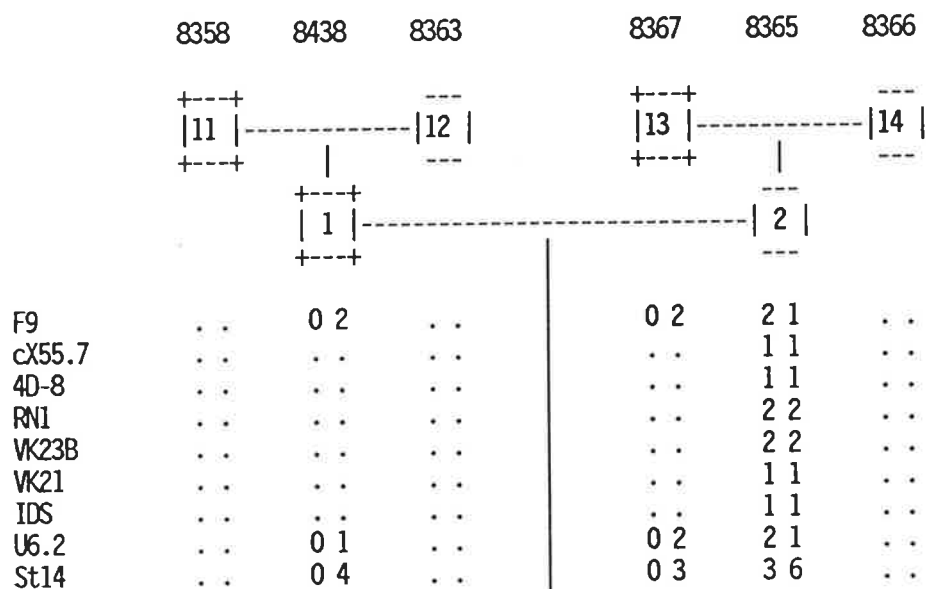
Pedigree No. : 1345



8335 8439 8332 8480 8325 8327 8331

F9	..	..	..	..	..	..	..
cX55.7	..	..	..	..	..	..	..
4D-8	..	..	..	..	..	..	..
RN1	..	..	..	..	..	..	..
VK23B	..	..	..	..	..	..	..
VK21	..	..	..	..	..	..	..
IDS	..	..	..	..	..	..	..
U6.2	..	..	..	..	..	..	..
St14	..	..	..	..	..	..	..

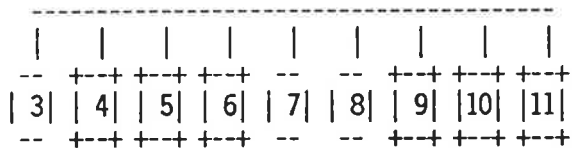
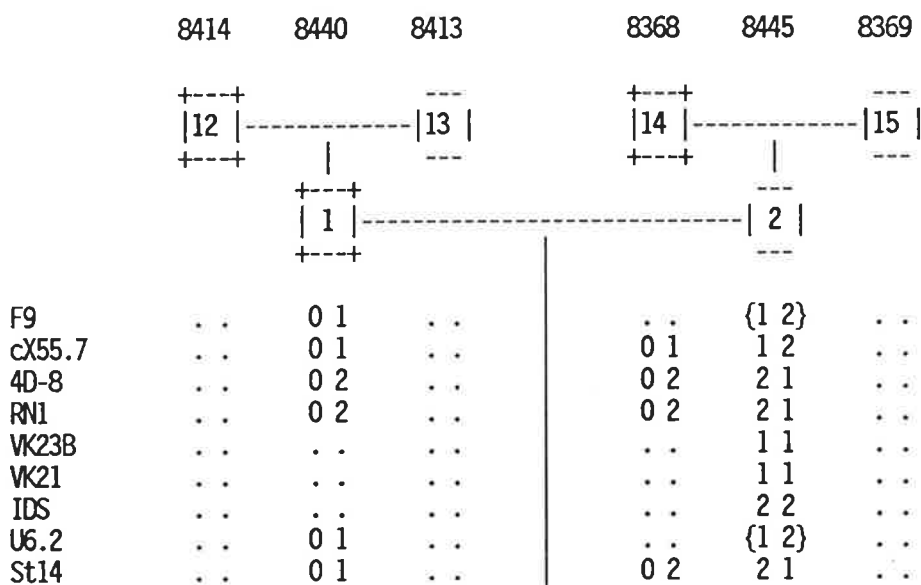
Pedigree No. : 1346



8362 8357 8603 8359 1370 8364 8360 8361

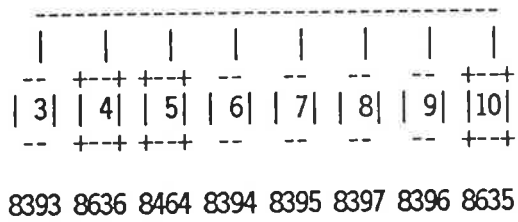
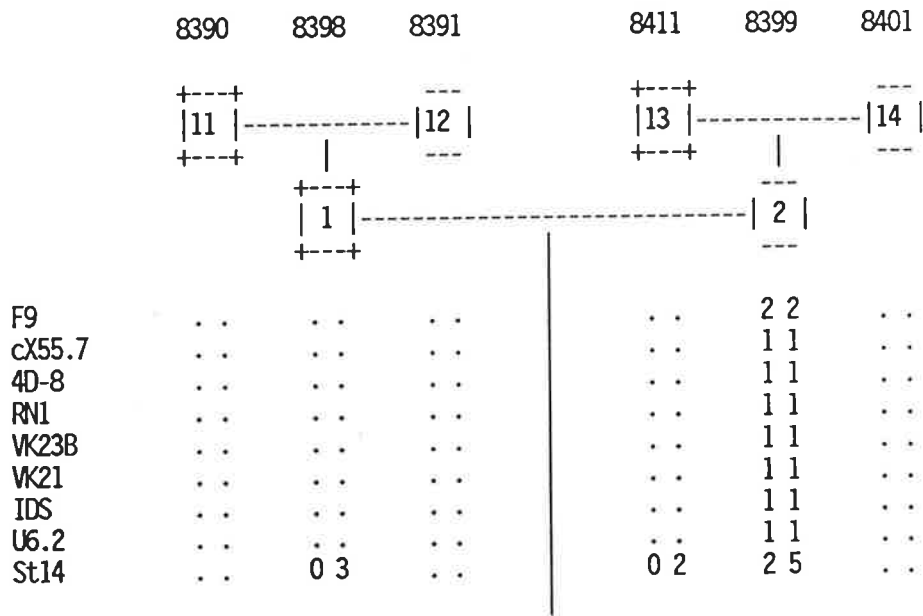
F9	0 2	0 1	< 0 1	0 2	< 0 1	2 2	2 2	..
cX55.7	..	..	..	..	..	..	..	..
4D-8	..	..	..	..	..	..	..	..
RNI	..	..	..	..	..	..	..	..
VK23B	..	..	..	..	..	..	..	..
VK21	..	..	..	..	..	..	..	..
IDS	..	..	..	..	..	..	..	..
U6.2	0 2	0 2	< 0 1	0 1	< 0 2	1 2	1 2	< 1 2
St14	0 3	0 3	0 6	0 6	0 3	4 3	4 6	< ..

Pedigree No. : 1347



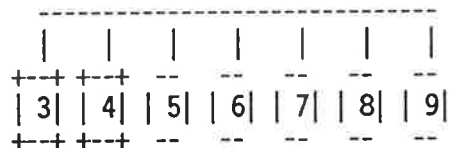
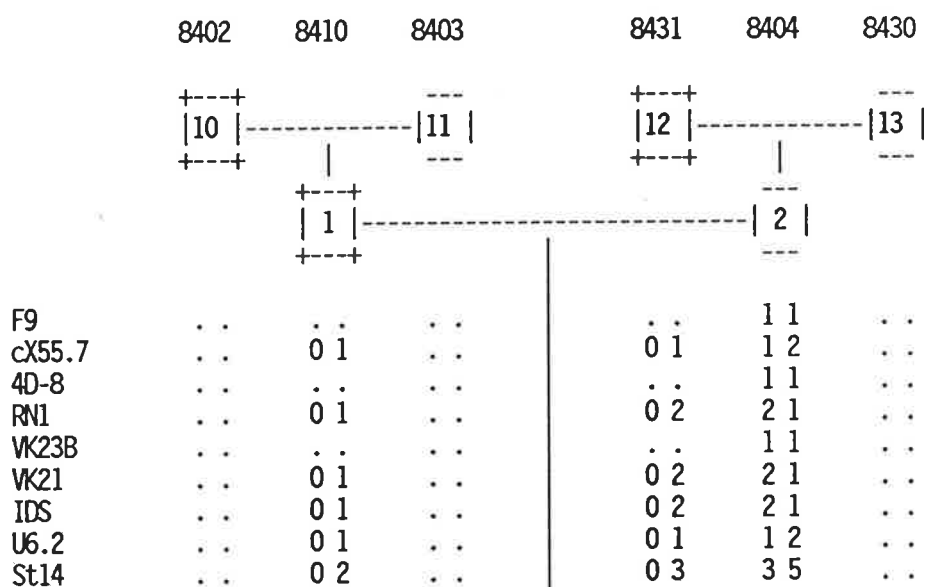
	8412	8436	8437	8442	8435	8441	8444	8443	8434
F9	1 1	0 1	0 2	0 2	1 2	1 2	0 1	..	0 1
cX55.7	1 1	0 1	0 2	0 2	1 2	1 2	0 1	0 2	0 1
4D-8	2 2	0 2	0 1	0 1	2 1	2 1	0 2	0 1	0 2
RN1	2 2	0 2	0 1	0 1	2 1	2 1	0 2	0 1	0 2
VK23B	..	..	..	..	..	..	..	..	..
VK21	..	..	..	..	..	..	..	..	..
IDS	..	..	..	..	..	..	..	..	..
U6.2	1 2	0 2	0 1	0 1	1 1	1 1	0 2<	..	0 2
St14	1 2	0 2	0 1	0 1	1 1	1 1	0 1<	0 1	0 2

Pedigree No. : 1349



	8393	8636	8464	8394	8395	8397	8396	8635
F9	..	..	..	..	..	..	..	..
cX55.7	..	..	..	..	..	..	..	..
4D-8	..	..	..	..	..	..	..	..
RN1	..	..	..	..	..	..	..	..
VK23B	..	..	..	..	..	..	..	..
VK21	..	..	..	..	..	..	..	..
IDS	..	..	..	..	..	..	..	..
U6.2	..	..	..	..	..	..	..	..
St14	3 2	0 5	0 5	3 2	3 2	3 5	3 2	0 5

Pedigree No. : 1350



8415 8408 8409 8407 8406 8405 8792

	8415	8408	8409	8407	8406	8405	8792
F9	..	..	..	..	..	..	..
cX55.7	0 2	0 1	1 1	1 1	1 1	1 1	1 2
4D-8	..	..	..	..	..	..	..
RN1	0 1	0 2	1 1	1 1	1 2	1 2	1 1
VK23B	..	..	..	..	..	..	..
VK21	0 1	0 2	1 1	1 1	1 2	1 0	1 1
IDS	0 1	0 2	1 1	1 1	1 2	1 2	1 1
U6.2	0 2	0 1	1 0	1 0	1 1	1 1	1 2
St14	0 3	0 3	2 5	2 5	2 3	2 0	2 5

Appendix B. CEPH Pedigrees. p. B-21

Pedigree No. : 1362

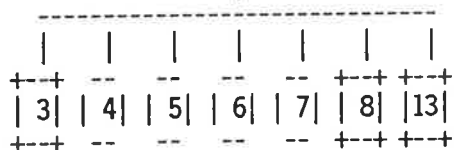
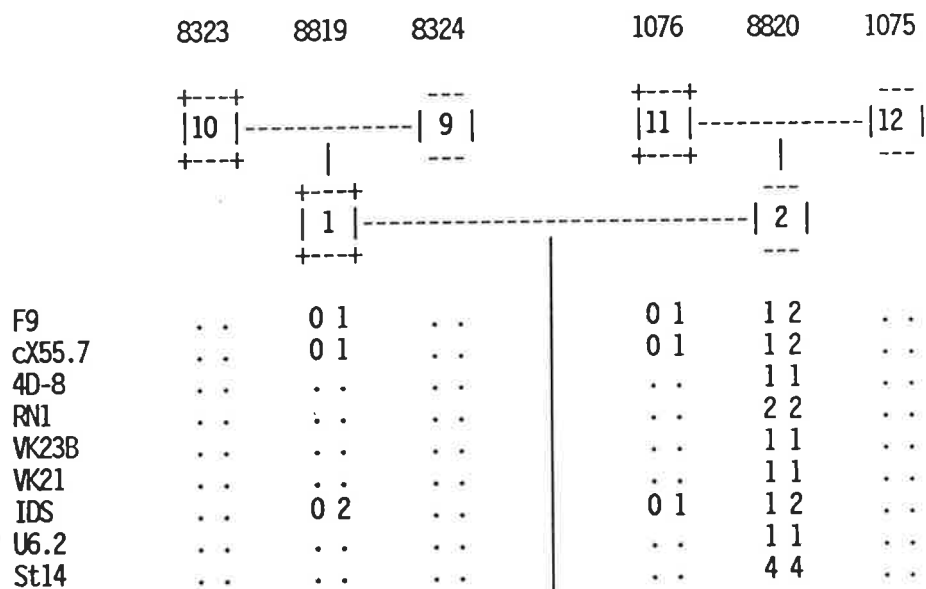
	8573	8563	8572	8575	8574	8589
	+	+	+	+	+	+
	13		14	15		16
	+	+	+	+	+	+
		+			+	
		1			2	
		+			+	
F9	..	..	..	..	1 1	..
cX55.7	..	..	..	..	1 1	..
4D-8	..	..	..	..	1 1	..
RN1	..	0 1	..	0 2	2 1	..
VK23B	..	..	..	..	1 1	..
VK21	..	0 1	..	0 1	1 2	..
IDS	..	0 2	..	0 2	2 1	..
U6.2	..	0 1	..	0 1	1 2	..
St14	..	0 3	..	0 6	6 5	..

	8568	8566	8564	8594	8215	8565	8567	8569	8570	8571	8929
	+	+	+	+	+	+	+	+	+	+	+
	3	4	5	6	7	8	9	10	11	12	17
	+	+	+	+	+	+	+	+	+	+	+
F9	..	..	..	..	..	..	..	..	..	..	..
cX55.7	..	..	..	..	..	..	..	..	..	..	..
4D-8	..	..	..	..	..	..	..	..	..	..	..
RN1	1 0	1 0	..	1 1	1 2	0 2	1 1	1 2	0 2	1 1	..
VK23B	..	..	..	..	..	..	..	..	..	..	..
VK21	1 2	1 0	0 2	1 0	1 1	0 1	1 2	1 1	0 2	1 2	0 2
IDS	2 1	2 1	0 1	2 1	2 2	0 2	2 1	2 2	0 1	2 1	0 1
U6.2	1 2	1 2	0 2	1 2	1 1	0 1	1 2	1 1	0 2	1 2	0 2
St14	3 5	..	..	3 5	3 6	0 6	3 5	3 6	0 5	3 5	0 5



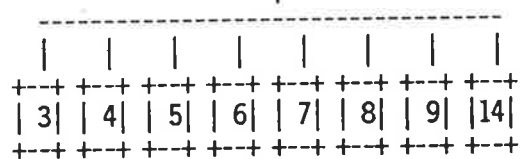
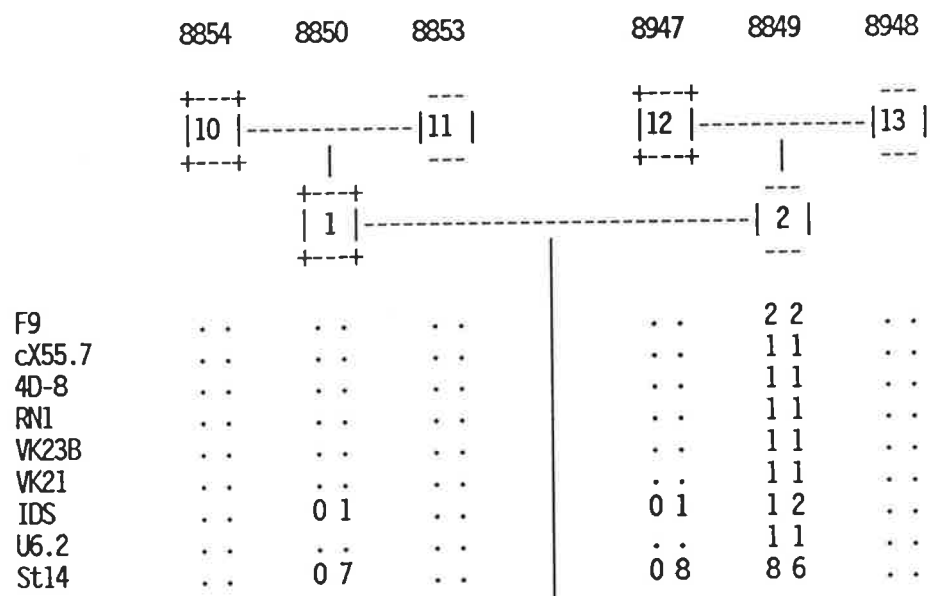
Pedigree No. : 1375



8796 8797 8800 8798 8799 8818 1441

F9	0 1	1 2	1 2	1 1	1 2	0 2	0 1
cX55.7	0 1	1 2	1 2	1 1	1 2	0 2	0 1
4D-8	..	..	..	..	..	..	..
RN1	..	..	..	..	..	..	..
VK23B	..	..	..	..	..	..	..
VK21	..	..	..	..	..	..	..
IDS	0 1	2 2	2 2	2 1	2 2	0 2	0 1
U6.2	..	..	..	..	..	..	..
St14	..	..	..	..	..	..	..

Pedigree No. : 1377

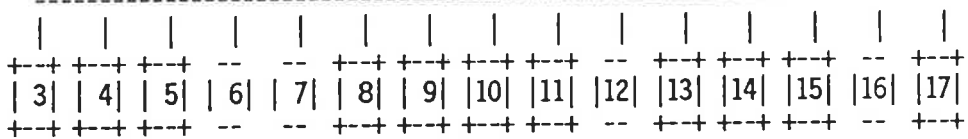
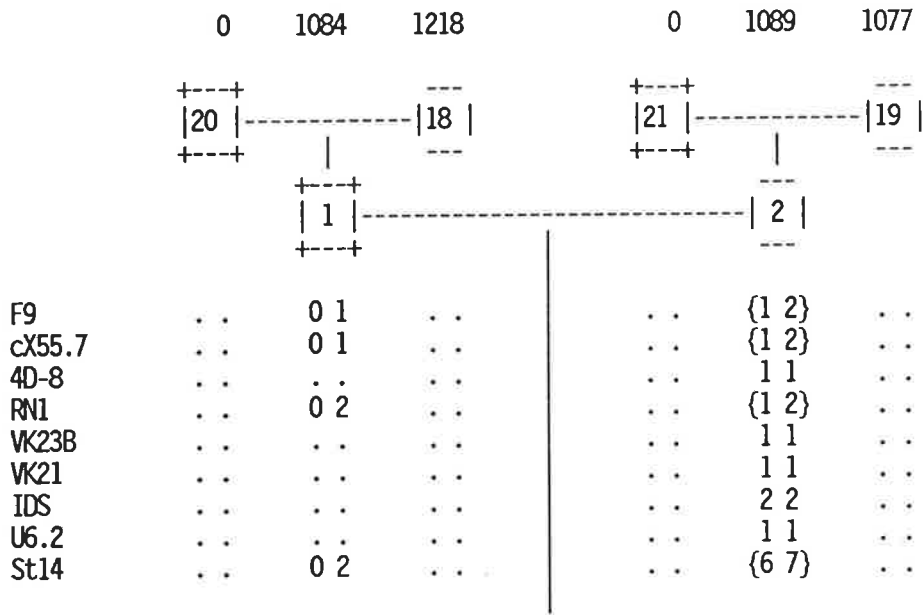


8847 8824 8823 8825 8846 8826 8848 1702

F9	..	..	..	..	..	..	..	..
cX55.7	..	..	..	..	..	..	..	..
4D-8	..	..	..	..	..	..	..	..
RN1	..	..	..	..	..	..	..	..
VK23B	..	..	..	..	..	..	..	..
VK21	..	..	..	..	..	..	..	..
IDS	0 1	0 2	0 1	0 1	0 2	0 2	0 2	0 2
U6.2	0 8	0 6	0 8	0 8	0 6	0 6	0 6	0 6
St14	0 8	0 6	0 8	0 8	0 6	0 6	0 6	0 6



Pedigree No. : 1413



1085 1083 1082 1092 1280 1091 1279 1081 1078 1079 1080 1086 1087 1088 1090

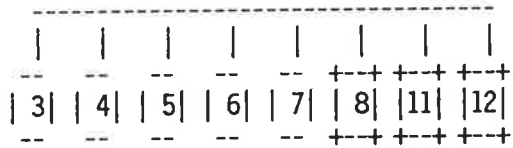
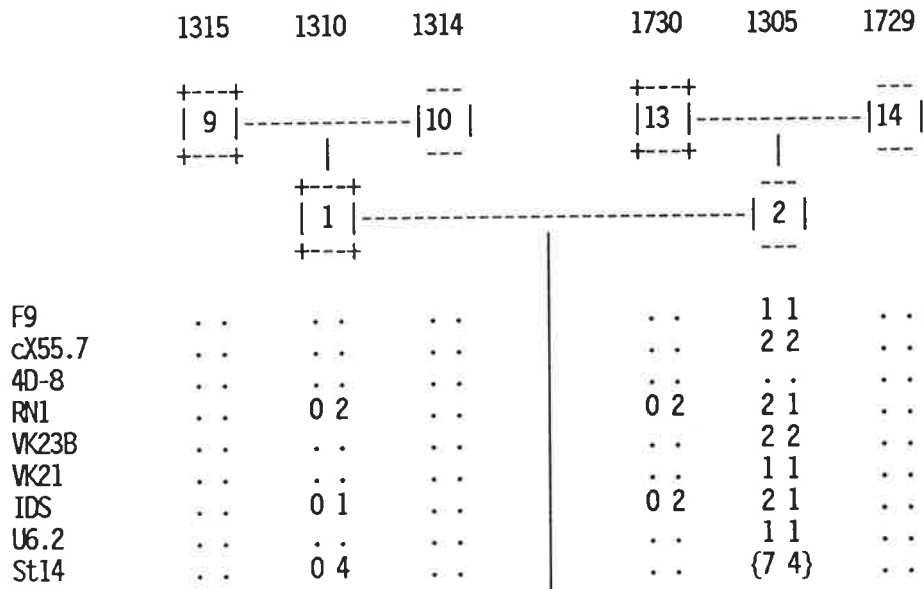
F9	0 1	0 1	0 2	1 1	1 0	0 2	0 2	0 2	0 2	0 2	< 1 1	0 1	0 1	0 2	1 2	0 2
CX55	0 1	0 1	0 1	< 1 1	1 0	0 2	0 2	0 2	0 1	< 1 1	0 1	0 1	0 2	< 1 2	0 2	
4D-8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
RN1	0 2	0 1	0 2	< 2 2	2 0	0 1	0 1	0 1	0 2	2 2	< 0 2	0 2	0 2	< 2 1	0 1	
VK23	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	
VK21	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	
IDS	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	
U6.2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
St14	0 6	0 7	0 6	2 6	2 0	0 7	0 7	0 7	0 6	2 7	< 0 6	0 6	0 6	0 6	2 7	0 7







Pedigree No. : 1421

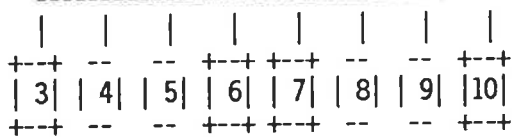
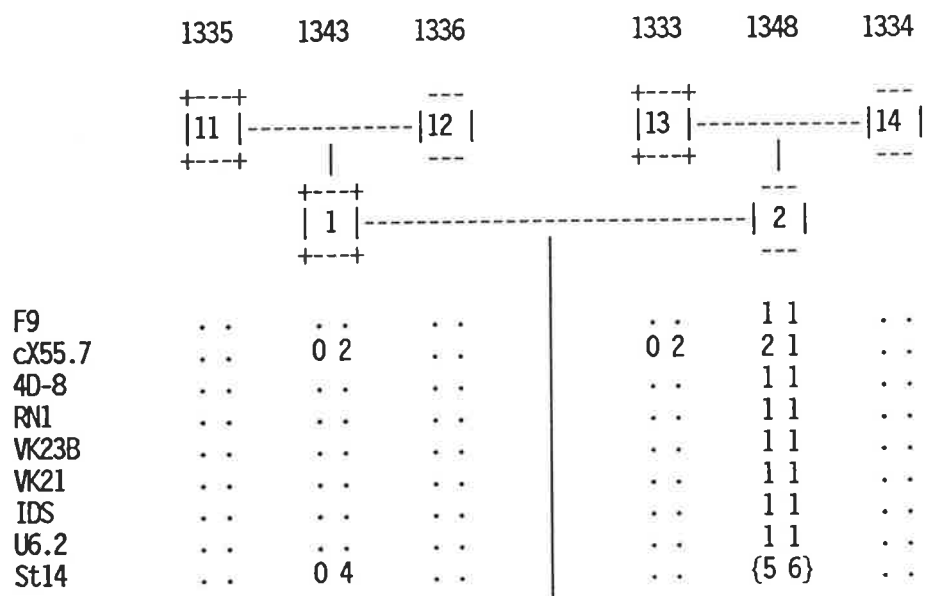


1309 1326 1308 1306 1307 1304 1560 1543

F9	..	..	..	..	..	..	..	..
cX55.7	..	..	..	..	..	..	..	..
4D-8	..	..	..	..	..	..	..	..
RN1	2 0	2 0	2 1	2 2	2 2	0 1	0 2	0 2<
VK23B	..	..	..	..	..	..	..	..
VK21	..	..	..	..	..	..	..	..
IDS	1 1	1 1	1 1	1 2	1 2	0 1	0 2	0 1<
U6.2	. .	. .	. .	. .	. .	. .	. .	. .
St14	4 4	4 4	4 4	4 7	4 7	0 4	0 7	0 4



Pedigree No. : 1423

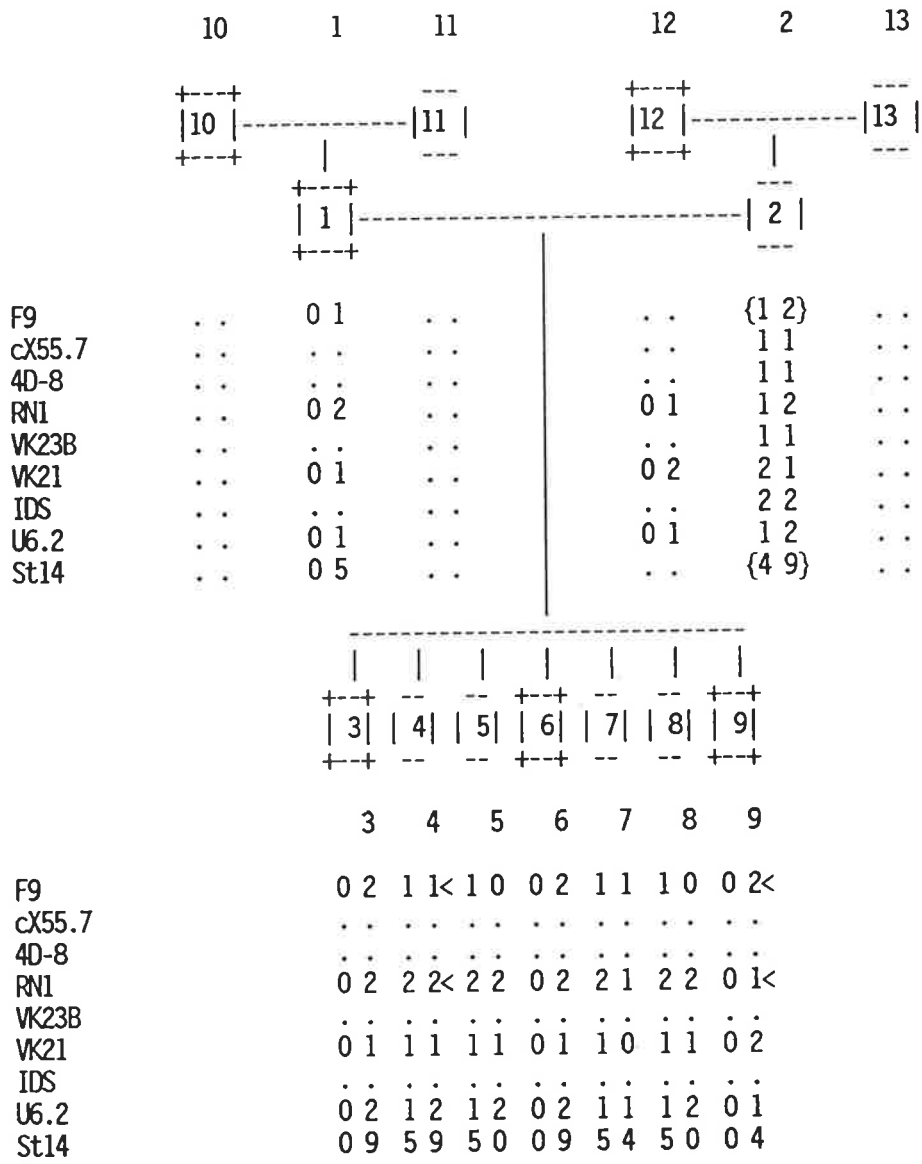


1347 1344 1342 1345 1339 1341 1340 1346

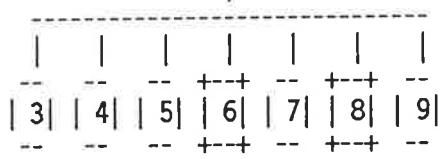
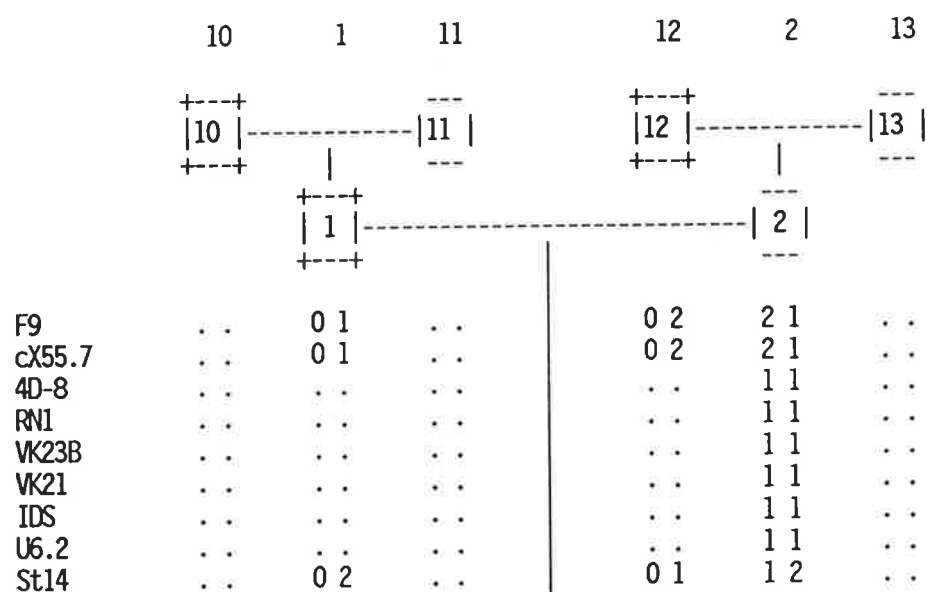
	1347	1344	1342	1345	1339	1341	1340	1346
F9	..	..	..	..	..	..	..	..
cX55.7	0 1	2 1	2 1	0 1	0 2	2 2	2 1	0 1
4D-8	..	..	..	..	..	..	..	..
RN1	..	..	..	..	..	..	..	..
VK23B	..	..	..	..	..	..	..	..
VK21	..	..	..	..	..	..	..	..
IDS	..	..	..	..	..	..	..	..
U6.2	..	..	..	..	..	..	..	..
St14	0 6	4 6	4 6	0 5	0 6	4 5	4 6	0 6



Pedigree No. : 66



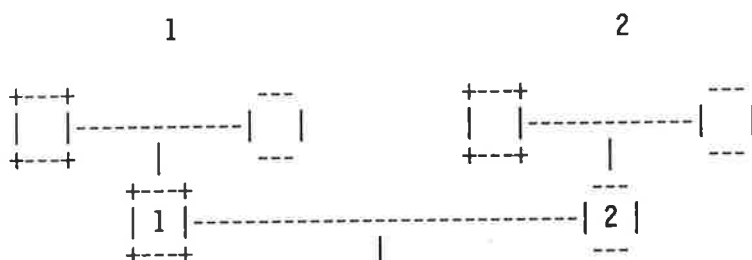
Pedigree No. : 12



3 4 5 6 7 8 9

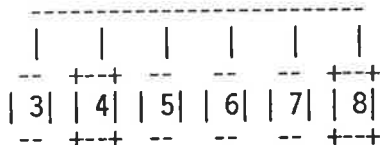
F9	1 2	1 1	1 1	0 2	1 2	..	1 1
cX55.7	1 2	1 1	1 1	0 2	1 2	0 1	1 1
4D-8	..	..	..	..	..	..	..
RN1	..	..	..	..	..	..	..
VK23B	..	..	..	..	..	..	..
VK21	..	..	..	..	..	..	..
IDS	..	..	..	..	..	..	..
U6.2	..	..	..	..	..	..	..
St14	2 1	2 2	2 1	0 1	2 1	..	2 2

Pedigree No. : 23



F9  
cX55.7  
4D-8  
RNI  
VK23B  
VK21  
IDS  
U6.2  
St14

0 1	(1 2)
..	1 1
0 1	(1 2)
..	..
..	1 1
..	1 1
0 2	(1 2)
..	1 1
..	4 4

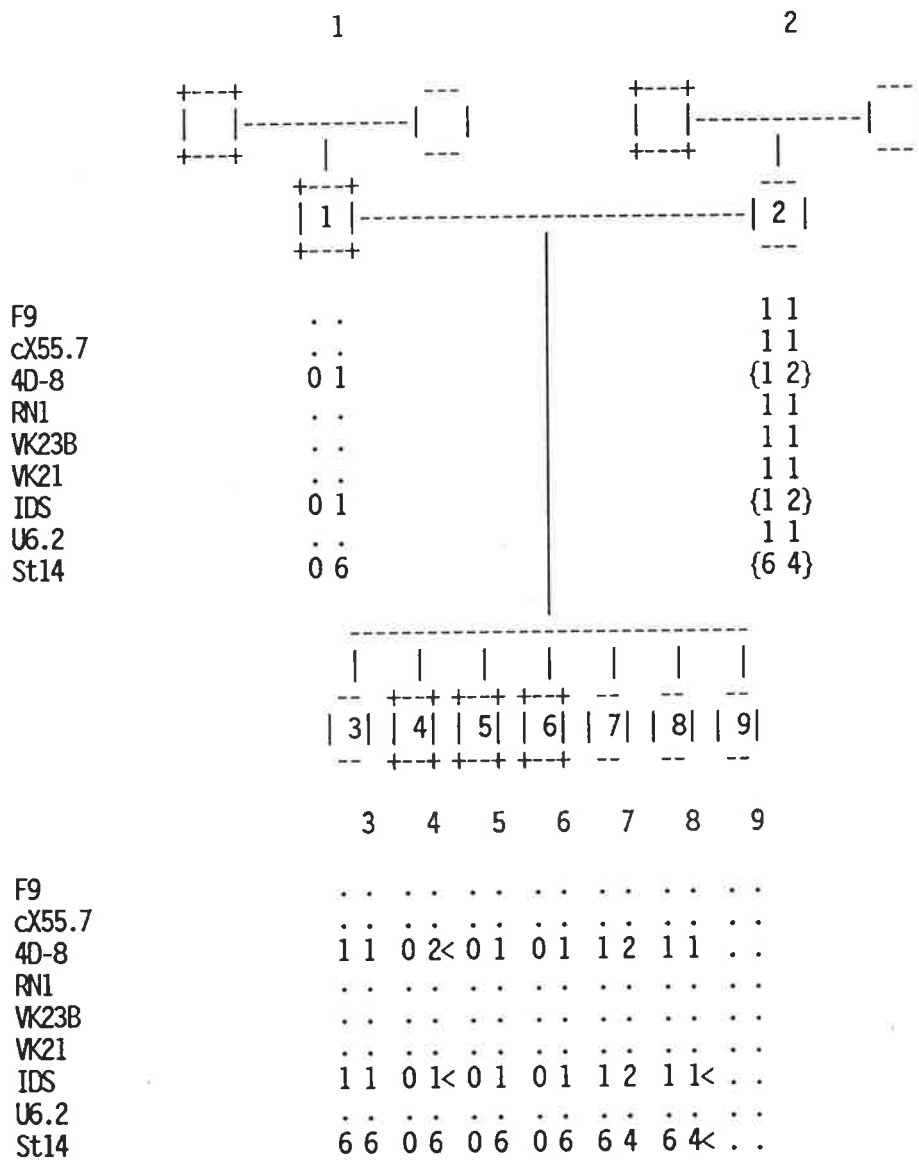


3 4 5 6 7 8

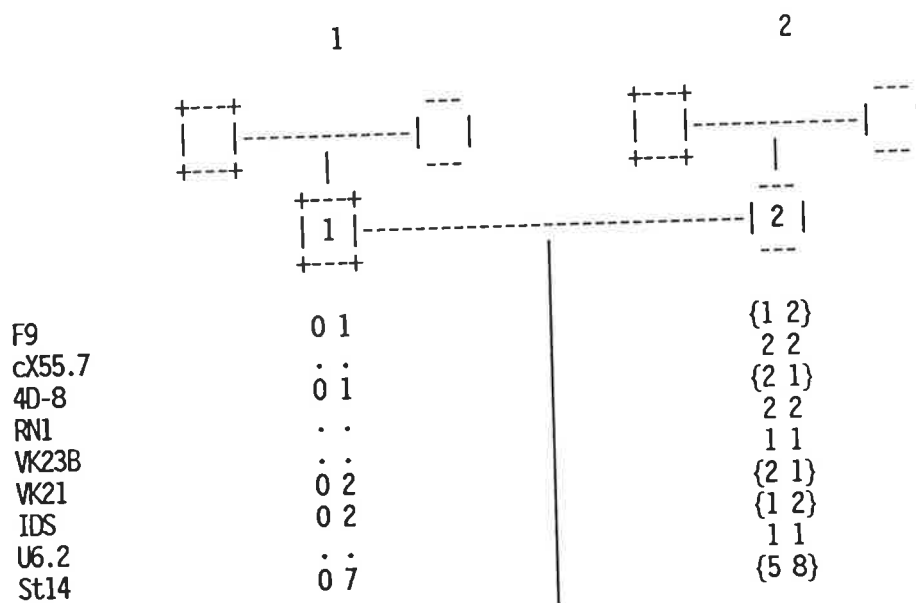
F9  
cX55.7  
4D-8  
RNI  
VK23B  
VK21  
IDS  
U6.2  
St14

1 1 < 0 2	1 1	1 1	1 1	1 1	0 1
..	..	..	..	..	..
1 2 < 0 2	1 1 <	1 1	1 1	0 1	
..	..	..	..	..	..
..	..	..	..	..	..
2 2 0 2	2 2 <	2 1	2 1	0 1	
..	..	..	..	..	..
..	..	..	..	..	..

Pedigree No. : 21

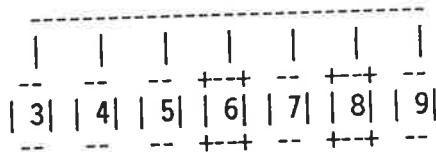


Pedigree No. : 2



F9  
 cX55.7  
 4D-8  
 RN1  
 VK23B  
 VK21  
 IDS  
 U6.2  
 St14

0 1	{ 1 2 }
· ·	2 2
0 1	{ 2 1 }
· ·	2 2
· ·	1 1
· ·	{ 2 1 }
0 2	{ 1 2 }
0 2	1 1
· ·	{ 5 8 }
0 7	



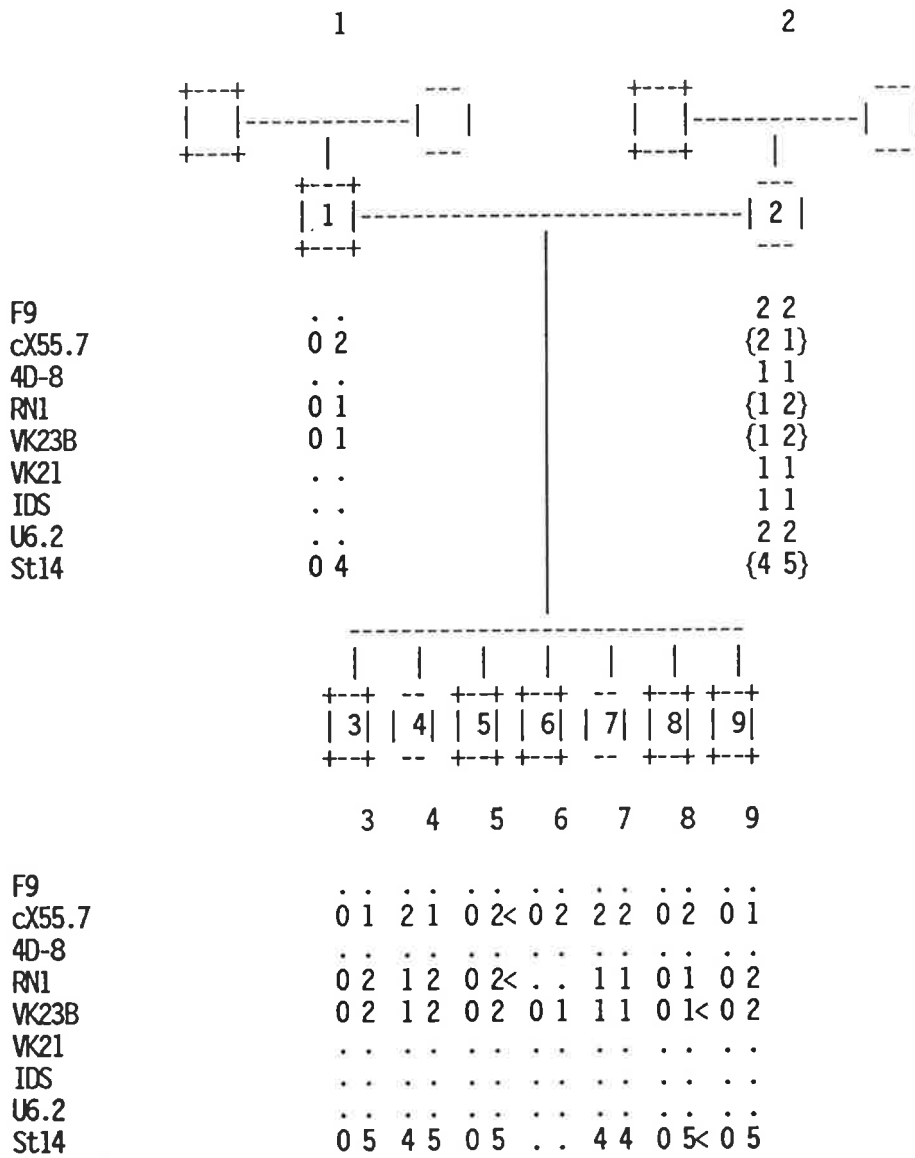
	3	4	5	6	7	8	9
F9	1 0	1 0	1 1	0 2	1 1	0 2	1 2
cX55.7	· ·	· ·	· ·	· ·	· ·	· ·	· ·
4D-8	1 1	1 1	1 2	0 1	1 2	0 1	1 1
RN1	· ·	· ·	· ·	· ·	· ·	· ·	· ·
VK23B	· ·	· ·	· ·	· ·	· ·	· ·	· ·
VK21	2 1	2 2	2 2	0 1	2 2	0 1	2 1
IDS	2 2	2 1	2 1	0 2	2 1	0 2	2 2
U6.2	· ·	· ·	· ·	· ·	· ·	· ·	· ·
St14	7 8	7 5	7 5	0 8	7 5	0 8	7 5

F9  
 cX55.7  
 4D-8  
 RN1  
 VK23B  
 VK21  
 IDS  
 U6.2  
 St14



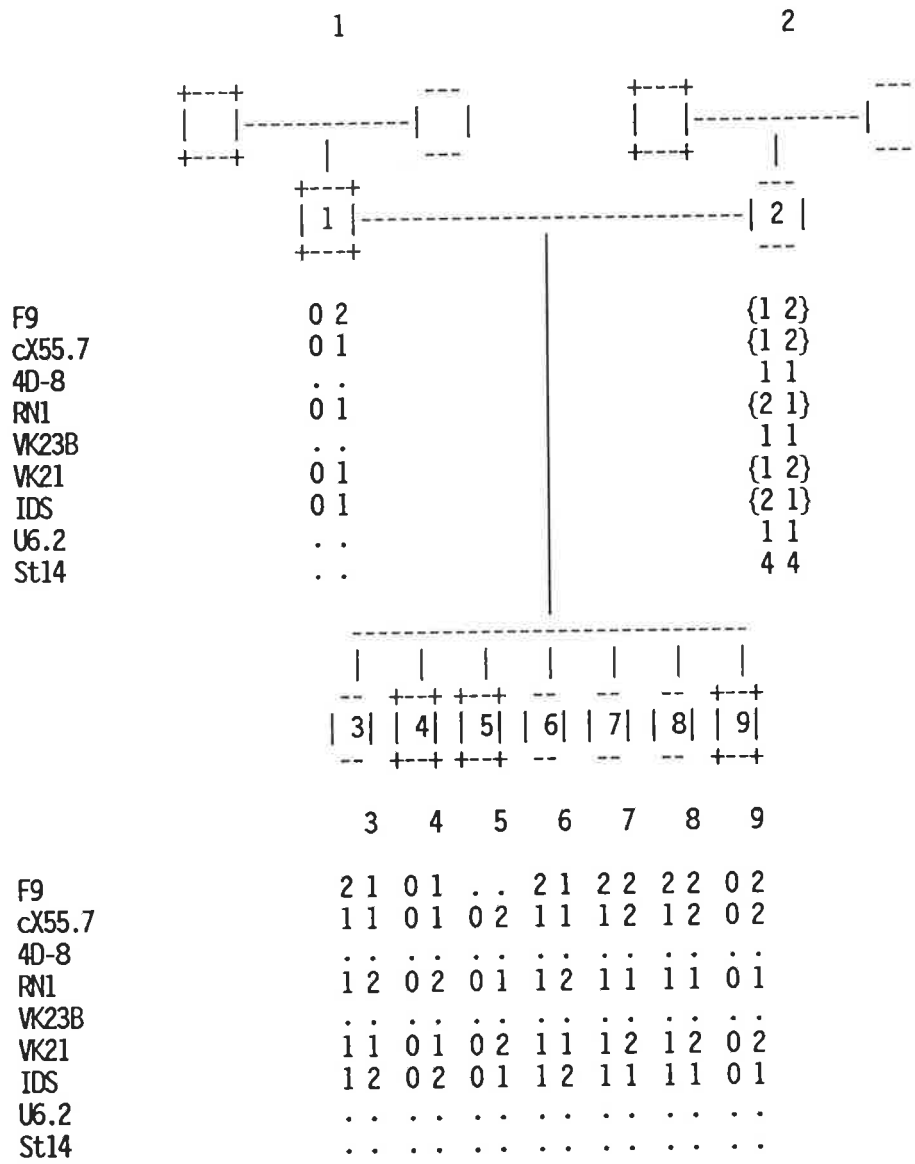


Pedigree No. : 37

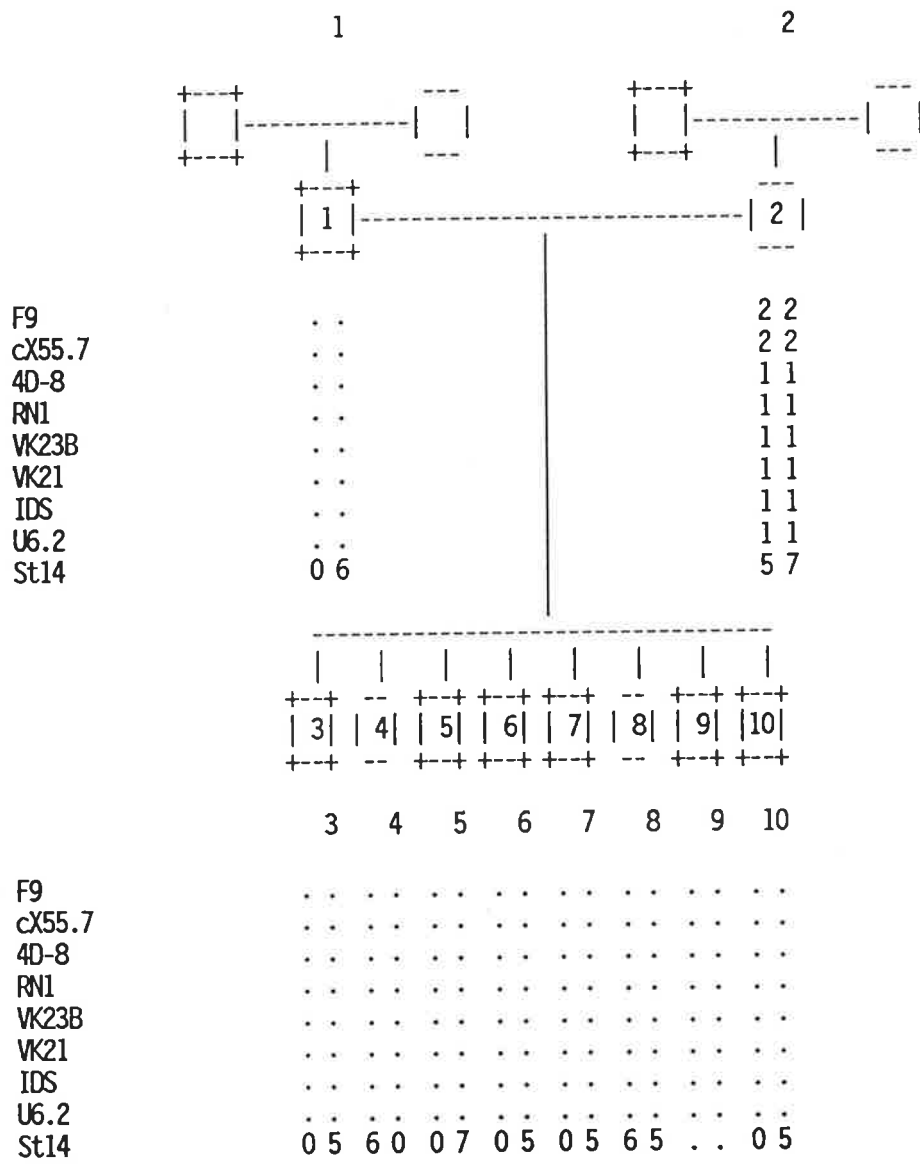




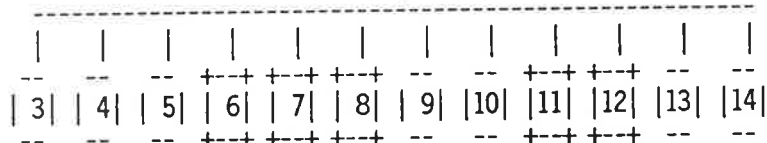
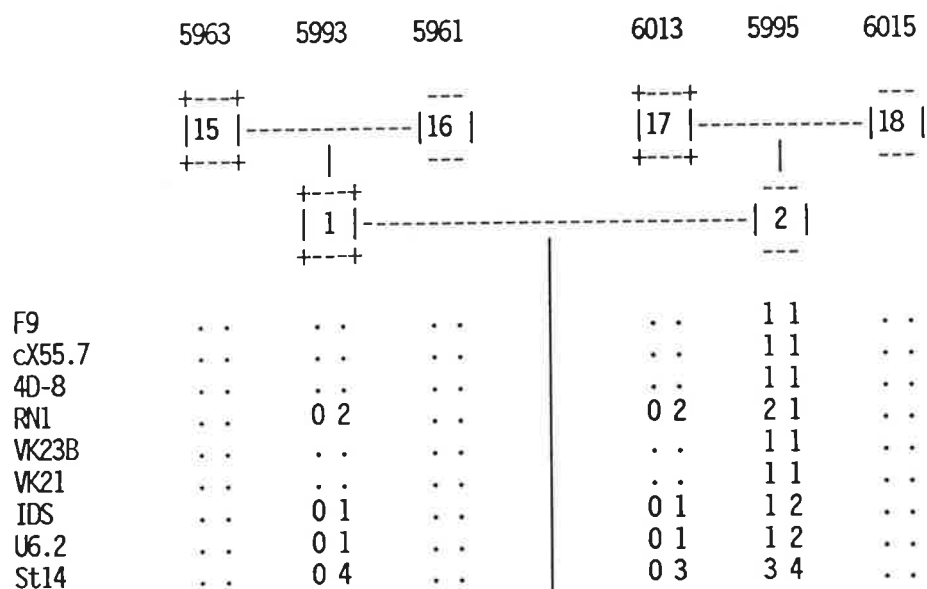
Pedigree No. : 28



Pedigree No. : 45



Pedigree No. : 884



5987 5901 5918 5989 5907 5914 5999 5991 6001 5997 6003 6005

	5987	5901	5918	5989	5907	5914	5999	5991	6001	5997	6003	6005
F9	..	..	..	..	..	..	..	..	..	..	..	..
cX55.7	..	..	..	..	..	..	..	..	..	..	..	..
4D-8	..	..	..	..	..	..	..	..	..	..	..	..
RN1	2 2	2 2	2 1	0 2	0 2	0 2	2 1	2 1	0 2	0 1	2 1	2 2<
VK23B	..	..	..	..	..	..	..	..	..	..	..	..
VK21	..	..	..	..	..	..	..	..	..	..	..	..
IDS	1 1	1 0	1 0	0 1	0 1	0 1	1 2	1 0	0 1	0 2	1 0	1 0
U6.2	1 1	1 1	1 2	0 1<	0 1	0 1	1 2	1 2	0 1	0 2<	1 2	1 2<
St14	4 3	4 3	4 4	0 4<	0 3	0 3	4 4	4 4	0 3	0 3<	4 4	4 4





**Appendix C**

**GENOTYPES IN THE FRAGILE X PEDIGREES.**



*Appendix C. Fragile X pedigrees. p. C-2*

This Appendix lists the genotypes at nine polymorphic loci at Xq26-q28 in 112 fragile X families. The analysis of these data is presented in Chapters 6 and 7.

The nine loci (and probes) listed are:

<u>F9</u>	pVIII
<u>DXS105</u>	cX55.7
<u>DXS98</u>	4D-8
<u>DXS369</u>	RN1
<u>DXS297</u>	VK23B
<u>DXS296</u>	VK21A and VK21C
<u>IDS</u>	pc2S15
<u>DXS304</u>	U6.2
<u>DXS52</u>	St14-1

Details of which RFLP was used at each locus are given in Chapter 6 (Table 6-A).

These families were contributed by 13 different centers around the world. Details of the source of the families are presented in the Table below. Analysis of the pedigree and genotype data of the Adelaide families has been published or submitted for publication, and these data are in the public domain. Permission to reproduce or utilize data of other families should be sought from the collaborators listed in the Table.

The selection criteria for the families are presented in Chapters 2 and 6. The data of each family were checked by hand and then entered in a computerized database. A data file containing all the pedigree and genotype data was then modified to remove redundant data (i.e. if all the women in a family were uninformative at a locus). For linkage analysis the genotypes had been entered as binary-factor systems (Chapter 2). These systems were changed to allele numbers for this listing as allele numbers are easier to read. The pedigrees were then plotted using the program TEXTPED (kindly provided by Dr M Badzioch, Houston). Note that some pedigrees spread over a number of pages (indicated by an arrow). If an individual had more than one spouse, each branch of the family is listed separately. Pedigree 208 had remote inbreeding; the inbreeding loop was 'broken' by duplicating one individual (id=1).

For each pedigree, the pedigree number is the reference number in the database. The 'id' numbers are database numbers for each individual. The two columns under each female list the alleles at each locus; these alleles are NOT shown in phase (cf. Appendix B). For males, the two columns of numbers are identical and indicate the allele at each locus. Where there is no data at a locus, the alleles have been shown as

*Appendix C. Fragile X pedigrees. p. C-4*

Individuals are shown as hemizygous or heterozygous for FRAXA if they expressed the fragile X or if they were obligate carriers on the basis of pedigree information. Note that some males were known to be transmitting males on the basis of pedigree information that is not shown below. These pedigree listings show the minimum data necessary for linkage analysis.

Recombination events are not identified on the pedigrees. The pedigrees in which recombination occurred between FRAXA and one or more loci are indicated in the Table.

Collaborative linkage study of the fragile X syndrome.

May 1990.

Coordinated by Graeme Suthers, Adelaide.

List of collaborators and pedigrees.

ADELAIDE

Dr John Mulley  
 Department of Cytogenetics & Molecular Genetics  
 Adelaide Children's Hospital  
 SA 5006 AUSTRALIA  
 tel 08-267 7333; fax 08-267 7342

Pedigree #.	Laboratory ref.	<sup>1</sup> Recombination with <u>FRAXA</u> <sup>a</sup>
3	same as ped #	F9 IDS
4		
6		
7		
8		
9		F9
10		
11		
13		
15		
16		
19		IDS ST14
21		
22		F9
23		F9 IDS
24		
26		VK23B ST14
28		F9
29		55.7 ST14
30		
31		
33		
34		
35		
36		

25 pedigrees

a VK23B (DXS297) and pc2S15 (IDS) were used in the Adelaide families only.

<sup>1</sup> For this table, recombination between a locus and FRAXA was defined as a two-point LOD score of < -1 at a recombination fraction of zero between the locus and FRAXA.

ULM

Dr Peter Steinbach  
 Universitat Ulm  
 Frauenstrasse 29  
 7900 Ulm (Donau)  
 Federal Republic of Germany  
 tel 0731-178 221; fax 49-731 69505

Pedigree #.	Laboratory ref.	Recombination with <u>FRAXA</u>
50	BU	
51	SCH	
52	KE	55.7 ST14
53	MON	
55	KOB	
56	GAR	
57	FUR	

7 pedigrees

BARCELONA

Dr Miguel Carballo  
 Genetica Molecular  
 Jorge Girona Salgado 18-26  
 08034 Barcelona  
 Spain  
 tel 93-204 0600; fax 3-204 5904

Pedigree #.	Laboratory ref.	Recombination with <u>FRAXA</u>
61	I	55.7 VK21 ST14
62	II	55.7 4D8

2 pedigrees

ZURICH

Prof. Dr Albert Schinzel  
 Institut fur Medizinische Genetik  
 Ramistrasse 74  
 (bei Tramhaltestelle Kantonsschule)  
 8001 Zurich  
 Switzerland  
 tel 01-257 2521

Pedigree #.	Laboratory ref.	Recombination with <u>FRAXA</u>
72	Fam 6	55.7
75	Fam 9	
78	Fam 13	

3 pedigrees

**ABERDEEN**

Dr Neva Haites  
Medical Genetics  
Medical School Buildings  
Foresterhill  
Aberdeen AB9 2ZD  
Scotland  
tel 0224-68 1818 ext 52120; fax 0224-68 5157

Pedigree #.	Laboratory ref.	Recombination with <u>FRAXA</u>
81	SM	
82	BR	ST14
83	BO	
84	AD	
85	WR	

5 pedigrees

**ROTTERDAM**

Dr Ben Oostra  
Clinical Genetics Department  
Erasmus University  
PO Box 1738  
3000 DR Rotterdam  
The Netherlands  
tel 010-408 7214; fax 10-408 7200

Pedigree #.	Laboratory ref.	Recombination with <u>FRAXA</u>
92	II	
93	III	ST14
94	IV	

3 pedigrees

**OULU**

Dr Marja-Leena Vaisanen  
 Oulu University Central Hospital  
 Department of Clinical Genetics  
 Kajaanintie 50  
 SF-90220 OULU  
 Finland  
 tel 981-33 2033

Pedigree #.	Laboratory ref.	Recombination with <u>FRAXA</u>
103	03	
108	08	55.7 U6.2
110	10	ST14
113	13	F9
115	15	
117	17	55.7
119	19	
123	23	F9 55.7
125	25	F9 55.7
127	27	
131	31	
132	32	

12 pedigrees

**BIRMINGHAM**

Dr Ian Glass  
 Clinical Genetics Unit  
 Birmingham Maternity Hospital  
 Edgbaston  
 Birmingham B15 2TG  
 England  
 tel 021-472 5199; fax 21-471 5017

Pedigree #.	Laboratory ref.	Recombination with <u>FRAXA</u>
140	2231	ST14
141	4669	
142	3475	55.7 ST14
143	8230	
144	3940	
145	Nottingham	
146	Birmingham	

7 pedigrees

**ROCHESTER**

Dr Stephen Thibodeau  
Molecular Genetics  
Mayo Clinic  
Rochester  
Minnesota 55905  
USA  
tel 507-284 2511; fax 507-284 0043

Pedigree #.	Laboratory ref.	Recombination with <u>FRAXA</u>
151	5	F9
152	18	F9 ST14
153	23	
155	64	
156	71	
157	86	F9 4D8

6 pedigrees

**NIJMEGEN**

Dr Bernard van Oost  
Department of Human Genetics  
University Hospital Nijmegen  
PO Box 9101  
6500 HB Nijmegen  
The Netherlands  
tel 080-51 9111; fax 080-54 0576

Pedigree #.	Laboratory ref.	Recombination with <u>FRAXA</u>
160	2661	
161	208	
162	210	
163	035	U6.2 ST14
164	032	ST14
165	111	

6 pedigrees



Appendix C. Fragile X pedigrees. p. C-10

MARSEILLE

Dr Marie-Antoinette Voelckel  
 C.R.E.B.I.O.P.  
 Hopital d'Enfants de la Timone  
 13385 Marseille CEDEX 5  
 France  
 tel 91-92 1379; fax 91-49 4194

Pedigree #.	Laboratory ref.	Recombination with <u>FRAXA</u>
201	GIU	RN1 VK21 ST14
202	SEI-FON	
203	LAL/DAU/PAN	F9 55.7 RN1
204	LES	
205	LAM	
208	SAL	
209	DI/CES/VIT	
210	SAN	55.7 RN1
212	NEV	F9 ST14
213	PAT/RUB/SAV/BEN	
214	TAU	
216	DON	F9 ST14
218	MON	ST14
219	LUC	F9 RN1
220	LAF	
222	WAC/PAS/GEN	
223	CAR	

17 pedigrees

Appendix C. Fragile X pedigrees. p. C-11

UPPSALA

Dr Niklas Dahl  
 Department of Medical Genetics  
 Biomedical Centre  
 Box 589  
 751 23 Uppsala  
 Sweden  
 tel 18-17 4580; fax 18-12 6849

Pedigree #.	Laboratory ref.	Recombination with <u>FRAXA</u>
303	3	
304	4	
305	5	ST14
306	7	ST14
309	11	RN1
310	12	
311	13	
313	17	RN1
314	19	F9 RN1
316	P3	F9 4D8 ST14
318	P7	
319	P12	
320	P8	
328	10 <sup>a</sup>	

14 pedigrees

a contains portion of pedigree 10 informative for RN1

GREENWOOD

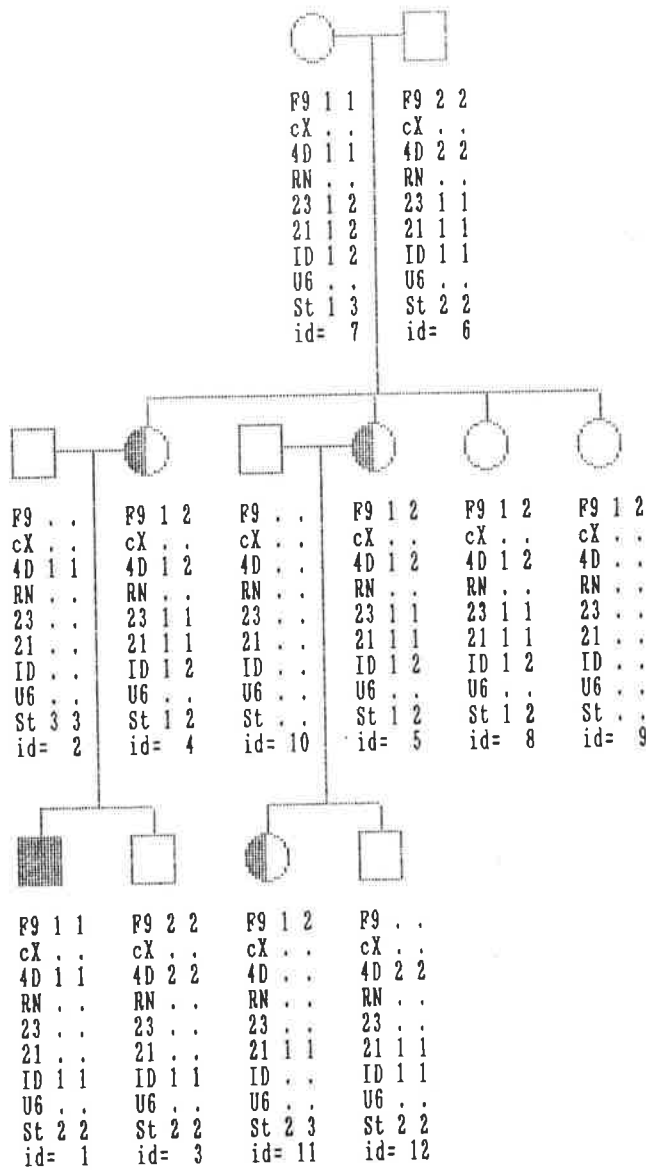
Dr Charles Schwartz  
 Department of Medical Genetics  
 Greenwood Genetics Center  
 1 Gregor Mendel Circle  
 Greenwood SC 29646  
 USA  
 tel 803-223 9411; fax 803-227 1614

Pedigree #.	Laboratory ref.	<sup>a</sup> Recombination with <u>FRAXA</u>
1050	same as ped #	VK21
1120		
1130		
1145		
1320		

5 pedigrees

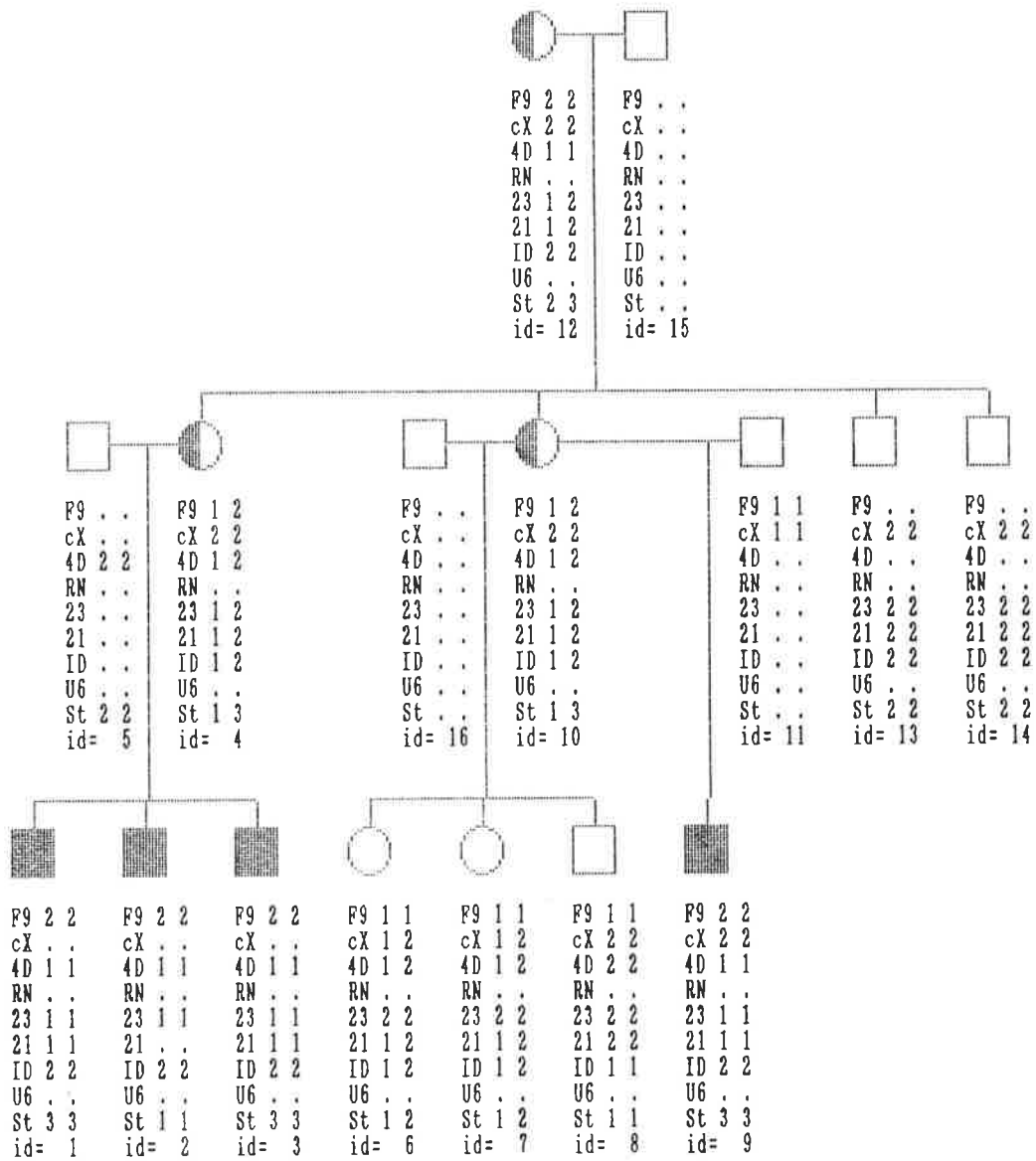
a VK21 genotypes only

Pedigree 3

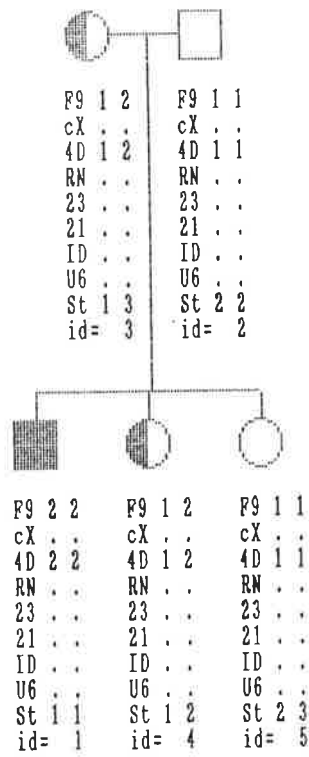


Appendix C. Fragile X pedigrees. p. C-13

Pedigree 4

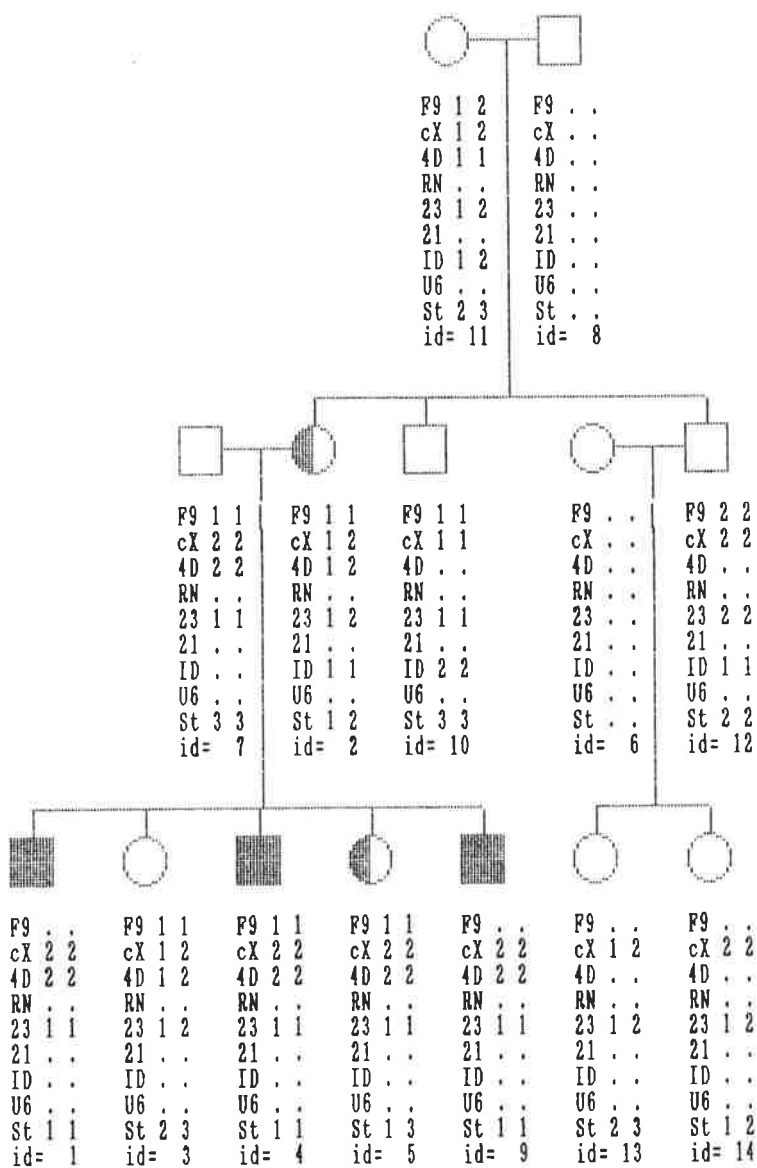


Pedigree 6

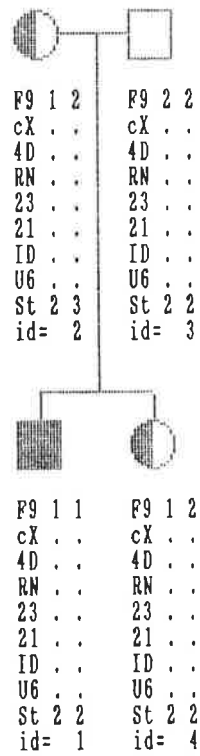


Appendix C. Fragile X pedigrees. p. C-15

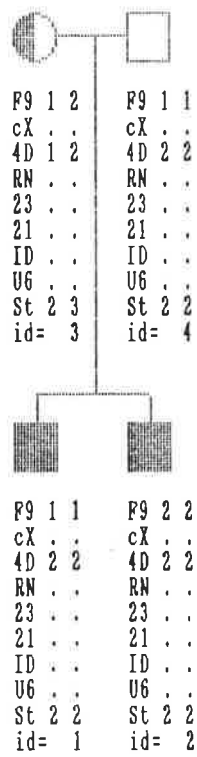
Pedigree 7



Pedigree 8

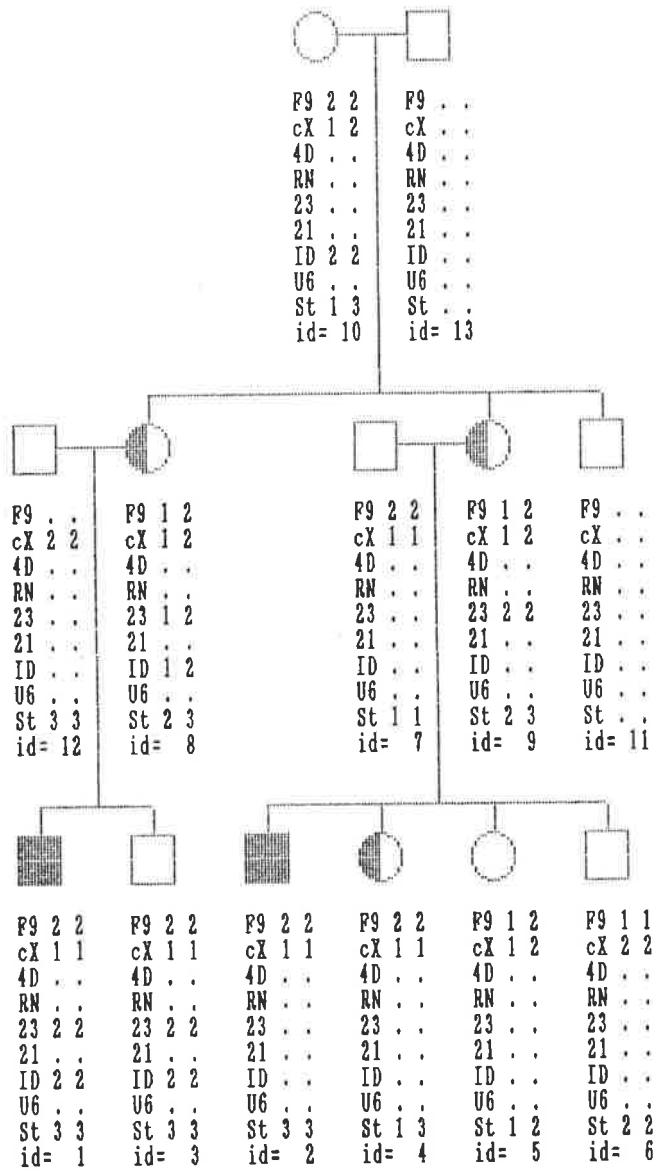


Pedigree 9



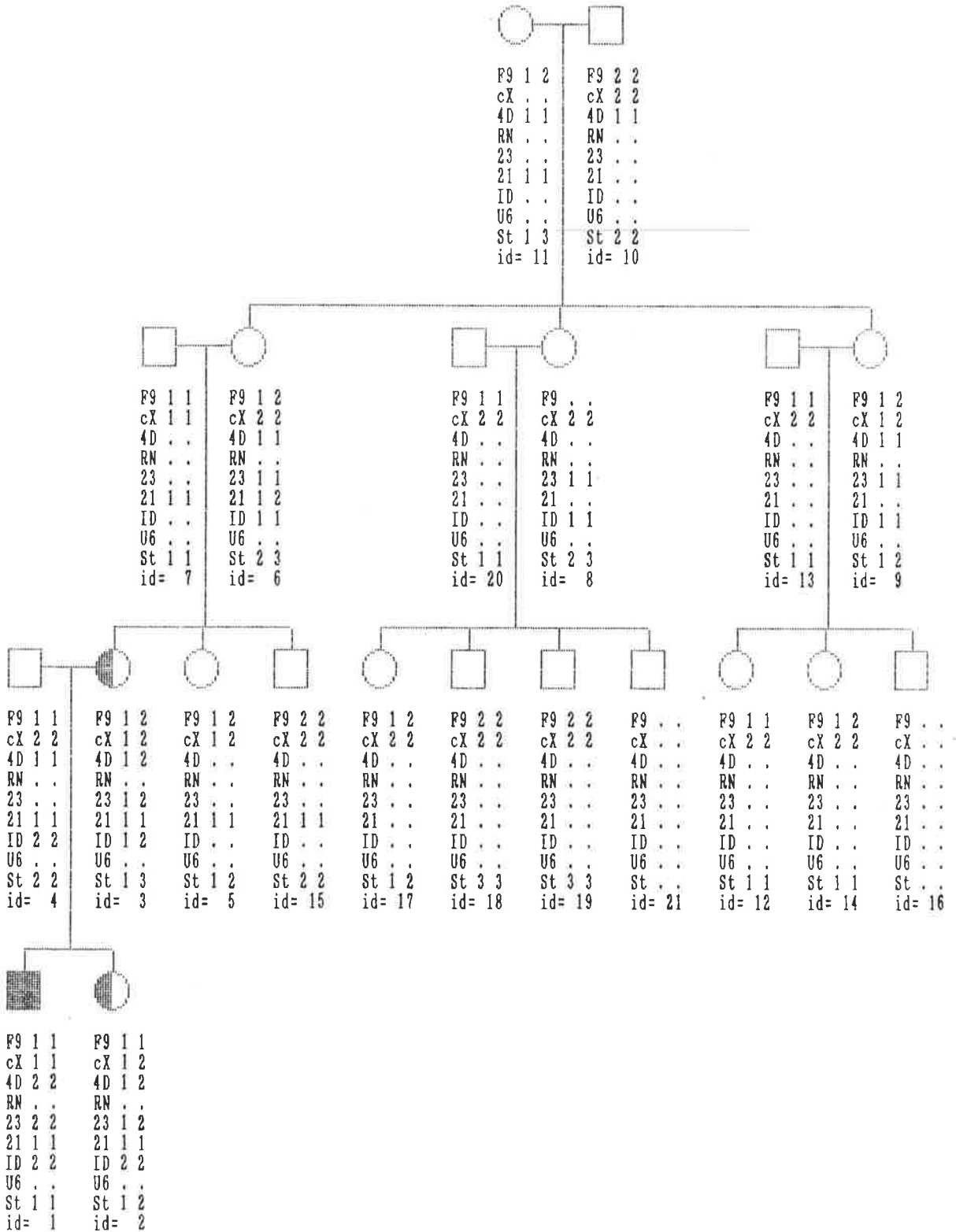


Pedigree 10

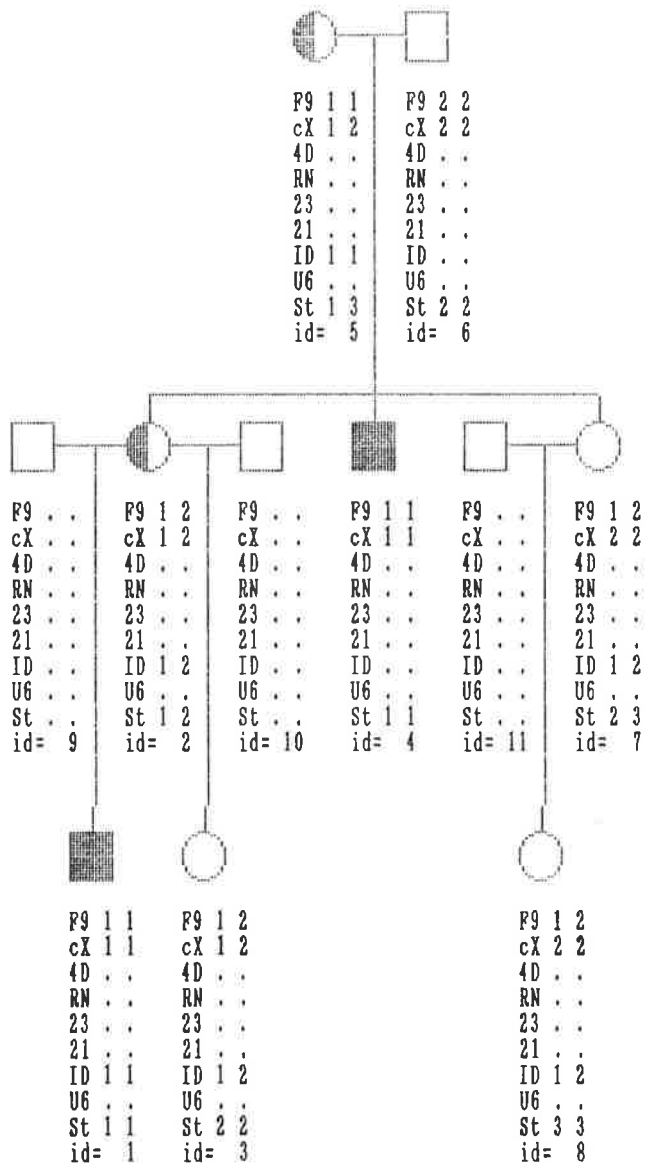


Appendix C. Fragile X pedigrees. p. C-19

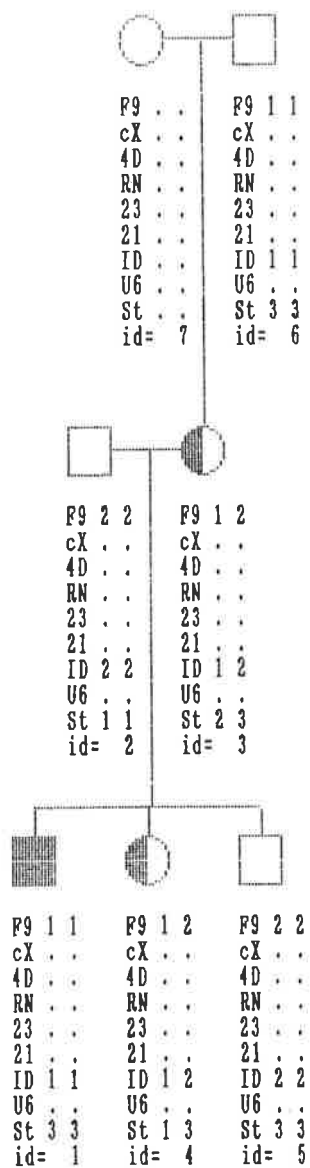
Pedigree 11



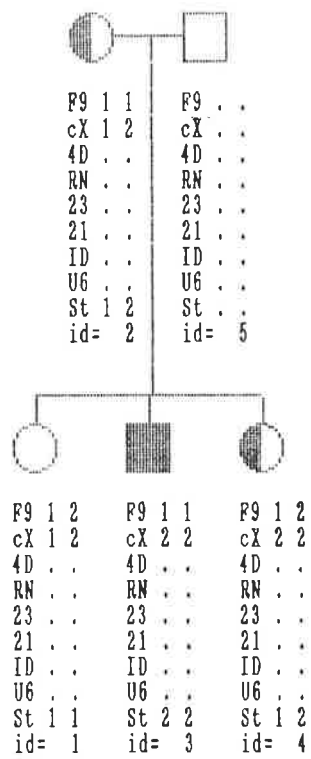
Pedigree 13



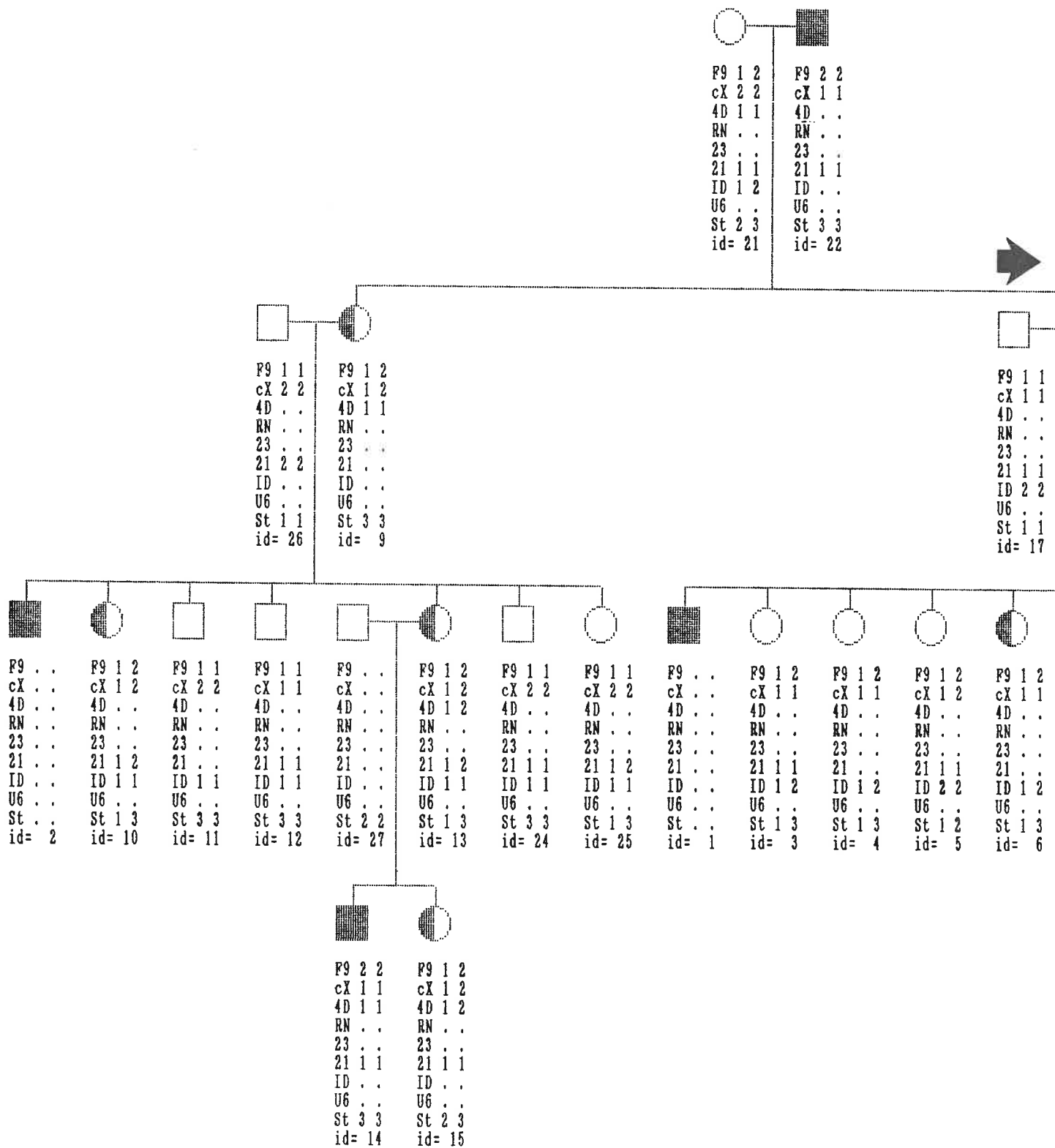
Pedigree 15



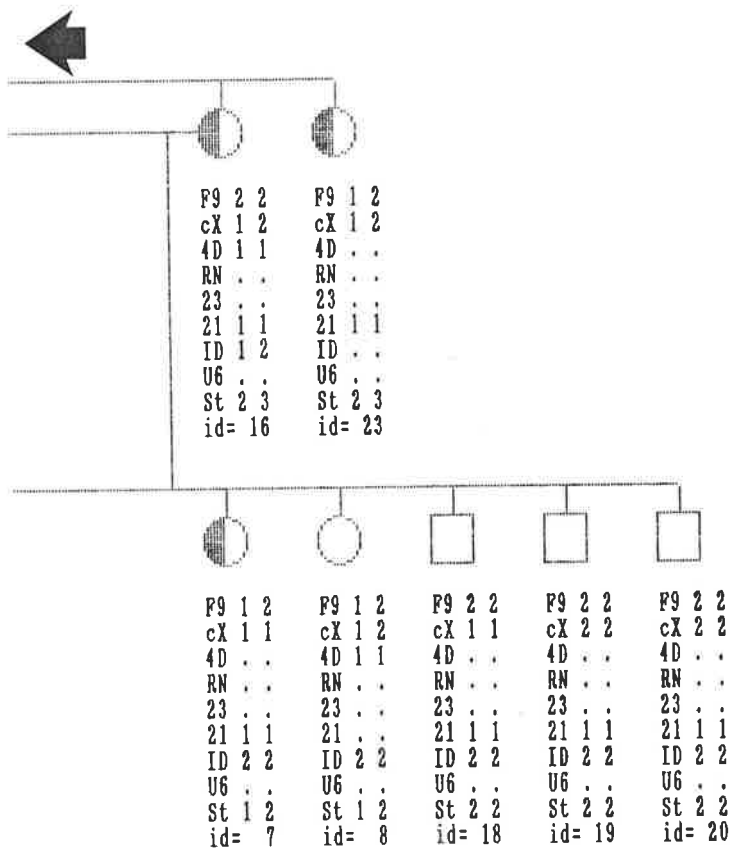
Pedigree 16



Pedigree 19

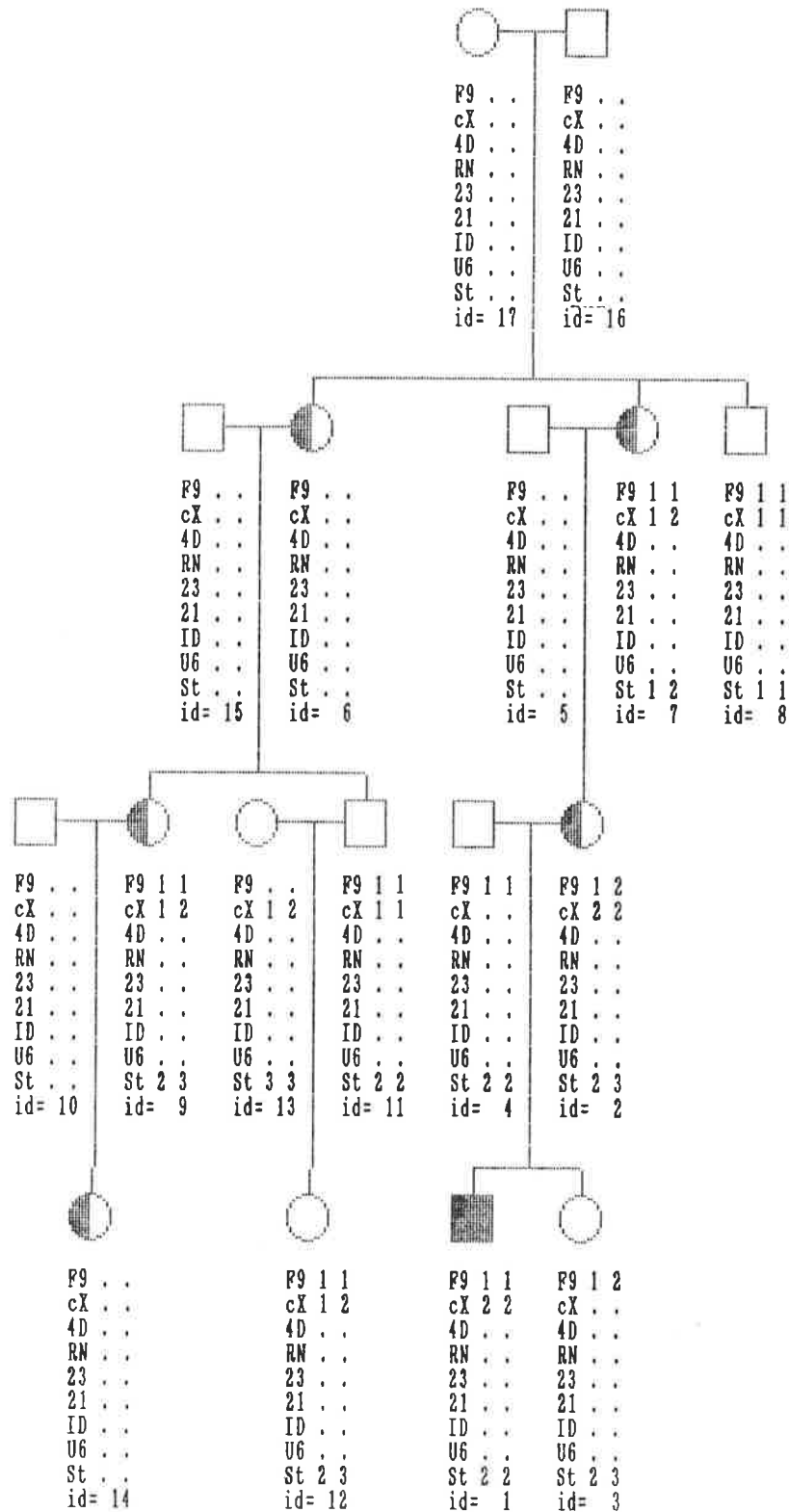


Appendix C. Fragile X pedigrees. p. C-24



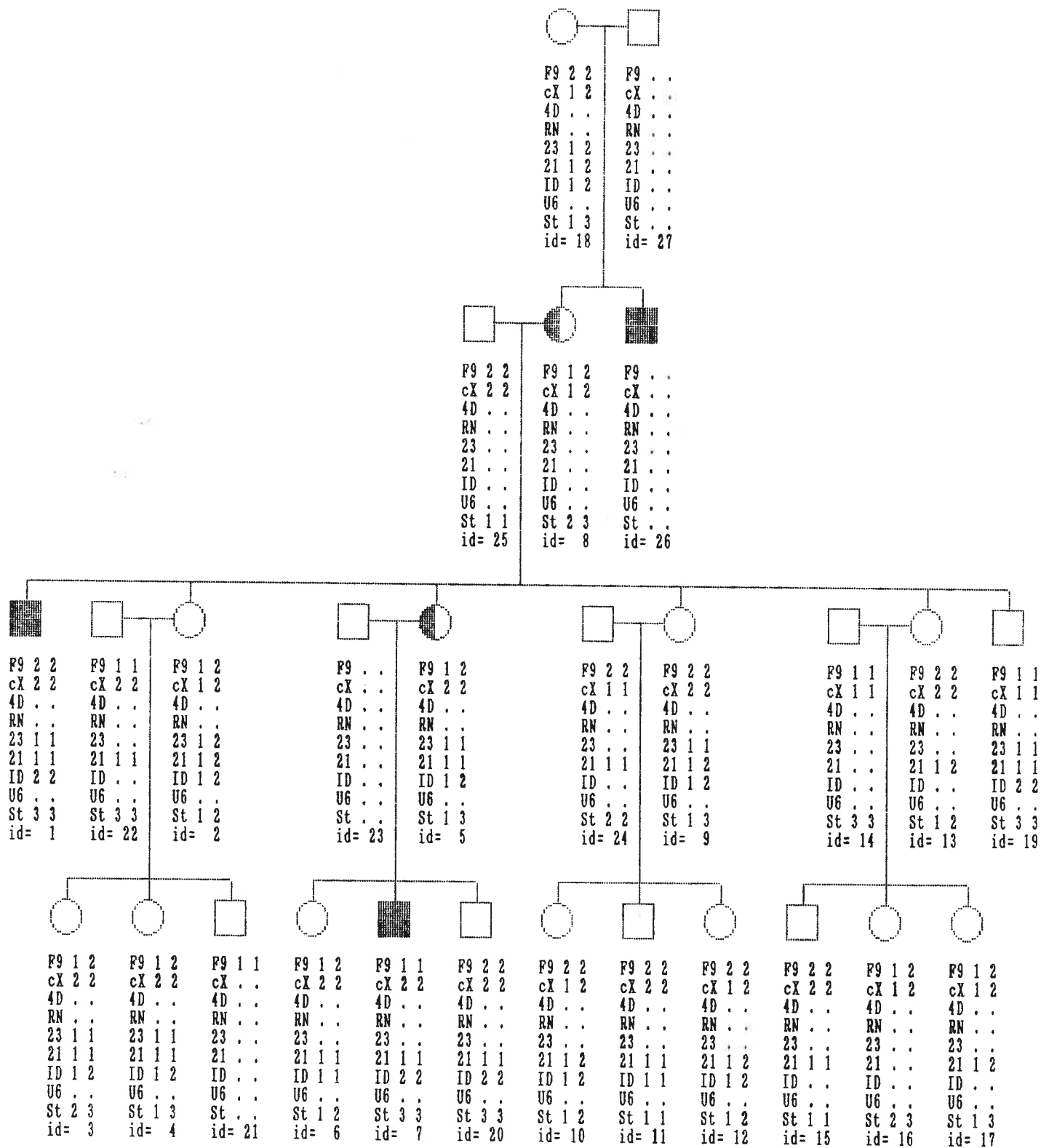
Appendix C. Fragile X pedigrees. p. C-25

Pedigree 21

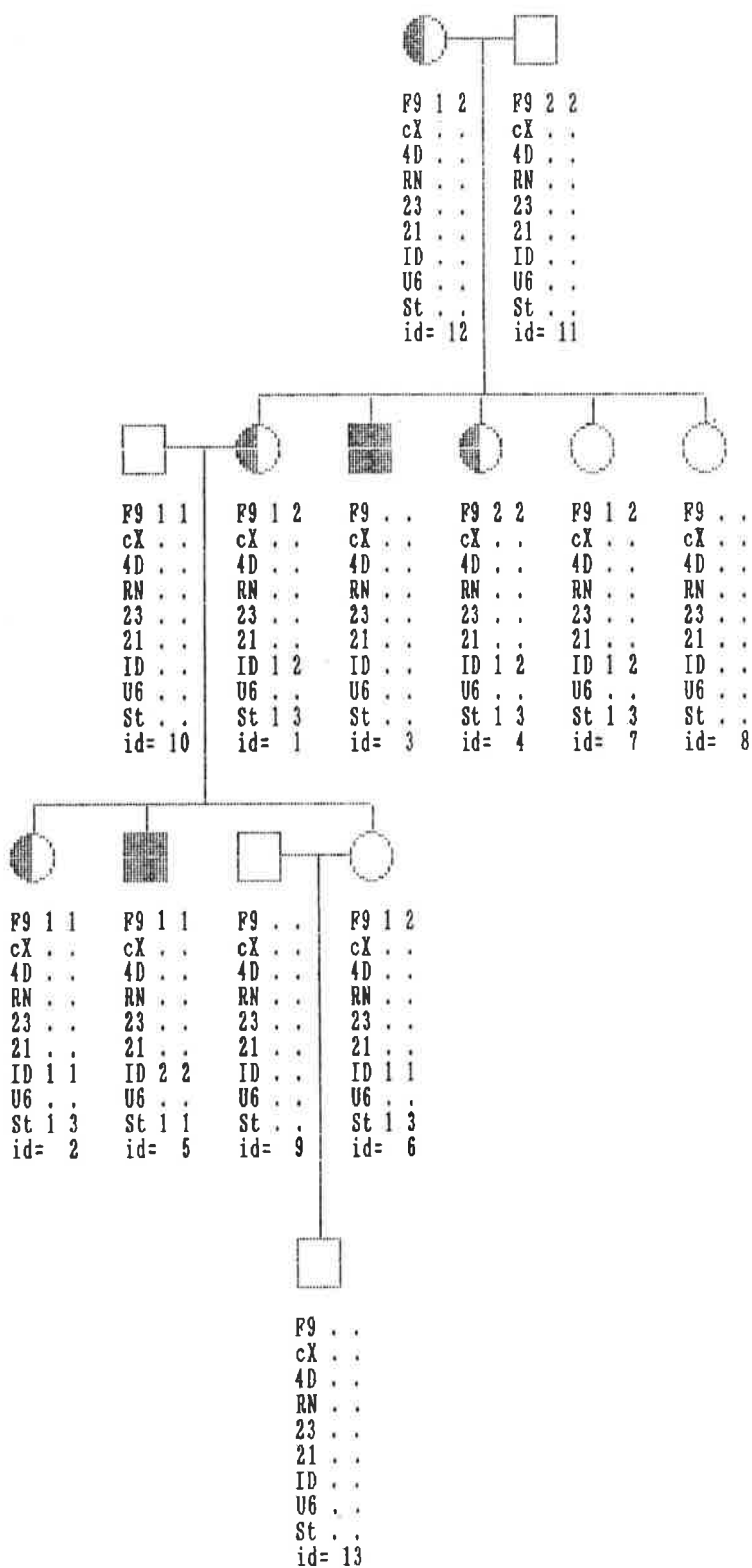




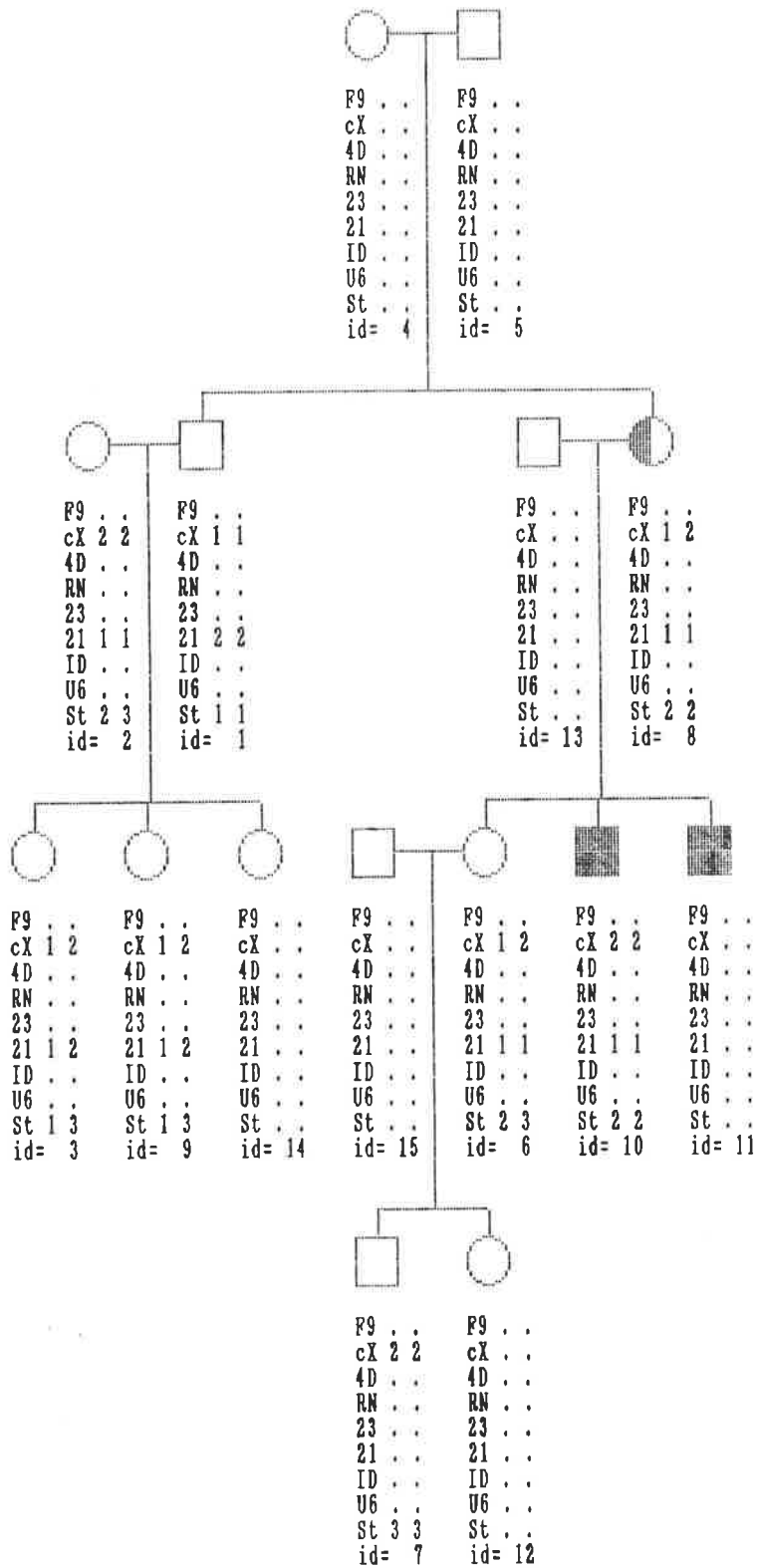
Pedigree 22



Pedigree 23

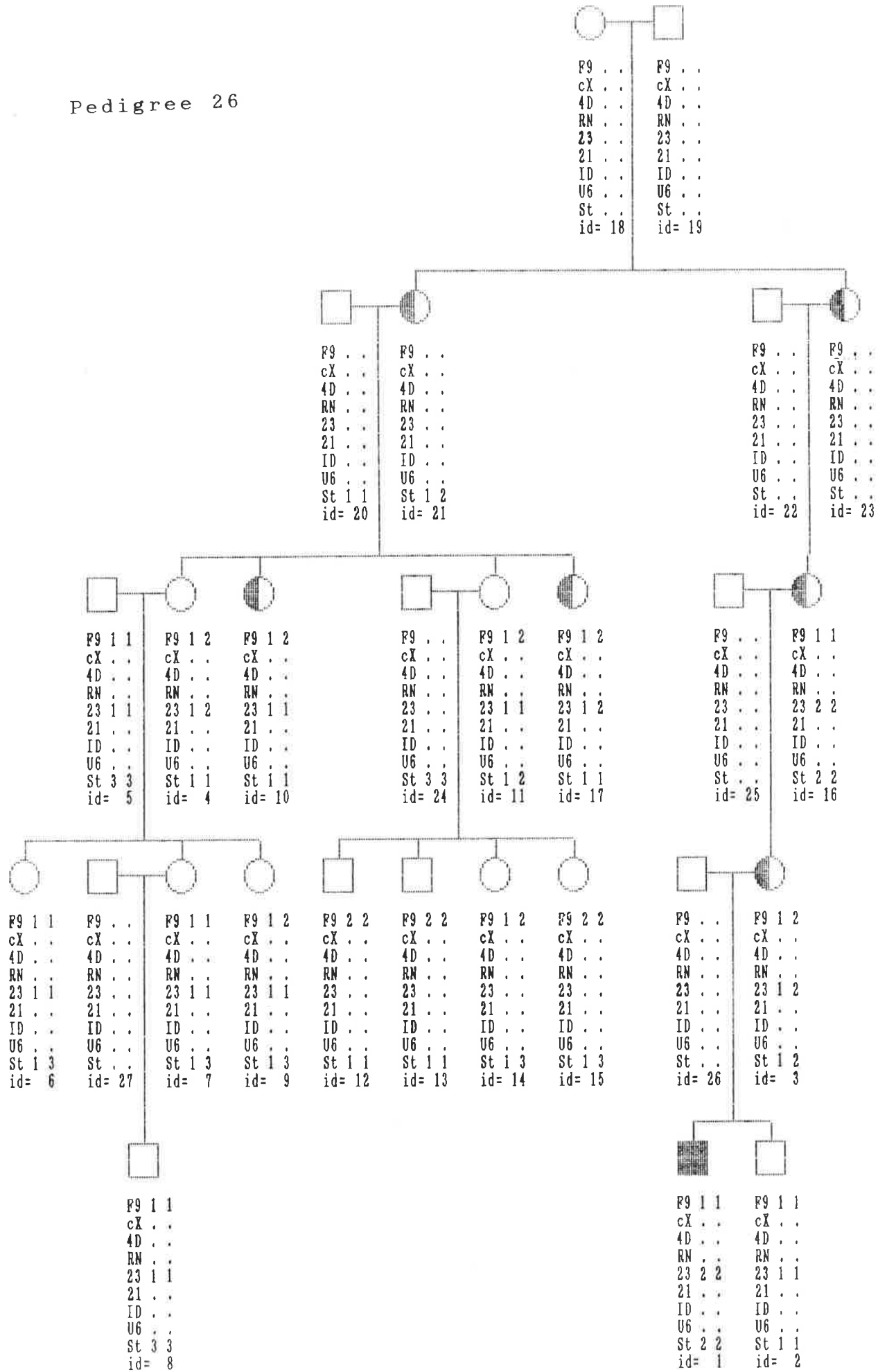


Pedigree 24

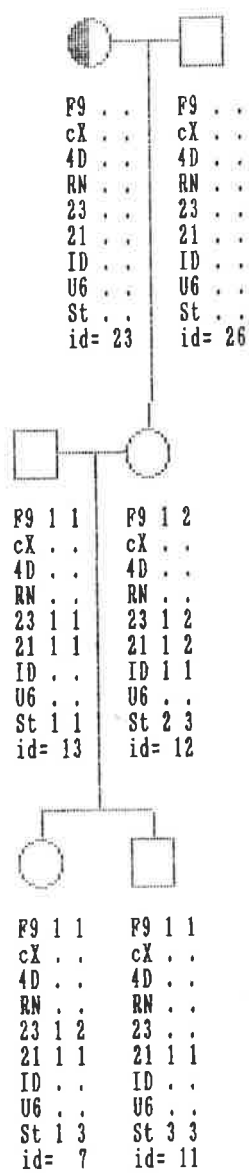


Appendix C. Fragile X pedigrees. p. C-29

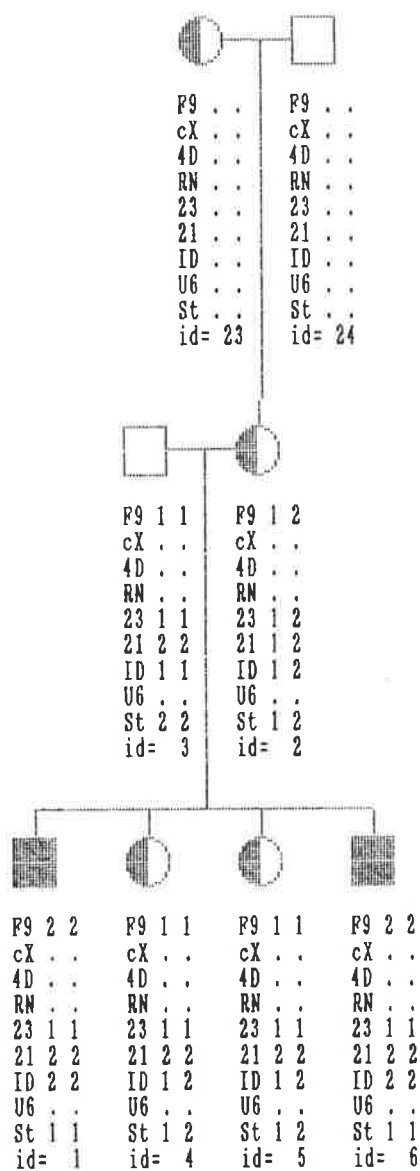
Pedigree 26



Pedigree 28

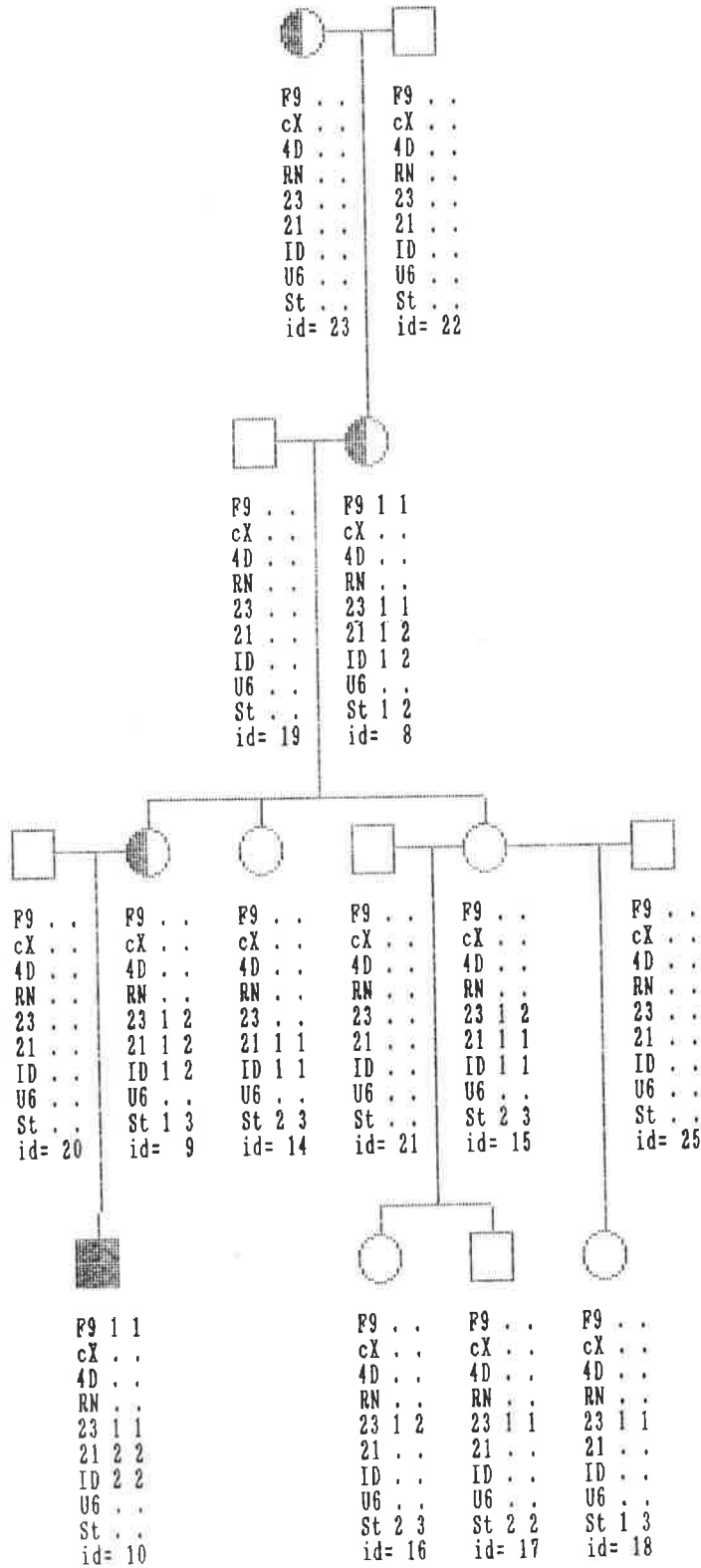


Pedigree 28

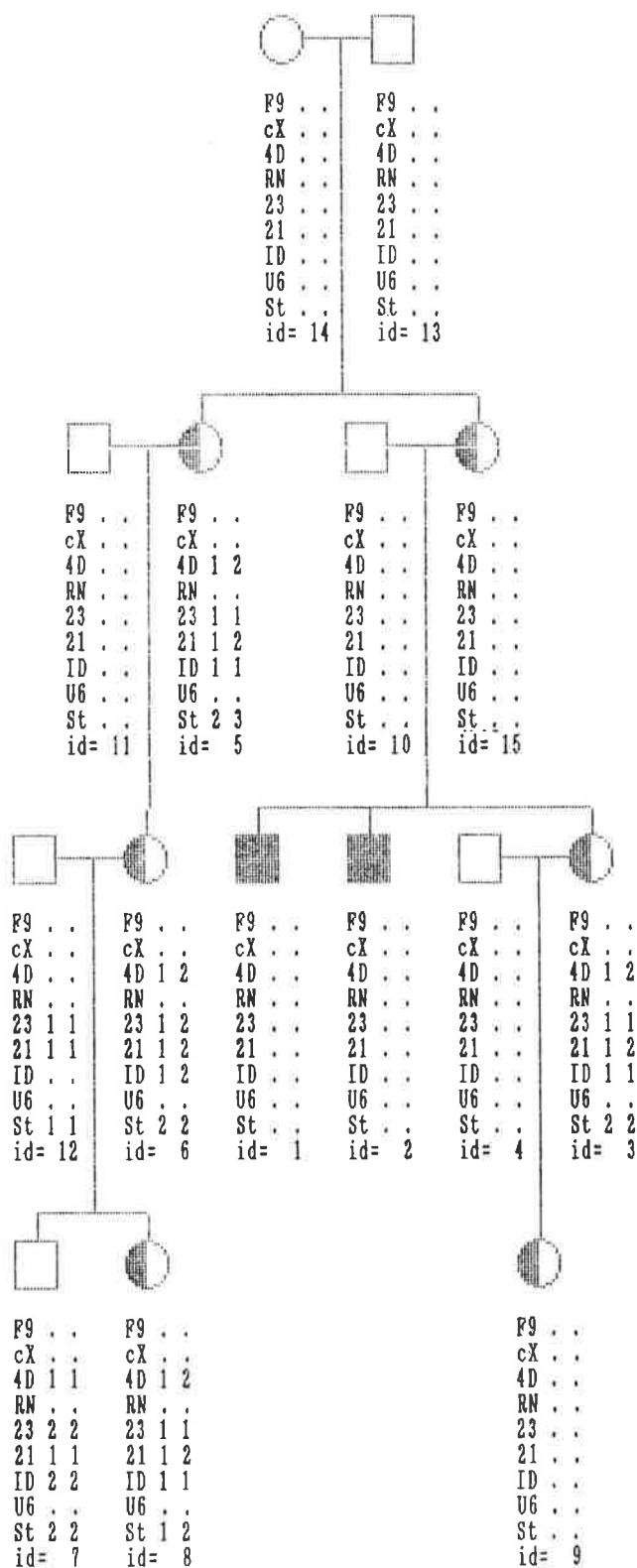




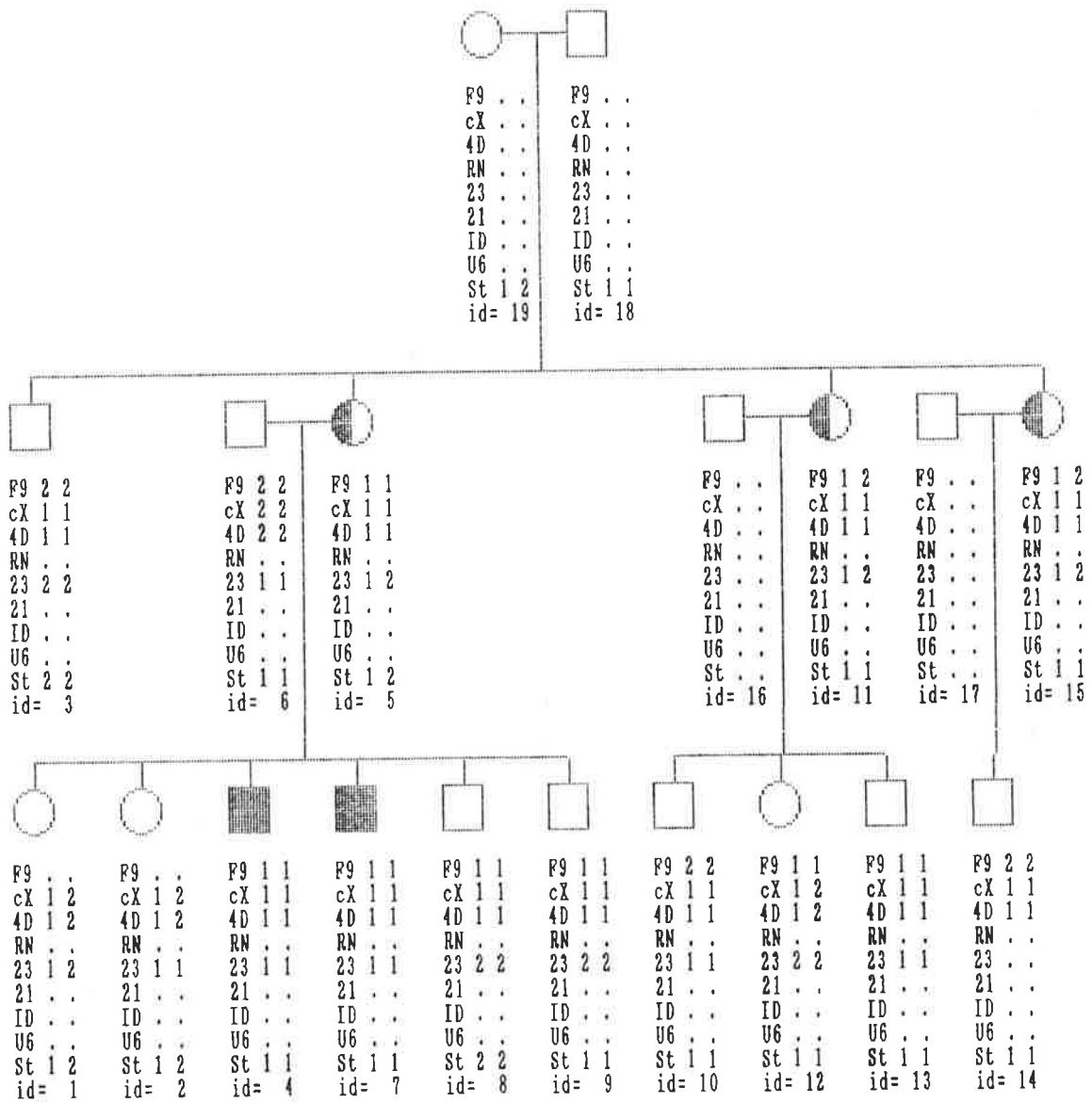
Pedigree 28



Pedigree 30

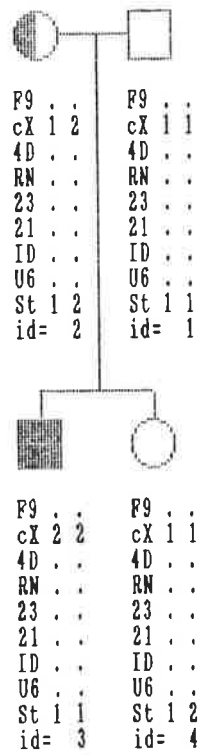


Pedigree 31

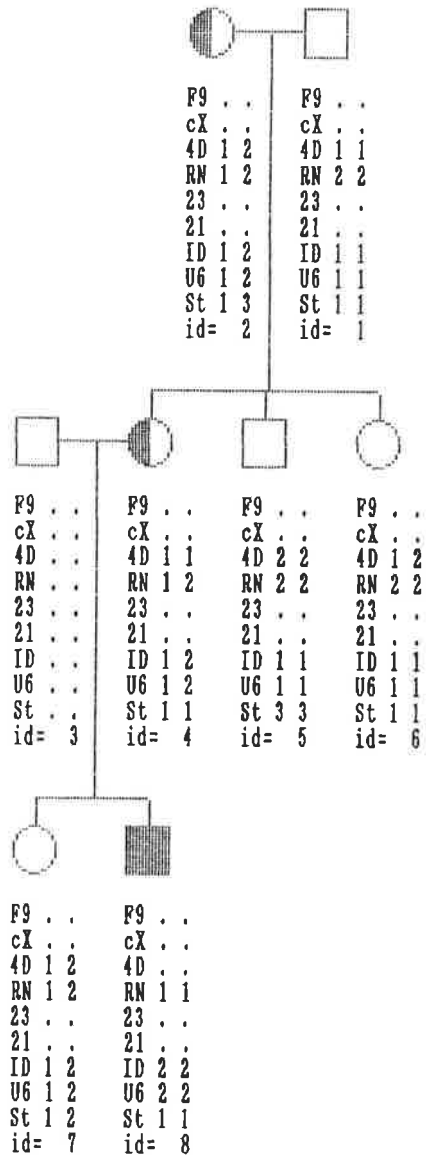




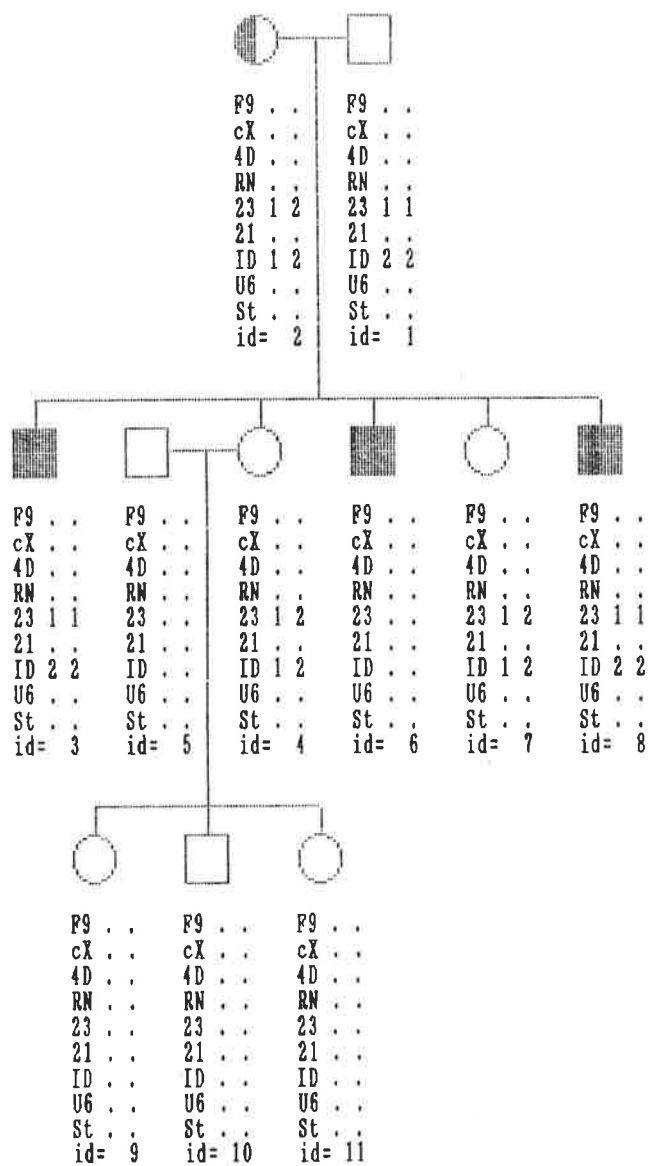
Pedigree 33



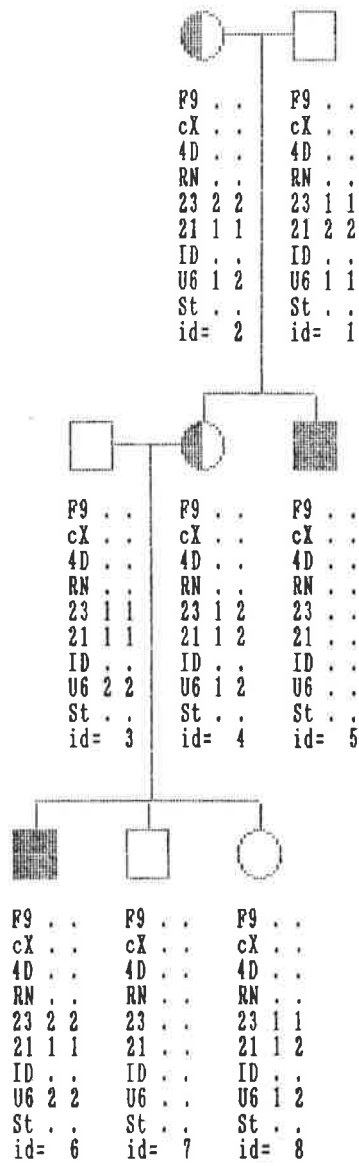
Pedigree 34



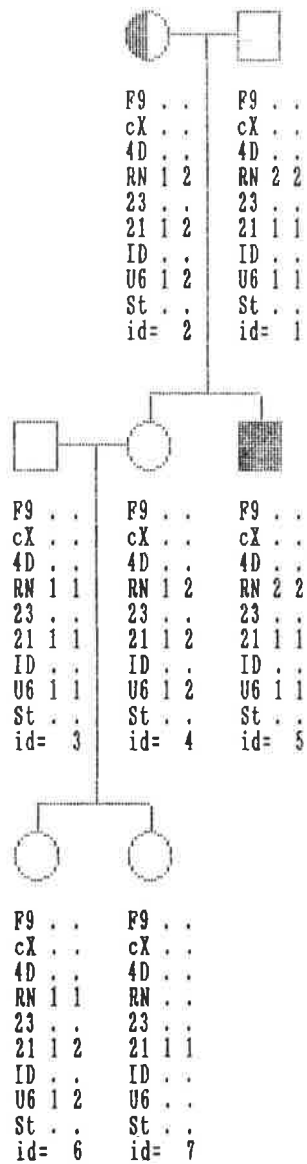
Pedigree 35



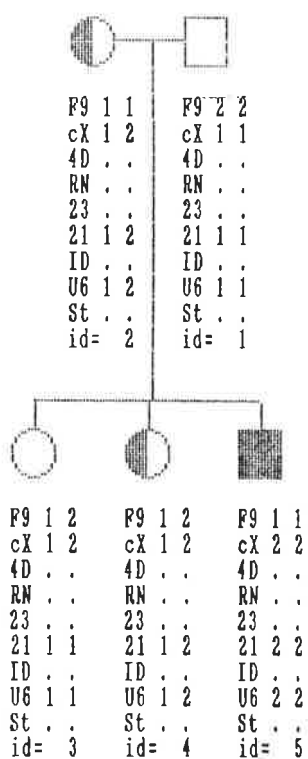
Pedigree 36



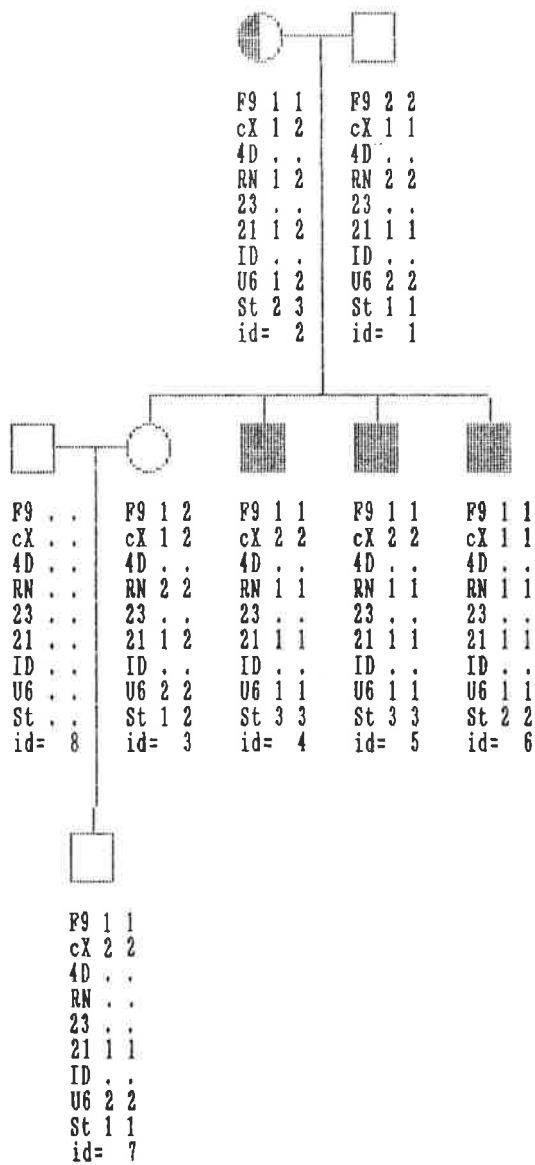
Pedigree 50



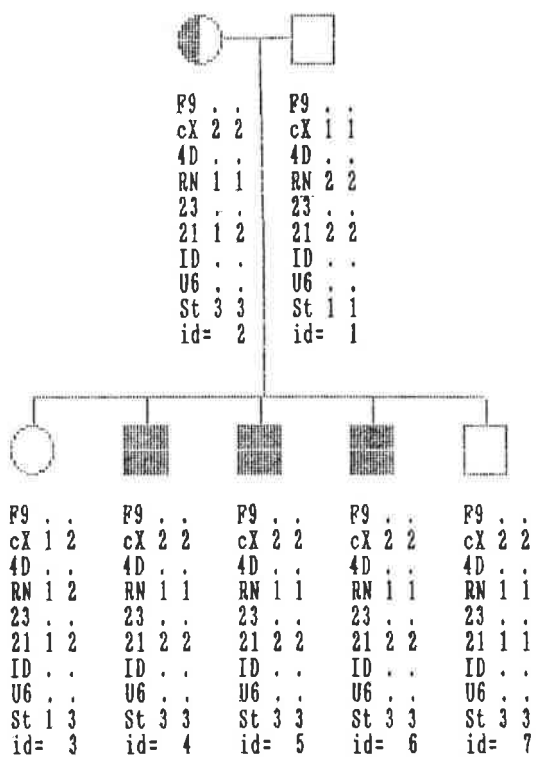
Pedigree 51



Pedigree 52

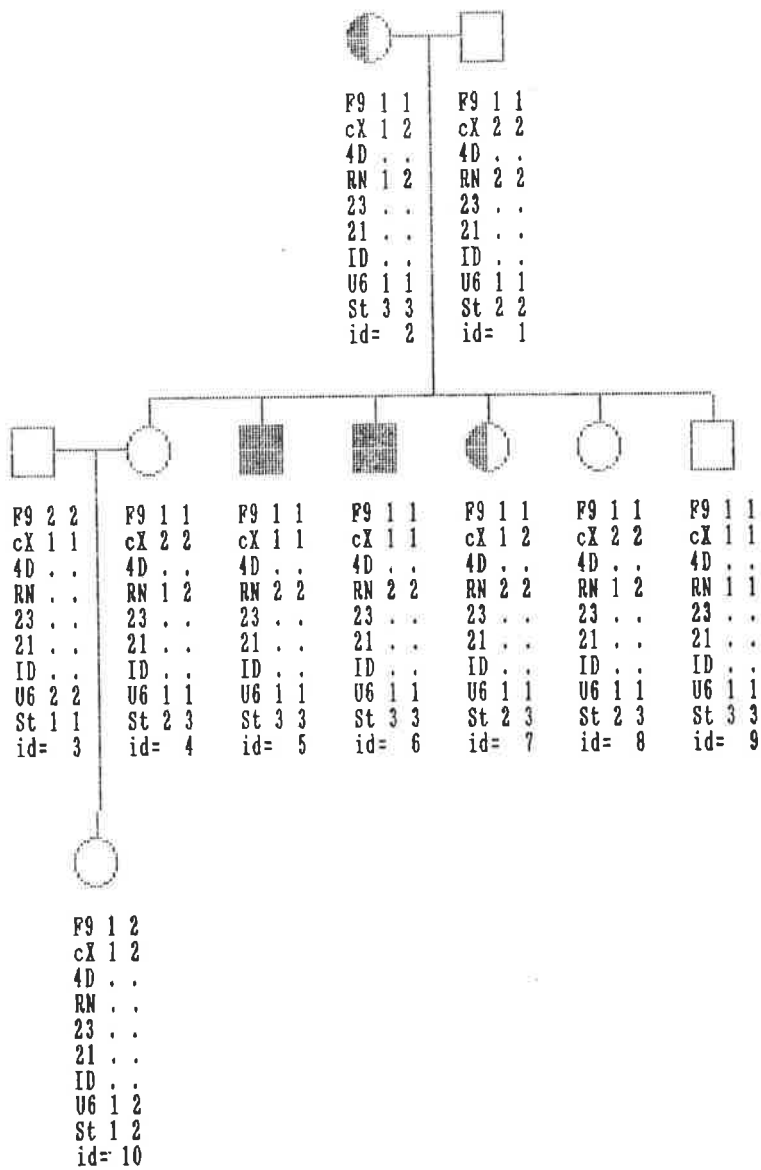


Pedigree 53

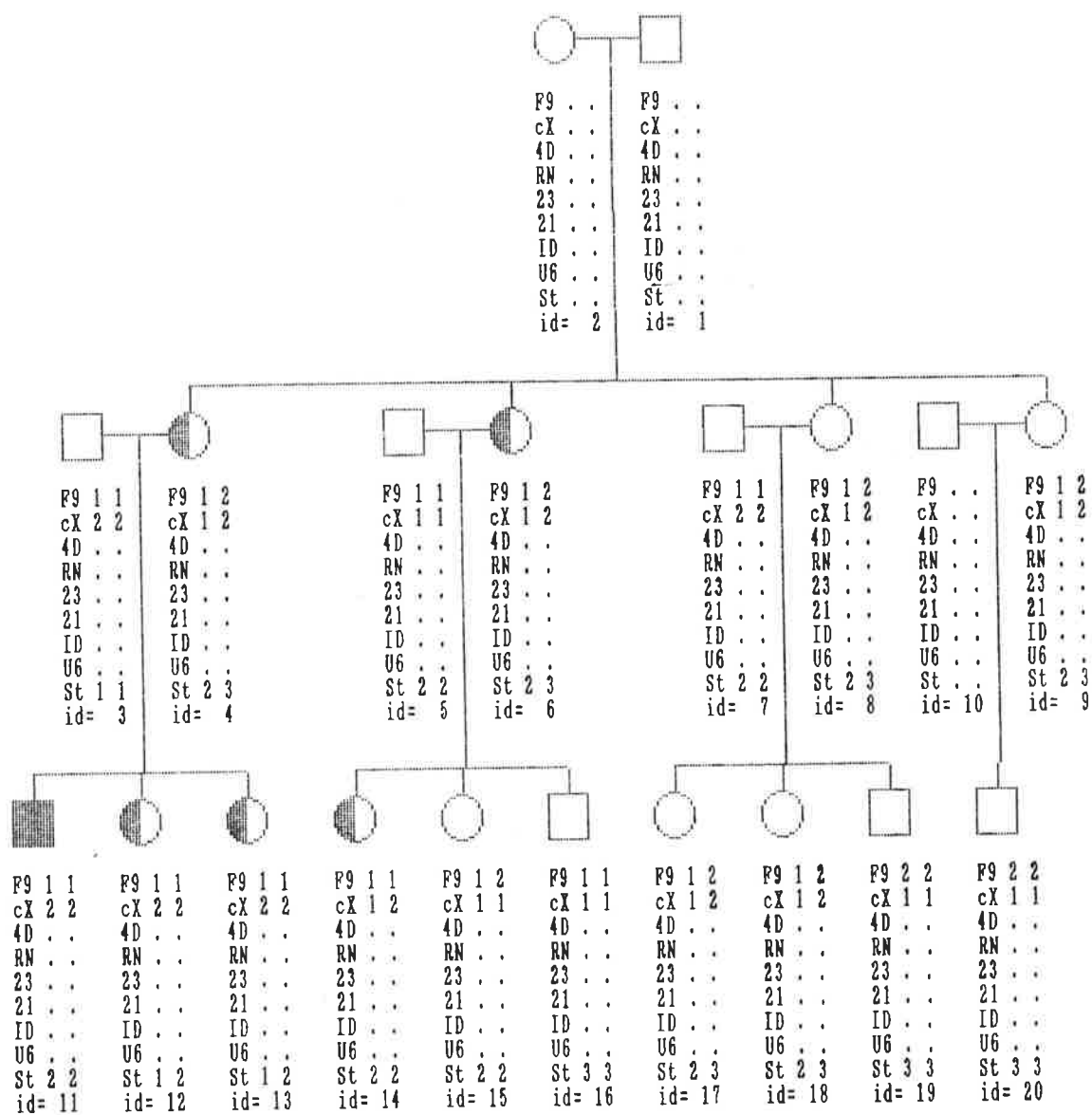




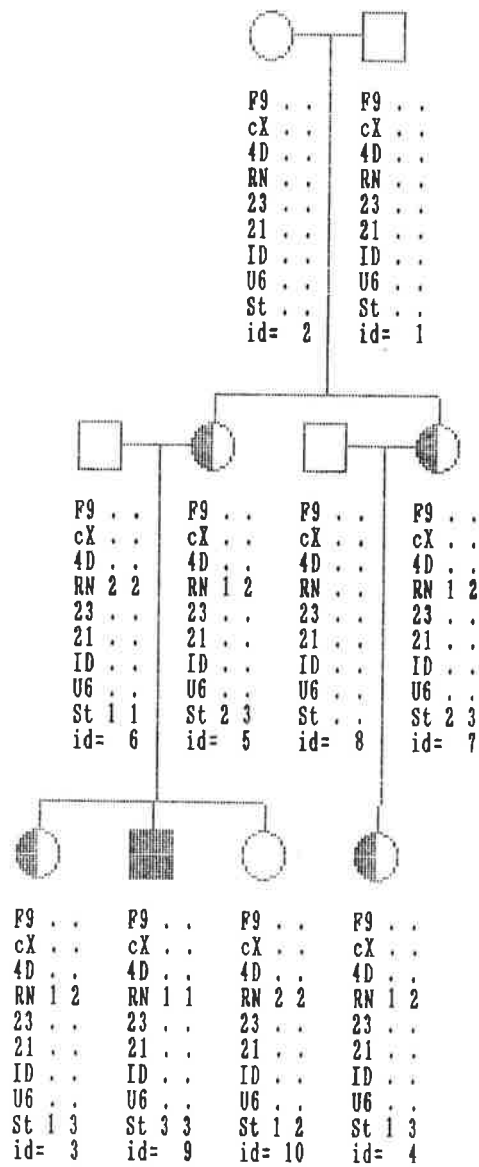
Pedigree 55



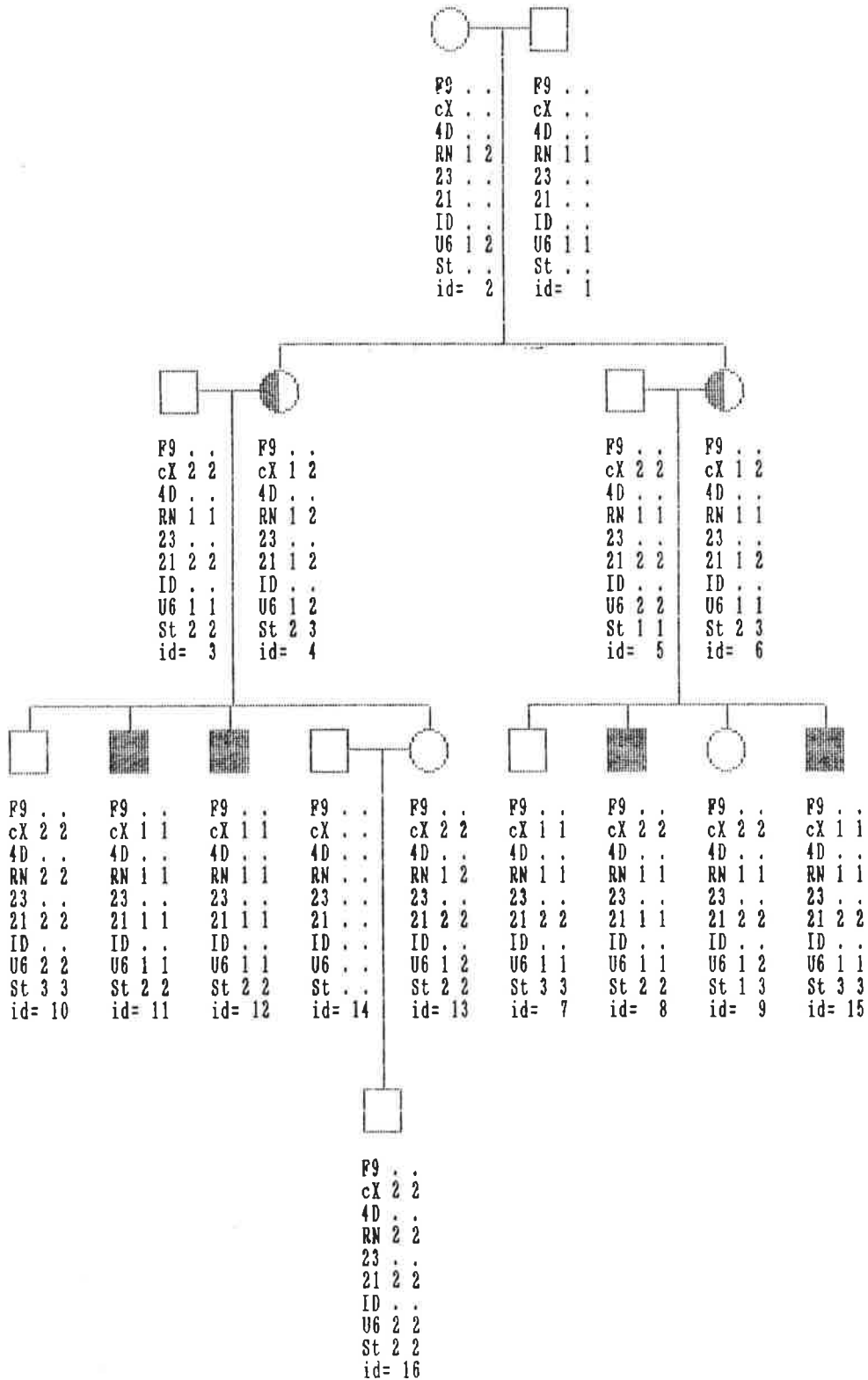
Pedigree 56



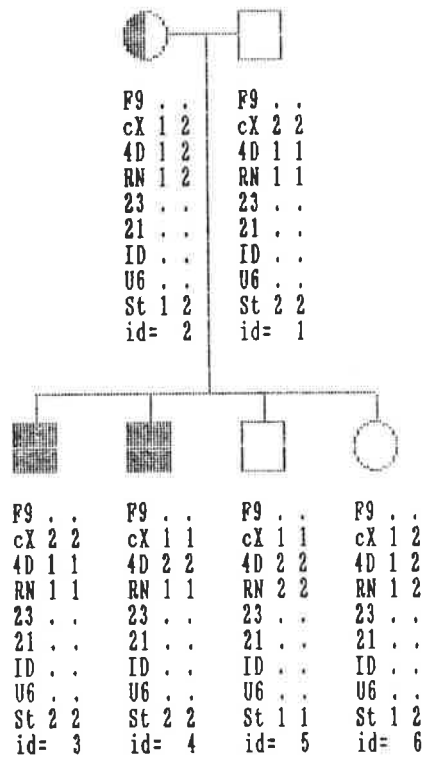
Pedigree 57



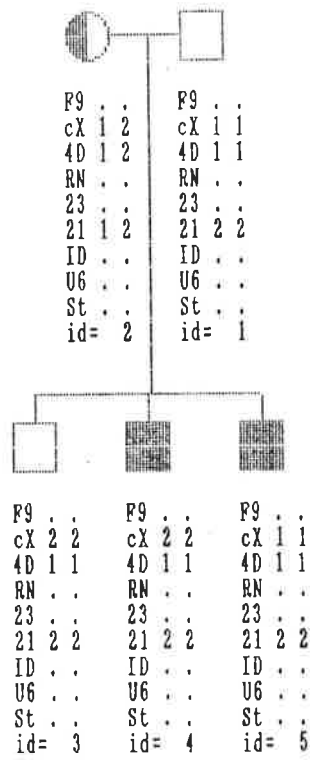
Pedigree 61



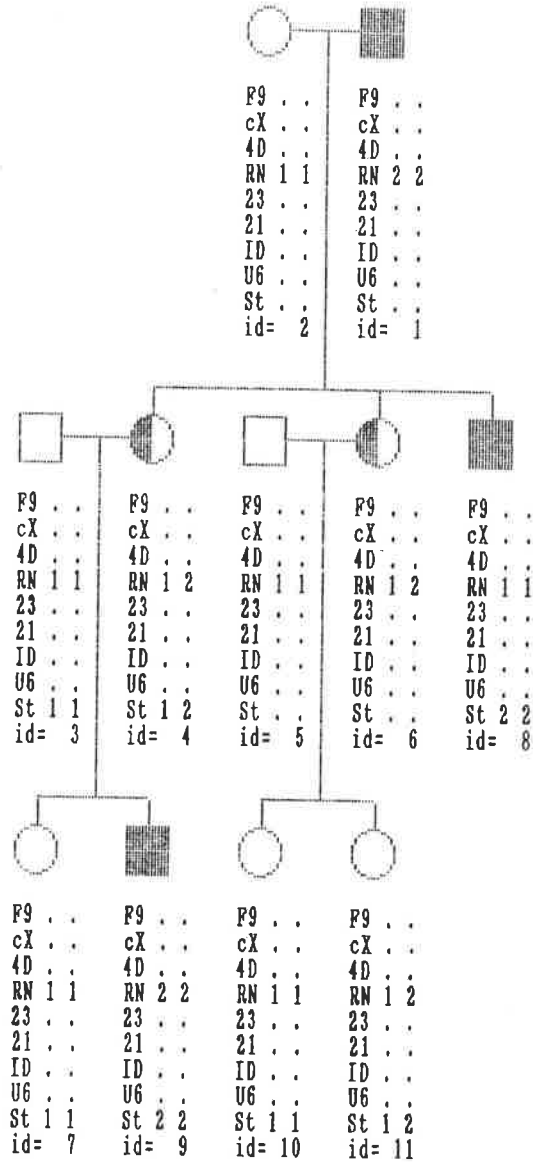
Pedigree 62



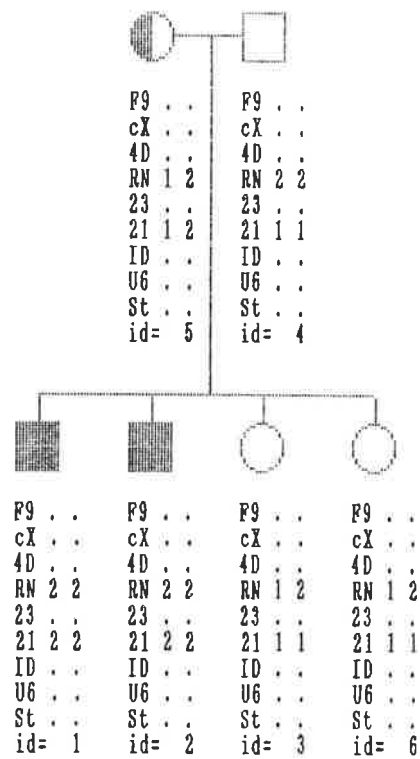
Pedigree 72



Pedigree 75

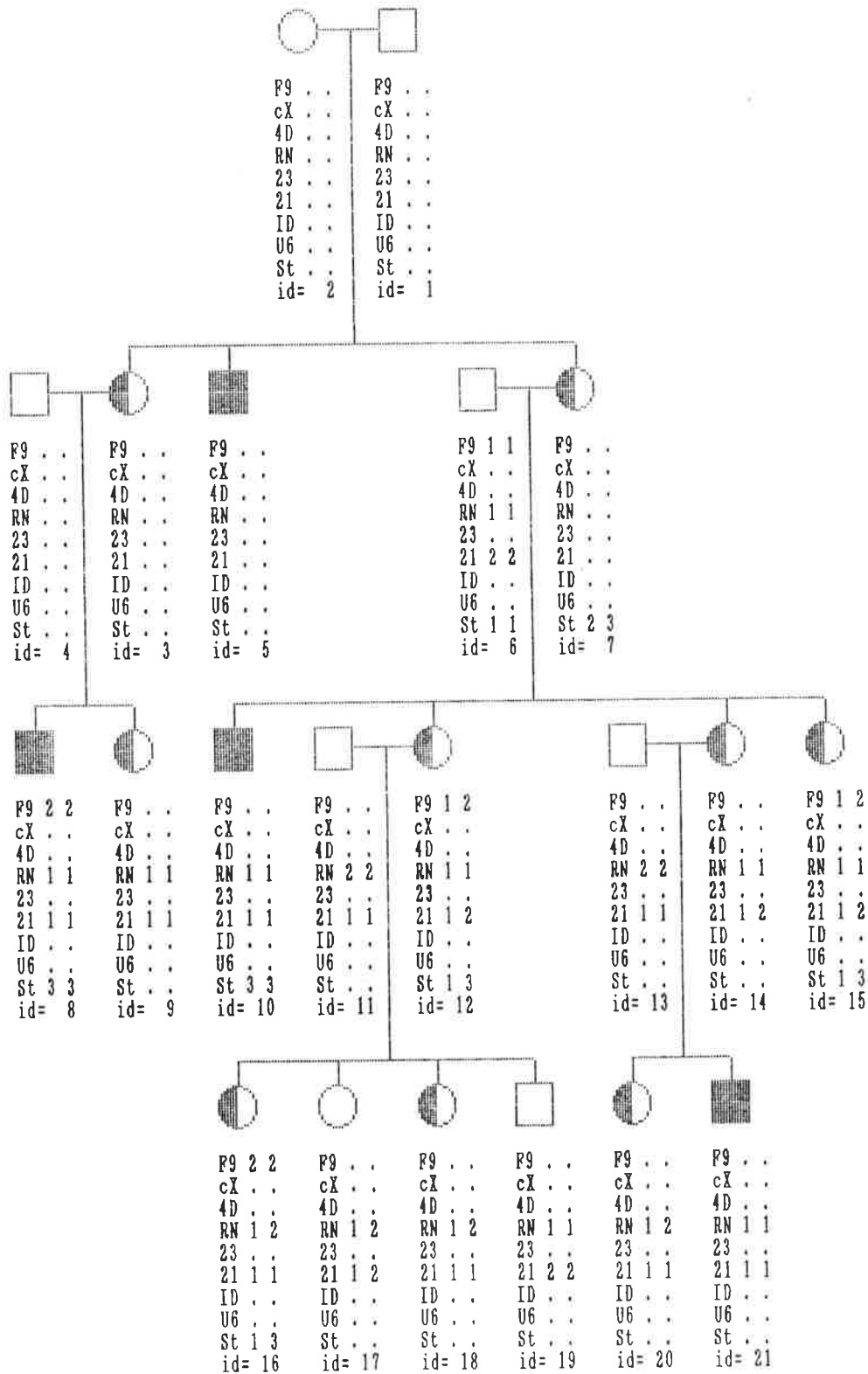


Pedigree 78

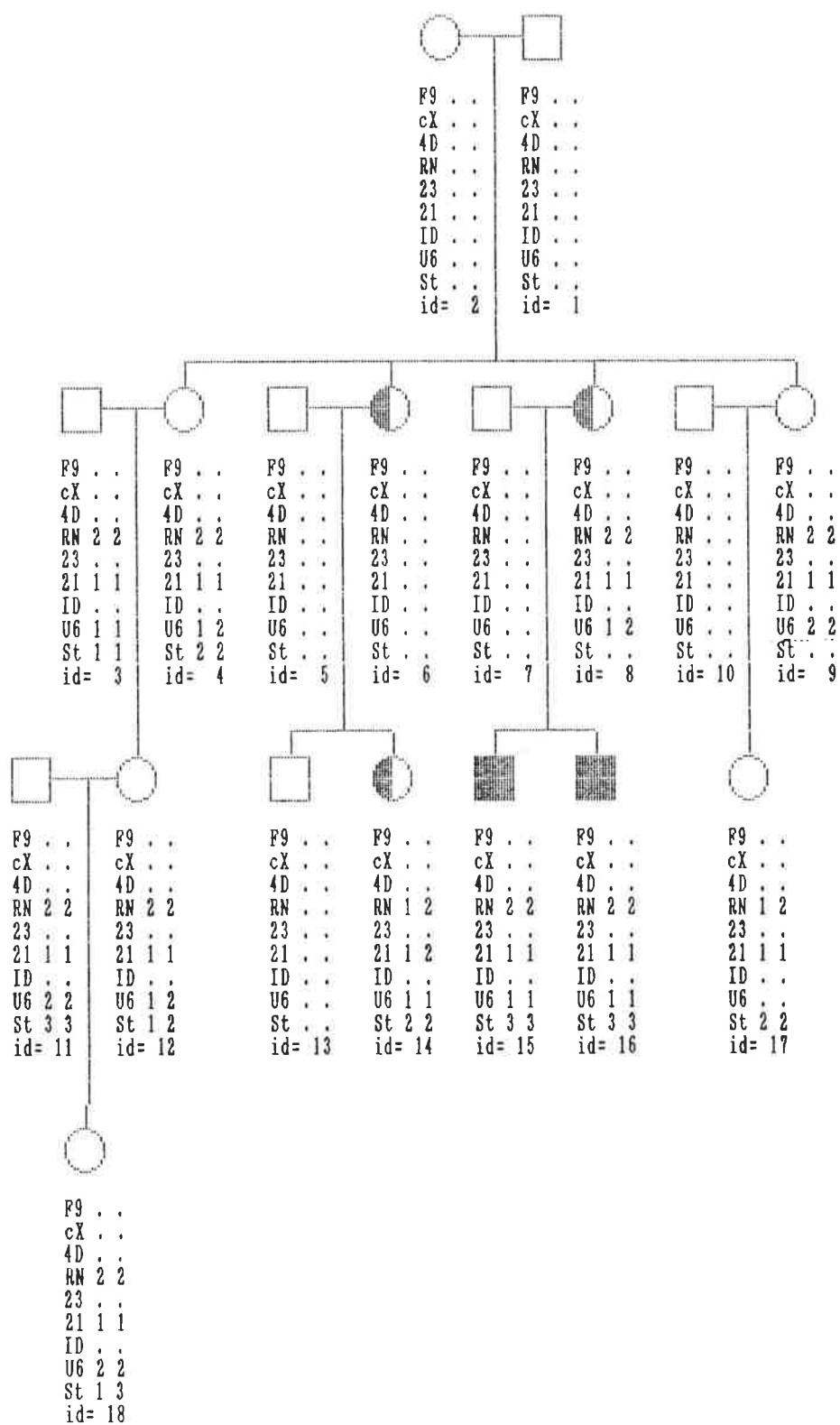




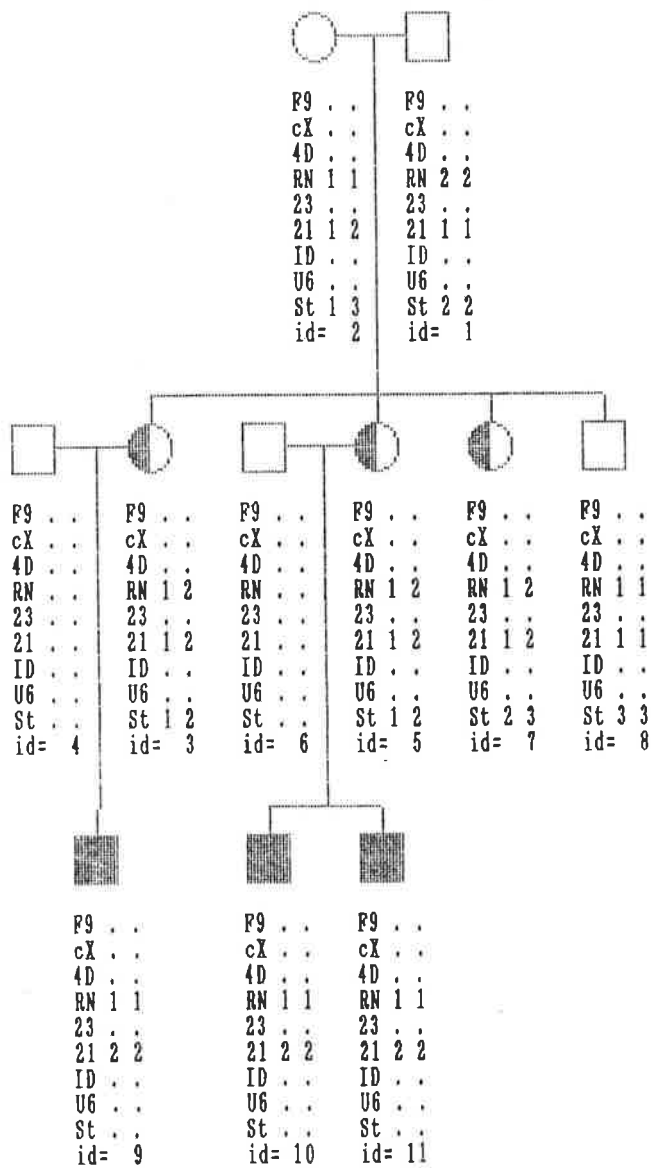
Pedigree 81



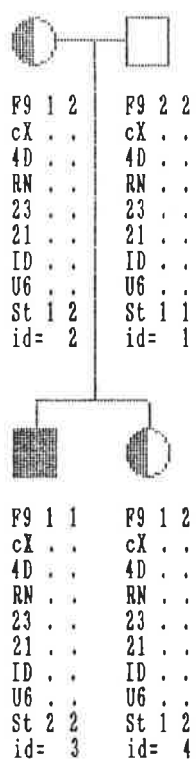
Pedigree 82



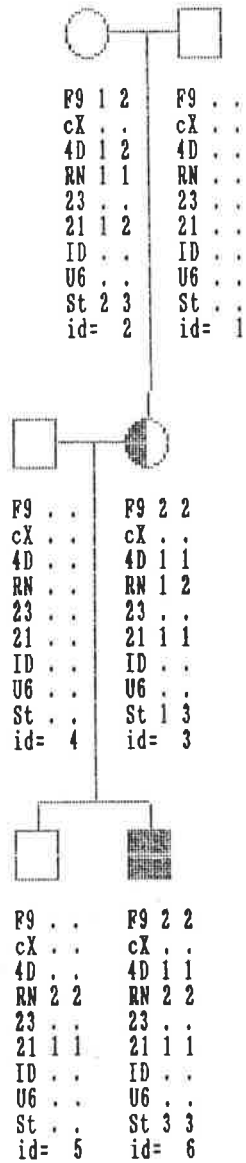
Pedigree 83



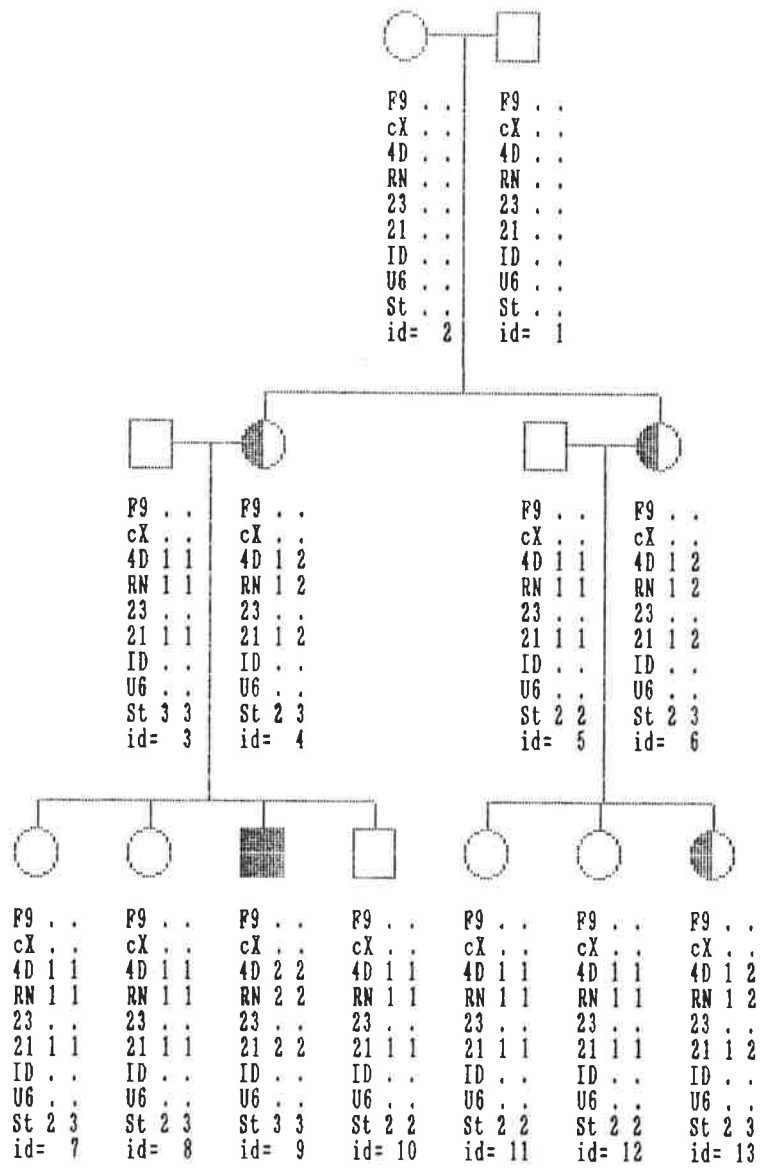
Pedigree 84



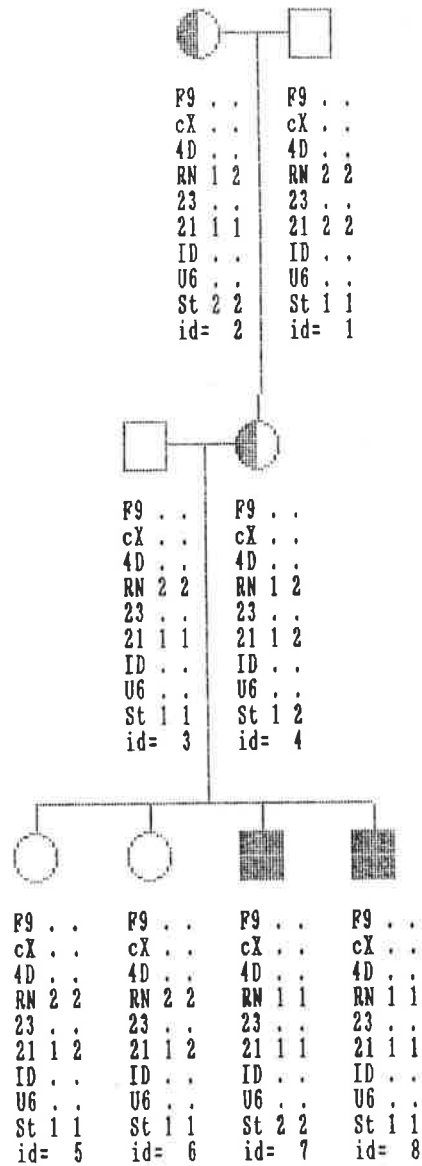
Pedigree 85



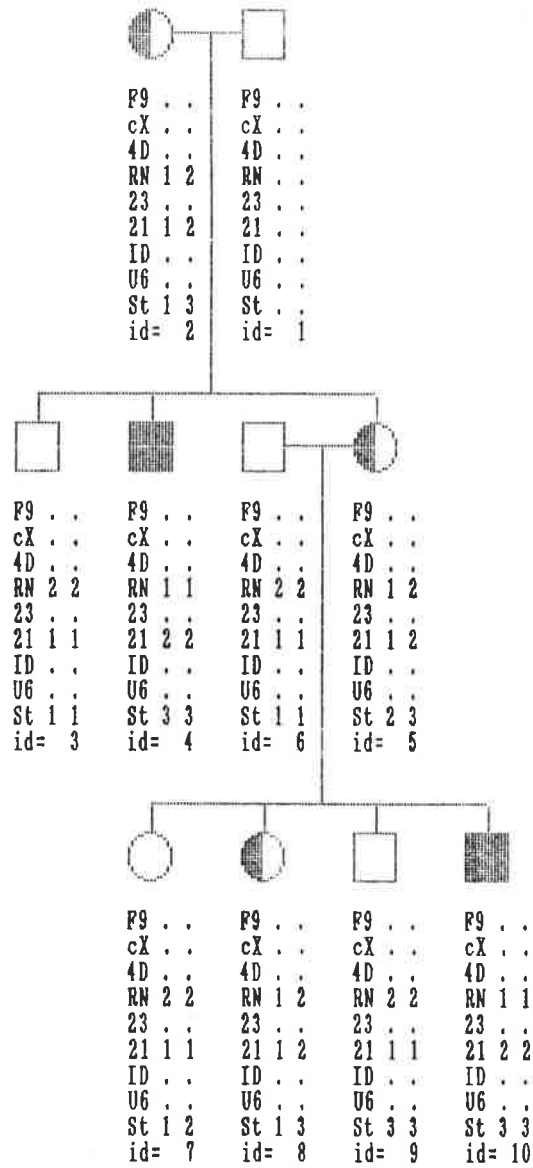
Pedigree 92



Pedigree 93



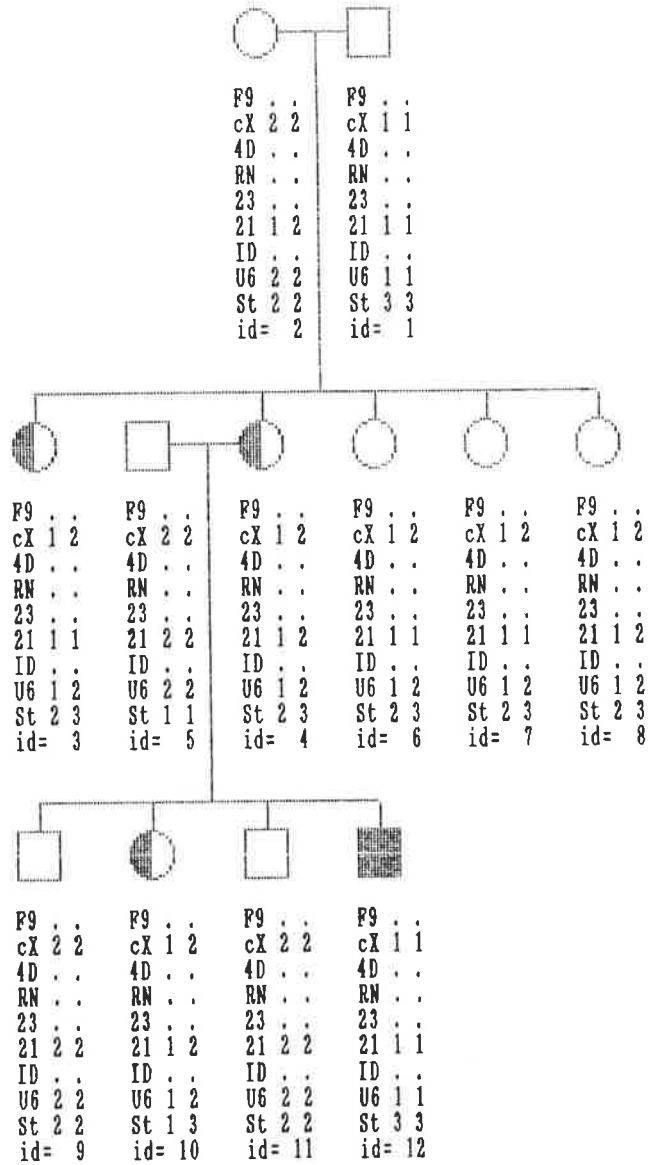
Pedigree 94





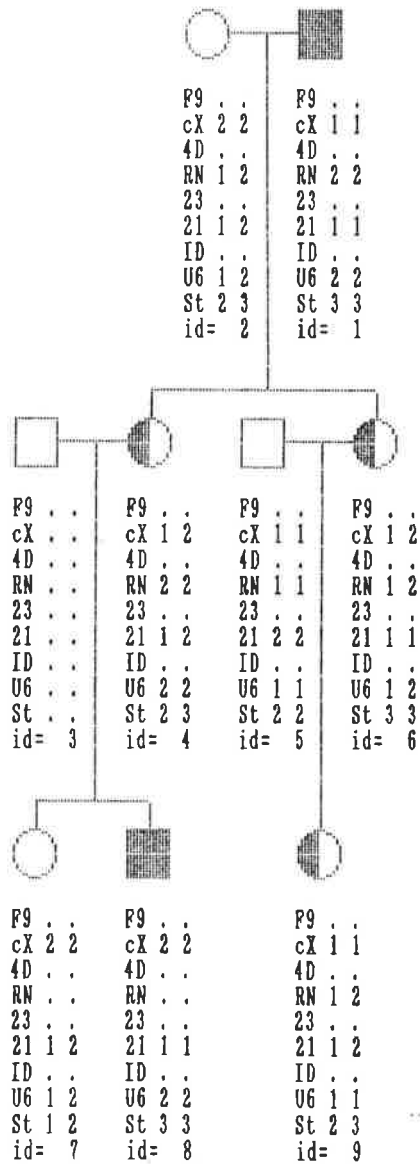
Appendix C. Fragile X pedigrees. p. C-58

Pedigree 103



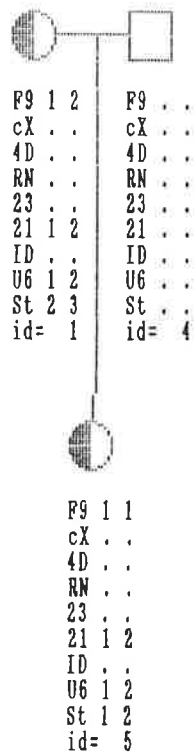
Appendix C. Fragile X pedigrees. p. C-59

Pedigree 108

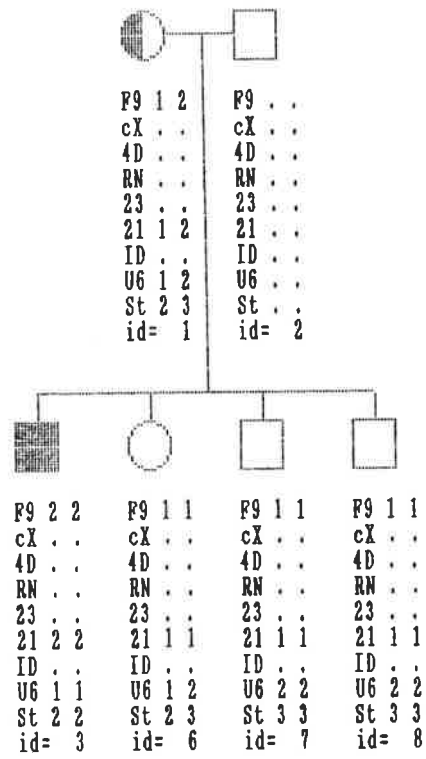




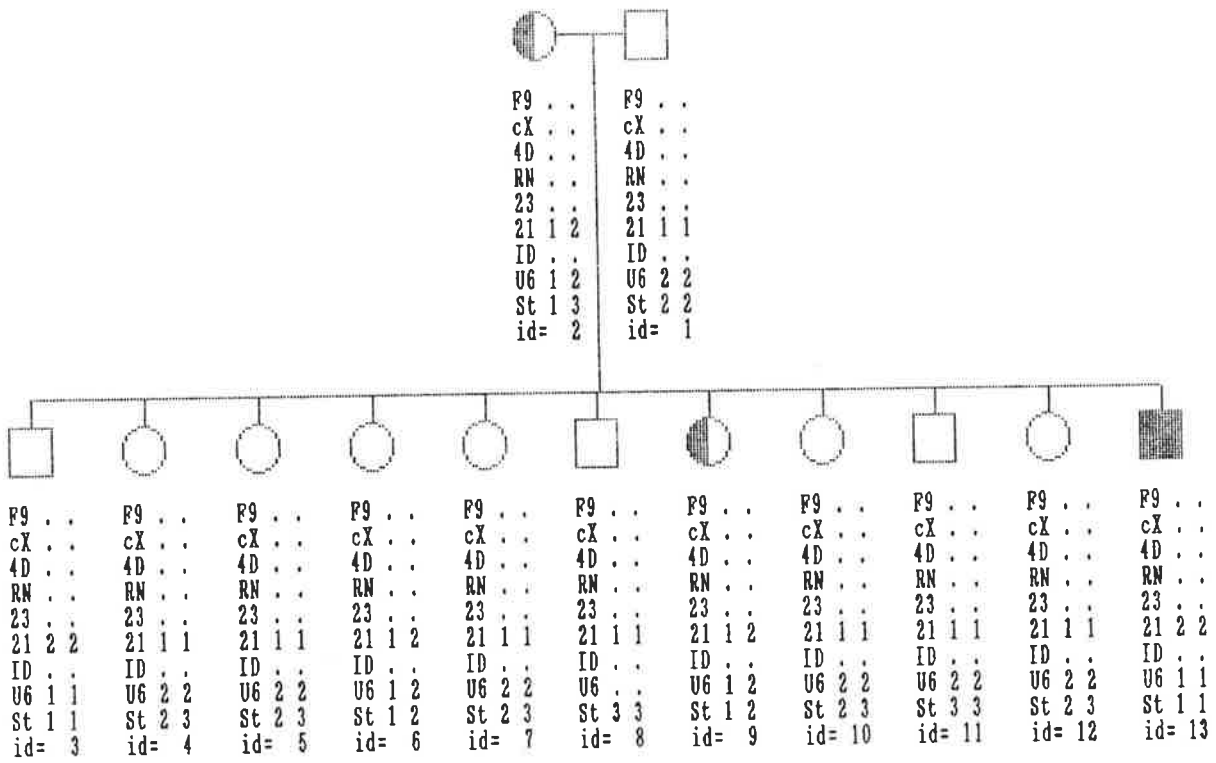
Pedigree 113



Pedigree 113

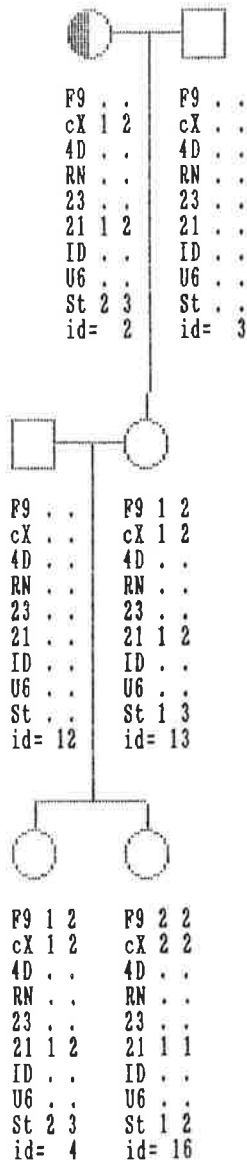


Pedigree 115

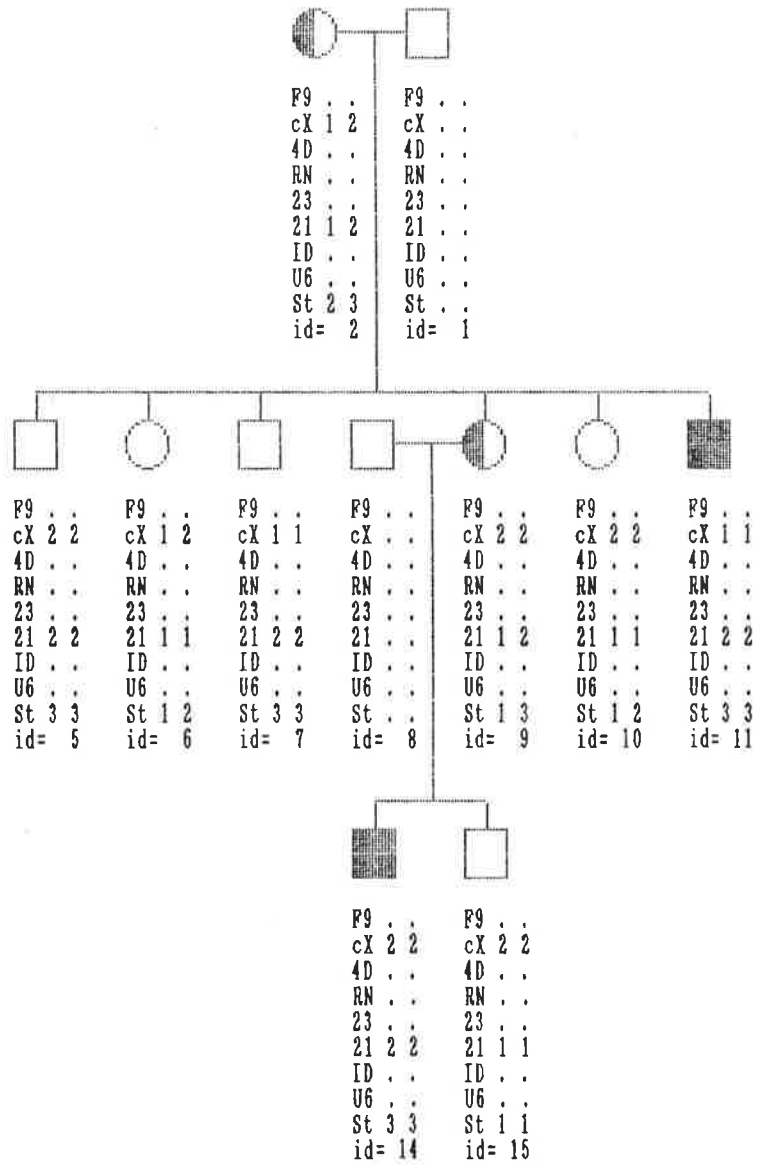


Appendix C. Fragile X pedigrees. p. C-64

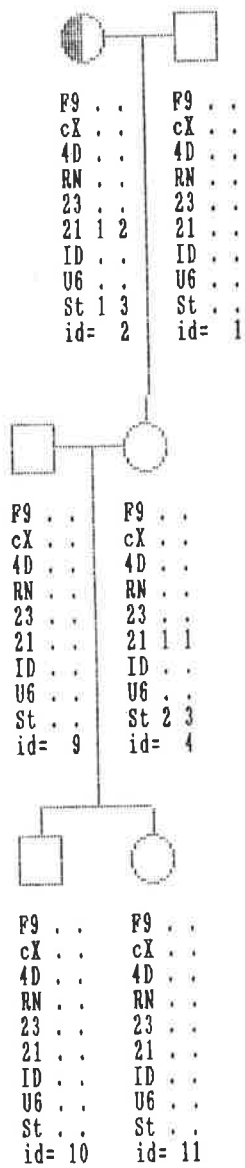
Pedigree 117



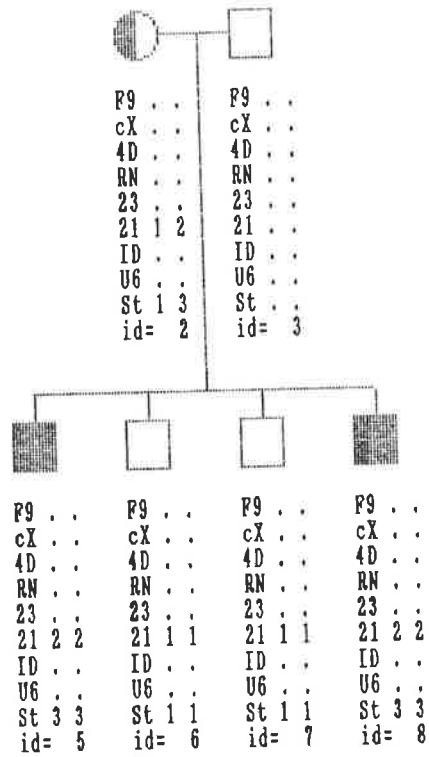
Pedigree 117



Pedigree 119

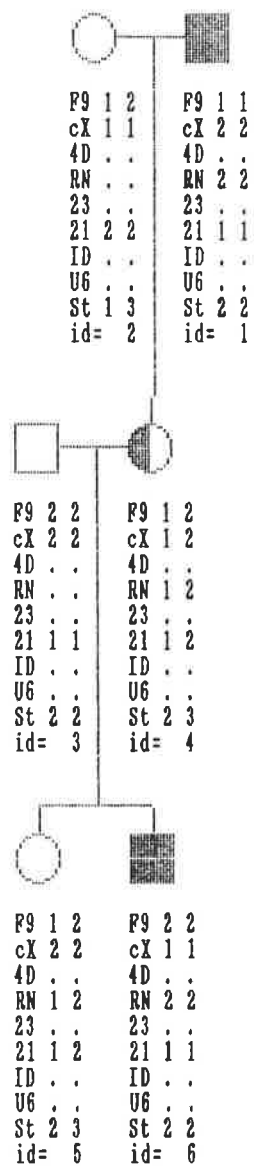


Pedigree 119

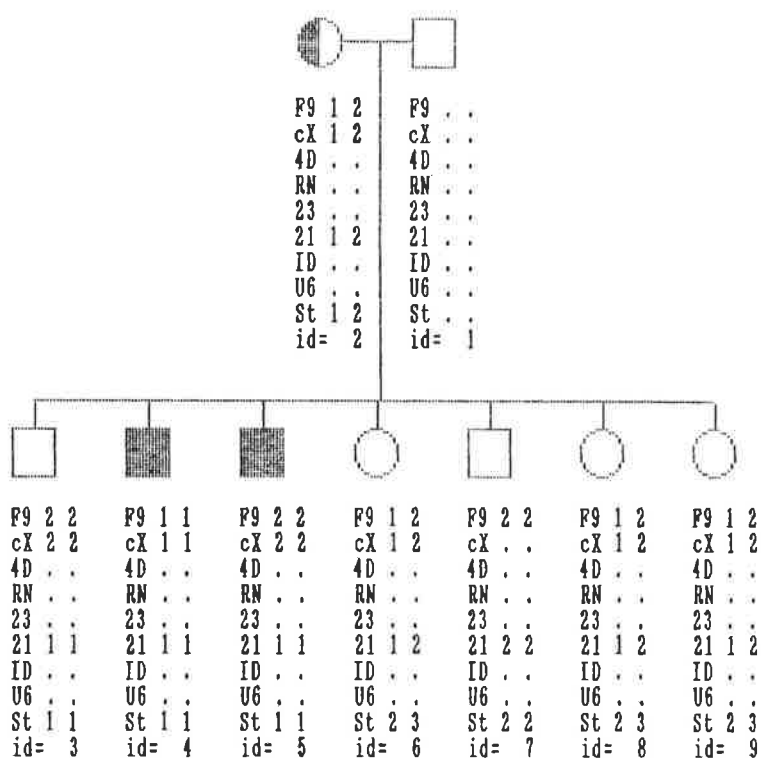




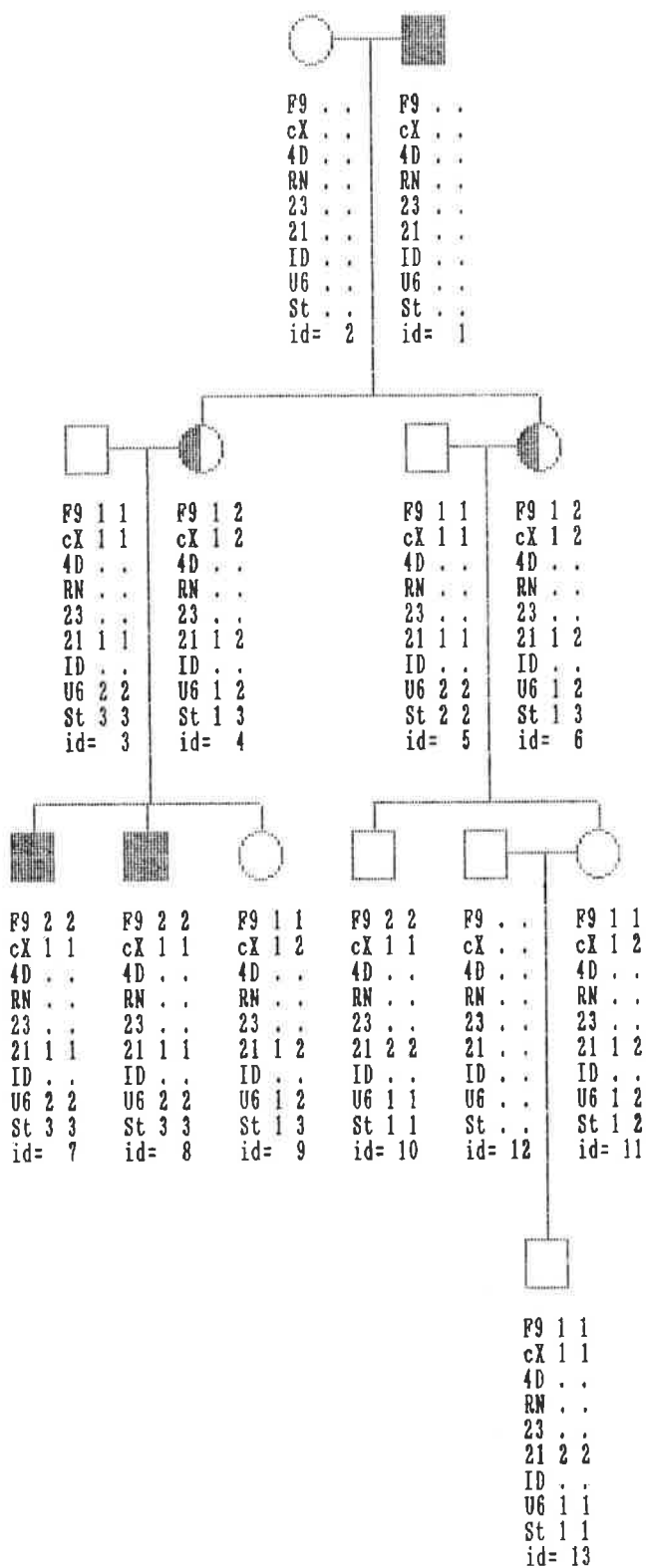
Pedigree 123



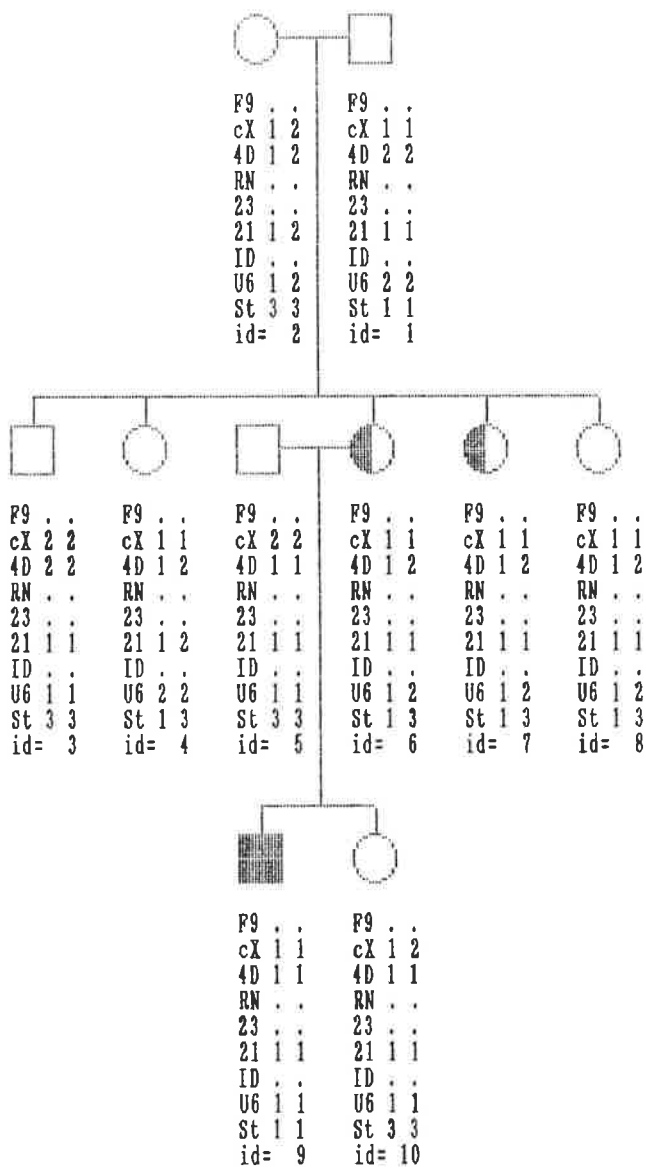
Pedigree 125



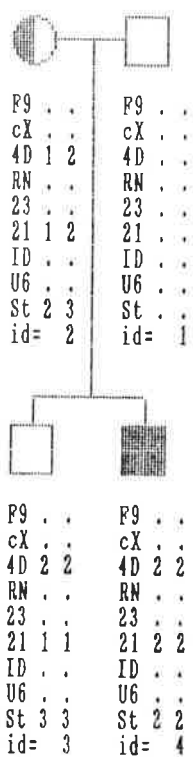
Pedigree 127



Pedigree 131

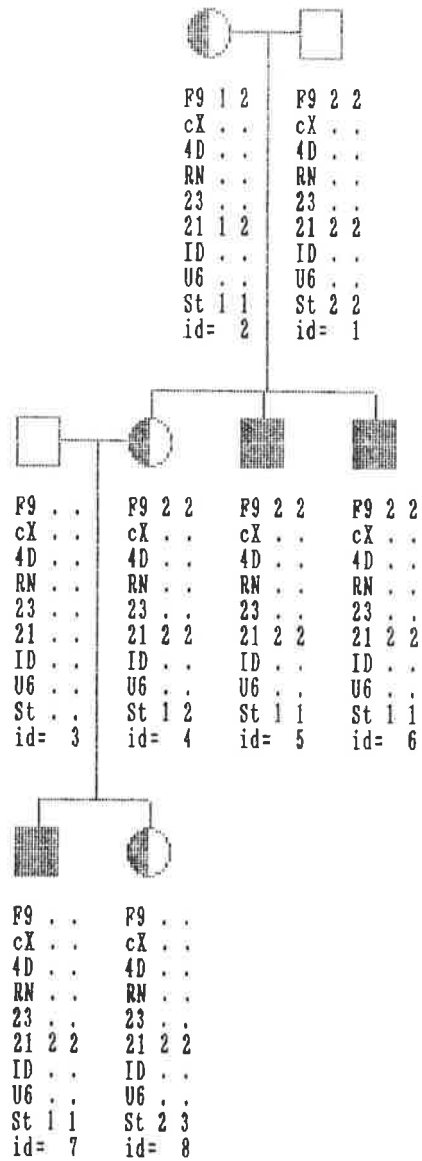


Pedigree 132

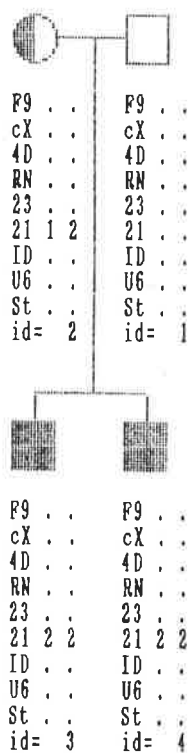


Appendix C. Fragile X pedigrees. p. C-71

Pedigree 140

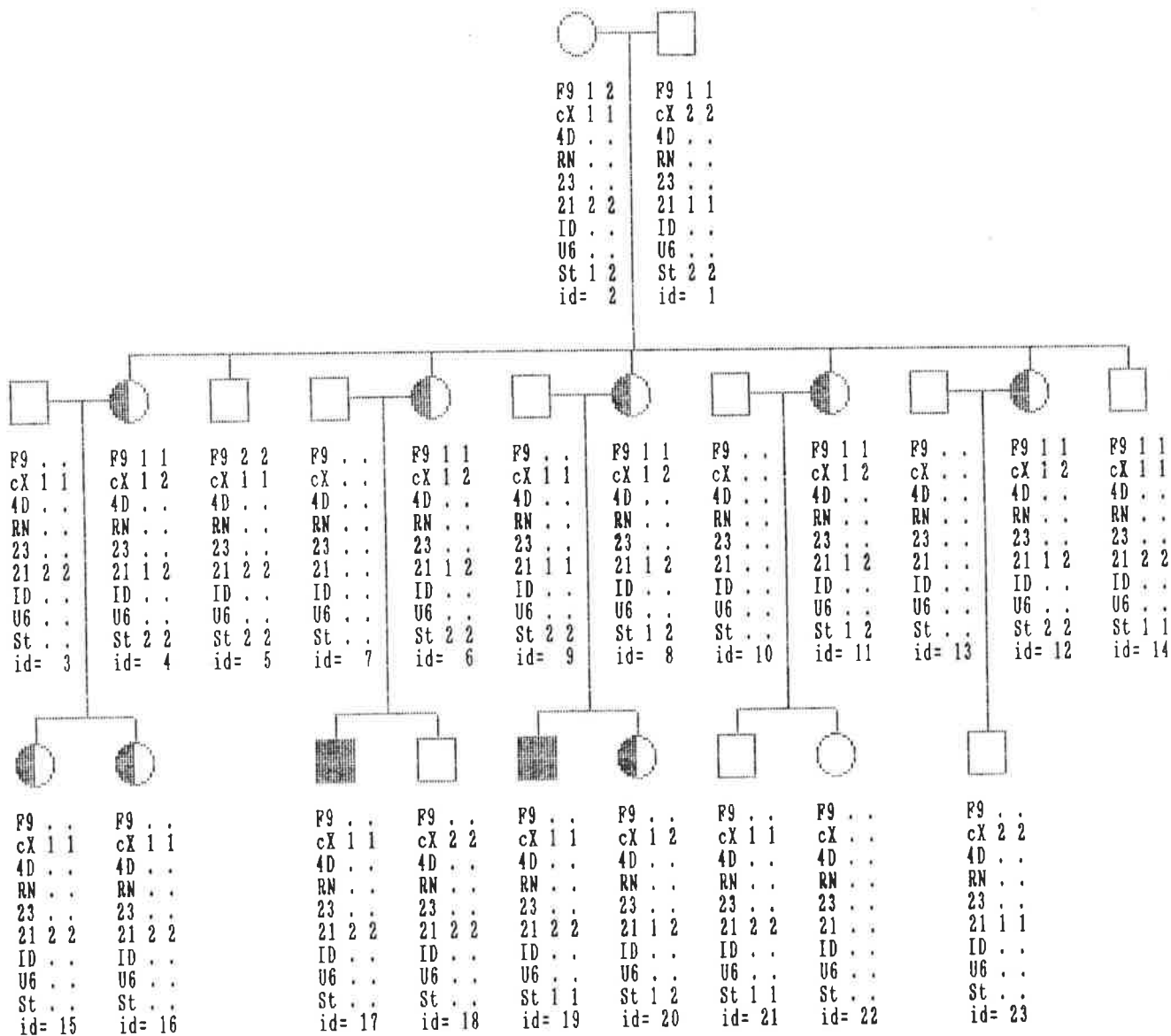


Pedigree 141



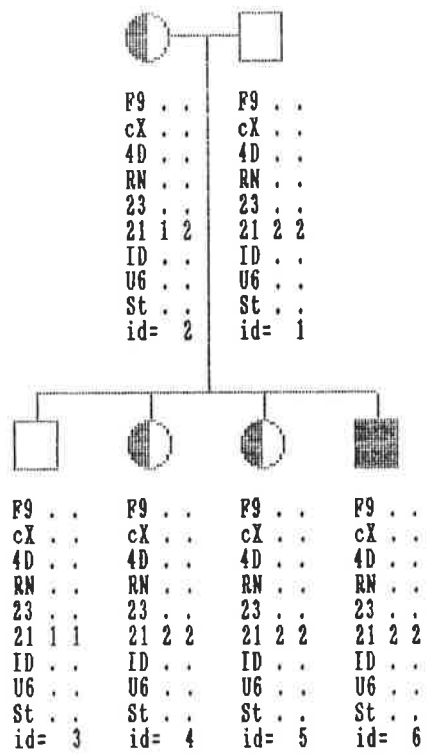
Appendix C. Fragile X pedigrees. p. C-73

Pedigree 142

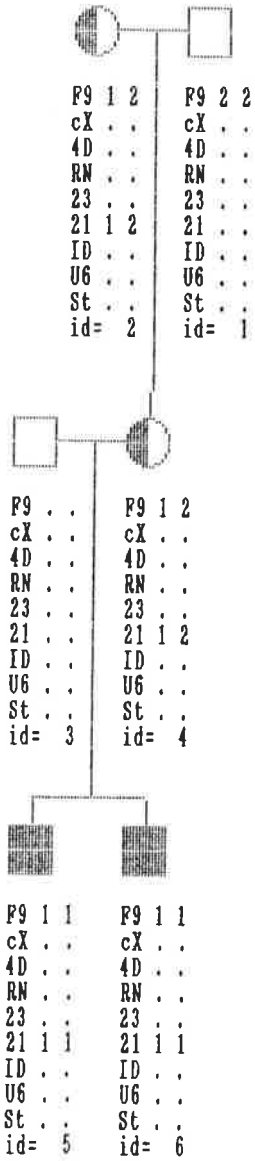




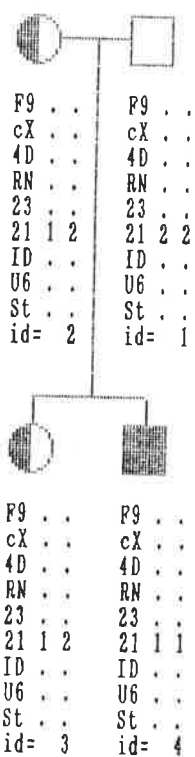
Pedigree 143



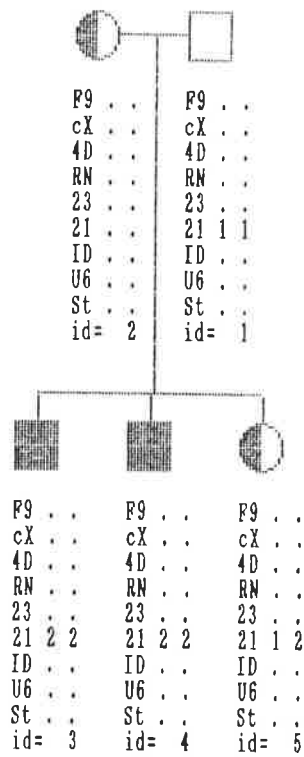
Pedigree 144



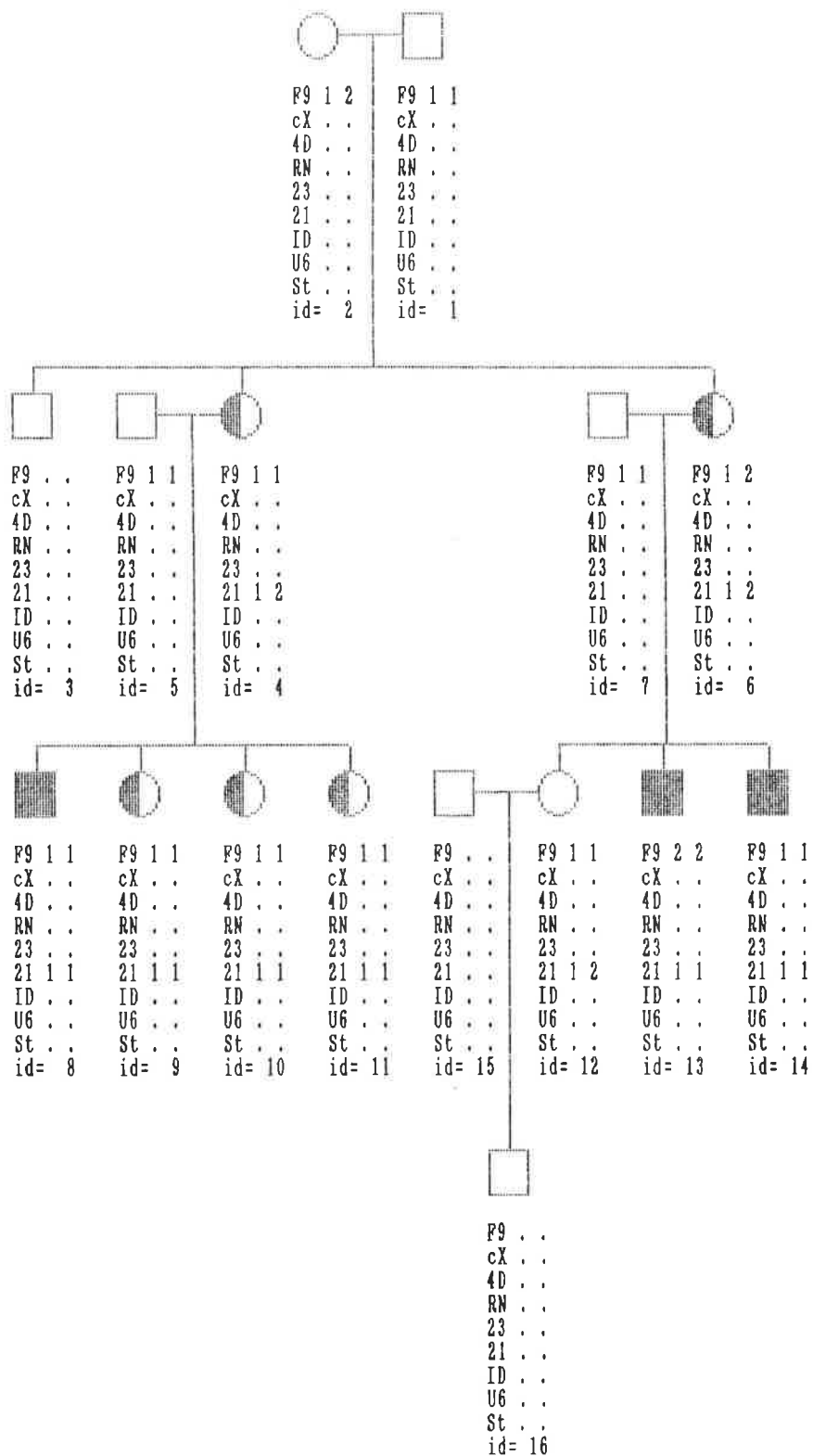
Pedigree 145



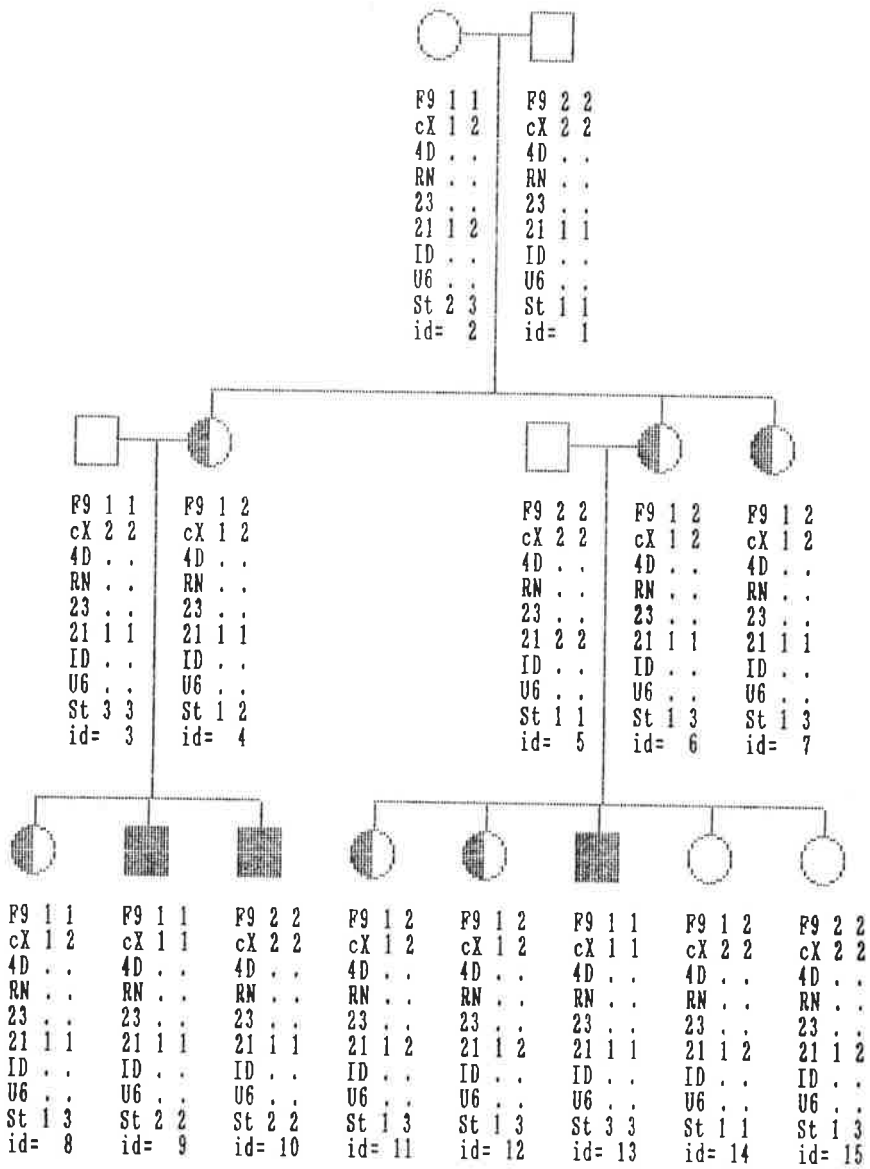
Pedigree 146



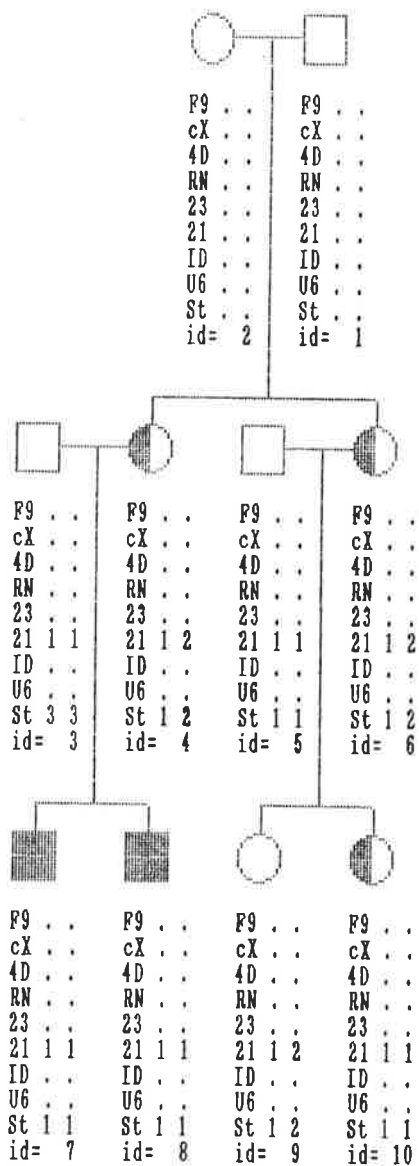
Pedigree 151



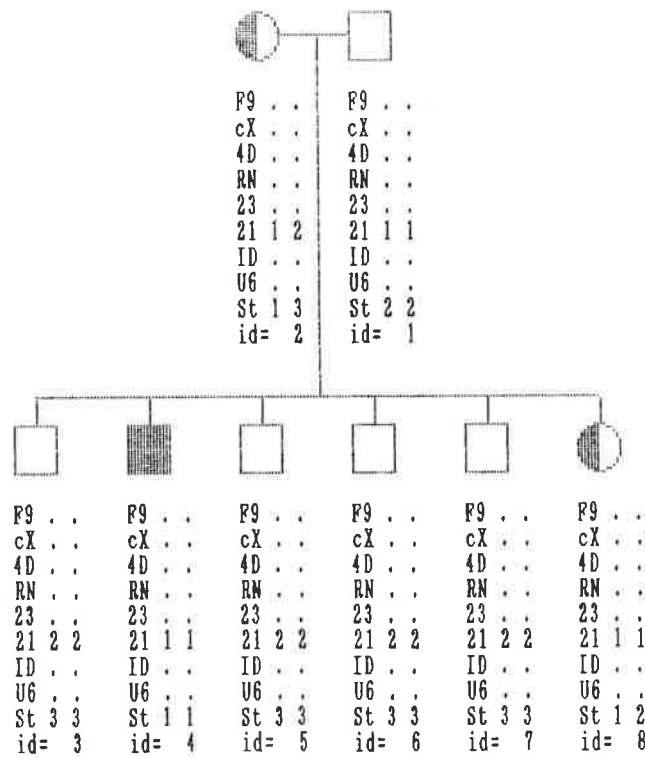
Pedigree 152



Pedigree 153

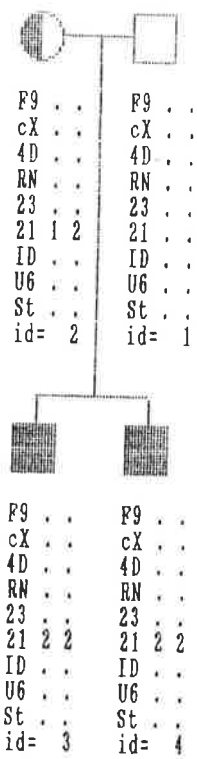


Pedigree 155



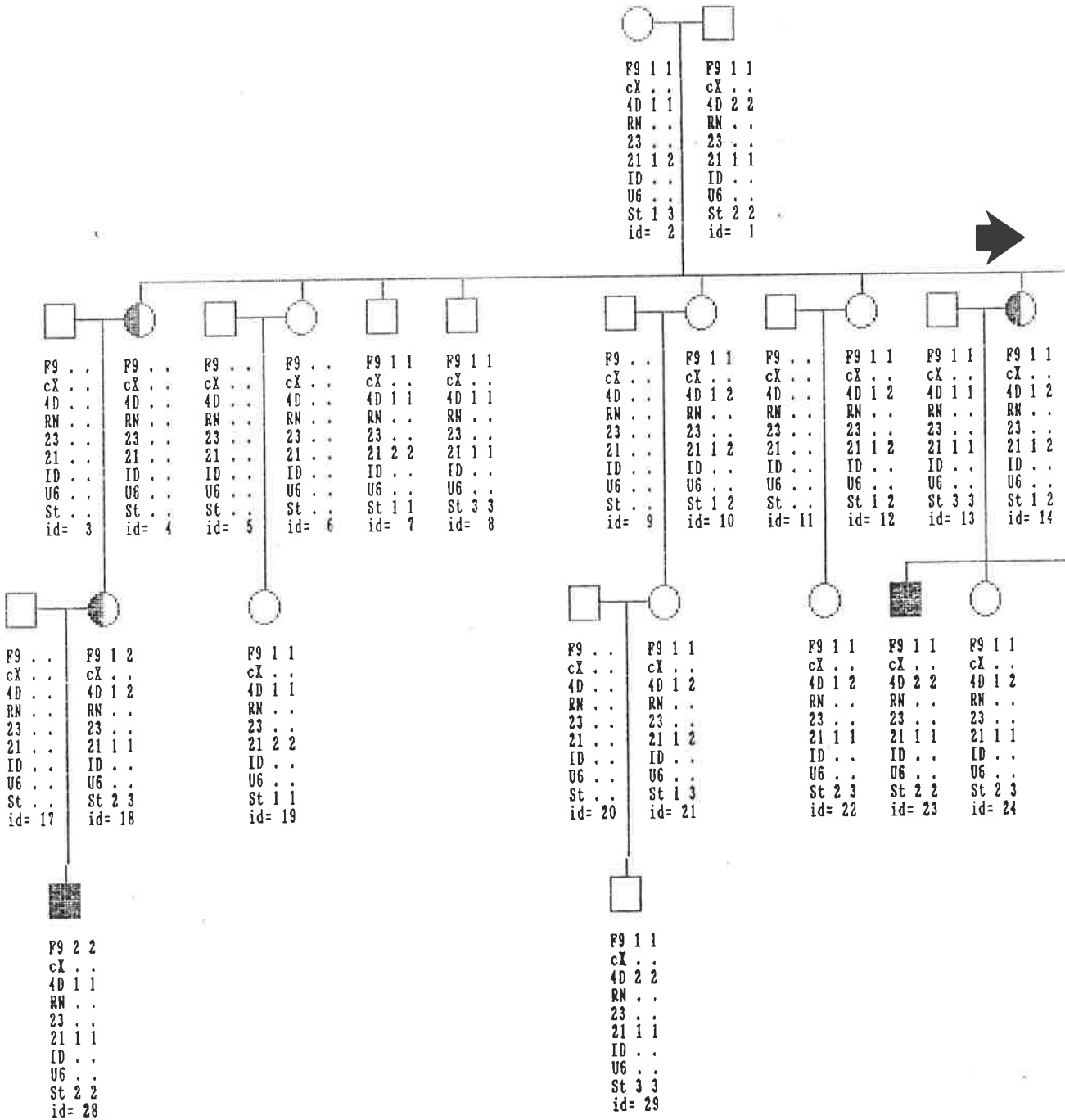


Pedigree 156

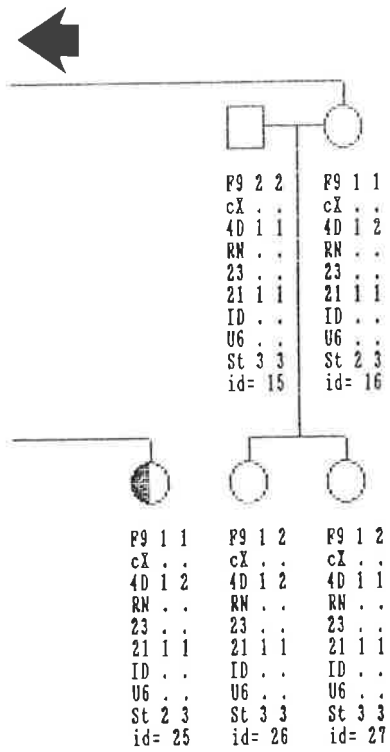


Appendix C. Fragile X pedigrees. p. C-83

Pedigree 157

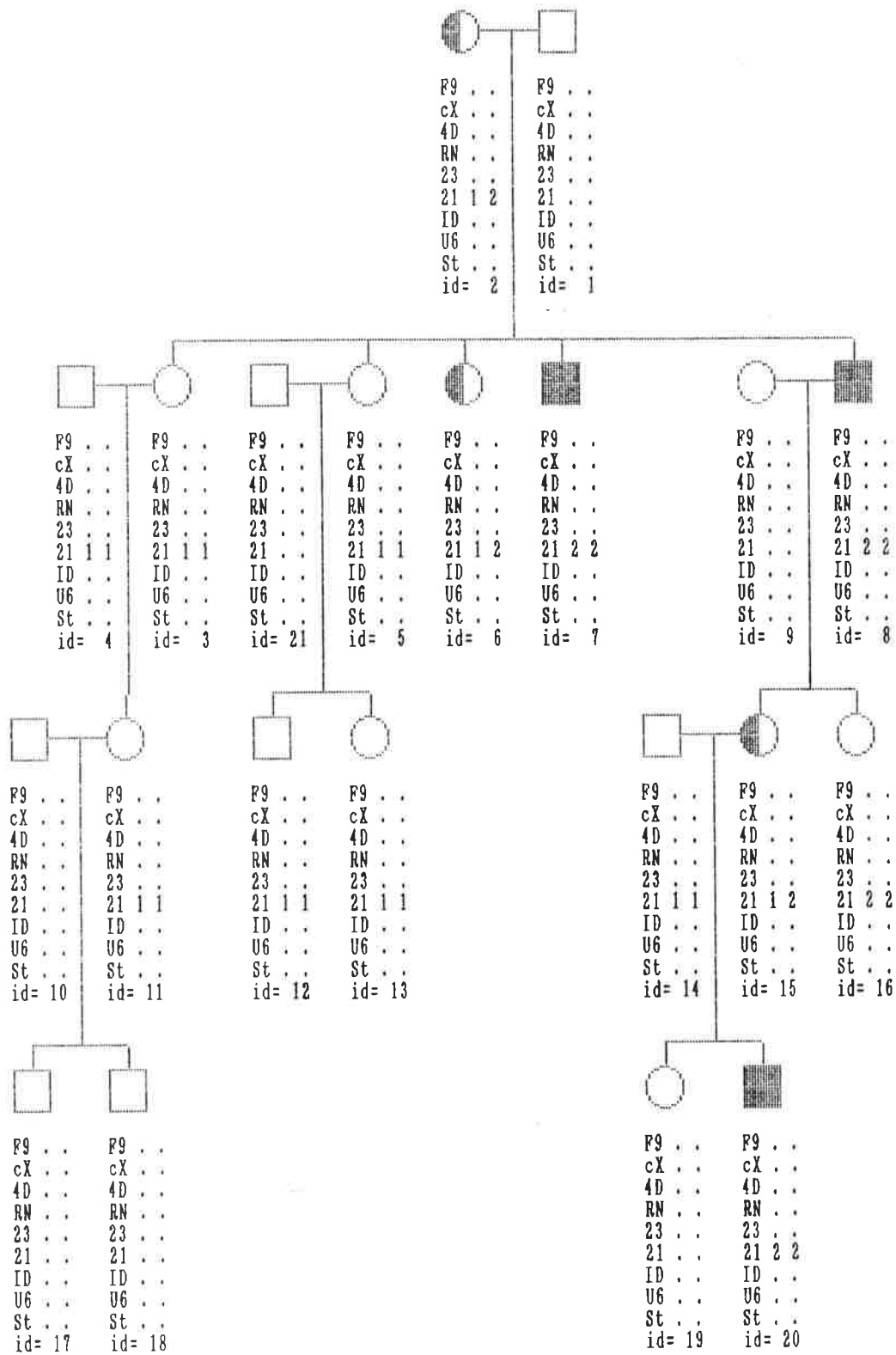


Appendix C. Fragile X pedigrees. p. C-84

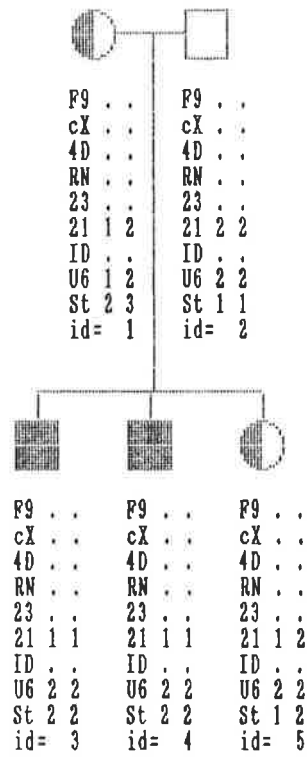


Appendix C. Fragile X pedigrees. p. C-85

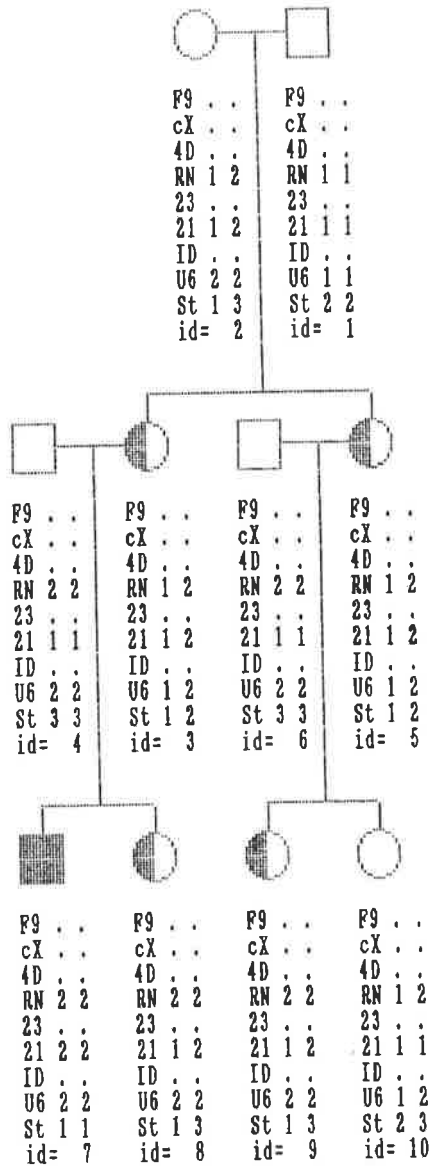
Pedigree 160



Pedigree 161

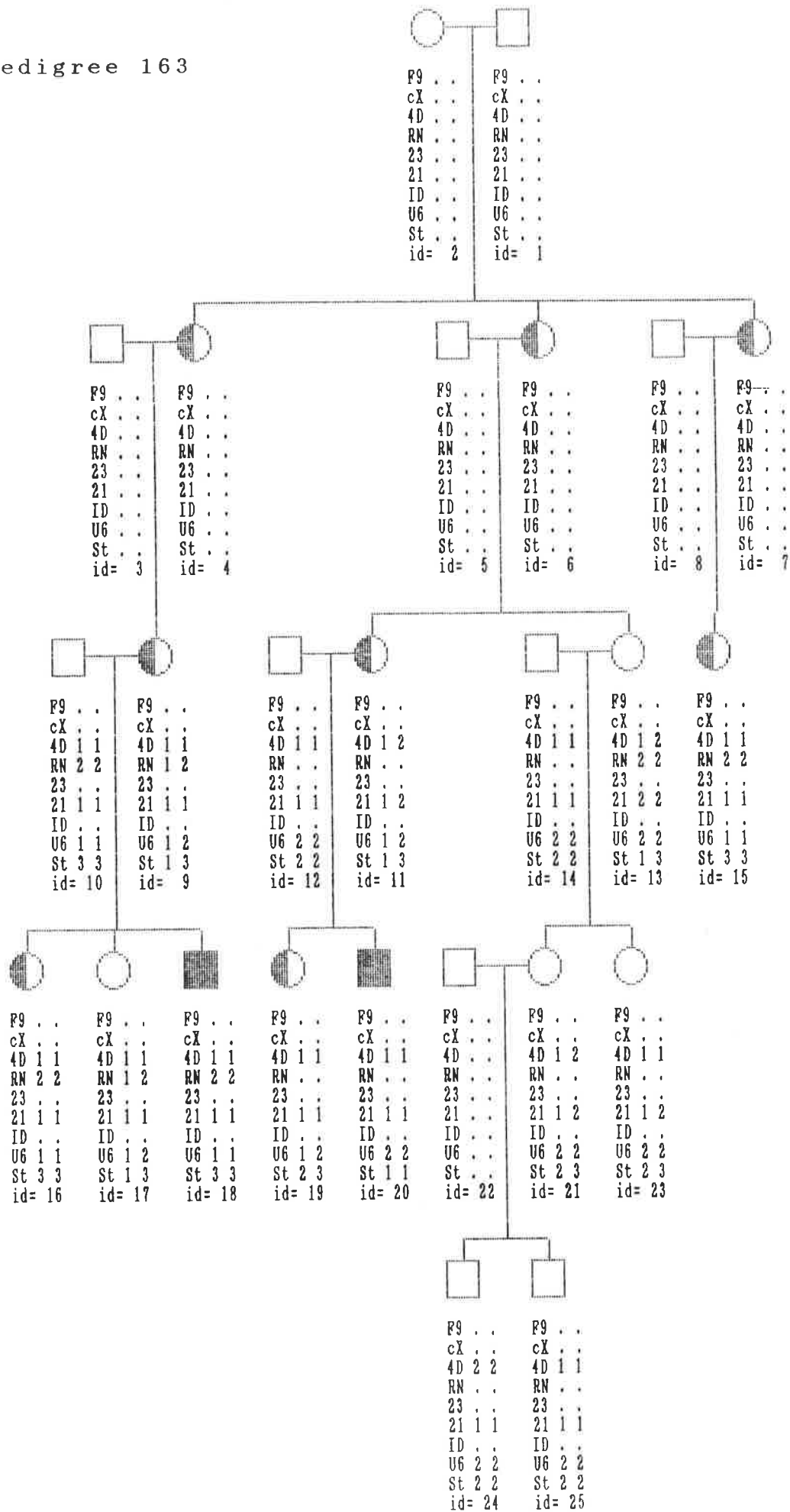


Pedigree 162



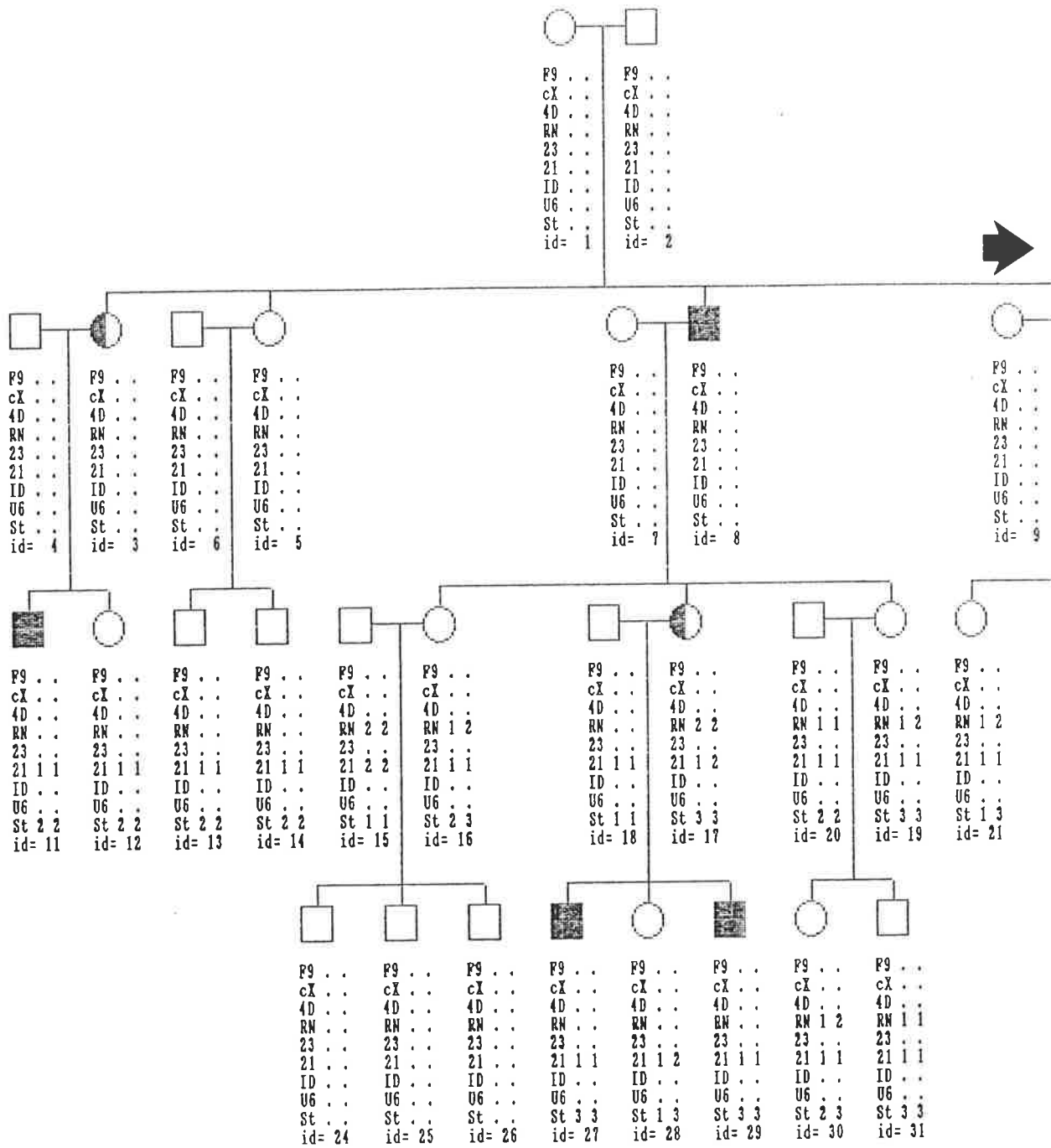
Appendix C. Fragile X pedigrees. p. C-88

Pedigree 163



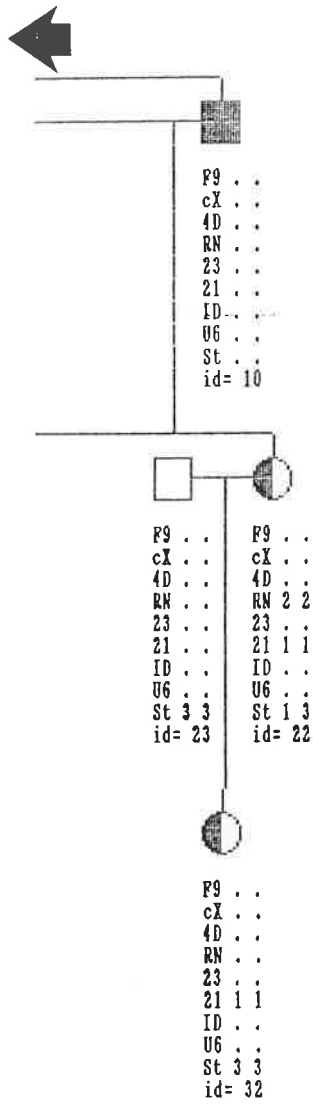
Appendix C. Fragile X pedigrees. p. C-89

Pedigree 164

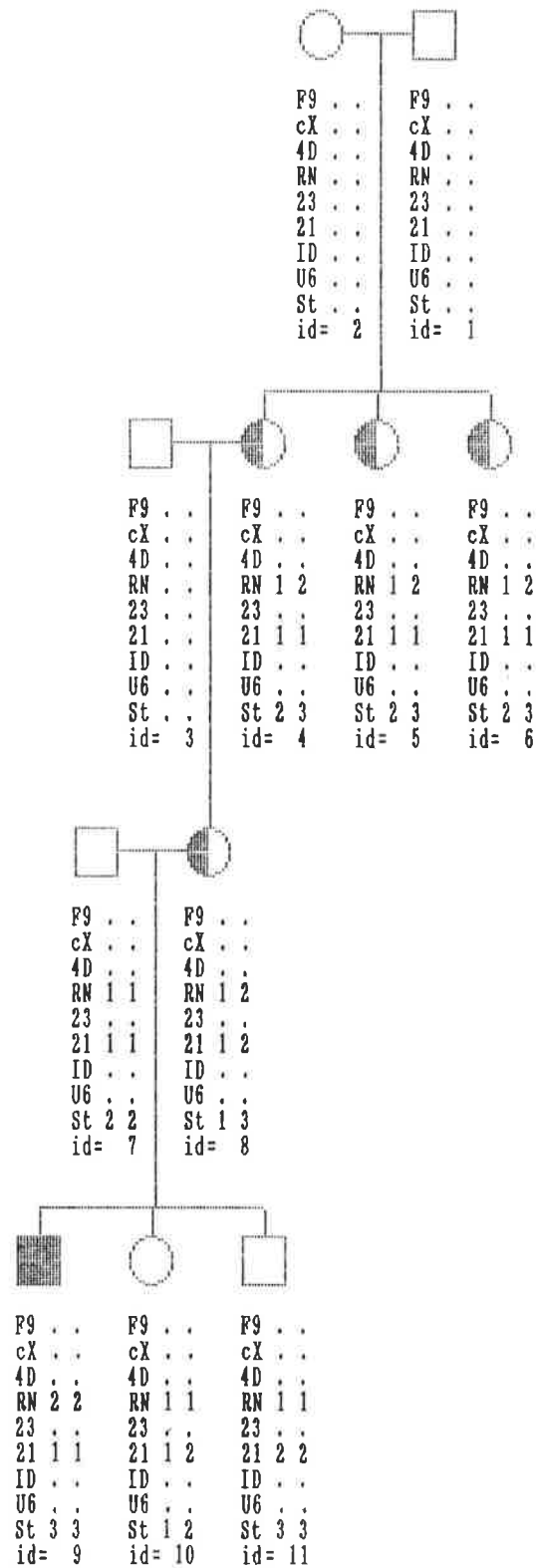




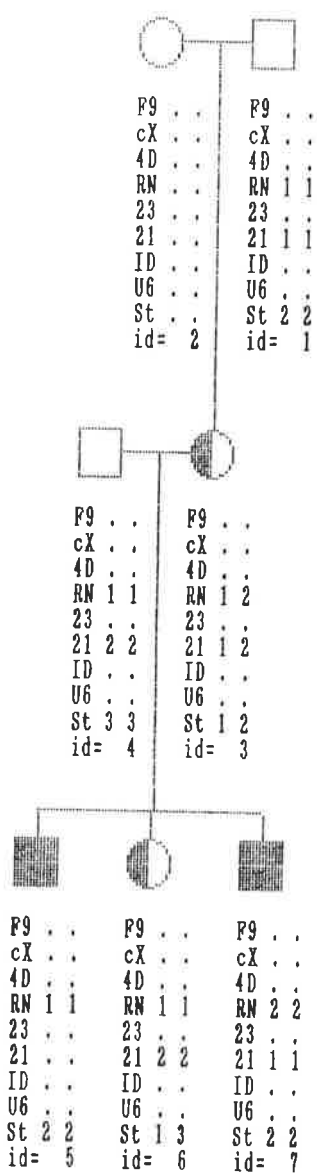
Appendix C. Fragile X pedigrees. p. C-90



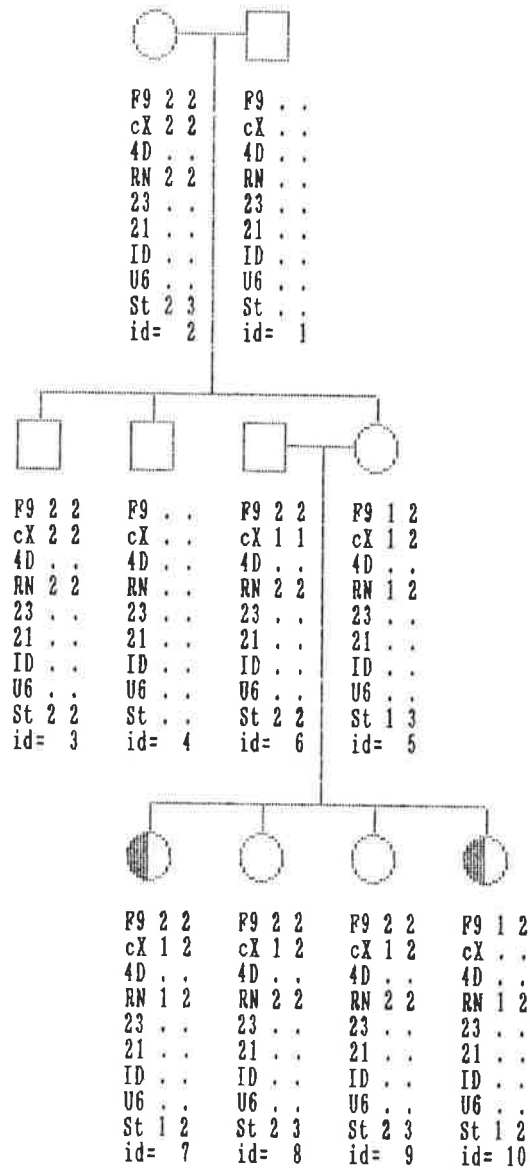
Pedigree 165



Pedigree 201

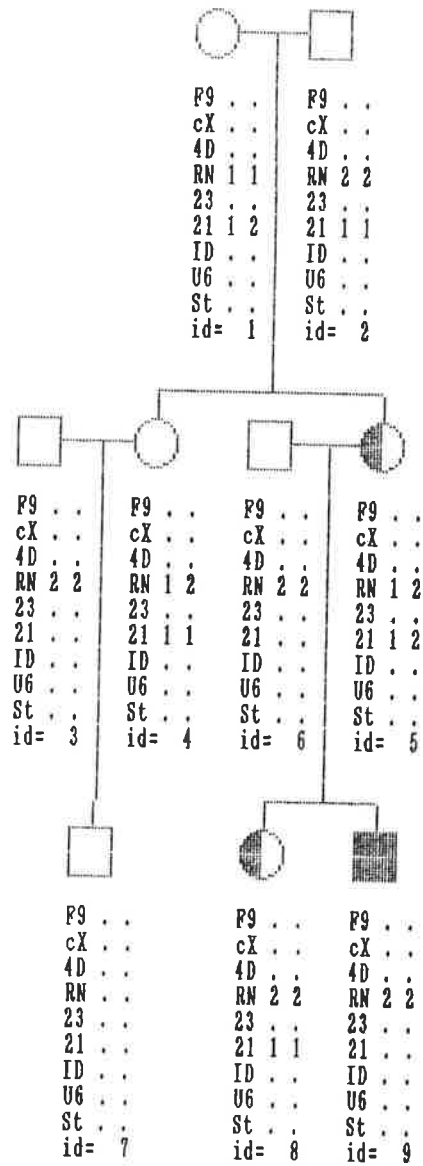


Pedigree 202



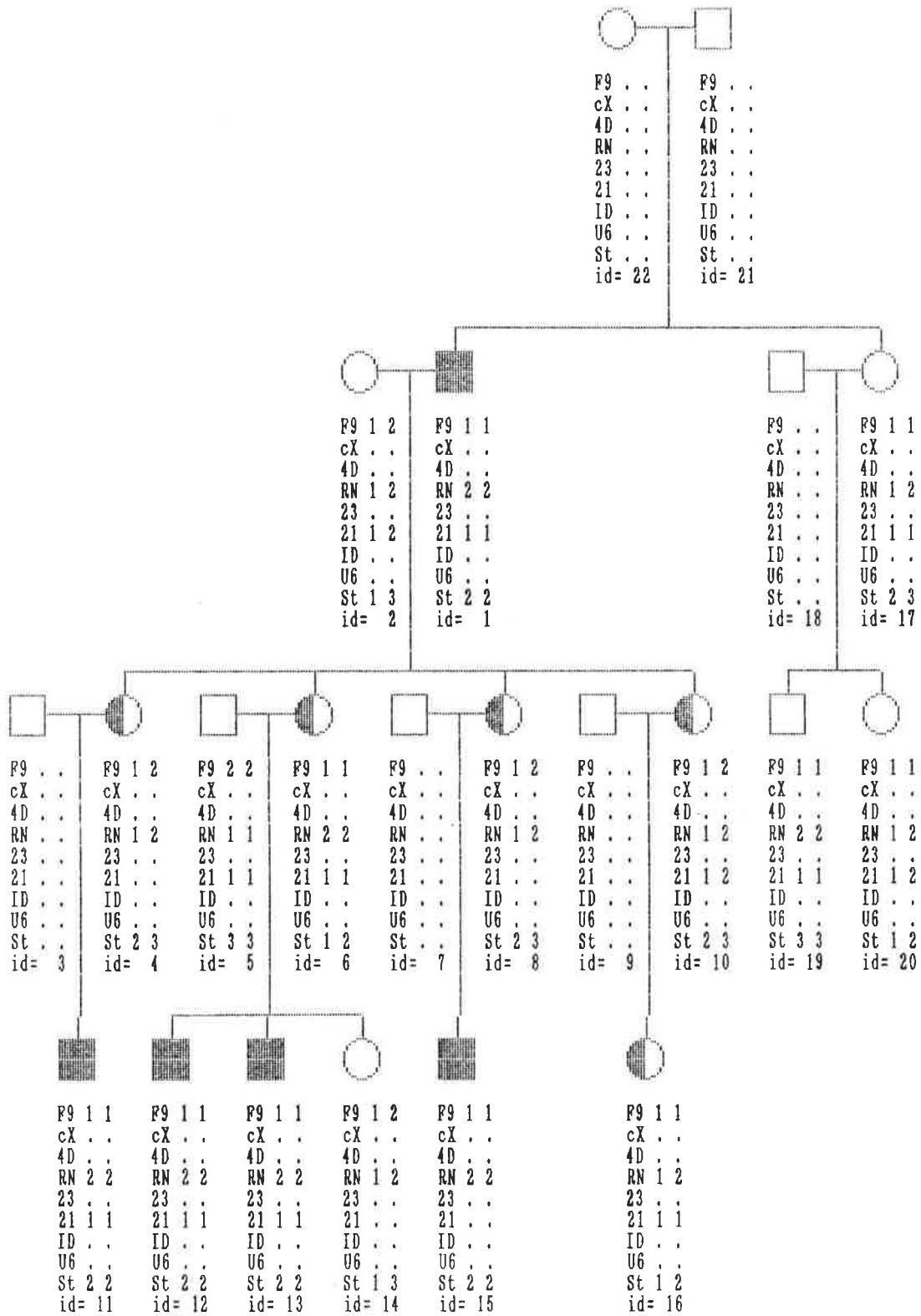


Pedigree 204

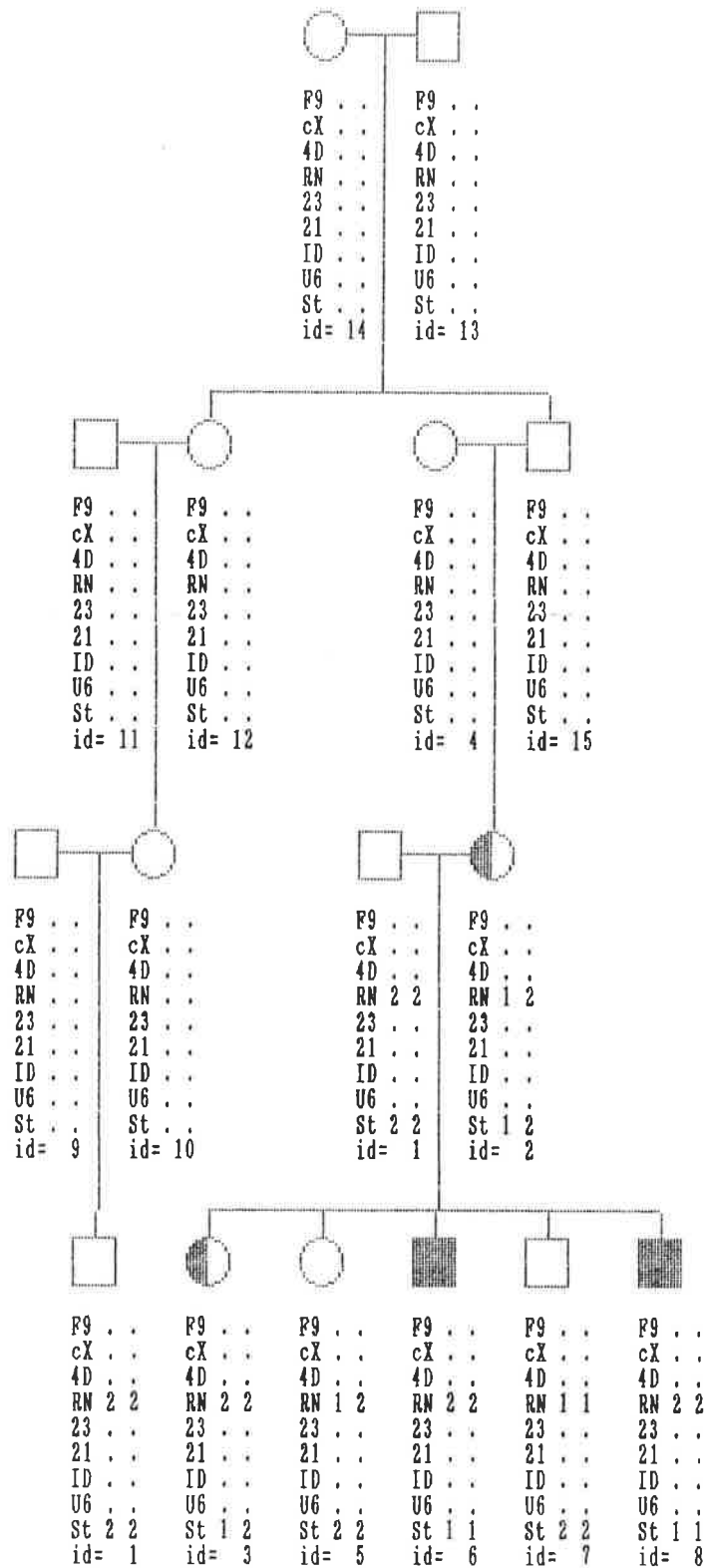


Appendix C. Fragile X pedigrees. p. C-96

Pedigree 205

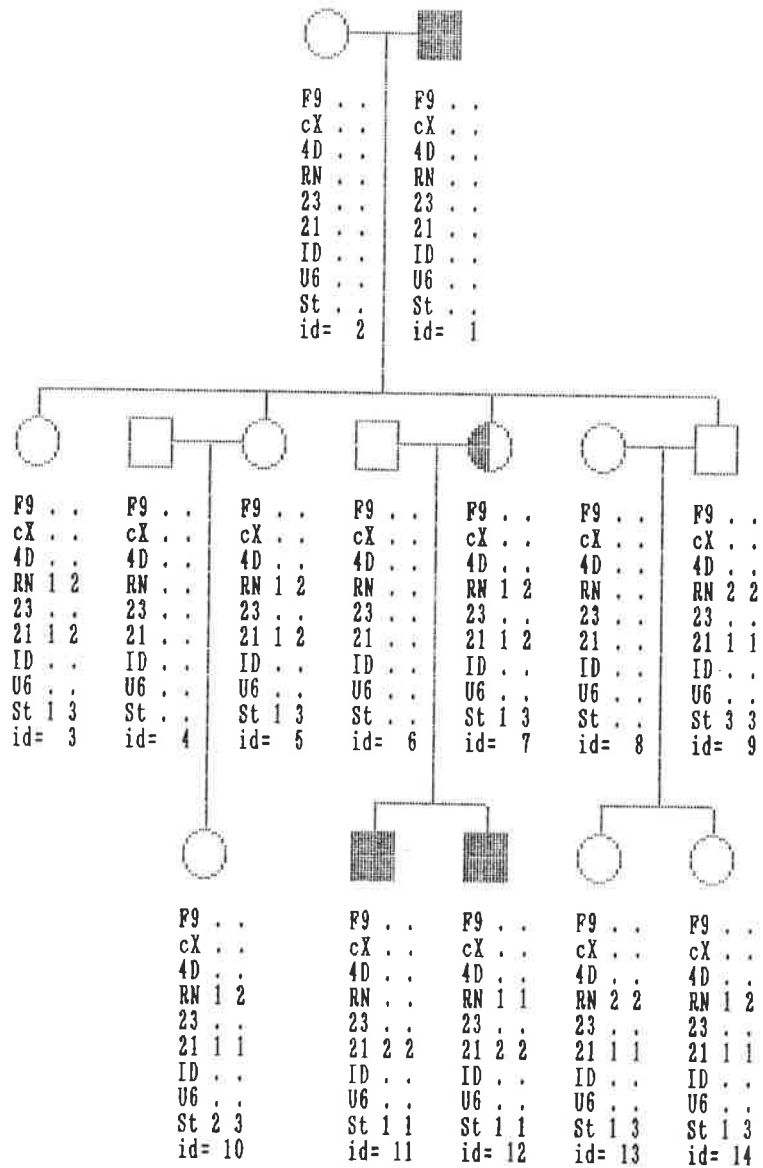


Pedigree 208

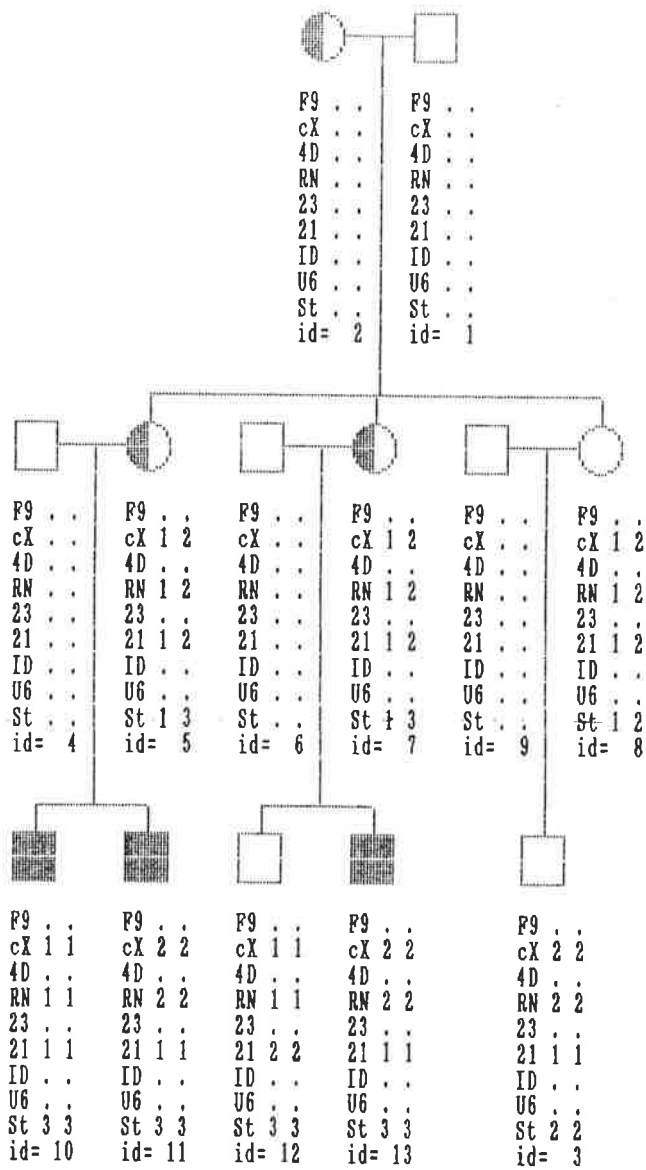




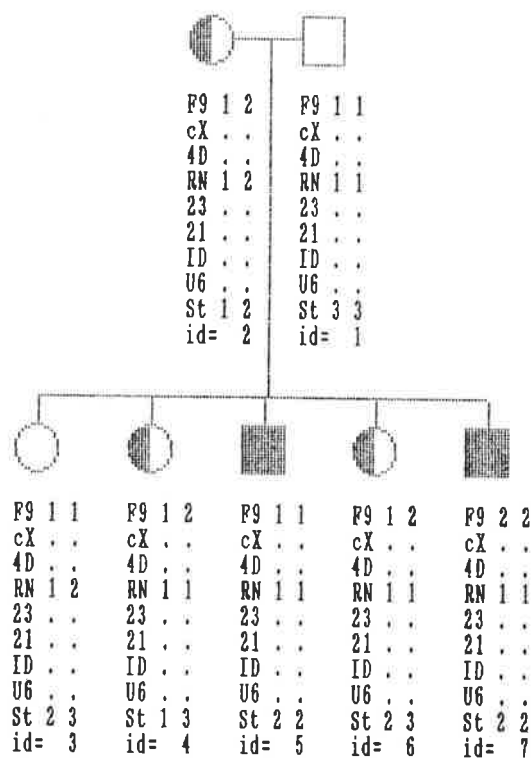
Pedigree 209



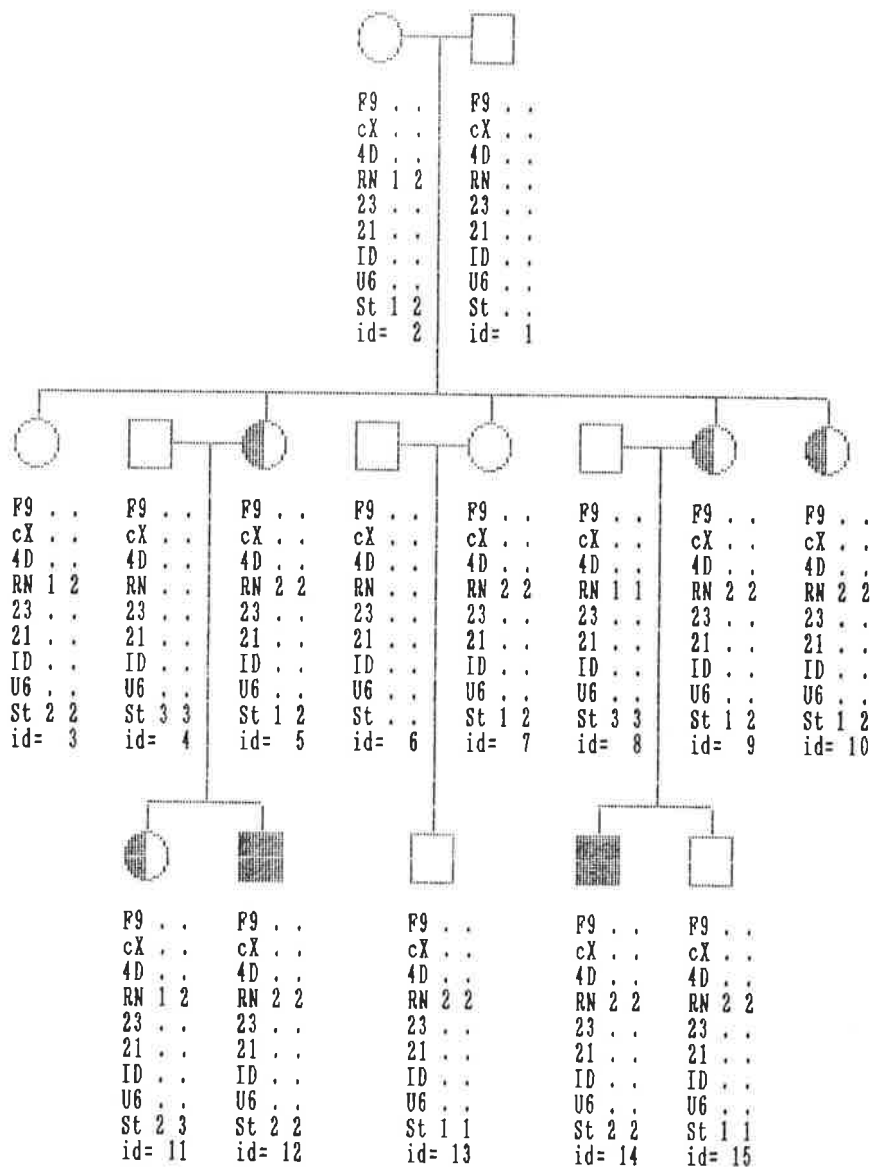
Pedigree 210



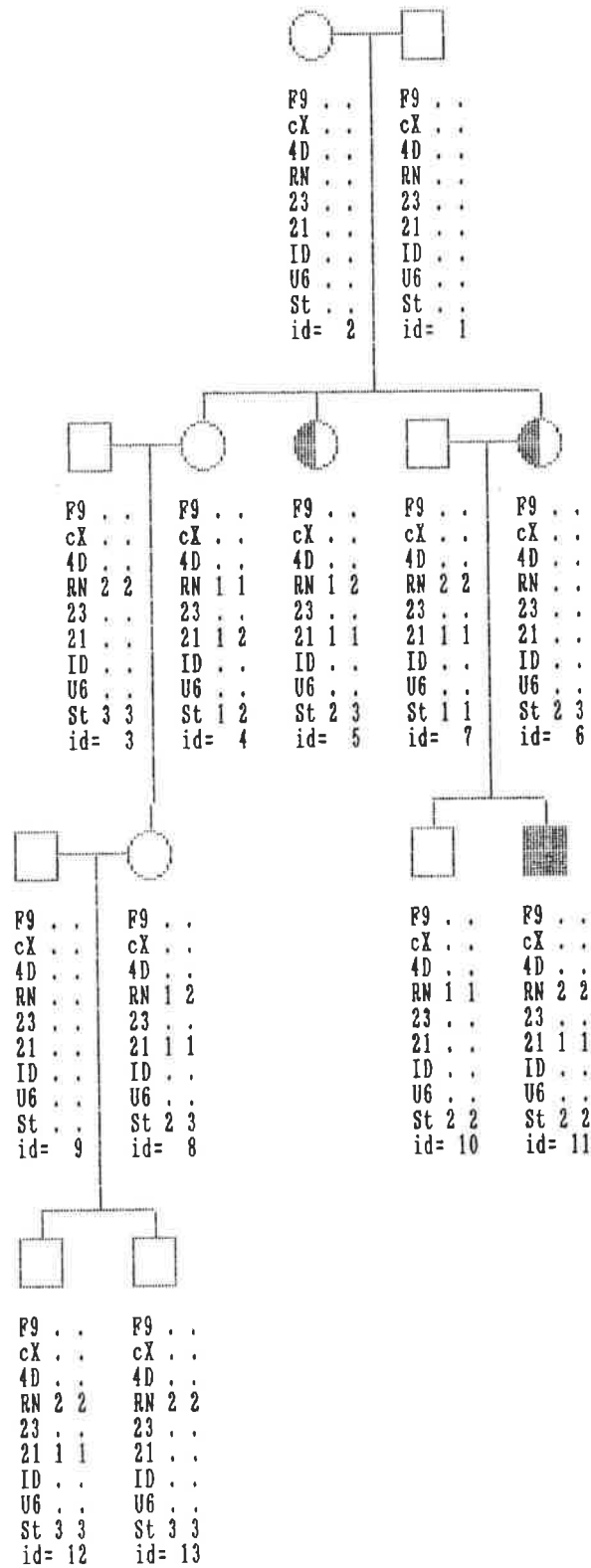
Pedigree 212



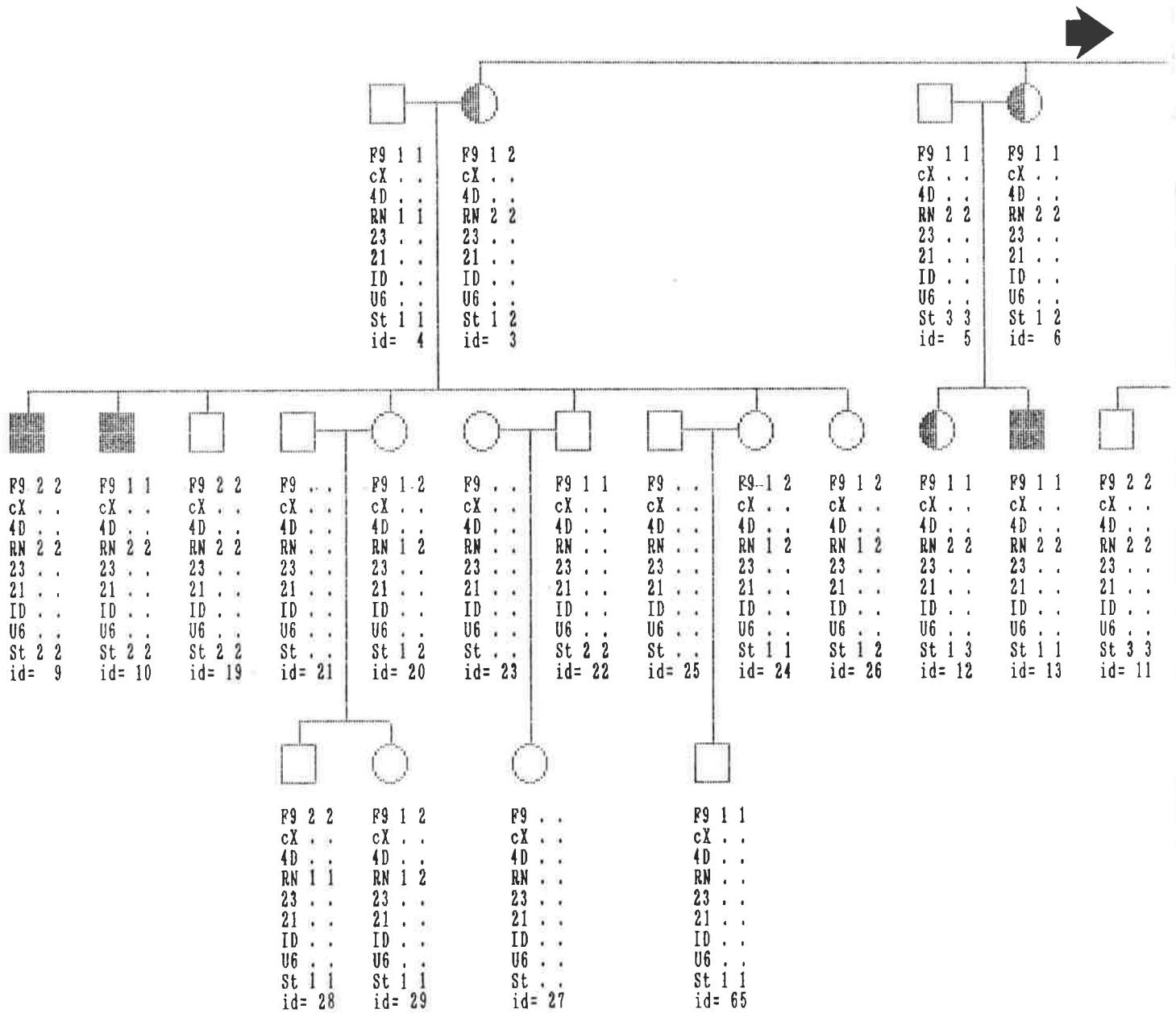
Pedigree 213



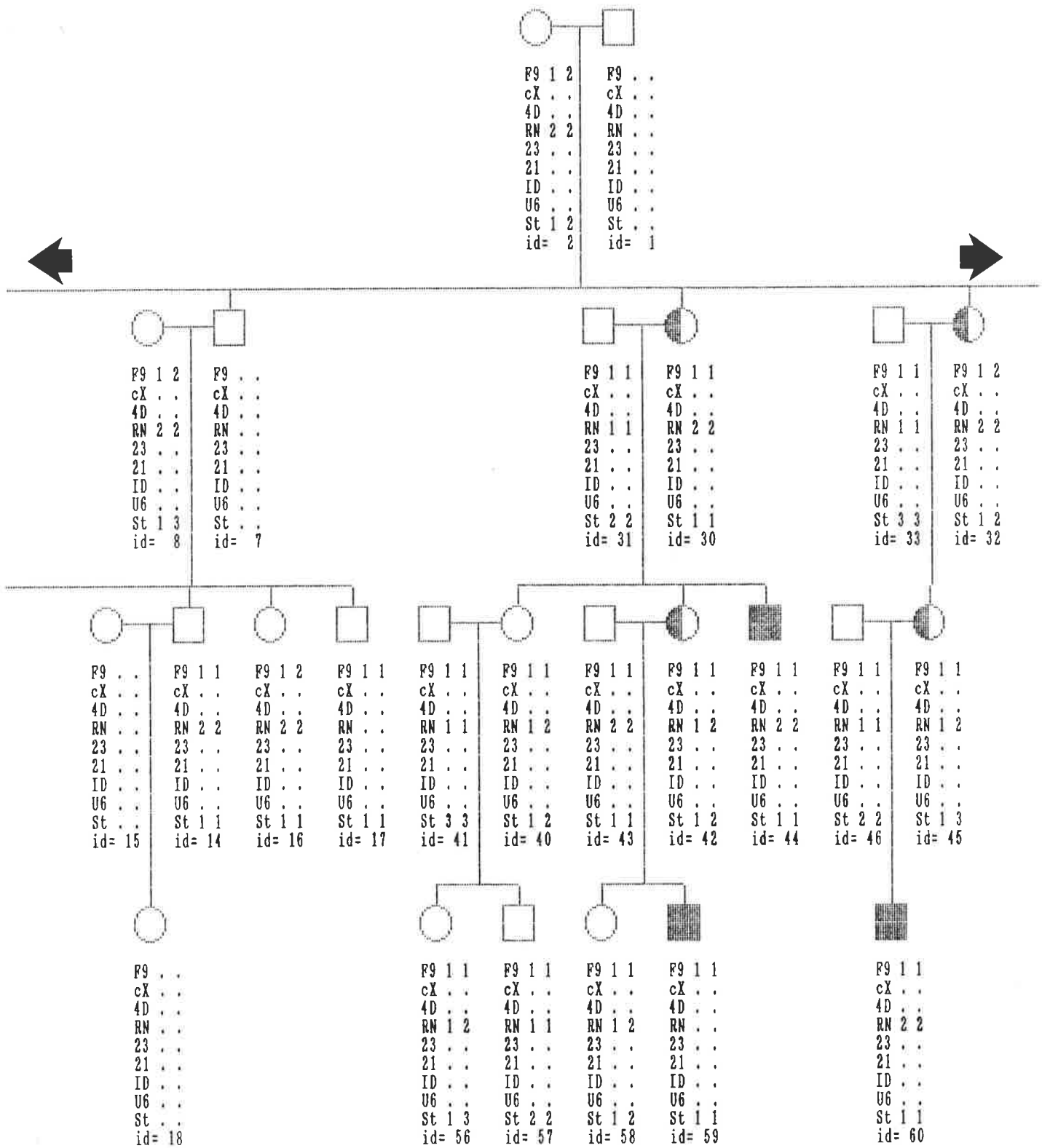
Pedigree 214

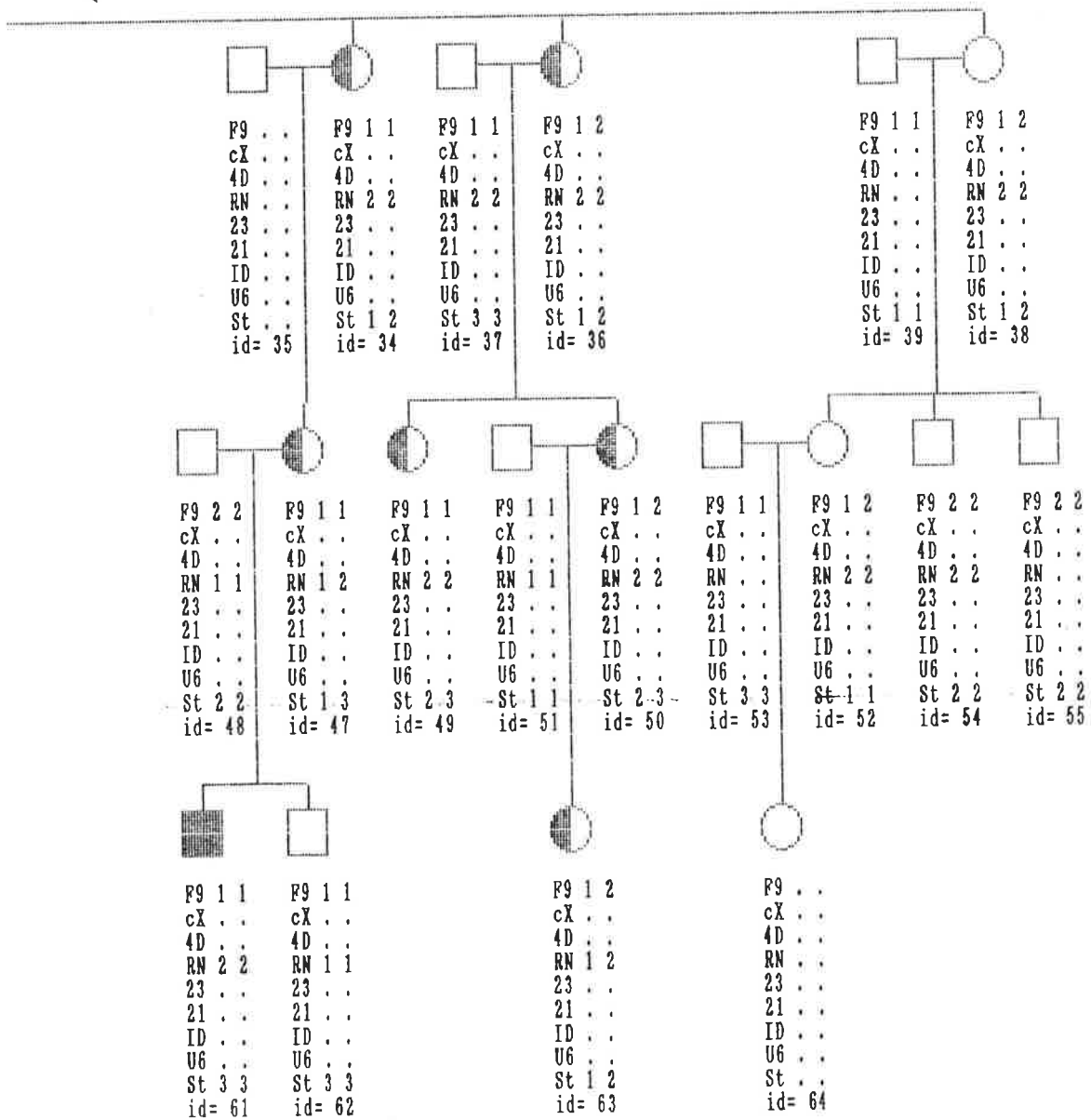


Pedigree 216



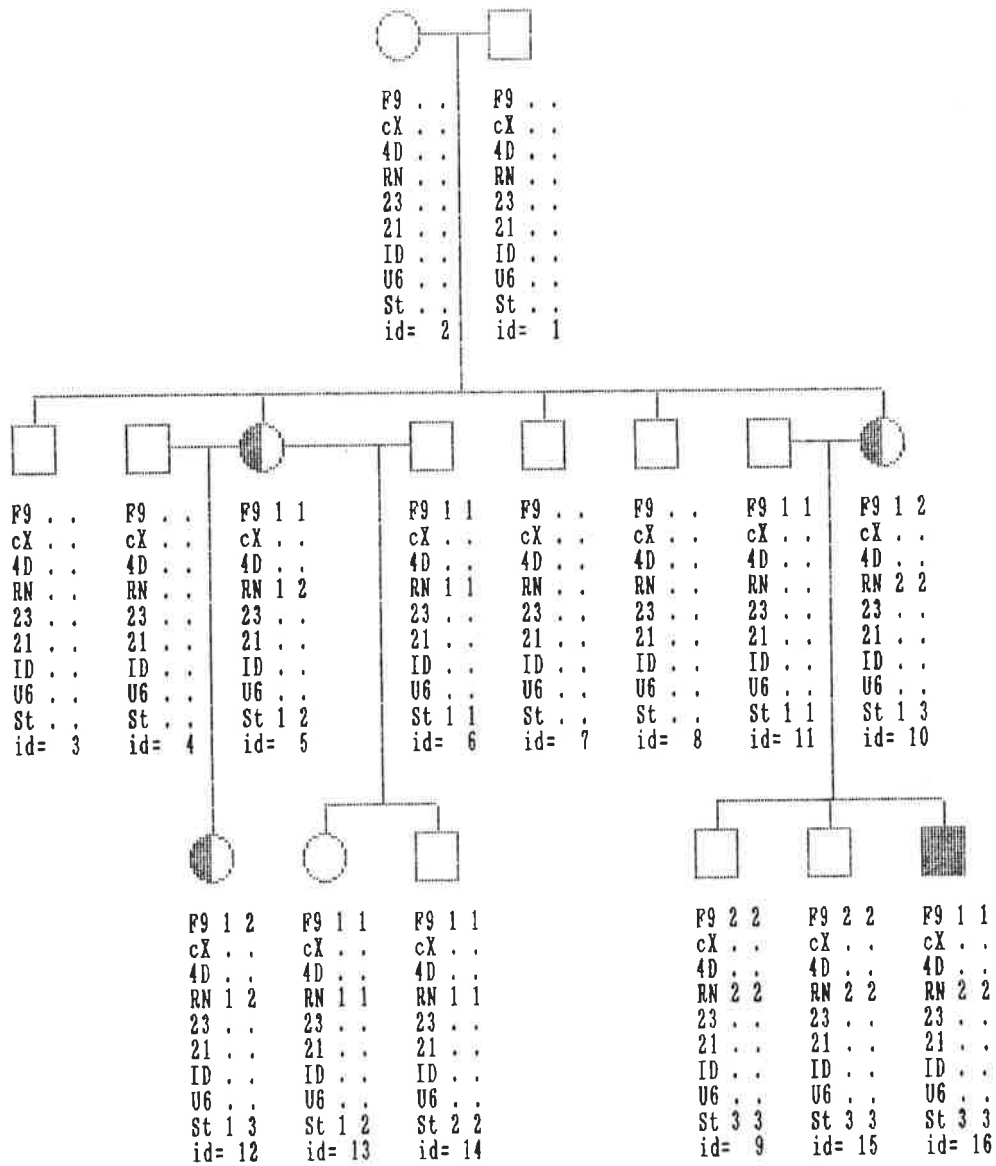
Appendix C. Fragile X pedigrees. p. C-104



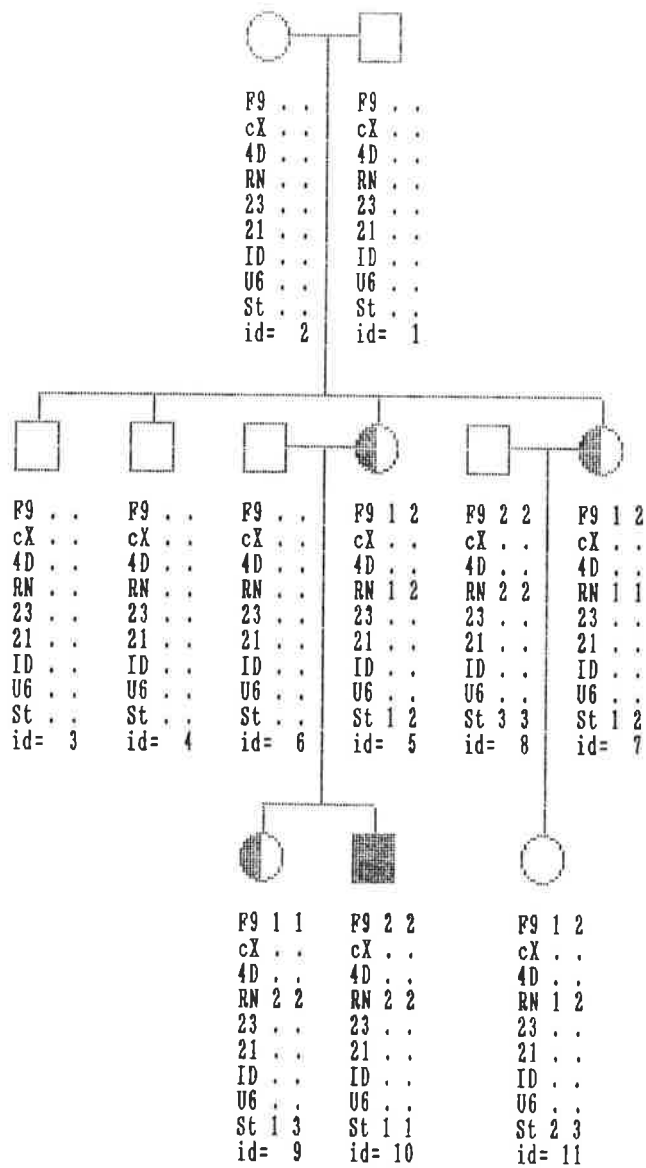




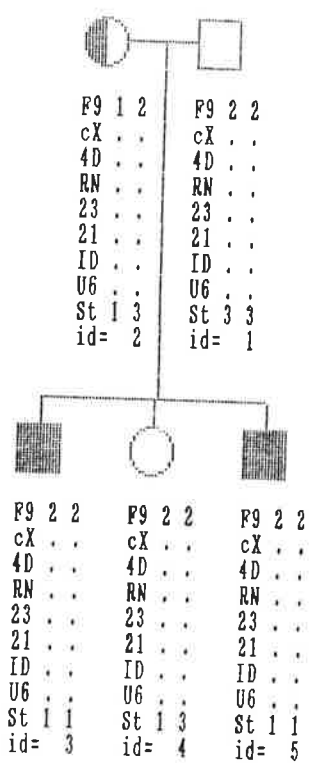
Pedigree 218



Pedigree 219

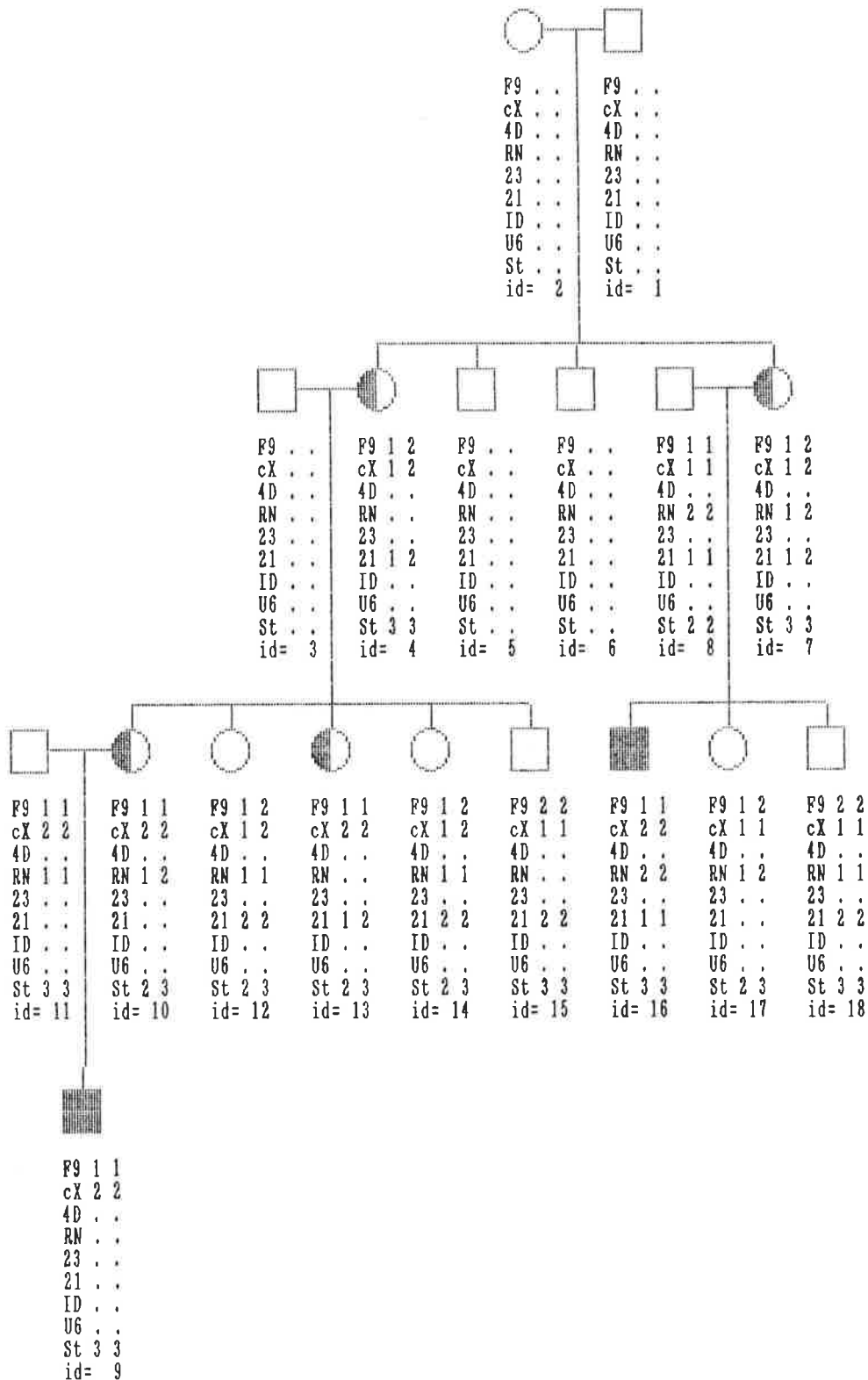


Pedigree 220

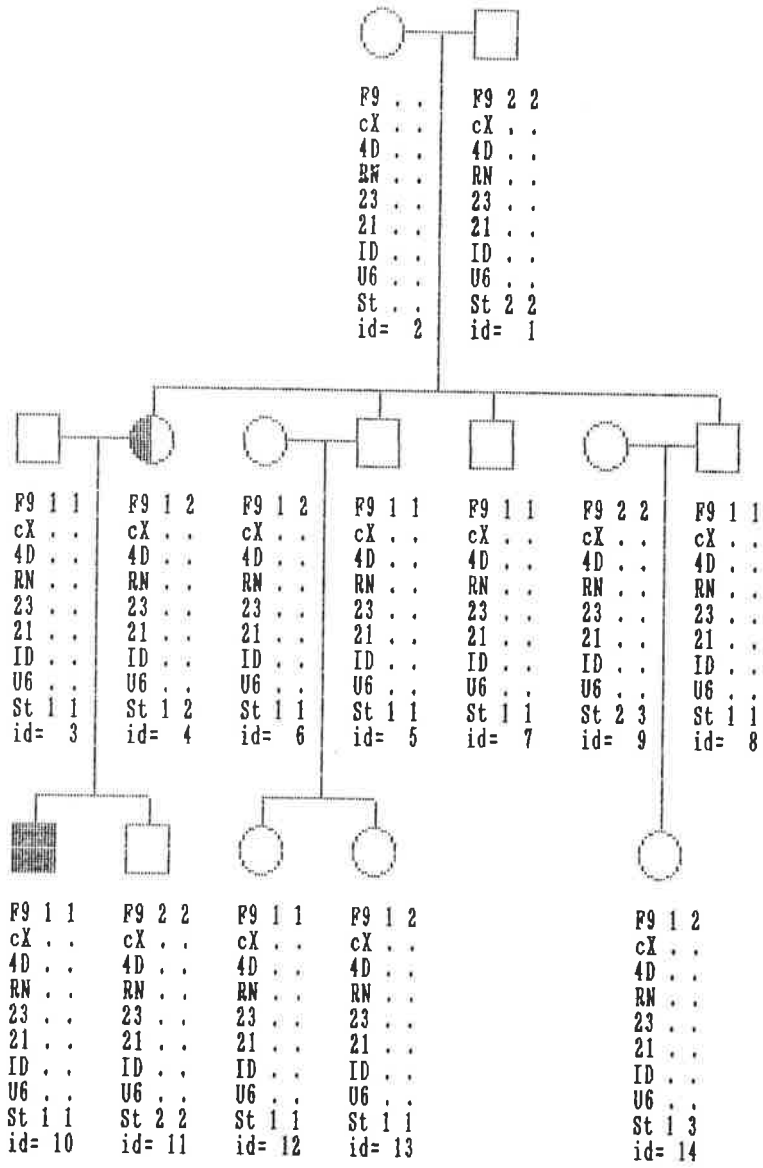


Appendix C. Fragile X pedigrees. p. C-109

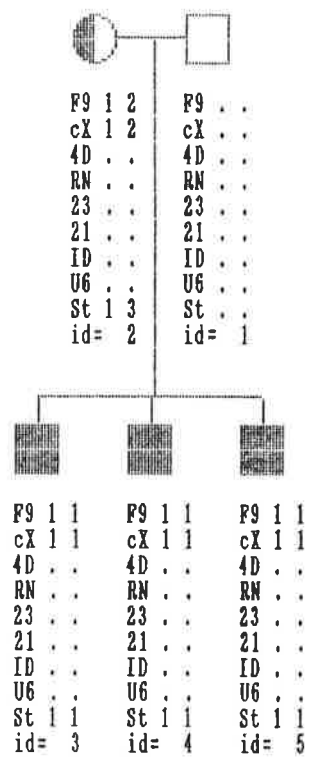
Pedigree 222



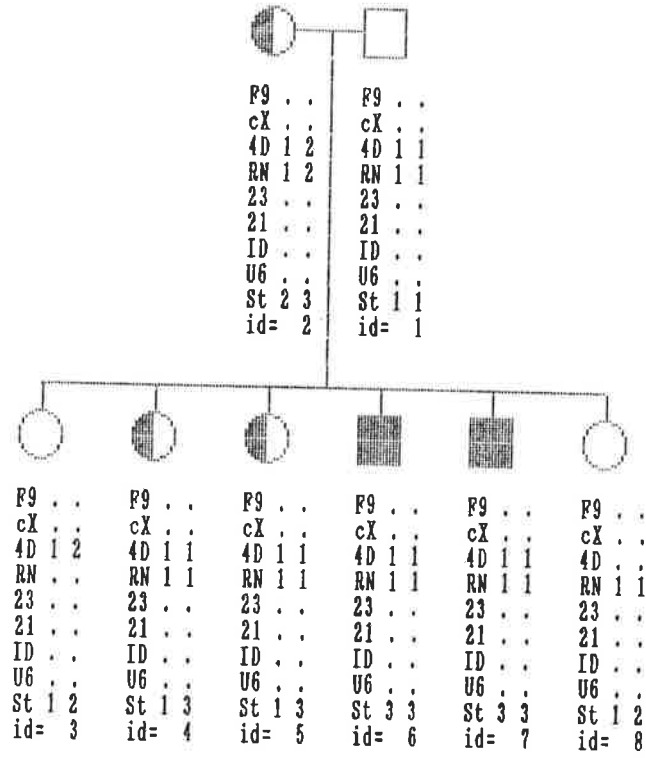
Pedigree 223



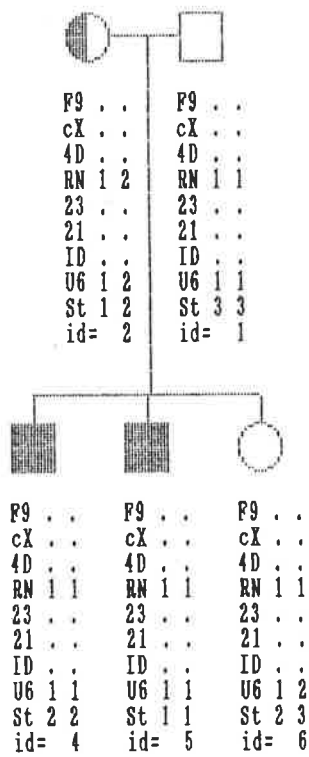
Pedigree 303



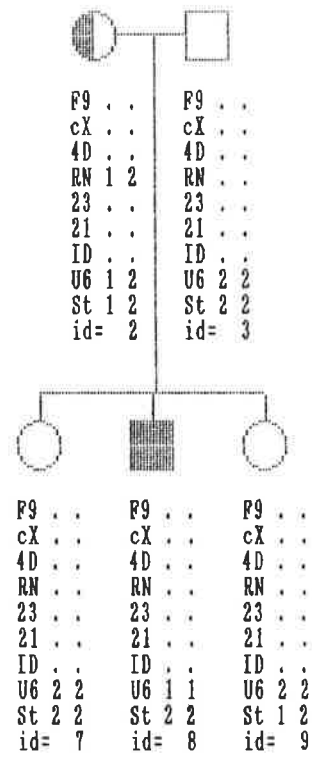
Pedigree 304



Pedigree 305

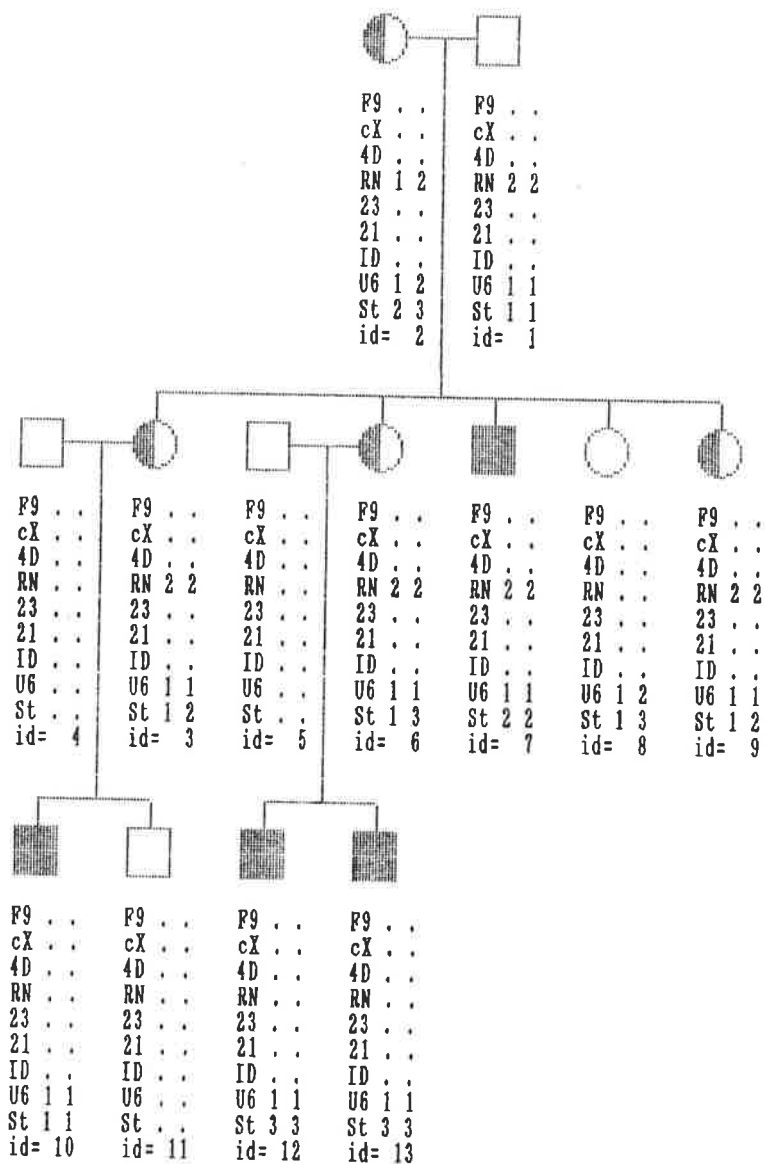


Pedigree 305

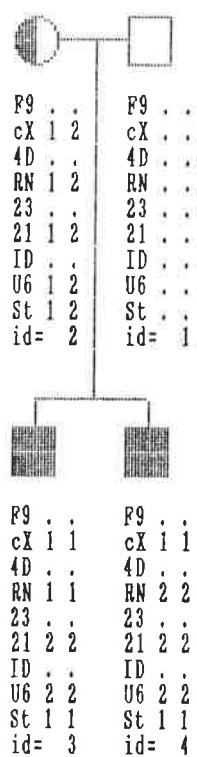




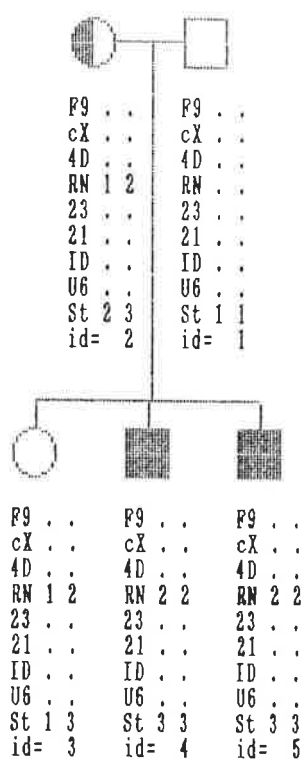
Pedigree 306



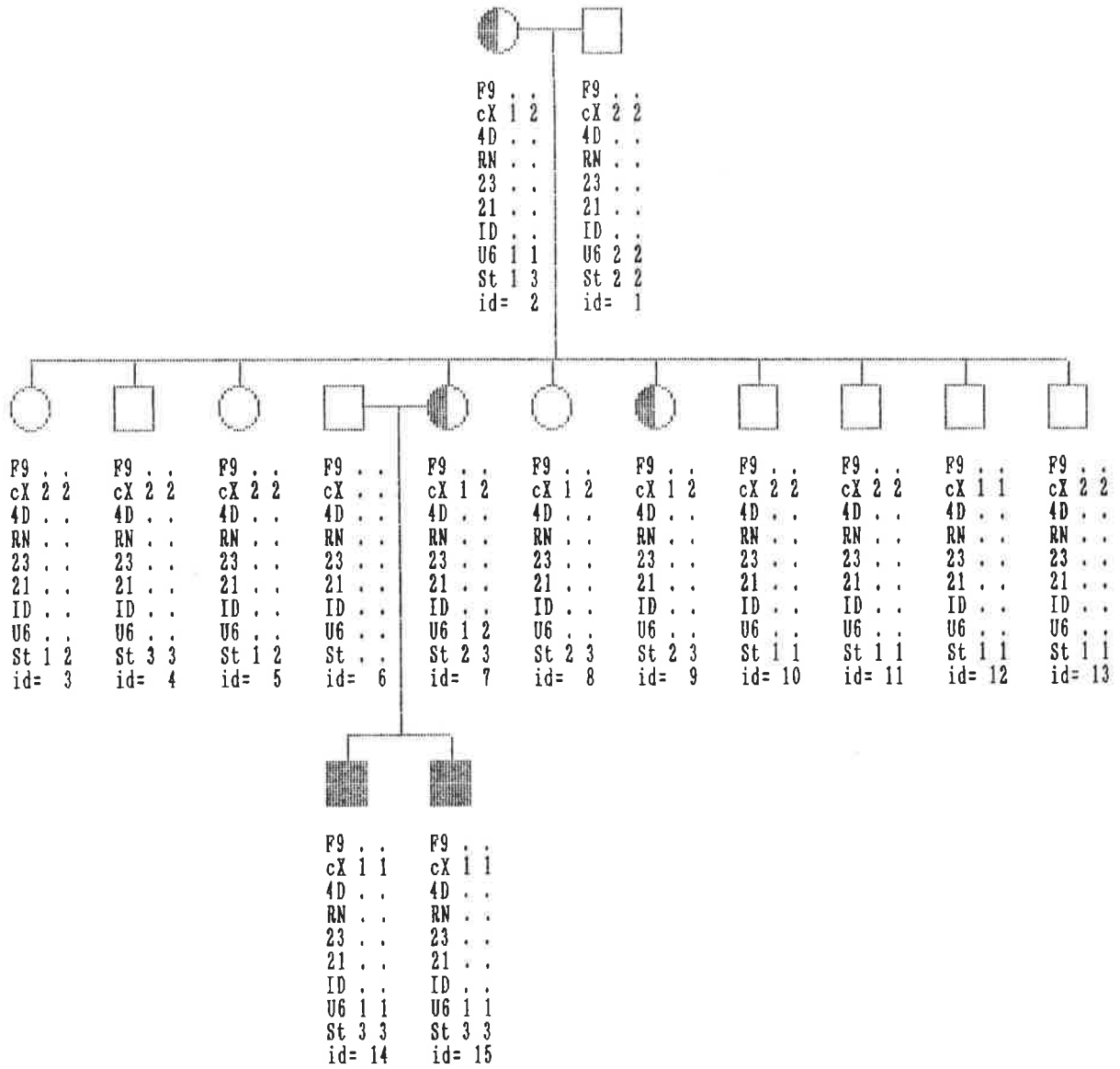
Pedigree 309



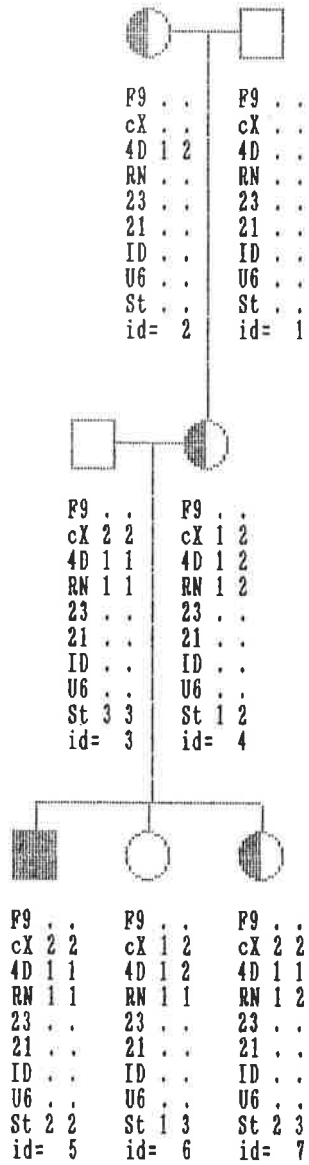
Pedigree 310



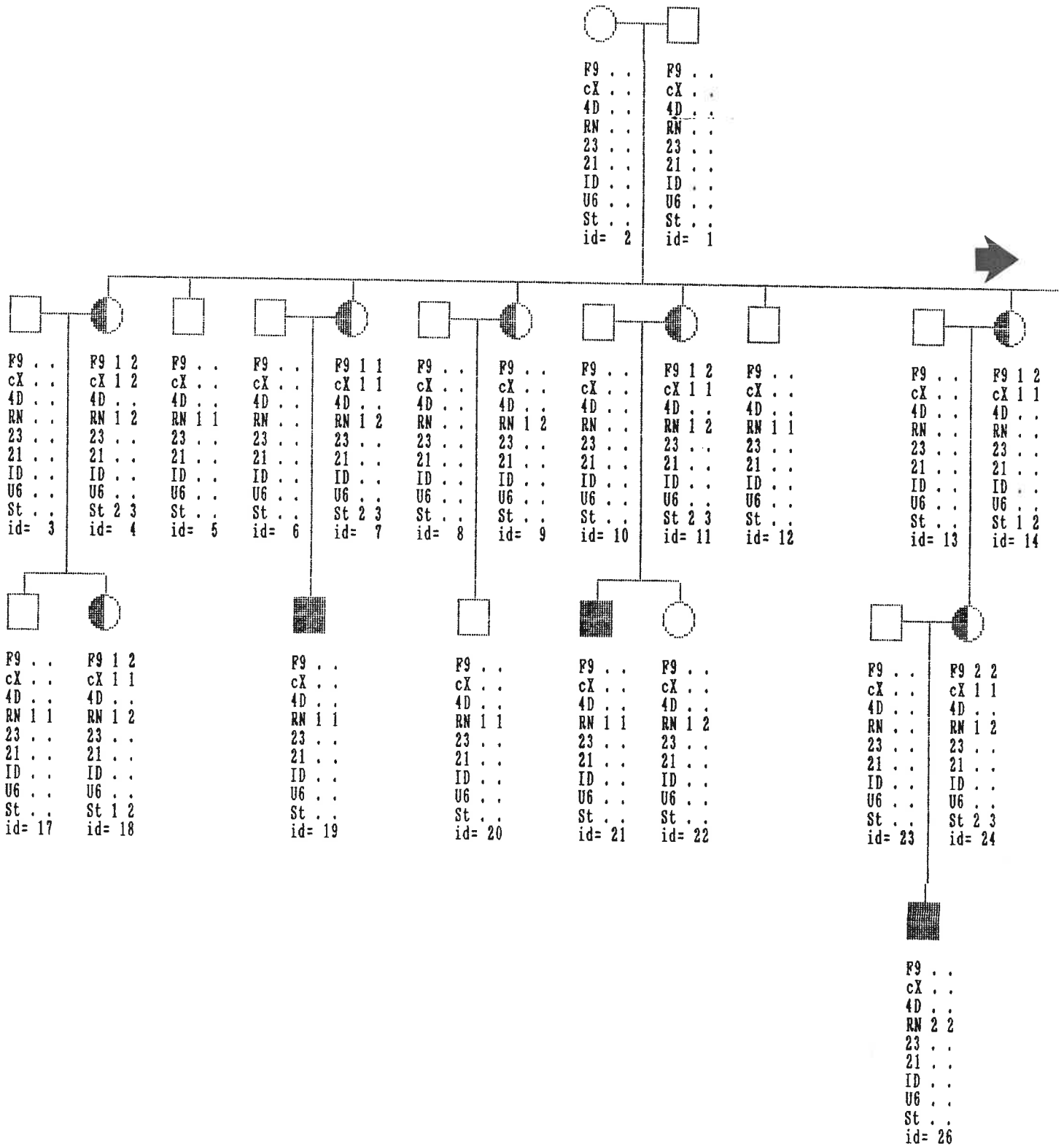
Pedigree 311



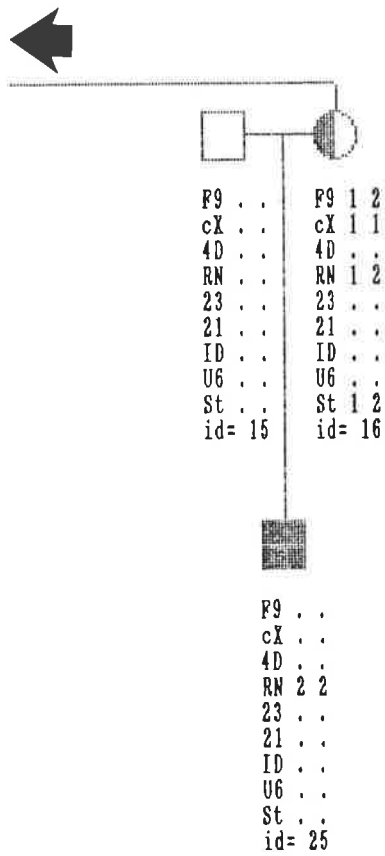
Pedigree 313



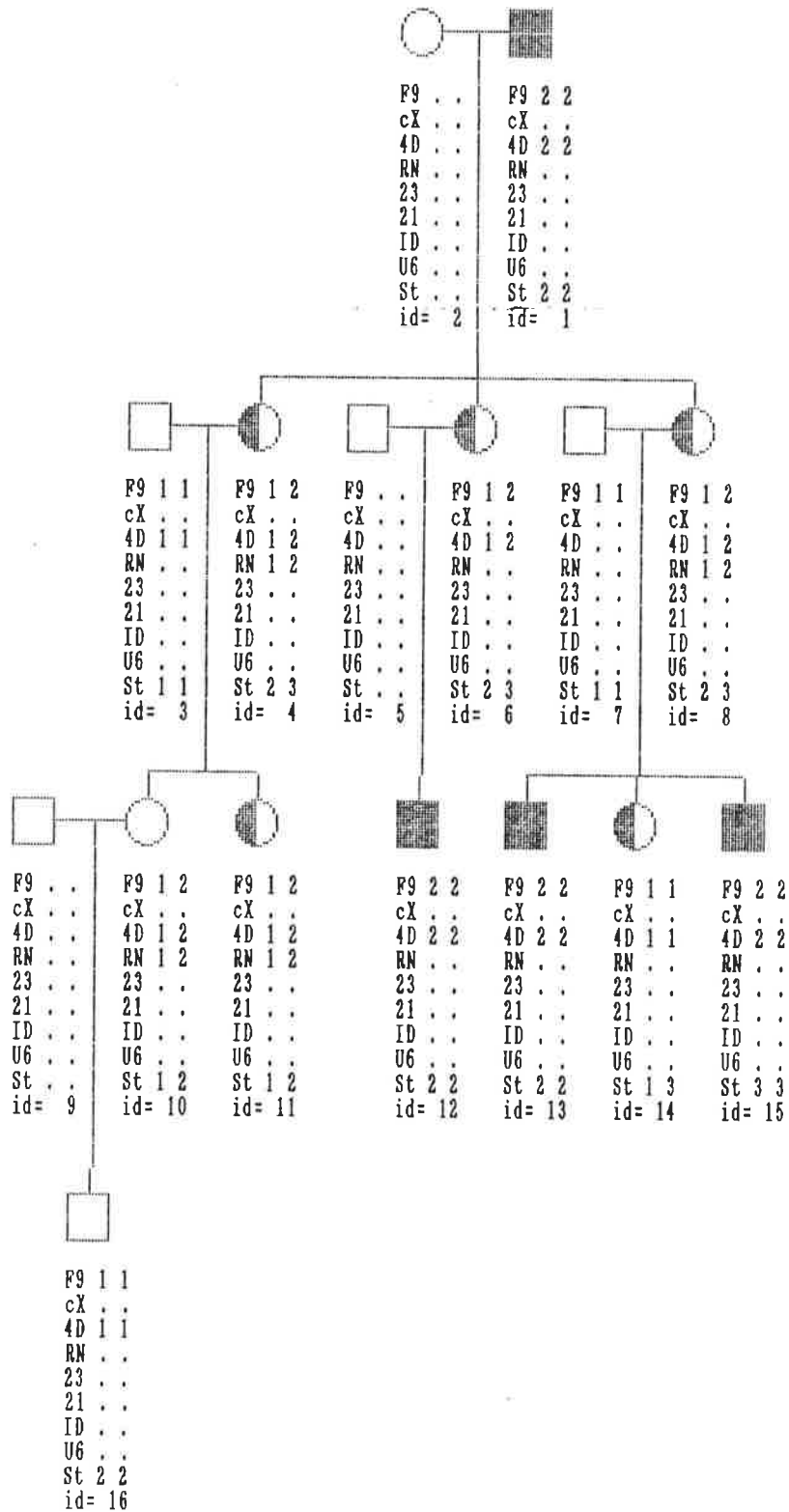
Pedigree 314



Appendix C. Fragile X pedigrees. p. C-120

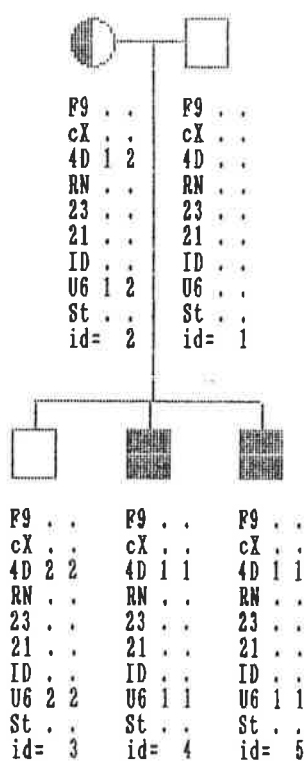


Pedigree 316

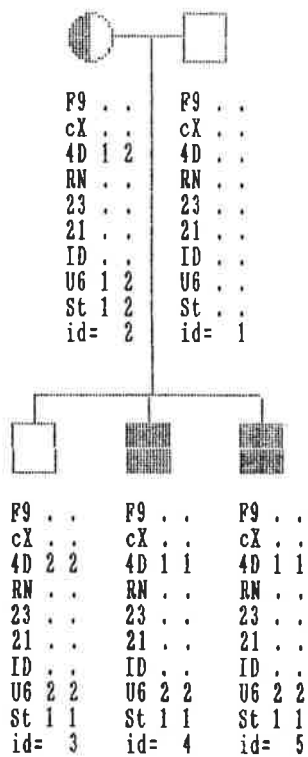




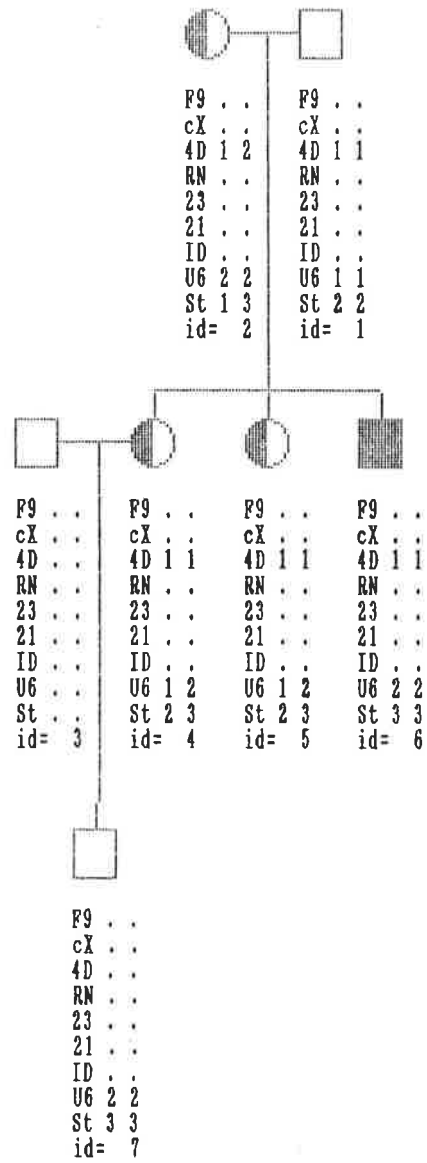
Pedigree 318



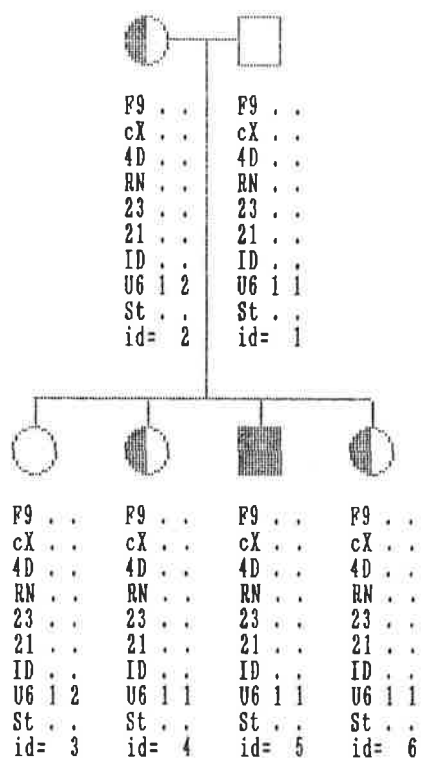
Pedigree 319



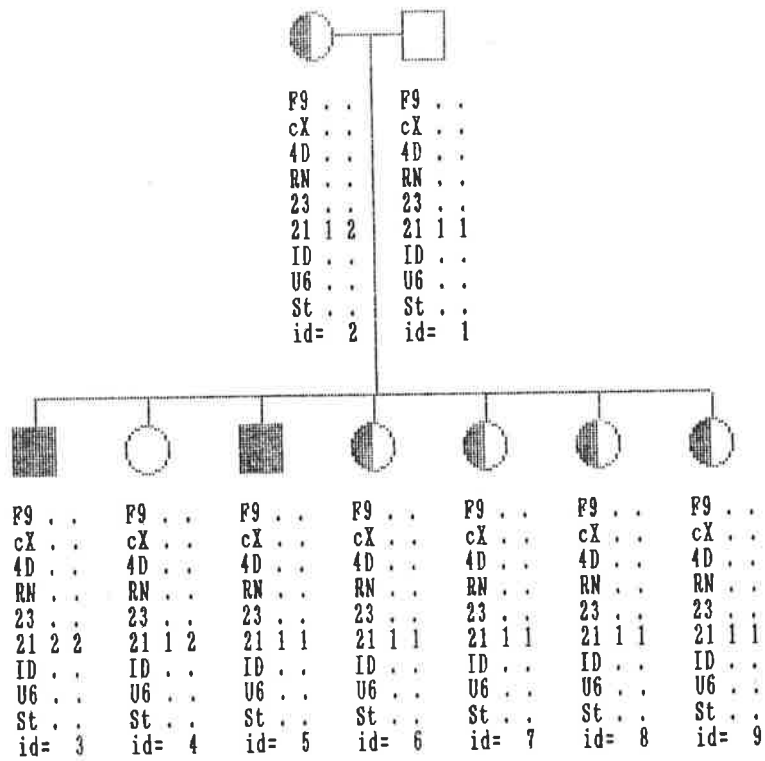
Pedigree 320



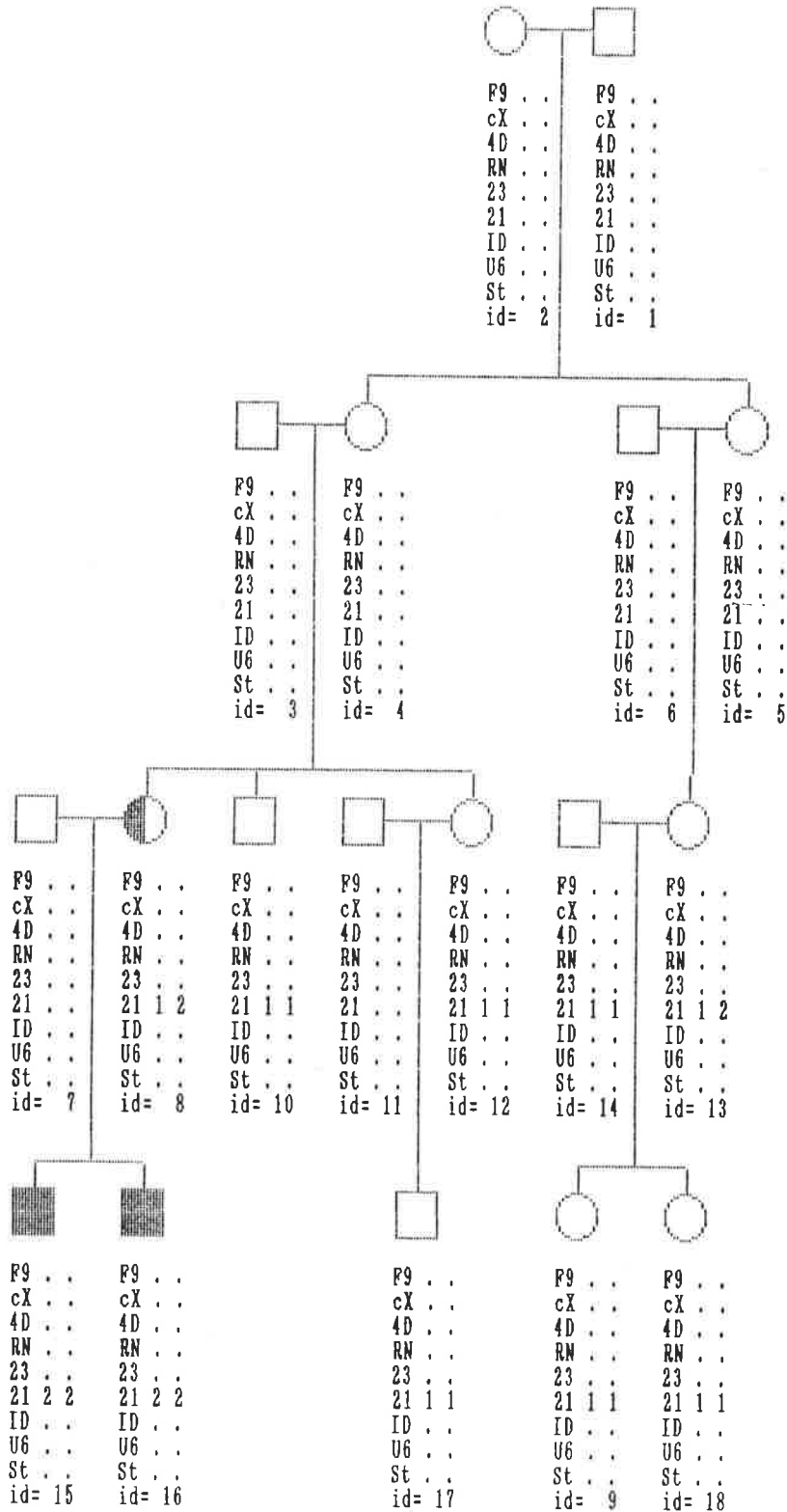
Pedigree 328



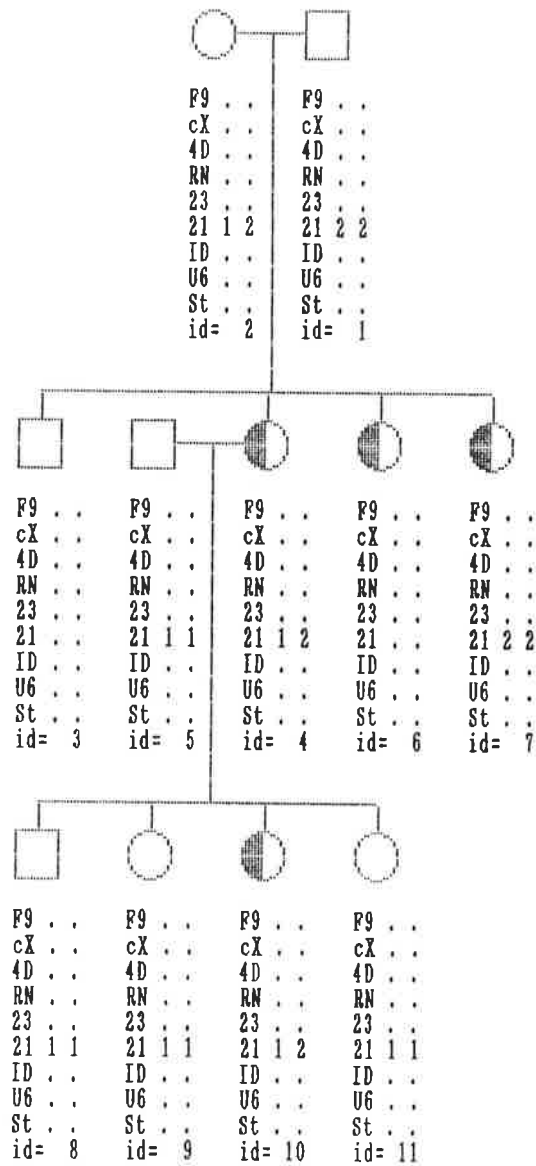
Pedigree 1050



Pedigree 1120

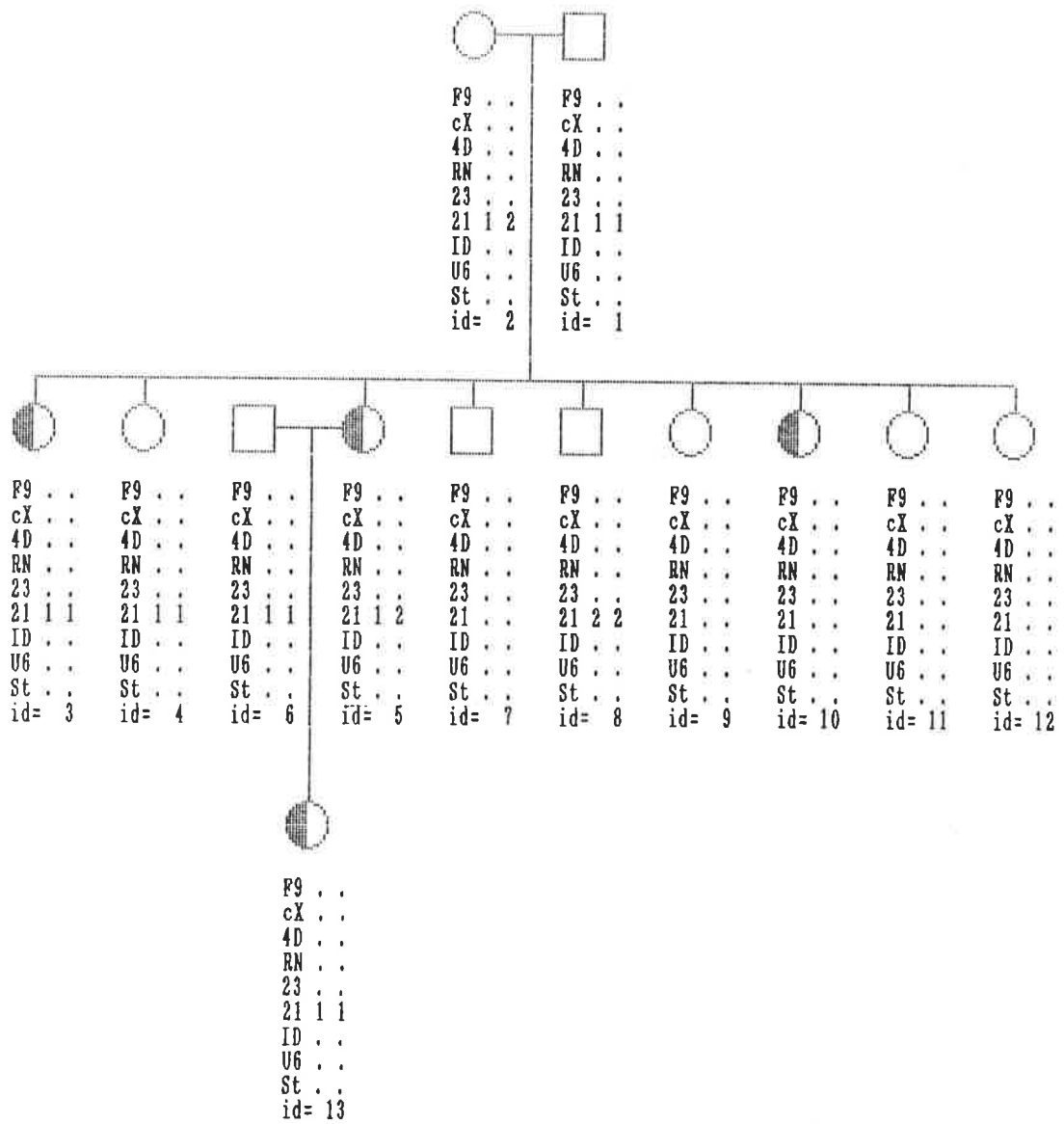


Pedigree 1130



Appendix C. Fragile X pedigrees. p. C-129

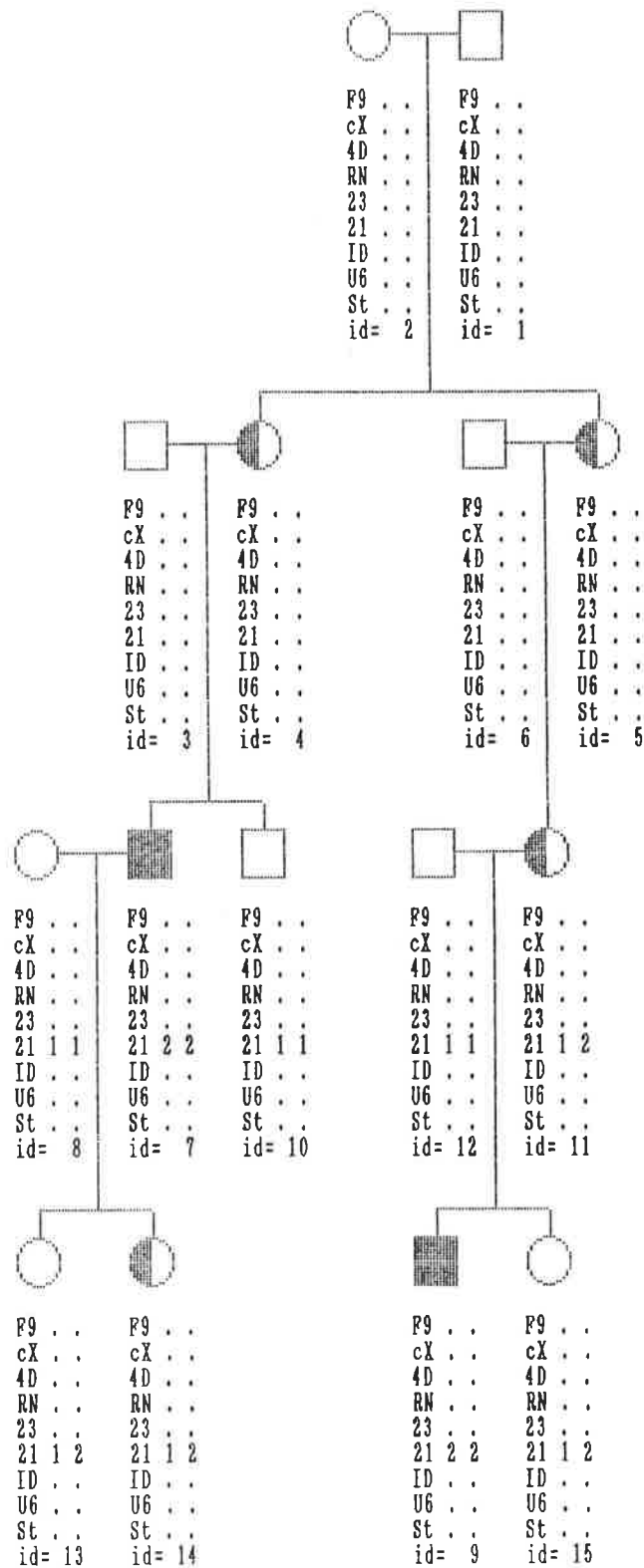
Pedigree 1145





Appendix C. Fragile X pedigrees. p. C-130

Pedigree 1320



**Appendix D**

**PUBLICATIONS AND MANUSCRIPTS.**

*Appendix D. Publications & manuscripts p. D-2*

Much of the material presented in this thesis has been published, is 'in press', or has been submitted. The papers and manuscripts are detailed below with acknowledgements to co-authors; copies of the papers and manuscripts are included in the remainder of this appendix. Reference to appropriate chapters in the thesis is indicated. Where no reference to a thesis chapter is made, the material in the article is not reproduced in the body of the thesis.

Suthers GK, Wilson SR (1990). Genetic counselling in rare syndromes: a resampling method for determining an approximate confidence interval for gene location with linkage data from a single pedigree. *Am J Hum Genet* (in press).

SR Wilson suggested and implemented the resampling protocol; the candidate performed the multipoint linkage analysis, calculated the multipoint risk estimates, and wrote the major part of the paper. (Chapters 2 & 3).

Suthers GK, Turner G, Mulley JC (1988). A non-syndromal form of X-linked mental retardation (XLMR) is linked to DXS14. *Am J Med Genet* 30:485-491.

The pedigree was identified and examined by G Turner; karyotypes of the affected males were checked at the Prince of Wales Children's Hospital, Randwick NSW; JC Mulley provided instruction in DNA methods and two-point

*Appendix D. Publications & manuscripts p. D-3*

linkage analysis; the candidate collected the blood samples, genotyped the pedigree, performed the linkage analysis, and wrote the paper. (Chapter 3).

Suthers GK, Callen DF, Hyland VJ, Kozman HM, Baker E, Eyre H, Harper PS, Roberts SH, Hors-Cayla MC, Davies KE, Bell MV, Sutherland GR (1989a). A new DNA marker tightly linked to the fragile X locus (FRAXA). *Science* 246:1298-1300.

DF Callen, PS Harper, SH Roberts, and MC Hors-Cayla provided cell lines; VJ Hyland isolated the new DNA probe; KE Davies and MV Bell provided an unpublished DNA probe; E Baker and H Eyre performed the *in situ* studies; HM Kozman genotyped the CEPH pedigrees; the candidate subcloned the new probe, defined the breakpoints in the cell lines and localized the new probe, defined the polymorphism, genotyped the fragile X pedigrees, performed the linkage analysis, and wrote the paper. (Chapters 4,5,& 6).

P.J. Wilson, G.K. Suthers, D.F. Callen, E. Baker, P.V. Nelson, A. Cooper, J.E. Wraith, G.R. Sutherland, C.P. Morris, & J.J. Hopwood. Frequent deletions at Xq28 indicate genetic heterogeneity in Hunter syndrome. (submitted).

PJ Wilson, CP Morris, and JJ Hopwood cloned the Hunter syndrome gene and provided patient DNA samples; DF Callen and JE Wraith provided cell line or patient DNA

*Appendix D. Publications & manuscripts p. D-4*

samples; E Baker performed the *in situ* studies; PV Nelson checked the patient samples for Hunter syndrome gene deletions; the candidate prepared and probed the cell line DNA, documented the extent of the deletions at the Hunter gene, and wrote a major part of the paper. (Chapter 4).

GK Suthers, VJ Hyland, DF Callen, I Oberle, M Rocchi, NS Thomas, CP Morris, CE Schwartz, M Schmidt, HH Ropers, E Baker, BA Oostra, N Dahl, PJ Wilson, JJ Hopwood, GR Sutherland (1990). Physical mapping of new DNA probes near the fragile X (FRAXA) with a panel of cell lines. *Am J Hum Genet* (in press).

VJ Hyland provided the new DNA probes; co-authors who provided cell lines are listed in Table 2-A; co-authors who provided DNA probes are listed in Table 2-B; CP Morris and JJ Hopwood supplied DNA from patients with Hunter syndrome; E Baker performed the *in situ* studies; the candidate arranged the collaboration, defined the breakpoints in the cell lines, mapped the new DNA probes in relation to the breakpoints, and wrote the paper. (Chapter 4).

S Yu, GK Suthers, JC Mulley (1990). A BclI RFLP for DXS296 (VK21) near the fragile X. *Nucl Acid Res* 18:690.

*Appendix D. Publications & manuscripts p. D-5*

S Yu and JC Mulley documented the RFLP and wrote the paper. The candidate subcloned the VK21 probes. (Chapter 5).

GK Suthers, VJ Hyland, Baker E, Fernandez KEW, Callen DF, Sutherland GR (1988). TaqI RFLP identified by probe VK17A (DXS294) at Xq26. *Nucl Acid Res* 16:11389.

VJ Hyland provided the probe VK17; the candidate subcloned the repeat-free fragment VK17A, documented the RFLP, and wrote the paper. (Chapter 5).

GK Suthers, I Oberle, J Nancarrow, JC Mulley, VJ Hyland, PJ Coulson, J McCure, CP Morris, JJ Hopwood, JL Mandel, GR Sutherland. Genetic mapping of new RFLPs at Xq27-q28. (submitted).

VJ Hyland isolated the probes VK16, VK18, and VK23. CP Morris and JJ Hopwood isolated the probe pc2S15. JC Mulley detected an RFLP with the probe VK23B. J Nancarrow searched for RFLPs with subclones of VK16 and VK18. J McCure genotyped the CEPH pedigrees at DXS297 and IDS. I Oberle provided the genotypes at other loci at Xq26-q28. The candidate subcloned VK23, defined RFLPs at DXS297 and IDS, performed the linkage analysis, and wrote the manuscript. (Chapter 5).

*Appendix D. Publications & manuscripts p. D-6*

GK Suthers, JC Mulley, MA Voelckel, N Dahl, ML Vaisanen, P Steinbach, IA Glass, CE Schwartz, BA van Oost, SN Thibodeau, NE Haites, BA Oostra, R Gine, M Carballo, CP Morris, JJ Hopwood, GR Sutherland. Genetic mapping of new DNA probes at Xq27 defines a strategy for DNA studies in the fragile X syndrome. (submitted).

CP Morris and JJ Hopwood provided the probe pc2S15. The other co-authors contributed pedigree and genotype data. The candidate genotyped the Adelaide families at DXS296, DXS297, and IDS, performed the analysis, and wrote the paper. (Chapter 6).

GK Suthers, JC Mulley, MA Voelckel, N Dahl, ML Vaisanen, P Steinbach, IA Glass, CE Schwartz, BA van Oost, SN Thibodeau, NE Haites, BA Oostra, A Schinzel, M Carballo, CP Morris, JJ Hopwood, GR Sutherland. Linkage homogeneity near the fragile X locus in normal and fragile X families: implications for genetic risk analysis. (submitted).

CP Morris and JJ Hopwood provided the probe pc2S15. The other co-authors contributed pedigree and genotype data. The candidate genotyped the Adelaide families at DXS296, DXS297, and IDS, performed the analyses, and wrote the paper. (Chapter 7).

Suthers GK, Turner G, Mulley JC (1988). Case report: fragile

*Appendix D. Publications & manuscripts p. D-7*

X syndrome and nephrogenic diabetes insipidus. *Am J Med Genet* 30:231-236.

The family was identified by G Turner; DNA studies were performed by JC Mulley; the candidate interpreted the DNA results and wrote the paper.

Suthers GK, Davies KE, Baker E, Sutherland GR (1989). TaqI RFLP identified by probe 1A1 (DXS374). *Nucl Acid Res* 17:8901.

KE Davies provided the DNA probe; E Baker performed the *in situ* hybridization studies; the candidate identified and documented the RFLP and wrote the article.

Suthers GK, Sutherland GR (1990). Letter to the Editor: Recombination and the fragile X. *Hum Genet* 85:141-142.

The candidate performed the analysis and wrote the paper.





*cpub1988*

13 July, 1990

*Due for publication in Am J Hum Genet  
(July 1990).*

**GENETIC COUNSELLING IN RARE SYNDROMES: A RESAMPLING METHOD FOR  
DETERMINING AN APPROXIMATE CONFIDENCE INTERVAL FOR GENE LOCATION WITH  
LINKAGE DATA FROM A SINGLE PEDIGREE**

Graeme K. Suthers\* and Sue R. Wilson#

\*Department of Molecular Genetics and Cytogenetics,  
Adelaide Children's Hospital, North Adelaide, South Australia; and

#Statistics, Institute of Advanced Studies,  
Australian National University, Canberra, A.C.T.

*Running Head:* Confidence interval for gene location.

*Address for Correspondence:*

Dr. Graeme Suthers,  
Department of Molecular Genetics and Cytogenetics,  
Adelaide Children's Hospital,  
North Adelaide, S.A. 5006  
Australia

## SUMMARY

Multipoint linkage analysis is a powerful method for mapping a rare disease gene on the human gene map despite limited genotype and pedigree data. However, there is no standard procedure for determining a confidence interval for gene location using multipoint linkage analysis. A genetic counsellor needs to know the confidence interval for gene location in order to determine the uncertainty of risk estimates provided to a consultand on the basis of DNA studies. We describe a resampling, or "bootstrap", method for deriving an approximate confidence interval for gene location with data from a single pedigree. This method was used to define an approximate confidence interval for the location of a gene causing non-syndromal X-linked mental retardation (*MRX1*) in a single pedigree. The approach seemed robust in that similar confidence intervals were derived using different resampling protocols. Quantitative bounds for the confidence interval were dependent on the genetic map chosen. Once an approximate confidence interval for gene location was determined for this pedigree it was possible to use multipoint risk analysis to estimate risk intervals for women of unknown carrier status. Despite the limited genotype data the combination of the resampling method and multipoint risk analysis had a dramatic impact on the genetic advice available to consultands.

## INTRODUCTION

Linkage analysis is widely used to locate disease loci in relation to anonymous polymorphic loci on the human gene map. A common approach is to first use two-point analyses to establish linkage since these are computationally fast. The linkage map is then refined using computationally intensive multipoint analyses which are statistically efficient in simultaneously using the information about all loci.

For two-point linkage analysis of data from simple defined pedigrees there are established techniques for estimating the confidence interval for the recombination fraction (Ott 1985). For data derived from pedigrees of arbitrary structure, an *ad hoc* "one-LOD-unit-down" method for approximating the 90% confidence interval for the recombination fraction is generally used (Conneally et al. 1985) provided that the peak LOD score exceeds 3.0 for autosomal data, or 2.0 in the case of X-linked data (Ott 1985). Multipoint linkage analysis of such data usually generates a complex likelihood function for gene location, and this function may have two or more maxima. There is no standard statistical technique for determining a confidence interval for gene location in such a situation (Lathrop et al. 1984). The "one-LOD-unit-down" method is often applied in multipoint linkage analysis but the significance level of such a confidence interval is unclear (Keats et al. 1989).

The confidence interval for gene location is an important factor in providing genetic counselling based on DNA studies. In the case of a common genetic disorder the confidence interval is usually narrow, and genetic risk estimates based on a single disease gene location provide sufficient information for consultands. However the genetic counsellor is often presented with a family having a rare condition that has been mapped in only a few pedigrees. In this situation the confidence interval for

gene location would usually be wide, and the risk estimates provided to consultands must reflect the uncertainty of gene location (Lange 1986). The uncertainty of risk estimates may be difficult to define because of the difficulty in determining a confidence interval for gene location.

Computer-intensive statistical techniques are becoming widely used in situations where complex problems elude formal analytical solution (Efron and Tibshirani 1986). In human linkage analysis, for example, Wilson and La Scala (1989) used the recently developed resampling (or "bootstrap") methodology for determining confidence intervals for the recombination fractions between a disease locus and marker loci, and for determining the evidence for locus order for data from a set of nuclear families. Further, Ott (1989) has proposed the use of Monte-Carlo tests for planning linkage studies, and for testing hypotheses in situations where there is a clear null hypothesis under which simulated data can be generated and against which the observed test statistic can be evaluated.

Here we briefly outline the application of the resampling method for determining an approximate confidence interval for disease gene location on a genetic map where the positions of the other loci are known. We became interested in this problem while studying a large family with non-syndromal X-linked mental retardation (*MRX1*). Two-point linkage analysis had indicated that *MRX1* was linked to *DXS14* (p58-1) which is at Xp11.21 (Suthers et al. 1988). With a peak LOD score of 2.12 at a recombination fraction of zero, the approximate 90% confidence interval for the recombination fraction between *DXS14* and *MRX1* was 0.00-0.22 (Conneally et al. 1985). It was not possible to narrow this confidence interval by pooling data from other families as non-syndromal X-linked mental retardation is genetically heterogeneous (Morton et al. 1977; Herbst and Miller 1980; Mandel et al. 1989). As a consequence of this wide confidence

interval, estimates of carrier risk based on *DXS14* genotypes alone would have wide risk intervals and be of little value in genetic counselling.

We used the resampling method to determine an approximate confidence interval for the location of *MRX1* with data from this single pedigree. Multipoint risk analysis was then used to calculate carrier risk intervals and this had a dramatic effect on the genetic advice available to this family.

## METHODS

### Family Study

The family studied (Fig. 1) has been described previously (Suthers et al. 1988). Males with non-syndromal X-linked mental retardation occurred in each of the 3 surviving generations. Blood for DNA extraction was collected from 8 affected males, 4 normal males, 6 carrier females, and 13 females of unknown carrier status. We were unable to collect blood from a further 3 affected males. Descriptions of the DNA probe p58-1 (*DXS14*) and flanking probes L1.28 (*DXS7*) and pDP34 (*DXYS1*) have been provided by Kidd et al. (1989).

### Linkage Analysis

Two-point linkage analysis was performed with the computer program LIPED (Ott 1974). Multipoint linkage analysis of *MRX1* in relation to the loci *DXS7*, *DXS14* and *DXYS1* was performed using the program LINKMAP (Versions 3.5 and 4.6) (Lathrop et al. 1985). The order of loci is pter - *DXS7* - *DXS14* - *DXYS1* - qter (Mandel et al. 1989). LINKMAP limited the calculation of location scores to discrete points along the genetic map defined by these three loci. (The location score is twice the natural logarithm of the odds for that location of *MRX1* on the genetic map versus no linkage.) The points at which location scores were initially calculated are indicated in Figure 2. Selected resamples were re-analysed in greater detail using a finer grid of location score calculations around points of interest. As LINKMAP does not support the inclusion of interference in the analysis, Haldane's mapping function was used.

When using LINKMAP to localise a gene in relation to a genetic map the genetic distances between the loci on the map are regarded as fixed. In practice these distances are not known exactly. For this reason we

repeated the analysis using different sets of recombination fractions between *DXS7*, *DXS14*, and *DXYS1* to determine whether our confidence interval for gene location was sensitive to changes in the background genetic map. For the first genetic map the recombination fractions were 0.16 (*DXS7-DXS14*) and 0.14 (*DXS14-DXYS1*). These values were derived from the Human Gene Mapping 10 summary genetic map of the X chromosome (Table 24 in Keats et al. 1989) using Rao's mapping function. For the second genetic map the recombination fractions were taken to be 0.20 (*DXS7-DXS14*) and 0.25 (*DXS14-DXYS1*); these values were estimated using published genotypes in the CEPH data base (Version 2) and the program ILINK (Version 4.6) (Lathrop et al. 1985) (unpublished observations). For the third map the recombination fractions chosen were 0.21 (*DXS7-DXS14*) and 0.20 (*DXS14-DXYS1*).

Two-point and multipoint risk estimates for females of unknown carrier status were calculated using the program MLINK (Version 4.6) (Lathrop et al. 1985). The penetrance of the disease gene in males was assumed to be 1.0 with the *MRX1* allele frequency being 0.0001 and a mutation rate of zero. Multipoint risk estimates were calculated using each of the three genetic maps. All calculations were performed using an IBM-compatible personal computer with a numerical co-processor.

### Statistical Analysis

Let  $d$  represent the unknown true position of the disease locus (with respect to an *a priori* determined origin). One type of resampling procedure for constructing a confidence region for  $d$  is as follows. From the observed data we have an estimate of  $d$ ,  $\hat{d}$ . The essence of the resampling approach is to take a random resample from our original data and to repeat the estimation procedure with this resample to obtain  $\hat{d}_1^*$ . This resampling procedure is repeated  $B$  times to give  $\hat{d}_1^*, \hat{d}_2^*, \dots, \hat{d}_B^*$ . Define  $\hat{G}(x)$  to be the parametric bootstrap cumulative density function of  $\hat{d}^*$ ,



$\hat{G}(x) = P^*\{\hat{d}^* < x\}$ , where  $P^*$  indicates probability computed according to the resampled distribution of  $\hat{d}^*$ . The simplest method of determining a confidence interval is the "percentile method" (Efron and Tibshirani 1986; Hinkley 1988; DiCiccio and Romano 1988). The  $1-2\alpha$  central interval for  $d$  is given by  $d \in [\hat{G}^{-1}(\alpha), \hat{G}^{-1}(1-\alpha)]$ . So the percentile method interval is just the interval between the  $100\alpha$  and  $100(1-\alpha)$  percentiles of the resampled distribution of  $\hat{d}^*$ . In the present study the approximate 98% confidence interval from 49 resampled values for  $\hat{d}^*$  was the range of the values.

The major difficulty in applying this method to the multipoint linkage analysis of a single pedigree is to choose an appropriate random resampling protocol. Although each meiosis in the pedigree is an independent event the information that can be obtained from each meiosis is dependent on other complex factors (such as pedigree structure) that enable, say, phase to be inferred. An essential point to keep in mind is that the resampling simulation should, implicitly or explicitly, simulate each component of variability.

We utilised two different resampling protocols to determine whether our conclusions concerning the disease gene location were robust to the exact form of resampling chosen. The first protocol (Protocol I) was based on a simulation method described by Lathrop et al. (1987). For this protocol the pedigree structure and genotypes for generations I and II in each resample were the same as in the original data. If an individual in generation II had had children then the original sibship in generation III was randomly resampled to create a sibship of the same size for generation III of the resample. Similarly resampled sibships were added to generation IV of the resample. In this way each resample had the same pedigree structure as the original data set, and the genotypes within each sibship in generations III and IV were randomly selected. The second resampling

protocol (Protocol II) was prompted by the Elston-Stewart algorithm (Elston and Stewart 1971) and exploited the conditional structure of the likelihood formulation. Each individual in the pedigree was regarded as being a branch end and these branch ends were taken to be independent for the resampling. If an individual was chosen in the resample his parents, grandparents etc. up the tree were included. If two or more individuals had a common ancestor then the branches were merged. The resampling stopped once the number of people in the resample (including parents, grandparents etc.) totalled 31 (the pedigree size). Using this protocol the pedigree size was fixed but the pedigree structure varied with each resampling. With both protocols females of unknown carrier status were included and the carrier status of a woman was determined by whether or not she had any affected sons or grandsons in that resample. Therefore each resample varied in the number of affected males, normal males, obligate carrier females and females of unknown carrier status.

The number of resamples taken,  $B$ , depends on the form of  $d^A$ , and will often be at least 100 (Hinkley 1988). For this pedigree we originally took 19 resamples under each protocol and then increased this number to 49. The qualitative conclusions were not changed by the increase in the number of resamples. In view of the considerable computing involved we saw no advantage to further increasing the number of resamples.

## RESULTS

### *MRX1* Location

The disease status and genotypes for *DXS7*, *DXS14*, and *DXYS1* for 17 members of the pedigree have been published previously (Table III in Suthers et al. 1988). The genotypes of the remaining members of the pedigree are listed in Table I.

The results of two-point linkage analysis of *MRX1* and the three loci are shown in Table II. The peak LOD score was 2.90 at zero recombination between *MRX1* and *DXS14*. This LOD score was higher than that reported previously as the earlier study did not include the normal males III-12 (no DNA collected) and IV-7. The approximate 90% confidence interval for the recombination fraction was 0.0 - 0.20 (Conneally et al. 1985).

The result of multipoint linkage analysis of *MRX1* in relation to *DXS7*, *DXS14*, and *DXYS1* using the first genetic map is shown in Figure 2(a). The peak location score was 18.25 with *MRX1* located at -5 centiMorgan (cM) relative to *DXS14*. The discontinuities of the likelihood function at *DXS7* and *DXYS1* indicated that *MRX1* was not located at those loci. Applying the "one-LOD-unit-down" method the confidence interval for gene location was -17 to +5cM from *DXS14*, i.e. the confidence interval extended into the intervals *DXS7-DXS14* and *DXS14-DXYS1*.

On resampling with both Protocol I and Protocol II and performing multipoint linkage analyses it was initially found that the peak location score occurred at *DXS7* in some of the resamples, suggesting that *MRX1* could be located at that point. Only two individuals in the pedigree (III-4 and III-21) had recombination between *MRX1* and *DXS7*. Inclusion of either individual at least once in each resample ensured that the discontinuity observed at *DXS7* in the original data was observed in each resample.

For the 49 resamples under Protocol I, 46 had maxima at 0cM relative to *DXS14*. The three resamples with maxima away from *DXS14* located *MRX1* at -1.6, -3.3 and -5.1cM from *DXS14*. The range of peak location scores was 11.25-27.46. For protocol II, 41 of the 49 resamples located *MRX1* at 0cM relative to *DXS14*. The remaining 8 resamples located *MRX1* at -1.6cM (5 resamples), -3.3cM (1 resample), -6.9cM (1 resample), and -8.7cM (1 resample) from *DXS14*. The range of peak location scores for 47 of the resamples was 11.81 - 24.09; the remaining two location scores were 6.93 and 5.77 (both resamples placed *MRX1* at 0cM). The resamples under both protocols which placed *MRX1* at 0cM relative to *DXS14* were examined in detail. In no resample was *MRX1* placed in the interval *DXS14-DXYS1*. The resamples that placed *MRX1* away from *DXS14* were not re-examined in detail. Under Protocol I the approximate 98% confidence interval for *MRX1* location was 0 to -5cM from *DXS14*. Under Protocol II the approximate 98% confidence interval was 0 to -9cM from *DXS14*.

The result of multipoint linkage analysis using the original data and the second genetic map is shown in Figure 2(b). The shape of the likelihood function was similar to that obtained with the first genetic map, and the most likely location of *MRX1* was at 0cM relative to *DXS14*. The distribution of resampled *MRX1* locations was similar to that noted with the first map. For the 49 resamples under Protocol I, 47 had maxima at 0cM relative to *DXS14*. The two resamples with maxima away from *DXS14* were analysed in detail and located *MRX1* at -2 and -3cM from *DXS14*. The range of peak location scores was 9.43 - 24.40. For Protocol II, 46 of the 49 resamples located *MRX1* at 0cM relative to *DXS14*. The resamples with maxima away from *DXS14* were re-analysed and placed *MRX1* at -3 (1 resample) and -6cM (2 resamples) from *DXS14*. The range of peak location scores was 7.51 - 23.82. The resamples under both protocols which placed *MRX1* at 0cM were examined in detail, and *MRX1* was not placed in the interval *DXS14-DXYS1* in

any resample. Under Protocol I the approximate 98% confidence interval for *MRX1* location was 0 to -3cM from *DXS14*. Under Protocol II the approximate 98% confidence interval was 0 to -6cM from *DXS14*.

Using the third genetic map the shape of the likelihood function (Figure 2(c)) was similar to that noted with the first genetic map, and the most likely location of *MRX1* was -4.5cM relative to *DXS14*. However, this modification of the genetic map altered the distribution of resampled *MRX1* locations quite markedly. Of 49 resamples under Protocol I, 28 located *MRX1* at 0cM relative to *DXS14* and the remainder located it up to -21cM from *DXS14* (location score range: 8.32 - 23.35). Under Protocol II, 34 resamples located *MRX1* at 0cM relative to *DXS14* and the remainder also located it up to -21cM from *DXS14* (location score range: 6.91 - 23.72). To further examine the region around *DXS14* we selected those resamples from the first 19 resamples under each protocol that had peak location scores at 0cM relative to *DXS14*. These resamples (8 under Protocol I; 14 under Protocol II) were re-analysed in greater detail and none had a peak location score in the interval *DXS14* - *DXYS1*. Those resamples that placed *MRX1* at -21cM relative to *DXS14* were not examined further; the discontinuity of the likelihood function at *DXS7* (-28cM from *DXS14*) placed a limit of approximately -27cM on the range of *MRX1* locations. Under both protocols the approximate 98% confidence interval for *MRX1* location was 0 to -27cM from *DXS14*.

Because the use of the third genetic map suggested a much wider confidence interval for *MRX1* location the contribution to the location score made by each individual in the pedigree was calculated at three location points on the third map (Table III). Based on these contributions to the likelihood function, we tried to select a non-random resample under Protocol I that would yield a location estimate for *MRX1* outside the range we found from our random resampling, but were not successful.

## Risk Estimates

There were 13 women in this pedigree who were of unknown carrier status (Table IV). On the basis of pedigree data alone the carrier risks for these women ranged between 0.17 and 0.50. The risk intervals estimated with pedigree information and *DXS14* genotypes were calculated for five values of the recombination fraction in the range 0.0 - 0.20 (the approximate 90% confidence interval for the recombination fraction between *DXS14* and *MRX1*). Two-point risk analysis modified the carrier risk significantly for 11 of the women but the risk intervals were wide for a number of them. The two women whose carrier risks were not modified by two-point risk analyses (II-6 and IV-15) had mothers who were uninformative for *DXS14*.

For multipoint risk analyses, the carrier risk intervals were calculated for five *MRX1* locations within the resampled approximate 98% confidence interval for gene location. For comparison, the risk intervals were calculated using each of the three genetic maps with the corresponding confidence interval. Where the confidence interval was wider with Protocol II than with Protocol I, the wider confidence interval was used for risk analysis.

Using the first genetic map the carrier risk intervals for all the women were narrowed. For eight of the women the carrier risk was less than 0.01. Although the mother of II-6 was uninformative for *DXS14*, she was informative for *DXS7* and *DXYS1* and multipoint risk analysis reduced the carrier risk significantly for II-6. Despite the mother of IV-15 being uninformative for both *DXS14* and *DXYS1*, the carrier risk for IV-15 was reduced with the carrier risk interval being narrow. The mother of the sisters IV-4, IV-5 and IV-6 was uninformative for *DXS7* and *DXYS1*. Despite this, the approximate confidence interval for *MRX1* location indicated much

narrower risk intervals for the sisters than had been estimated on the basis of two-point linkage analysis. The risk intervals calculated using Map 2 were very similar to those obtained with Map 1. The use of Map 3 (with a much wider confidence interval for *MRX1* location) resulted in wider risk intervals for many of the women.

Risk intervals were also calculated using the first map and the "one-LOD-unit-down" confidence interval for *MRX1* location. The "one-LOD-unit-down" confidence interval extended over two intervals, *DXS7-DXS14* and *DXS14-DXYS1*. None of the women were recombinants in the interval *DXS7 - DXS14*. Two women, III-16 and IV-18, were recombinants in the interval *DXS14-DXYS1*. As expected, the use of the "one-LOD-unit-down" confidence interval for multipoint risk analysis significantly widened the risk intervals for these two women.

## DISCUSSION

We have described a resampling or "bootstrap" method for estimating an approximate confidence interval for gene location on a known genetic map using genotype data from a single pedigree. In using this resampling method to determine the approximate 98% confidence interval for *MRX1* location, *MRX1* was consistently located in the interval *DXS7* - *DXS14*. In contrast the "one-LOD-unit-down" method suggested that the confidence interval was -17 to +5cM relative to *DXS14* (using the first genetic map) i.e. extending over the intervals *DXS7* - *DXS14* and *DXS14* - *DXYS1*. However, none of the approximate 98% confidence intervals obtained by resampling with three different genetic maps extended into the interval *DXS14*-*DXYS1*. Furthermore, using the genetic map which provided the widest confidence interval for *MRX1* location (Map 3), we were unable to resample the pedigree (using Protocol I) in such a way as to place *MRX1* in the interval *DXS14*-*DXYS1*. We conclude that in this family the resampling method provided a more accurate approximate confidence interval for *MRX1* location than the "one-LOD-unit-down" method.

For a given background genetic map the resampled estimate of an approximate confidence interval for gene location appeared robust i.e. the same estimate was obtained using different resampling protocols. However, three cautions should be noted. Firstly, the resampling protocol must be adapted for each situation being analysed. Moreover, the appropriate number of resamples (B) must be determined for each situation. In analysing this pedigree the results were not qualitatively altered by increasing the number of resamples from 19 to 49. This contrasts with the situation examined by Wilson and La Scala (1989) where there was a change in the bounds of the 1-2% confidence interval as the resample size increased from 19 to 49.



Secondly, changes in the background genetic map may radically alter the size of the estimated confidence interval. An accurate quantitative estimate of an approximate confidence interval for gene location requires an accurate genetic map. These maps will become available as the CEPH consortium maps for each chromosome are published. It was not clear why the approximate confidence interval for *MRX1* location determined with Map 3 was so different from those determined with Map 1 and Map 2. The critical factor may have been the relative sizes of the *DXS7-DXS14* and *DXS14-DXYS1* intervals. Whatever the reason, it is evident that the size of an estimated confidence interval for gene location may be critically dependent on the background genetic map chosen.

Thirdly, sophisticated data analysis cannot replace appropriate data collection. Genotypic data from other affected males in the pedigree or for other markers in the interval *DXS7-DXS14* would alter the confidence interval for *MRX1* location. The highly polymorphic locus *DXS255* (Kidd et al. 1989) lies in the interval *DXS7-DXS14* (Keats et al. 1989) but unfortunately the pedigree was uninformative for this marker.

The ability to define an approximate confidence interval for gene location has implications for genetic counselling. For each consultand the estimate of carrier risk is a likelihood function which varies with gene location. This function need not be monotonic, as Krawczak (1987) demonstrated for two-point data. He showed that it is not sufficient to calculate two-point risk estimates for just the two confidence limits for the recombination fraction since there may be intermediary higher or lower risk estimates. In the case of multipoint risk analysis, the approximate confidence interval for gene location indicates the range of gene location values for which risks should be determined. The combination of multipoint risk analysis and an appropriate confidence interval for gene location can change significantly the estimated risk intervals provided to a consultand.

Incorporating interference in multipoint risk analysis may narrow the risk interval even further by reducing the possibility of double recombination. In the pedigree we described none of the women of unknown carrier status were definite recombinants between the polymorphic loci which flanked *MRX1*, and the inclusion of interference would have had little effect on the width of the carrier risk intervals.

The arbitrary structure of human pedigrees and the complexity of multipoint linkage analysis usually make it impossible to provide numerical estimates of the reliability of the results obtained. The resampling method is a powerful non-analytical approach to estimating the reliability of linkage results, and - as indicated in this paper - can be applied in the study of very rare genetic conditions where there is a limited amount of data.

**ACKNOWLEDGEMENTS**

Barbara La Scala gave invaluable programming assistance and Athalie Nation and Norah Burton provided excellent secretarial assistance. The linkage analysis programs were kindly provided by Dr. J. Ott. The CEPH database was provided by the Centre d'Etude du Polymorphisme Humain. G.K.S. was supported by the National Health and Medical Research Council of Australia.

## REFERENCES

- Conneally PM, Edwards JH, Kidd KK, Lalouel JM, Morton NE, Ott J, White R (1985) Report of the committee on methods of linkage analysis and reporting. *Human Gene Mapping 8. Cytogenet Cell Genet* 40:356-359
- DiCiccio TJ, Romano JP (1988) A review of bootstrap confidence intervals. *J Roy Stat Soc B* 50:338-354
- Efron B, Tibshirani R (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci* 1:54-77
- Elston RC, Stewart J (1971) A general model for the analysis of pedigree data. *Hum Hered* 21:523-542
- Herbst DS, Miller JR (1980) Non-specific X-linked mental retardation II: the frequency in British Columbia. *Am J Med Genet* 7:461-469
- Hinkley DV (1988) Bootstrap methods. *J Roy Stat Soc B* 50:321-337
- Keats B, Ott J, Conneally M (1989) Report of the committee on linkage and gene order. *Human Gene Mapping 10 (1989): Tenth International Workshop on Human Gene Mapping. Cytogenet Cell Genet* 51:459-502
- Kidd KK, Bowcock AM, Schmidtke J, Track RK, Ricciuti F, Hutchings G, Bale A, Pearson P, Willard HF, Gelernter J, Giuffra L, Kubzdela K (1989) Report of the DNA committee and catalogs of cloned and mapped genes and DNA polymorphisms. *Human Gene Mapping 10 (1989): Tenth International Workshop on Human Gene Mapping. Cytogenet Cell Genet* 51:622-947
- Krawczak M (1987) Genetic risk and recombination fraction - an example of non-monotonic dependency. *Hum Genet* 75:189-190

- Lange K (1986) Approximate confidence intervals for risk prediction in genetic counseling. *Am J Hum Genet* 38:681-687
- Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 81:3443-3446
- Lathrop GM, Lalouel JM, Julier C, Ott J (1985) Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. *Am J Hum Genet* 37:482-498
- Lathrop GM, Chotai J, Ott J, Lalouel JM (1987) Tests of gene order from three-locus linkage data. *Ann Hum Genet* 51:235-249
- Mandel JL, Willard HF, Nussbaum RL, Romeo G, Puck JM, Davies KE (1989) Report of the committee on the genetic constitution of the X chromosome. *Human Gene Mapping 10 (1989): Tenth International Workshop on Human Gene Mapping. Cytogenet Cell Genet* 51:384-437
- Morton NE, Rao DC, Lang-Brown H, MacLean CJ, Bart RC, Lew R (1977) Colchester revisited: A genetic study of mental defect. *J Med Genet* 14:1-9
- Ott J (1974) Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. *Am J Hum Genet* 26:588-597
- Ott J (1985) Analysis of human genetic linkage. The Johns Hopkins University Press, Baltimore.
- Ott J (1989) Computer-simulation methods in human linkage analysis. *Proc Natl Acad Sci USA* 86:4175-4178

Suthers GK, Turner G, Mulley JC (1988) A non-syndromal form of X-linked mental retardation (XLMR) is linked to DXS14. Am J Med Genet 30:485-491

Wilson SR, La Scala B (1989) Multipoint linkage : statistical evaluation. p87-92. in Elston RC, Spence MA, Hodge SE, MacClure JW, eds. Multipoint mapping and linkage based upon affected pedigree members: Genetic Analysis Workshop 6. Progress in Clinical and Biological Research. Vol 329. Alan R. Liss, New York.

**TABLE I** Genotypes of 14 individuals in the pedigree at the loci DXS7, DXS14, and DXYS1 (a)

Individual	Status (b)	<u>DXS7</u>	<u>DXS14</u>	<u>DXYS1</u>
II-6	U	A2	A1/A2	A2
III-3	U	A1/A2	A2	A2
III-6	U	A1/A2	A2	A2
III-15	U	A2	A1/A2	A2
III-16	U	A2	A1/A2	A1/A2
IV-4	U	A1	A1/A2	A2
IV-5	U	A1	A1/A2	A2
IV-6	U	A1	A2	A2
IV-7	N	A2	A2	A2
IV-8	U	A1	A1/A2	A1/A2
IV-9	U	A1	-	A2
IV-10	U	A1/A2	A2	A2
IV-15	U	A1/A2	A1	A1/A2
IV-18	U	A1/A2	A1	A1

(a) Genotypes for the remaining 17 individuals are provided in Suthers et al. (1988).

(b) Disease Status: U female of unknown carrier status

N normal male

**TABLE II** LOD scores ( $z$ ) from two-point linkage analysis of *MRX1* and 3 loci.

	Recombination Fraction ( $\theta$ )							$\hat{z}$	$\hat{\theta}$
	0.001	0.01	0.05	0.10	0.20	0.30	0.40		
<b><u>MRX1</u> vs.</b>									
<b><u>DXS7</u></b>	-1.24	0.70	1.82	2.06	1.89	1.38	0.68	2.07	0.12
<b><u>DXS14</u></b>	2.89	2.85	2.67	2.43	1.93	1.37	0.74	2.90	0.00
<b><u>DXYS1</u></b>	-11.27	-6.30	-2.93	-1.61	-0.48	-0.03	0.10	0.10	0.40



**TABLE III** Impact on the total location score of removing an individual (with descendants).

Individual	Location points (third map)		
	-40.8cM	-4.5cM	3.1cM
II-2	-2.7	-7.6	-7.4
II-4	-1.1	-0.3	0.2
II-5	-1.1	-0.7	-0.4
II-6	-6.0	-4.9	-4.1
II-7	-3.4	-3.5	-2.9
II-8	-1.6	-5.0	-4.6
II-9	-1.1	-0.3	0.2
III-1	-1.1	-1.3	-1.3
III-2	-1.9	-2.7	-2.4
III-3	-0.4	-0.6	-0.6
III-4	3.0	-1.0	-1.4
III-5	-1.1	-1.4	-1.4
III-12	-0.5	-0.3	-0.2
III-13	-1.1	-1.4	-1.1
III-14	-1.1	-1.4	-1.4
III-17	-1.1	-1.4	-1.3
III-18	-1.1	-1.4	-1.3
III-19	-1.1	-1.0	-0.9
III-21	3.0	-1.0	-1.3
IV-1	-0.7	-1.3	-1.3
IV-7	-0.4	-0.5	-0.6

Legend to TABLE III

The table lists the impact on the total location score of removing each individual (with descendants) from the analysis. These calculations were performed at three points on the third genetic map : -40.8cM, -4.5cM, and 3.1cM relative to *DXS14*. In the intact pedigree the total location scores at these points were 10.3, 17.8, and 16.1 respectively. All values are rounded. The 11 women of unknown carrier status who did not have sons made no contribution to the total location score and are not included.

**TABLE IV** Carrier risks for women of unknown carrier status

Individual	Pedigree alone (a)	Two-point risk analysis (b)	Multi-point risk analysis (c)			
			Map 1	Map 2	Map 3	Map 1 (one-LOD-unit-down)
II-6	0.33	0.35 - 0.44	0.05 - 0.13	0.11 - 0.16	0.01 - 0.17	0.01 - 0.14
III-3	0.33	0.0 - 0.11	<0.01	<0.01	<0.01	<0.01
III-6	0.50	0.0 - 0.20	<0.01	<0.02	<0.02	<0.01
III-15	0.50	0.0 - 0.20	<0.01	<0.02	<0.02	<0.01
III-16	0.50	0.0 - 0.20	<0.01	<0.02	<0.02	0.00 - 0.32
IV-4	0.50	0.80 - 1.00	0.92 - 1.00	0.94 - 1.00	0.79 - 1.00	0.85 - 1.00
IV-5	0.50	0.80 - 1.00	0.92 - 1.00	0.94 - 1.00	0.79 - 1.00	0.85 - 1.00
IV-6	0.50	0.00 - 0.20	0.00 - 0.08	0.00 - 0.06	0.00 - 0.21	0.00 - 0.15
IV-8	0.17	0.00 - 0.06	<0.01	<0.01	<0.01	<0.01
IV-9	0.25	0.00 - 0.10	<0.01	<0.01	<0.01	<0.01
IV-10	0.25	0.00 - 0.10	<0.01	<0.01	<0.02	<0.01
IV-15	0.50	0.50	0.35 - 0.38	0.38 - 0.39	0.32 - 0.40	0.33 - 0.39
IV-18	0.25	0.00 - 0.04	<0.01	<0.01	<0.01	0.00 - 0.10

Legend to TABLE IV

- (a) Risks estimated using pedigree data alone;
- (b) Risk intervals estimated using *DXS14* genotypes for 5 values of the recombination fraction in the range 0.00 - 0.20.
- (c) Risk intervals estimated using *DXS7*, *DSX14*, and *DXYS1* genotypes with *MRX1* located at 5 points within the approximate 98% confidence interval for *MRX1* location. Risk intervals were calculated using each of the background genetic maps (with the corresponding confidence interval). In addition, risk intervals are presented using Map 1 and the "one-LOD-unit-down" confidence interval.

**Figure 1** Pedigree of the family studied (Reproduced with permission from Suthers et al. 1988).

**Figure 2** The multipoint likelihood function for the location of *MRX1* is shown for the first (a), second (b), and third (c) genetic maps. The arrows (▲) indicate the points on each map at which location scores were initially calculated for each resample. The symbols within the figures show the range of resampled *MRX1* locations under Protocol I (●) and Protocol II (■).

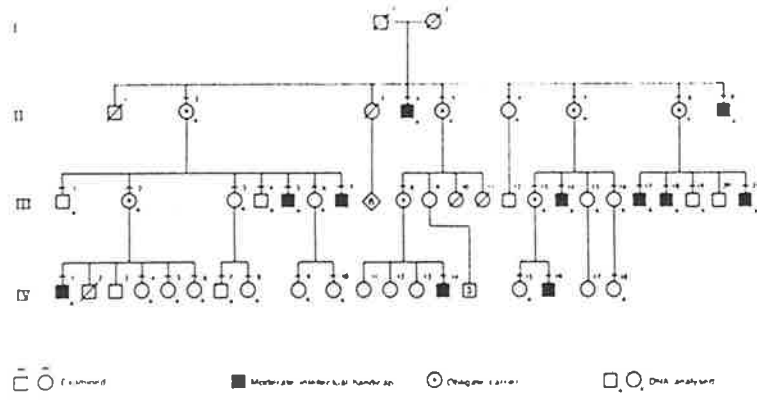
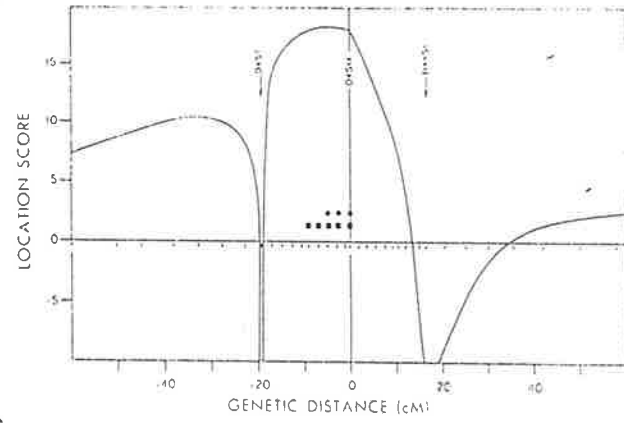
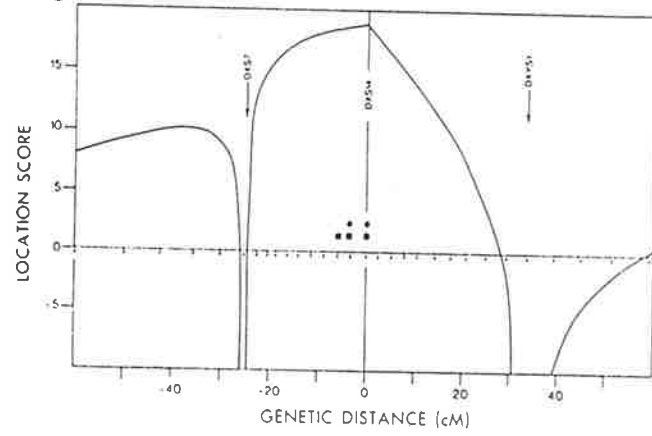


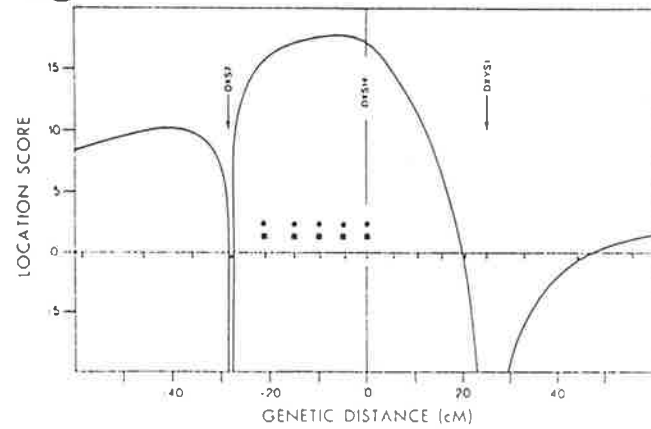
Fig 2(a)



2(b)



2(c)





Suthers, G.K., Turner, G. and Mulley, J.C. (1988). A non-syndromal form of X-linked mental retardation (XLMR) is linked to DXS14. *American Journal of Medical Genetics*, 30(1/2), 485-491.

NOTE: This publication is included in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

<http://dx.doi.org/10.1002/ajmg.1320300151>





Suthers, G.K., Callen, D.F., Hyland, V.J. et al. (1989). A new dna marker tightly linked to the fragile-x locus (FRAXA). *Science*, 246(4935), 1298-1300.

NOTE: This publication is included in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

<http://dx.doi.org/10.1126/science.2573953>



Submitted - June 1990

b10pw002  
phil10  
28/2/89 (pk)  
27.6.90 (sl)

Frequent deletions at Xq28 indicate genetic heterogeneity in Hunter syndrome

Peter J. Wilson<sup>1</sup>, Graeme K. Suthers<sup>2</sup>, David F. Callen<sup>2</sup>, Elizabeth Baker<sup>2</sup>, Paul V. Nelson<sup>1</sup>, Alan Cooper<sup>3</sup>, J. Ed Wraith<sup>3</sup>, Grant R. Sutherland<sup>2</sup>, C. Phillip Morris<sup>1</sup>  
& John J. Hopwood\*<sup>1</sup>.

---

Lysosomal Diseases Research Unit, Department of Chemical Pathology<sup>1</sup> and Department of Cytogenetics & Molecular Genetics<sup>2</sup>, Adelaide Children's Hospital, North Adelaide, SA 5006 Australia

<sup>3</sup>Willink Biochemical Genetics Unit, Royal Manchester Children's Hospital, Pendlebury, Manchester M27 1HA England

---

\* to whom correspondence should be addressed.

Abbreviations: IDS, iduronate-2-sulphatase; MPS II, mucopolysaccharidosis type II; FRAXA, fragile X mutation

## SUMMARY

Hunter syndrome is a human X-linked disorder caused by deficiency of the lysosomal exohydrolase iduronate-2-sulphatase (IDS). The consequent accumulation of the mucopolysaccharides dermatan sulphate and heparan sulphate in the brain and other tissues often results in death before adulthood. There is, however, a broad spectrum of severity that has been attributed to different mutations at the Hunter syndrome gene. We have used an IDS cDNA clone to localise the IDS gene to Xq28, distal to the Fragile X mutation (*FRAXA*). One-third of Hunter syndrome patients had various deletions or rearrangements of their IDS gene proving that different mutations are common in this condition. Deletions of the IDS gene can include a conserved locus that is tightly linked to *FRAXA*, suggesting that deletion of nearby genes could also contribute to the variable clinical severity noted in Hunter syndrome. The cDNA clone was also shown to span the X chromosome breakpoint in a female Hunter syndrome patient with an X;autosome translocation.

## INTRODUCTION

The mucopolysaccharidoses (MPS) are a group of lysosomal storage disorders caused by deficiency of individual enzymes responsible for the degradation of mucopolysaccharides (Neufeld & Muenzer, 1989). The MPS show a high degree of clinical variability between and within subtypes which has been attributed to different alleles at genes in the pathway involved in the lysosomal degradation of the mucopolysaccharides (Neufeld & Muenzer, 1989; <sup>used in this</sup> Hopwood & Morris, 1990). Hunter syndrome (MPS II), an MPS archetype, is the only X-linked MPS. This, together with the very wide spectrum of severity in Hunter syndrome, raises difficulties in carrier detection and providing prognostic and genetic advice to the families of affected children.

An MPS II clinical phenotype results when there is a deficiency of the exosulphatase iduronate-2-sulphatase (IDS) leading to the lysosomal storage of the mucopolysaccharides dermatan sulphate and heparan sulphate. Human IDS has been purified to homogeneity (Bielicki *et al.* 1990), and peptide sequence data used to isolate a cDNA clone (Wilson *et al.* submitted). We now report the use of this cDNA clone to study a group of MPS II patients and report that one-third of these patients had various deletions or rearrangements of the IDS gene.

## MATERIALS AND METHODS

DNA was isolated from patient blood or cultured fibroblasts and analysed using the standard Southern blotting procedure. A cDNA clone (pc2S15), that contained more than 90% of the coding region of IDS mRNA, was isolated from a human endothelial cDNA library (Wilson *et al.* submitted). All patients diagnosed as MPS II had less than 2% of normal IDS activity in cultured skin fibroblasts or peripheral blood

leucocytes (Lim *et al.* 1974; Hopwood 1979). The somatic cell hybrid CY34, containing Xpter-q28, was derived from a girl with the clinical and biochemical manifestations of Hunter syndrome and a t(X;5) reciprocal translocation (Suthers *et al.* 1989). CY34A, containing only Xq24-q28, was a subclone of CY34.

The pc2S15 probe was labelled and used in Southern blots of *Pst*I or *Hind*III-digested DNA (Robertson *et al.* 1988; Nelson *et al.* 1989). The pc2S15 probe was <sup>3</sup>H-labelled and hybridised (Simmers *et al.* 1988) to metaphases from a normal female and from two males expressing the fragile (*FRAXA*) site at Xq27.3.

## RESULTS AND DISCUSSION

The chromosomal location of the IDS gene has been inferred by analysis of the translocation breakpoint in a female patient with Hunter syndrome due to a t(X;5) translocation and consistent inactivation of the normal X chromosome (Mossman *et al.* 1983; Roberts *et al.* 1989; Suthers *et al.* 1989). The X-chromosome breakpoint was at the Xq27/Xq28 boundary (Roberts *et al.*, 1989). The translocation was postulated to disrupt the function of the Hunter syndrome gene, although it was not clear that the translocation breakpoint lay within the IDS gene. Suthers *et al.* (1989) analysed a somatic cell hybrid, CY34, which contained the derived X-chromosome from this patient and lacked the Xq28-qter fragment. The breakpoint was mapped in relation to DNA markers at Xq28; the locus order at Xq28 was centromere-*DXS98-DXS369-FRAXA-DXS296-DXS304-DXS374*-telomere (Suthers *et al.* 1989; Mandel *et al.* 1989) and the breakpoint was between *DXS296* and *DXS304*.

Figure 1 shows that in normal DNA pc2S15 detected 8 fragments (7.5, 5.5, 4.1, 4.0, 2.5, 2.3, 1.3, and (very faintly) 0.76 kb), whereas in CY34 and CY34A DNA four fragments were not detected (5.5, 2.5, 1.3 and 0.76 kb) and a new 1.0 kb fragment

was faintly visible (Fig. 1). A very faint 14 kb fragment was visible in the A9 lane. pc2S15 did not detect polymorphic *Hind*III fragments in DNA samples from 16 normal X chromosomes. This demonstrated that the translocation disrupted the IDS gene and that the IDS gene lay between *DXS296* and *DXS304*. *In situ* hybridisation of pc2S15 to the chromosomes of a normal female established that there were no sequences homologous to pc2S15 elsewhere in the genome (Fig. 2). As expected, the IDS gene was distal to the fragile X mutation (*FRAXA*) site at Xq27.3. The silver grains which appeared to touch a chromosome were scored in Figure 2a, showing 150 grains from 30 metaphases of a normal female. pc2S15 only detected sequences at distal Xq; sequences homologous to pc2S15 were not detected elsewhere on the X chromosome or on the autosomes. There was no significant hybridisation to the Y chromosomes of two normal males (data not shown). An additional 40 metaphases with high resolution chromosome banding of Xq (600-1000 bands per metaphase) were scored in this female and indicated that the IDS gene is located at Xq28 (Figure 2b). When pc2S15 was hybridised to chromosomes expressing *FRAXA*, 79 grains were scored relative to the fragile site (data not shown). 45 grains lay between *FRAXA* and the telomere, 21 grains lay within a similar distance proximal to *FRAXA*, and 13 grains lay centrally over the fragile site. The difference between the number of proximal versus distal grains was significant ( $\chi^2 = 8.73$ ;  $p < 0.005$ ) giving a location for the IDS gene distal to *FRAXA*. A genomic clone which contained only the 5'-end of the IDS gene (P.J. Wilson, unpublished observations) detected a 1.3 kilobase (kb) fragment in normal female DNA but did not detect any fragments in CY34, indicating that the IDS gene was orientated with the 5'-end on the telomeric side of the CY34 breakpoint.

The pc2S15 probe was used to analyse genomic DNA from 23 unrelated British and Australian males who had Hunter syndrome (Fig. 3). Seven individuals had structural alterations of the IDS gene; two (03-1 and 04-1) had deletions of the



entire pc2S15 coding region while five had various partial deletions or rearrangements. pc2S15 did not hybridise to DNA samples from patients 03-1 and 04-1 (Figure 3, lanes 2 and 5 respectively). Other probes demonstrated the presence of similar amounts of DNA in each lane when hybridised to the same filter. The DNA samples in lanes 4,7,8 and 10 (Figure 3) each had a novel pattern of DNA fragments indicating partial deletions or rearrangements of the IDS gene. pc2S15 did not detect polymorphic *Pst*I fragments in DNA from 16 normal X chromosomes. The mothers of three of these seven patients had their carrier status determined by hair root analysis (Hopwood *et al.* 1982). One mother was not a carrier suggesting that the mutation had occurred during oogenesis. The remaining sixteen males had Southern blot patterns identical to those found in normal controls.

The filter shown in Fig. 3 was re-probed to determine whether markers near the IDS gene were deleted in 03-1 and 04-1. *DXS296* was absent in 04-1 indicating that his deletion extended proximally from the IDS gene. The deletion in 03-1 did not extend to *DXS296* and neither deletion extended to include *DXS98*, *DXS369*, *DXS304* or *DXS374*. Pulsed field gel electrophoresis studies have demonstrated that *DXS296* approximately 800 kb proximal to the IDS gene (Y. Sui, personal communication). The probe which detects *DXS296* (VK21; Suthers *et al.* 1989) also detects a single conserved *Hind*III fragment in both mouse (1.9 kb) and hamster DNA (1.8 kb) (data not shown).

Both 03-1 and 04-1 have extremely severe features of Hunter syndrome. They presented in their first year of life with developmental delay and by the second year had developed hernias, curvature of their spines, enlarged livers and spleens and developmental regression. In contrast with other patients, neither attained speech and both were troubled by epileptic seizures from an early age. Patient 03-1 also had a congenital abnormality of the eyelids causing excessive drooping

(ptosis). They did not have any other congenital malformations, although the severity of their Hunter phenotype made it difficult to exclude other minor congenital physical or intellectual abnormalities attributable to their deletions extending beyond the IDS gene.

The frequency of various deletions or rearrangements at the IDS gene proves that Hunter syndrome is genetically heterogeneous. The patients who did not have alterations of their IDS gene detected by pc2S15 presumably had more subtle mutations. This genetic heterogeneity may explain much of the phenotypic variability noted in this condition. The probe pc2S15 will be of value in providing genetic advice to families with detectable alterations of the IDS gene.

The observation of one patient with an extremely severe phenotype and a deletion extending from the IDS gene to an adjacent conserved sequence (*DXS296*) suggests that some Hunter patients could have additional symptoms due to deletions of genes other than for IDS. Mental retardation is the dominant feature of the Fragile X syndrome and *FRAXA* is adjacent to the IDS gene. *DXS296* shows no recombination with *FRAXA* (Suthers *et al.* 1989) and deletions extending from the IDS gene to *DXS296* could conceivably include *FRAXA*. Furthermore, deletions extending distally from the IDS gene could encompass genes in Xq28.

Despite the severity of their clinical presentations the boys 03-1 and 04-1 did not have specific features to suggest that a gene other than IDS had been deleted. With the development of a large scale restriction map of this region it may be possible to correlate the extent of deletions in Hunter syndrome with the phenotype. Conversely, the identification of patients with deletions of the IDS gene that extend towards *FRAXA* will assist in localising new DNA markers near *FRAXA* and ultimately in characterising the Fragile X mutation.

**ACKNOWLEDGEMENTS**

We thank the following colleagues for kindly providing DNA probes: Dr B.A. Oostra (RN1/DXS369); Dr N. Dahl (U6.2/DXS304); Dr K.E. Davies (1A1.1/DXS374). This research was supported by the National Health and Medical Research Council of Australia and the Adelaide Children's Hospital Research Foundation.

## REFERENCES

Bielicki J, Freeman C, Clements PR, Hopwood JJ (1990) Human liver iduronate-2-sulphatase: purification, characterization and catalytic properties. *Biochem J*  
In press

Hopwood JJ (1982)  $\alpha$ -L-Iduronidase,  $\beta$ -D-glucuronidase and 2-sulfo-L-iduronate 2-sulphatase preparation and characterization of radioactive substrates from heparin. *Carbohydr Res* 69: 203-216

Hopwood JJ (1989) Enzymes that degrade heparin and heparan sulphate. In: *Heparin: Clinical and Biological Properties, Clinical Applications* Lane D & Lindahl U (eds.) Edward Arnold, London, pp 191-288

Hopwood JJ, Muller V, Harrison JR, Carey WF, Elliott H, Robertson EF, Pollard AC (1982) Enzymatic diagnosis of the mucopolysaccharidoses: experience of 96 cases diagnosed in a 5 year period. *Med J Aust* 1: 257-260.

Lim TW, Leder IG, Bach G, Neufeld EF (1974) An assay for iduronate sulfatase (Hunter corrective factor). *Carbohydr Res* 37: 103-109

Mandel JL, Willard, HF, Nussbaum RL, Romeo G, Puck JM, Davies KE (1989) Report of the committee on genetic constitution of the X chromosome. *Cytogenet Cell Genet* 51: 384-437

Mossman J, Blunt S, Stephens R, Jones EE, Pembrey M (1983) Hunter disease in a girl: association with X:5 chromosomal translocation disrupting the Hunter gene. *Arch Dis Child* 58: 911-915

Nelson PV, Carey WF, Morris CP, Pollard AC (1989) Cystic fibrosis: prenatal diagnosis and carrier detection by DNA analysis. *Med J Aust* 151: 126-131

Neufeld EF, Muenzer J (1989) The Mucopolysaccharidoses. In: *The Metabolic Basis of Inherited Disease* Scriver CR, Beaudet AL, Sly WS, Valle D, (eds.) 6th Edition, McGraw Hill, New York, pp 1565-1588

Roberts SH, Upadhyaya M, Sarfarazi M, Harper PS (1989) Further evidence localising the gene for Hunter's syndrome to the distal region of the X chromosome long arm. *J Med Genet* 26: 309-313

Robertson DA, Callen DF, Baker EG, Morris CP, Hopwood JJ (1988) Chromosomal localization of the gene for human glucosamine-6-sulphatase to 12q14. *Hum Genet* 79: 175-178

Simmers RN (1988)

Hum Genet 78: 134-

Sutherland GR, Hecht F (1985) *Fragile Sites on Human Chromosomes*, Oxford University Press, New York

Suthers GK, Callen DF, Hyland VJ, Kozman HM, Baker E, Eyre H, Harper PS, Roberts SH, Hors-Cayla MC, Davies KE, Bell MV, Sutherland GR (1989) A new DNA marker tightly linked to the Fragile X locus (FRAXA). *Science* 246: 1298-1300

Wilson PJ, Morris CP, Anson DS, Occhiodoro T, Bielicki J, Clements PR, Hopwood JJ (1990) Hunter syndrome: isolation of an iduronate-2-sulphatase cDNA clone. Manuscript submitted

Figure 1

Southern blot of *Hind*III-digested DNA samples from a normal female (lane 1), normal male (lane 2), the human/mouse cell lines CY34 (lane 3) and CY34A (lane 4), and the mouse cell line A9 (lane 5). CY34 contained Xpter-q28 while CY34A was a subclone of CY34 having just Xq24-q28; both cell lines had a mouse (A9) background (Suthers *et al.* 1989). Approximate DNA fragment size indicators are shown on the left of the figure.

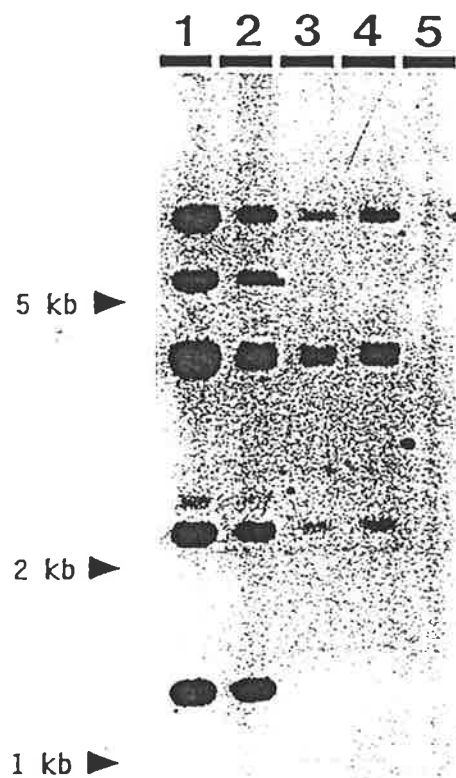


Figure 2

Ideograms of G-banded human chromosomes showing the distribution of silver grains after hybridisation of pc2S15 to (a) individual human metaphase chromosomes, and (b) Xq.

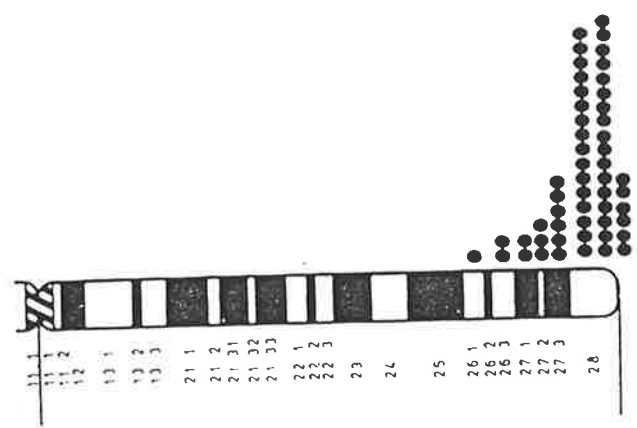
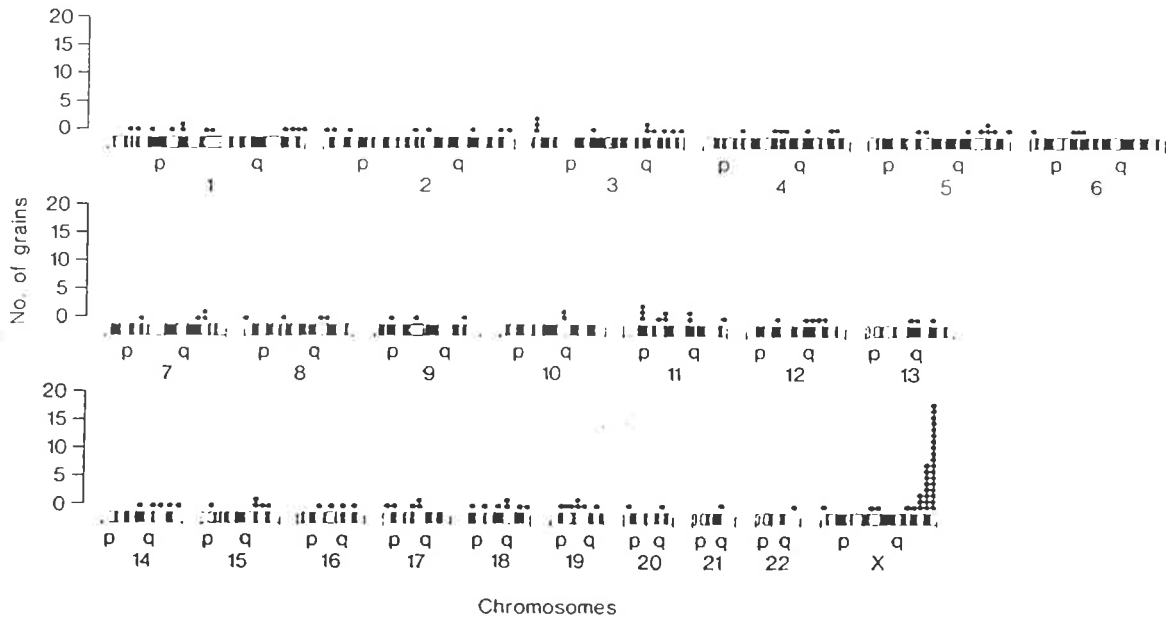
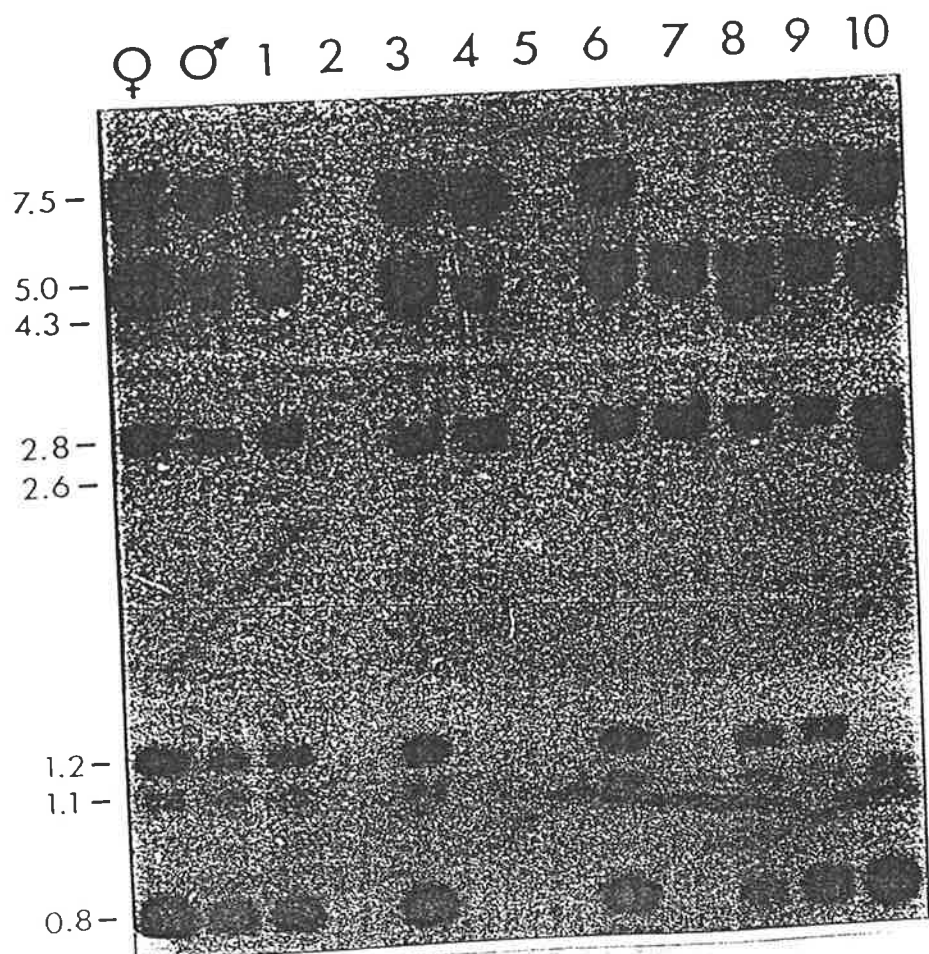


Figure 3

pc2S15 was used to probe a Southern blot of *Pst*I-digested DNA samples from a normal female and normal male (indicated by symbols above the lanes) and from 23 Hunter syndrome patients (10 samples are shown in lanes 1-10). Each lane was loaded with similar amounts of DNA. The pattern of DNA fragments from 13 other MPS II patients (data not shown) and in lanes 1,3,6 and 9 was the same as in the control samples. The positions of DNA size markers (in kb) are shown on the left of the figure.







Am J Hum Genet (in press).

due August 1990

Suthers et al. Page 1.

March 1, 1990

PHYSICAL MAPPING OF NEW DNA PROBES NEAR THE FRAGILE X (FRAXA)  
WITH A PANEL OF CELL LINES

GK Suthers<sup>1</sup>, VJ Hyland<sup>1\*</sup>, DF Callen<sup>1</sup>, I Oberle<sup>2</sup>, M Rocchi<sup>3</sup>, NS Thomas<sup>4</sup>, CP Morris<sup>5</sup>, CE Schwartz<sup>6</sup>, M Schmidt<sup>7</sup>, HH Ropers<sup>8</sup>, E Baker<sup>1</sup>, BA Oostra<sup>9</sup>, N Dahl<sup>10</sup>, PJ Wilson<sup>5</sup>, JJ Hopwood<sup>5</sup>, and GR Sutherland<sup>1</sup>.

---

1 Department of Cytogenetics & Molecular Genetics, Adelaide Children's Hospital, North Adelaide, South Australia 5006, Australia.

2 L.G.M.E./C.N.R.S., INSERM U184, Faculte de Medecine, 11 Rue Humann, 67085 Strasbourg Cedex, France.

3 Laboratorio di Genetica Molecolare, Ist. Gaslini, Largo Gaslini 5, 16134, Genova, Italy.

4 Institute of Medical Genetics, University of Wales College of Medicine, Heath Park, Cardiff CF4 4XN, Wales.

- 5 Lysosomal Diseases Research Unit, Department of Chemical Pathology, Adelaide Children's Hospital, North Adelaide, South Australia 5006, Australia.
- 6 Greenwood Genetic Center, 1 Gregor Mendel Circle, Greenwood, South Carolina SC 29646, U.S.A.
- 7 The Murdoch Institute, Royal Children's Hospital, Parkville, Victoria 3052, Australia.
- 8 Department of Human Genetics, University of Nijmegen, Radboud Hospital, P.O. Box 9101, 6500HB Nijmegen, The Netherlands.
- 9 Department of Cell Biology, Erasmus University, P.O. Box 1738, 3000DR Rotterdam, The Netherlands.
- 10 Department of Medical Genetics, Biomedical Centre, Box 589, S-751 23, Uppsala, Sweden.

\*Current Address: Molecular Genetics Unit, Kolling Institute of Medical Research, Royal North Shore Hospital, St. Leonards, NSW 2065, Australia.

---

Address for correspondence and reprints: Dr GR Sutherland, Department of Cytogenetics and Molecular Genetics, Adelaide Children's Hospital, North Adelaide, South Australia 5006, Australia.

---

Running heading: Mapping of DNA probes near FRAXA

## SUMMARY

The Fragile X syndrome is a very common disorder but there has been little progress towards isolating the Fragile X mutation (FRAXA). We describe a panel of 14 somatic cell hybrid lines, lymphoblastoid cell lines, and peripheral lymphocytes with X chromosome translocation or deletion breakpoints near FRAXA. The locations of the breakpoints were defined with 16 established probes between pX45d (DXS100) and St14-1 (DXS52). Seven of the cell lines had breakpoints between the probes RN1 (DXS369) and U6.2 (DXS304) which flank FRAXA at a distances of 3-5 centimorgans. The panel of cell lines was used to localize 16 new DNA probes in this region. Six of the probes, VK16, VK18, VK23, VK24, VK37, and VK47, detected loci near FRAXA and it was possible to order both the X chromosome breakpoints and the probes in relation to FRAXA. The order of probes and loci near FRAXA is

cen-RN1, VK24-VK47-VK23-VK16, FRAXA-VK21A-VK18-IDS-VK37-U6.2-qter.

The breakpoints near FRAXA are sufficiently close together that probes localized with this panel can be linked on a large scale restriction map by pulsed field gel electrophoresis. This panel of cell lines will be valuable in rapidly localizing other probes near FRAXA.

## INTRODUCTION

The Fragile X syndrome is the most common cause of familial mental retardation and consumes significant health care resources in Western societies (Sutherland & Hecht 1985; Turner et al. 1986). It is characterized by a folate-sensitive fragile site at Xq27.3. The locus responsible for the Fragile X syndrome (FRAXA) is located at or very near the fragile site but the development of a precise genetic map in this region has been hampered by a lack of closely linked polymorphic loci. Until recently the closest probes to FRAXA, cX55.7 (DXS105), 4D-8 (DXS98), and St14-1 (DXS52), lay more than 10 centimorgans (cM) from FRAXA (Mandel et al. 1989).

In 1989 three probes that detected restriction fragment length polymorphisms (RFLPs) within 5 cM of FRAXA were reported, RN1 (DXS369), VK21A (DXS296), and U6.2 (DXS304) (Hupkes et al. 1989; Suthers et al. 1989a; Vincent et al. 1989). The established order of probes and genes near FRAXA was

cen-F9-cX55.7-4D8-RN1-FRAXA-VK21A-IDS-U6.2-1A1.1-St14-qter (Mandel et al. 1989). A key factor in the localization of RN1, VK21A, and U6.2 was the mapping of the corresponding locus using somatic cell hybrids before the search for RFLPs was undertaken. A panel of cell lines with well-defined X chromosome breakpoints near FRAXA would allow for the rapid identification of more probes close to FRAXA and the further development of precise genetic and physical maps around FRAXA.

In this paper we describe a panel of somatic cell hybrid lines, lymphoblastoid cell lines, and peripheral lymphocytes that contain human X chromosomes with deletion or translocation breakpoints near FRAXA. This represents a collaborative effort that was initiated at the Fourth International Workshop on the Fragile X Syndrome and X-linked Mental Retardation held in New York in July, 1989. This panel of cell lines was used to physically localize a series of new DNA probes near FRAXA. The mapping of one of these probes proximal to FRAXA by in situ hybridization defined the locations of the X chromosome breakpoints and the other probes in relation to FRAXA. Six of the new probes and the breakpoints in seven of the cell lines were located close to FRAXA in the interval RN1-U6.2.

## MATERIAL AND METHODS

CELL LINES

The cell type, cytogenetic description, contributing co-author, and references for each cell line are listed in Table 1. Those cell lines that have not been described elsewhere are presented below.

## LL556

This is a lymphoblastoid line from one of two brothers with hemophilia B (Factor 9 deficiency) and mental retardation. The boys had cytogenetically visible deletions extending from Xq26.2 to Xq27.2. The probes cX38.1 (DXS102) and RN1 and intervening probes have been shown to be deleted in both boys (Mandel et al., manuscript in preparation).

## TC4.8

Peripheral lymphocytes from a male expressing the fragile site at Xq27.3 were fused (Davidson & Gerald 1976) with HPRT<sup>-</sup>/G6PD<sup>-</sup> hamster cells (YH.21) (Rosenstraus & Chasin 1975). After selection for HPRT in HAT medium, one clone (HY.84P11) containing a human X chromosome and no other human chromosomes was isolated. HY.84P11 was treated with FUdR and caffeine to induce expression of the fragile site (Abruzzo et al. 1986) and breakage of the X chromosome at that point. Clones retaining HPRT were selected

against with 6-thioguanine; surviving clones were selected for retention of G6PD by treatment with diamide (D'Urso et al. 1983). TC4.8 was a homogeneous clone that retained G6PD. Subsequent analysis (see below) demonstrated that the breakpoint was not at the fragile site but had occurred more proximally on the X chromosome.

#### Y.162.Aza

Y.162.Aza was derived from a somatic cell hybrid line (Y.162.SE1T4) which contained an intact late-replicating human X chromosome in a hamster background. On cytogenetic screening of Y.162.SE1T4, a proportion of clones were noted to contain an elongated human X chromosome with an additional early-replicating fragment attached at Xqter. A subclone homogeneous for this rearrangement (Y.162.Aza) was studied with in situ hybridization of labelled total human or total hamster DNA. The early-replicating fragment attached to the human X chromosome was shown to be of hamster origin, and no other human chromosomal material was detected (Rocchi et al. 1989).

#### APC-5

A girl with mild mental retardation and no major dysmorphic features was found to have a balanced X;autosome translocation: 46,X t(X;19)(Xpter->Xp11.2;19q13.3->19pter). Skin fibroblasts were fused (Davidson & Gerald 1976) with an HPRT<sup>-</sup> hamster cell line (Wg3h) and clones containing the derived X chromosome were selected for HPRT in HAT medium. The human chromosomal material



in the cell line was fragmented and the cell line has been only partially characterised cytogenetically. Subsequent analysis (see below) demonstrated a breakpoint at Xq27 which was presumably due to fragmentation in the hybrid cell line.

#### HUNTER SYNDROME DNA

Two boys, 03-1 and 04-1, with Hunter syndrome (iduronate-2-sulfatase [IDS] deficiency; mucopolysaccharidosis type II) due to complete deletions of IDS have been reported (Wilson et al. 1990). DNA was extracted from peripheral lymphocytes of these boys. For the sake of brevity in the text, the peripheral lymphocytes will be included in the term "cell lines".

#### DNA PROBES

Table 2 lists the locus names and locations of 15 established DNA probes used to localize the breakpoints in the cell lines. The breakpoint in cell line CY34 lies between the probes VK21A and U6.2 and is within the Hunter syndrome gene (Suthers et al. 1989). The probe pc2S15 is an IDS cDNA clone (Wilson et al., manuscript in preparation) which spans the breakpoint in CY34 and detects deletions and rearrangements at IDS in individuals with Hunter syndrome (Wilson et al. 1990).

The isolation of the series of 16 new DNA probes (all with the

prefix "VK") has been described (Hyland et al. 1989). The locus names for these probes are listed in Table 3.

#### DNA METHODS

DNA was extracted from peripheral lymphocytes using the phenol/chloroform method (Maniatis et al. 1982). Cell line DNA was extracted using the high-salt extraction method (Miller et al. 1988) and treated with RNase A (Boehringer Mannheim; 0.15 ug/ml at 37° for 4 hours) followed by proteinase K (Boehringer Mannheim; 0.07 ug/ml at 37° overnight). The DNA was precipitated by the addition of 1/3 the volume of saturated NaCl and 1/5 the volume of 50% PEG 6000, gently mixed, and allowed to stand at 4° overnight; the DNA pellet was recovered by spinning at 2500 rpm at 4° for 15 minutes, washed with 70% ethanol, dessicated, and resuspended in 10mM TrisHCl/1mM EDTA.

DNA from cell lines was digested with HindIII, EcoRI, or TaqI (New England Biolabs). Digested DNA was electrophoresed in 0.8% agarose and transferred to nylon filters (Gene Screen Plus) by Southern blotting. DNA from normal human lymphocytes, mouse (A9) cells (Callen 1986), and hamster (RJK88) cells (Fusco et al. 1983) were used as positive and negative controls for physical mapping studies. The probe VK21A is known to detect conserved sequences in man, mouse, and hamster (Wilson et al. 1990). The new probe VK25 also detected a single fragment in mouse and

hamster DNA that was easily distinguished from the human fragments.

Probes were radio-labelled by random primer extension to incorporate  $^{32}\text{P}$ -dCTP (Amersham). Probes containing repeated DNA sequences were pre-reassociated with an excess of unlabelled human DNA (Sealey et al. 1985). Nylon filters were prehybridized with 5xSSC/50% formamide/1% SDS/7% dextran at  $42^{\circ}$  for 1 hour. After addition of the labelled probe the filters were hybridized at  $42^{\circ}$  overnight and then washed in 2xSSC/0.5% SDS and 0.1xSSC/0.1% SDS (30 minutes each at  $65^{\circ}$ ). The labelled filters were exposed to X-Omat film (Kodak) at  $-70^{\circ}$  for 1 to 14 days.

For in situ hybridization studies,  $^3\text{H}$ -labelled probes were hybridized (Simmers et al. 1988) to prephotographed metaphase chromosomes at concentrations of 0.01-0.1  $\mu\text{g}/\text{ml}$  for 19-27 days. Probes containing repeats were pre-reassociated with unlabelled total human DNA (Sealey et al. 1985). Silver grains that appeared to touch the X chromosome were scored.

## RESULTS

Mapping translocation and deletion breakpoints

The established probes listed in Table 2 were used to probe samples of cell line DNA and DNA from the two boys with complete deletions of IDS. The probes used for each cell line were selected so as to delineate the breakpoints on the X chromosome. The presence or absence of the respective loci in each sample are listed in Table 3. Four of the cell lines, LL556, LC12K15, 03-1, and 04-1, had interstitial deletions (Table 3); the suffix 'p' or 'd' will be used to indicate the proximal or distal breakpoints in these cell lines.

Four cell lines, PeCH-N, LL556d, APC-5, and 04-1p, had breakpoints between the probes which flank FRAXA, RN1 and VK21A. Using the established probes it was not possible to determine the order of the breakpoints either in relation to each other or to FRAXA. Three cell lines had breakpoints between 4D-8 and RN1, proximal to FRAXA. Four cell lines had breakpoints between VK21A and U6.2, distal to FRAXA.

There were no inconsistencies in these data to suggest complex rearrangements in the cell lines in the region Xq26-28. The human chromosomal component of the cell line APC-5 is known to be fragmented but the data in Table 3 do not indicate an interstitial deletion or rearrangement at Xq26-28. Assuming that

complex rearrangements were not present, the locations of translocation breakpoints and the locations and extent of interstitial deletions in the cell lines are summarized in Figure 1.

#### Mapping new DNA probes using the cell line panel

Sixteen VK probes were mapped using the panel of cell lines (Fig. 2). The presence or absence of the respective loci in each cell line is listed in Table 3 and summarized in Fig. 1.

Five of the VK probes were located close to FRAXA between the probes RN1 and U6.2. One probe was located in the same interval as RN1. Eight probes were located proximal to RN1, one was in the same interval as U6.2 and 1A1.1, and one mapped distal to 1A1.1.

The VK probes indicated the order of some of the breakpoints that could not be resolved using the established probes. VK14 (DXS292) lay between 4D-8 and RN1 and separated the TC4.8 breakpoint from the 908K1B17 and 2384-A2 breakpoints. VK37 (DXS302) lay between IDS and U6.2 and separated the LC12K15d breakpoint from the 03-1d and 04-1d breakpoints.

Of particular interest were the results with VK probes that detected loci close to FRAXA. The probes VK16 (DXS293), VK23 (DXS297), and VK47 (DXS308) were located between RN1 and VK21A

and resolved the location of the breakpoints of the four cell lines with breakpoints near FRAXA. The order of breakpoints between RN1 and VK21A down the chromosome was LL556d/PeCH-N/APC-5/04-1p. A fourth probe, VK18 (DXS295), was in the same interval as VK21A.

The location of VK23 in relation to FRAXA was determined by in situ hybridization. A 2.6 kb EcoRI-PstI fragment of VK23 (VK23B1) containing few repeated sequences was isolated. VK23B1 was hybridized to chromosomes of a male expressing the fragile site at Xq27.3. Of the 42 silver grains that could be scored relative to the fragile site, 8 grains lay between the fragile site and the telomere, 31 grains lay within a similar distance proximal to the fragile site, and 3 grains were located centrally over the chromosome gap at the fragile site. The difference in the number of proximal versus distal grains was significant ( $\chi^2 = 13.56$ ;  $p < 0.0005$ ) indicating that VK23 was proximal to FRAXA. This in turn indicated that VK47 and the PeCH-N and LL556d breakpoints were proximal to FRAXA.

The location of the probe VK11 (DXS291) in relation to the breakpoint in cell line GM08121 was also determined by in situ hybridization. There was no significant difference in the number of grains on the normal versus deleted X chromosomes (normal/deleted chromosomes=24/23 grains;  $\chi^2 = 0.02$ ,  $p > 0.10$ ) indicating that VK11 was proximal to the breakpoint.

The probes VK29 (DXS300), VK34 (DXS301), and VK41 (DXS310) detected an interstitial deletion in the cell line CY34A, a subclone of CY34 which contained DNA from Xq24-26 (Suthers et al. 1989a). These probes detected the corresponding loci in cell line CY34 (Table 3) but did not detect the loci in CY34A (data not shown). On the basis of results from other cell lines the three probes could be localized to Xq26. This indicated that CY34A had an interstitial deletion at Xq26. Two other probes in this region, VK10 (DXS290) and VK17 (DXS294), detected loci in both CY34 (Table 3) and CY34A (data not shown).

The gene probe for a cerebellar-degeneration-related protein, CDR-9, has been localised to the interval between the GM08121 and 2384-A2 breakpoints (Hirst et al. 1990). CDR-9 detected sequences in the cell line LC12K15 (Table 3) thus localising CDR proximal to the LC12K15p breakpoint in the same interval as the probe cX55.7.

## DISCUSSION

Somatic cell hybrids containing human X chromosomes with translocation or deletion breakpoints have been used to physically map DNA probes in various regions of the long arm of the human X chromosome (Wieacker et al. 1984; Oberle et al. 1986; Hofker et al. 1987; Cremers et al. 1988). Few cell lines have been described with breakpoints close enough to FRAXA to be useful in localising new probes near this locus. We have presented a number of cell lines with precisely defined breakpoints close to FRAXA and have localised a series of DNA probes close to FRAXA (Fig. 1).

The availability of these cell lines makes it feasible to rapidly localise clones from a DNA library and to identify those clones derived from regions physically close to FRAXA. A total of 13 breakpoints in 10 cell lines are now defined between cX55.7 and U6.2. Four of these cell lines had breakpoints between RN1 and VK21A. The cell line with the closest breakpoint known to be proximal to FRAXA was PeCH-N; the closest breakpoint known to be distal to FRAXA was 03-1p. The APC-5 and 04-1p breakpoints were within the interval defined by the PeCH-N and 03-1p breakpoints, but the locations of the APC-5 and 04-1p breakpoints in relation to FRAXA are unknown.

Two intervals between RN1 and U6.2 contained more than one probe, RN1 with VK24 (DXS298), and VK21A with VK18. Probes that detect



RFLPs can be ordered by genetic linkage studies, but if the probes are very close together it becomes increasingly unlikely that recombination between the loci will be observed. It was possible to order the probes near FRAXA by pulsed field gel electrophoresis. VK18 has been shown to be distal to VK21A using pulsed field gel electrophoresis and VK21A, VK18, IDS, and U6.2 have been linked together on a large scale restriction map (Hyland et al. 1990). Since VK21A detects no recombination with FRAXA (Suthers et al. 1989a) and U6.2 lies 3 cM distal to FRAXA (Mandel et al. 1989) this map encompasses five probes (including VK37) and five breakpoints in a region immediately distal to FRAXA. It may be possible to generate a similar map of the region proximal to FRAXA. The probe RN1 lies 5 cM proximal to FRAXA (Oostra et al. 1990b) and the region between RN1 and VK21A now encompasses five probes (and possibly a sixth, VK24), four breakpoints, and FRAXA. The development of the large scale restriction map around FRAXA would be further enhanced by the use of probes from linking or jumping libraries that had been localized with the cell panel.

The order of probes and genes near FRAXA is now  
cen-RN1, VK24-VK47-VK23-VK16, FRAXA-VK21A-VK18-IDS-VK37-U6.2-qter.  
The fact that these new VK probes could be easily localised and ordered near FRAXA without any genetic linkage studies demonstrates the value of this panel of cell lines in further investigation of FRAXA. The established probes RN1, VK21A, and U6.2 detect RFLPs (Hupkes et al. 1989; Suthers et al. 1989a;

Vincent et al. 1989). If RFLPs are detected by the VK probes they will be valuable in developing the fine scale genetic linkage map around FRAXA. A polymorphism detected by VK16 would be particularly useful in both analyzing the unusual segregation of the Fragile X mutation (Sherman et al. 1985) and in providing genetic advice in affected pedigrees.

The boy 04-1 had a deletion that included IDS and extended towards FRAXA (Table 3). It is conceivable that this deletion encompassed FRAXA or other genes. Patients with Hunter syndrome may have complete deletions of IDS (Wilson et al. 1990) and it will be important to carefully correlate the extent of large deletions around IDS with the patients' phenotypes to define any contiguous gene syndromes (Schmickel 1986). The general availability of lymphoblastoid or fibroblast lines from such patients would be a valuable resource for mapping other new probes near IDS and FRAXA. At present there is no evidence to suggest that FRAXA is itself a deletion (Sutherland et al. 1985; Laird et al. 1987) and DNA from patients with Hunter syndrome is currently the best potential source of interstitial deletions near FRAXA.

The utility of these cell lines is not limited to studies of the Fragile X syndrome. An increasing number of disorders are being localized to Xq26-28 (Mandel et al. 1989). The rapid mapping of DNA probes within a small region of the human genome with a panel of cell lines opens up the possibility of isolating yeast

artificial chromosomes (YACs) that overlap in the region, developing a large scale restriction map, localizing conserved sequences or cDNA probes, and ultimately isolating the gene of interest.

## ACKNOWLEDGEMENTS

We would like to thank the following colleagues for advice and assistance: John Mulley, Rob Richards, and Julie Nancarrow (Adelaide); Paul Kalitsis (Melbourne); Suzanne Sauer and Angie Brown (Greenwood); S Castagnola and I Giambarrasi (Genova). The following people kindly provided DNA probes or cell lines: Dr M Siniscalco (CDR-9); Dr KE Davies (IA1.1); Dr MC Hors-Cayla (PeCH-N and PeCH-A); Dr JE Wraith (DNA from 03-1 and 04-1). This project was supported by the National Health and Medical Research Council of Australia, the Adelaide Children's Hospital Research Foundation, and Progetto Strategico Genoma Humano (Rome).

REFERENCES

- 1 Abruzzo MA, Pettay D, Mayer M, Jacobs PA (1986). The effect of caffeine on fragile X expression. Hum Genet 73:20-22.
- 2 Callen DF (1986). A mouse-human hybrid cell panel for mapping human chromosome 16. Ann Genet 29:235-239.
- 3 Cremers FPM, van de Pol TJR, Wieringa B, Hofker MH, Pearson PL, Pfeiffer RA, Mikkelsen M, Tabor A, Ropers HH (1988). Molecular analysis of male-viable deletions and duplications allows ordering of 52 DNA probes on proximal Xq. Am J Hum Genet 43:452-461.
- 4 Davidson RL, Gerald PS (1976). Improved techniques for the induction of mammalian cell hybridization by polyethylene glycol. Somat Cell Genet 2:165-176.
- 5 D'Urso M, Marenzi C, Toniolo D, Piscopo M, Schlessinger D, Luzzatto L (1983). Regulation of glucose-6-phosphate dehydrogenase expression in CHO-human fibroblast somatic cell hybrids. Somat Cell Genet 9:429-443.
- 6 Hirst MC, Bell MV, MacKinnon RN, Watson JEV, Callen D, Sutherland G, Dahl N, Patterson MN, Schwartz C, Ledbetter D, Ledbetter S, Davies KE (1990). Mapping a cerebellar

degeneration related protein and DXS304 around the fragile site. Am J Med Genet (in press).

- 7 Hofker MH, Bergen AAB, Skraastad MI, Carpenter NJ, Veenema H, Connor JM, Bakker E, van Ommen GJB, Pearson PL (1987). Efficient isolation of X chromosome-specific single-copy probes from a cosmid library of human X/hamster hybrid-cell line: mapping of new probes close to the locus for X-linked mental retardation. Am J Hum Genet 40:312-328.
- 8 Hupkes PE, van Oost BA, Perdon LF, van Bennekom CA, Bakker E, Halley DJJ, Schmidt M, Smits A, Wieringa B, Oostra BA (1989). New polymorphic DNA marker (DXS369) close to the fragile site FRAXA. Human Gene Mapping 10 (1989): Tenth International Workshop on Human Gene Mapping. Cytogenet Cell Genet 51:1016.
- 9 Hyland VJ, Fernandez KEW, Callen DF, MacKinnon RN, Baker EG, Friend K, Sutherland GR (1989). Assignment of anonymous DNA probes to specific intervals of human chromosomes 16 and X. Hum Genet 83:61-66.
- 10 Hyland VJ, Nancarrow J, Callen DF, Suthers GK, Wilson PJ, Morris CP, Hopwood JJ, Dahl NP, Sutherland GR (1990). Physical linkage of four loci distal to the fragile X locus. (Submitted).

- 11 Keats B, Ott J, Conneally M (1989). Report of the committee on linkage and gene order. Human Gene Mapping 10 (1989): Tenth International Workshop on Human Gene Mapping. Cytogenet Cell Genet 51:459-502.
- 12 Kidd KK, Bowcock AM, Schmidtke J, Track RK, Ricciuti F, Hutchings G, Bale A, Pearson P, Willard HF, Gelernter J, Giuffra L, Kubzdela K (1989). Report of the DNA committee and catalogs of cloned and mapped genes and DNA polymorphisms. Human Gene Mapping 10 (1989): Tenth International Workshop on Human Gene Mapping. Cytogenet Cell Genet 51:622-947.
- 13 Laird C, Jaffe E, Karpen G, Lamb M, Nelson R (1987). Fragile sites in human chromosomes as regions of late-replicating DNA. Trend Genet 3:274-281.
- 14 Ledbetter SA, Schwartz CE, Davies KE, Ledbetter DH (1990). New somatic cell hybrids for physical mapping in distal Xq and the fragile X region. Am J Med Genet (in press).
- 15 Mandel JL, Willard HF, Nussbaum RL, Romeo G, Puck JM, Davies KE (1989). Report of the committee on the genetic constitution of the X chromosome. Human Gene Mapping 10 (1989): Tenth International Workshop on Human Gene Mapping. Cytogenet Cell Genet 51:384-437.

- 16 Maniatis T, Fritsch EF, Sambrook J (1982). Molecular Cloning: a Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- 17 Miller SA, Dykes DD, Pulesky HF (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucl Acid Res 16:1215.
- 18 Fuscoe JC, Fenwick RG, Ledbetter DH, Caskey CT (1983). Deletion and amplification of the HPRT locus in Chinese hamster cells. J Mol Cell Biol 3:1086-1096.
- 19 Oberle I, Camerino G, Kloepper C, Moisan JP, Grzeschik KH, Hellkuhl B, Hors-Cayla MC, Van Cong N, Weil D, Mandel JL (1986). Characterization of a set of X-linked sequences and of a panel of somatic cell hybrids useful for the regional mapping of the human X chromosome. Hum Genet 72:43-49.
- 20 Oostra BA, Majoor-Krakauer DF, van Hemel JO, Bakker E, Callen DF, Schmidt M, van Oost BA (1990a). Mapping of a new RFLP marker RN1 (DXS369) close to the fragile site FRAXA on Xq27-q28. Am J Med Genet (in press).
- 21 Oostra BA, Hupkes PE, Perdon LF, van Bennekom CA, Bakker E, Halley DJJ, Schmidt M, Du Sart D, Smits A, Wieringa B, van Oost BA (1990b). New polymorphic DNA marker close to the fragile site FRAXA. Genomics 6:129-132.



- 22 Patterson M, Schwartz C, Bell M, Sauer S, Hofker M, Trask B, van den Engh G, Davies KE (1987). Physical mapping studies on the human X chromosome in the region Xq27-Xqter. *Genomics* 1:297-306.
- 23 Rocchi M, Romeo G, Giambarrasi I, Castagnola S, Archidiacono N (1989). Physical dissection of the human X chromosome, with special reference to the region Xq27->qter. *Human Gene Mapping 10* (1989): Tenth International Workshop on Human Gene Mapping. *Cytogenet Cell Genet* 51:1067.
- 24 Rosenstraus M, Chasin LA (1975). Isolation of mammalian cell mutants deficient in glucose-6-phosphate dehydrogenase activity: linkage to hypoxanthine phosphoribosyl transferase. *Proc Natl Acad Sci USA* 72:493-497.
- 25 Schmickel RD (1986). Contiguous gene syndromes: a component of recognizable syndromes. *J Pediatr* 109:231-241.
- 26 Schmidt M, Certoma A, Du Sart D, Kalitsis P, Leversha M, Fowler K, Sheffield L, Jack I, Danks DM (1990). Unusual X chromosome inactivation in a mentally retarded girl with an interstitial deletion Xq27. Implications for the fragile X syndrome. *Hum Genet* (in press).

- 27 Schonk D, Coerwinkel-Driessen M, van Dalen I, Oerlemans F, Smeets B, Schepens J, Hulsebos T, Cockburn D, Boyd Y, Davis M, Rettig W, Shaw D, Roses A, Ropers H, Wieringa B (1989). Definition of subchromosomal intervals around the myotonic dystrophy gene region at 19q. *Genomics* 4:384-396.
- 28 Sealey PG, Whittaker PA, Southern EM (1985). Removal of repeated sequences from hybridization probes. *Nucl Acid Res* 13:1905-1922.
- 29 Sherman SL, Jacobs PA, Morton NE, Froster-Iskenius U, Howard-Peebles PN, Nielsen KB, Partington MW, Sutherland GR, Turner G, Watson M (1985). Further segregation analysis of the fragile X syndrome with special reference to transmitting males. *Hum Genet* 69:289-299.
- 30 Simmers RN, Smith J, Shannon MF, Wong G, Lopez AF, Baker E, Sutherland GR, Vadas MA (1988). Localization of the human G-CSF gene to the region of a breakpoint in the translocation typical of acute promyelocytic leukemia. *Hum Genet* 78:134-136.
- 31 Sutherland GR, Baker E, Fratini A (1985). Excess thymidine induces folate sensitive fragile sites. *Am J Med Genet* 22:433-443.

- 32 Sutherland GR, Hecht F (1985). Fragile Sites on Human Chromosomes. Oxford University Press, New York.
- 33 Suthers GK, Callen DF, Hyland VJ, Kozman HM, Baker E, Eyre H, Harper PS, Roberts SH, Hors-Cayla MC, Davies KE, Bell MV, Sutherland GR (1989a). A new DNA marker tightly linked to the fragile X locus (FRAXA). Science 246:1298-1300.
- 34 Suthers GK, Davies KE, Baker E, Sutherland GR (1989b). Iaql RFLP identified by probe 1A1 (DXS374) at Xq28. Nucl Acid Res 17:8901.
- 35 Turner G, Robinson H, Laing S, Purvis-Smith S (1986). Preventive screening for the fragile X syndrome. N Engl J Med 315:607-609.
- 36 Vincent A, Dahl N, Oberle I, Hanauer A, Mandel JL, Malmgren H, Pettersson U (1989). The polymorphic marker DXS304 is within 5 centimorgans of the fragile X locus. Genomics 5:797-801.
- 37 Wieacker P, Davies KE, Cooke HJ, Pearson PL, Williamson R, Bhattacharya S, Zimmer J, Ropers HH (1984). Toward a complete linkage map of the human X chromosome: regional assignment of 16 cloned single-copy DNA sequences employing a panel of somatic cell hybrids. Am J Hum Genet 36:265-276.

- 38 Wilson PJ, Suthers GK, Callen DF, Baker E, Nelson PV, Cooper A, Wraith JE, Sutherland GR, Morris CP, Hopwood JJ (1990). Frequent deletions at Xq28 indicate genetic heterogeneity in Hunter syndrome. (Submitted)

Table 1

List of the cell lines used in this study.

	Cell type <sup>a</sup>	Co-author	Reference
CY2	SCH/mouse	D.F.C.	2
CY3	SCH/mouse	D.F.C.	2,33
CY34	SCH/mouse	D.F.C.	33
PeCH-N <sup>b</sup>	SCH/hamster	D.F.C.	33
LL556	lymphoblastoid (XY)	I.O.	this paper
TC4.8	SCH/hamster	M.R.	23, this paper
Y.162.Aza	SCH/hamster	M.R.	23, this paper
APC-5	SCH/hamster	N.S.T.	this paper
04-1	lymphocytes (XY) <sup>c</sup>	J.J.H.	38
03-1	lymphocytes (XY) <sup>c</sup>	J.J.H.	38
2384-A2	SCH/hamster	C.E.S.	14
LC12K15	SCH/mouse	M.S.	26
908K1B17	SCH/hamster	H.H.R.	20, 27
GM08121	lymphoblastoid (XX)	N.I.G.M.S. <sup>d</sup>	22

a SCH indicates a somatic cell hybrid line; the background cell type is shown. The sex chromosome content of lymphocytes or lymphoblastoid lines is indicated.

b A somatic cell hybrid (PeCH-A) containing the reciprocal translocation product i.e. Xq26-qter, is available from D.F.C.

c Fibroblast cell lines have been established from these patients.

d U.S. National Institute of General Medical Sciences Human Genetic Mutant Cell Repository; a somatic cell hybrid line, 8121-A1, containing just the deleted X chromosome has been described (Ledbetter et al. 1990).

Table 2.

Locus name and location for each of the established probes used in the study.

Probe	Locus	Location	Reference
pX45d	<u>DXS100</u>	Xq25	12
pHPT30	<u>HPRT</u>	Xq26	12
52A	<u>DXS51</u>	Xq26.2-q26.3	12
cX38.1	<u>DXS102</u>	Xq26.2-q27.1	12
pVIII	<u>F9</u>	Xq26.3-q27.1	12
cX55.7	<u>DXS105</u>	Xq27.1-q27.2	12
4D-8	<u>DXS98</u>	Xq27.2	12
RN1	<u>DXS369</u>	Xq27.2-q27.3	12, 21
*			
VK21A	<u>DXS296</u>	Xq27.3-q28	12, 33
pc2S15	<u>IDS</u>	Xq28	33, 38
U6.2	<u>DXS304</u>	Xq28	12, 36
1A1.1	<u>DXS374</u>	Xq28	12, 34
p482.6	<u>F8C</u>	Xq28	12
pKSB	<u>G6PD</u>	Xq28	12
MN12	<u>DXS33</u>	Xq28	12
St14-1	<u>DXS52</u>	Xq28	12

The probes are listed in order down the X chromosome (Keats et al. 1989; Mandel et al. 1989). FRAXA (\*) is located between RN1 and VK21A. pc2S15 is a probe for IDS (Wilson et al., manuscript in preparation).

Table 3.  
Location of breakpoints in cell lines determined by Xq probes.

	CY2	CY3	LL556	GM08121	LC12K15	2384-A2	B17	TC4.8	PeCH-M	APC-5	O4-1	O3-1	CY34	Y162.A
px45d	+	-							+				+	
HPRT	-	+				+				+				+
52A			+							+				
VK10 (DXS290)	-	+	+		+	+							+	+
VK17 (DXS294)	-	+	+		+	+			+					
VK29 (DXS300)	-	+	+		+	+			+				+	
VK34 (DXS301)	-	+	+		+	+			+				+	
VK41 (DXS310)	-	+	+		+	+			+				+	
cX38.1														
F9	-	+		+	+	+				+				+
VK11 (DXS291)	-	+		+	+	+								
cX55-7	-	+			+	+			+	+			+	+
CDR-9					+	+							+	+
4D-8	-	+				+			+	+	+	+	+	+
VK7 (DXS288)	-	+				+			+				+	+
VK14 (DXS292)	-	+					+		+				+	
RM1	-	+					+	+	+	+	+	+	+	+
VK24 (DXS298)	-	+					+	+	+				+	
VK47 (DXS308)	-	+	+				+	+	+				+	
VK23 (DXS297)	-	+	+				+	+		+	+	+	+	+
VK16 (DXS293)	-	+	+				+	+			+	+	+	+
VK21A	-	+	+					+					+	+
VK18 (DXS295)	-	+	+				+	+					+	+
10S	-	+	+				+	+					+/-	+
VK37 (DXS302)	-	+					+	+			+	+		
U6.2	-	+	+		+		+	+			+	+		+
1A1	-	+	+		+		+	+			+	+		+
VK25 (DXS299)	-	+			+									+
F8C					+		+	+						+
G6PD					+			+						+
MN12							+							+
St14-1	-	+	+		+		+	+						+
VK9 (DXS289)	-	+	+		+		+	+						+

Legend to Table 3.

DNA probes were used to determine whether the corresponding loci were present (+) or absent (-) in the specified cell lines. This composite Table lists results using both the established probes and the new VK probes. The established probes and cloned genes are listed in order down the X chromosome. The VK probes (and CDR-9) are listed in the order defined by the locations of adjacent breakpoints on the X chromosome. This table includes both data published on individual cell lines (references in Table 1) and previously unpublished data. The cell line "B17" refers to cell line 908K1B17. The breakpoint in CY34 is located at IDS (shown as +/-).



LEGENDS TO THE FIGURES

Figure 1

The locations of cell line breakpoints, established DNA probes, cloned genes, and the VK probes at Xq26-28. The cell line translocation and deletion breakpoints are indicated to the left of the X chromosome. For each cell line the arrow indicates on which side of the breakpoint lies the X chromosome material retained in the cell line. In the case of cell lines with interstitial deletions the proximal (p) and distal (d) breakpoints are shown. The established probes and cloned genes are shown to the right of the chromosome in order down the X chromosome. The locations of the VK probes are shown on the far right. The locations of all the probes and the cell line breakpoints in relation to chromosome banding is approximate. The Fragile X is represented as a hatched region at distal Xq27; the locations of the APC-5 and 04-1p breakpoints and of the probe VK16 relative to the Fragile X are unknown.

CELL LINES

Xq

DNA PROBES

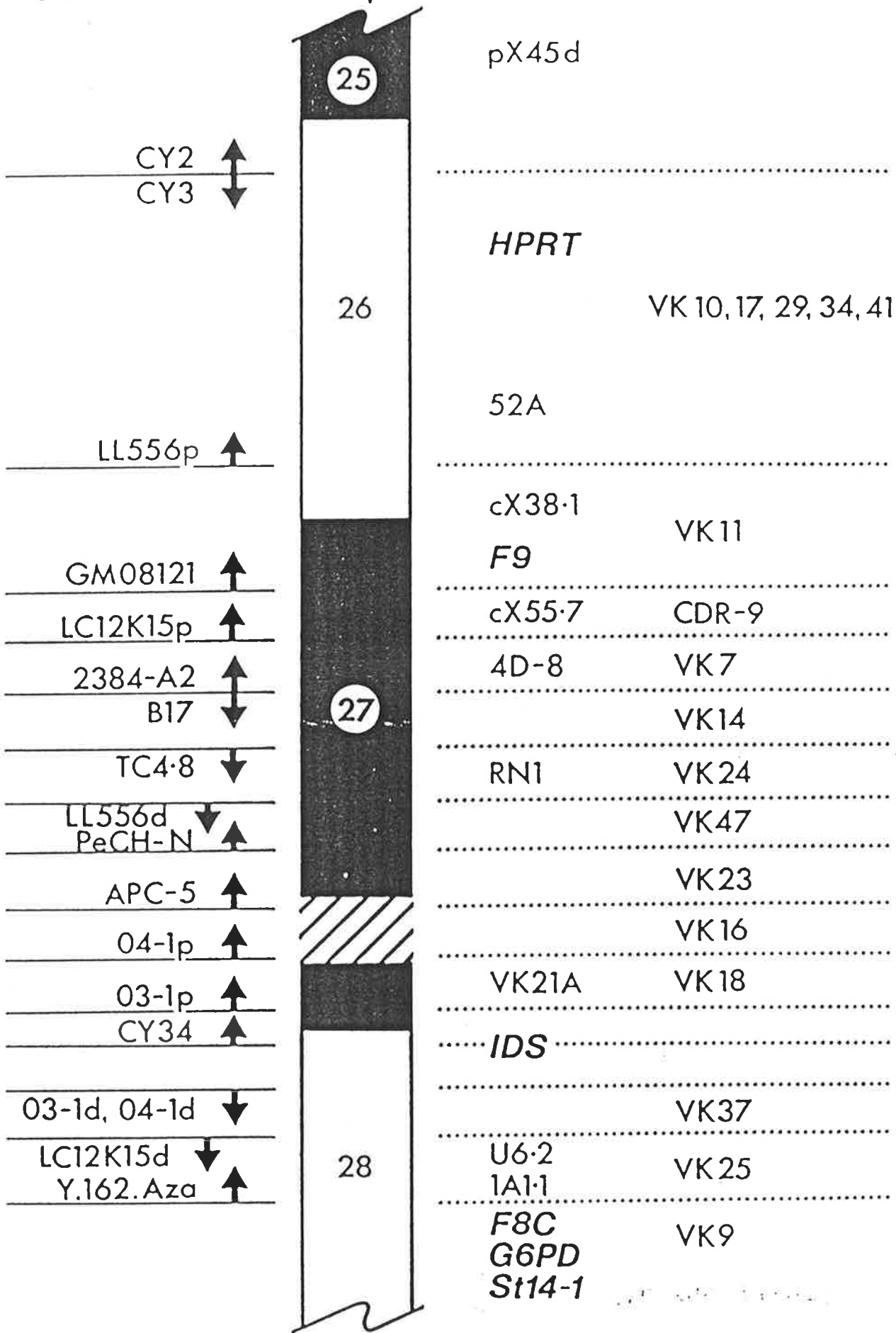


Figure 2

Radio-labelled VK18 was hybridised to HindIII-digested DNA from 6 sources. VK18 hybridized to DNA from normal females and from cell lines CY3 and CY34; VK18 did not hybridize to DNA from cell lines CY2 and PeCH-N. This indicated that VK18 detected a locus between the breakpoints in the cell lines PeCH-N and CY34.

1 kb ▼

2 kb ▼

5 kb ▼



46, XX

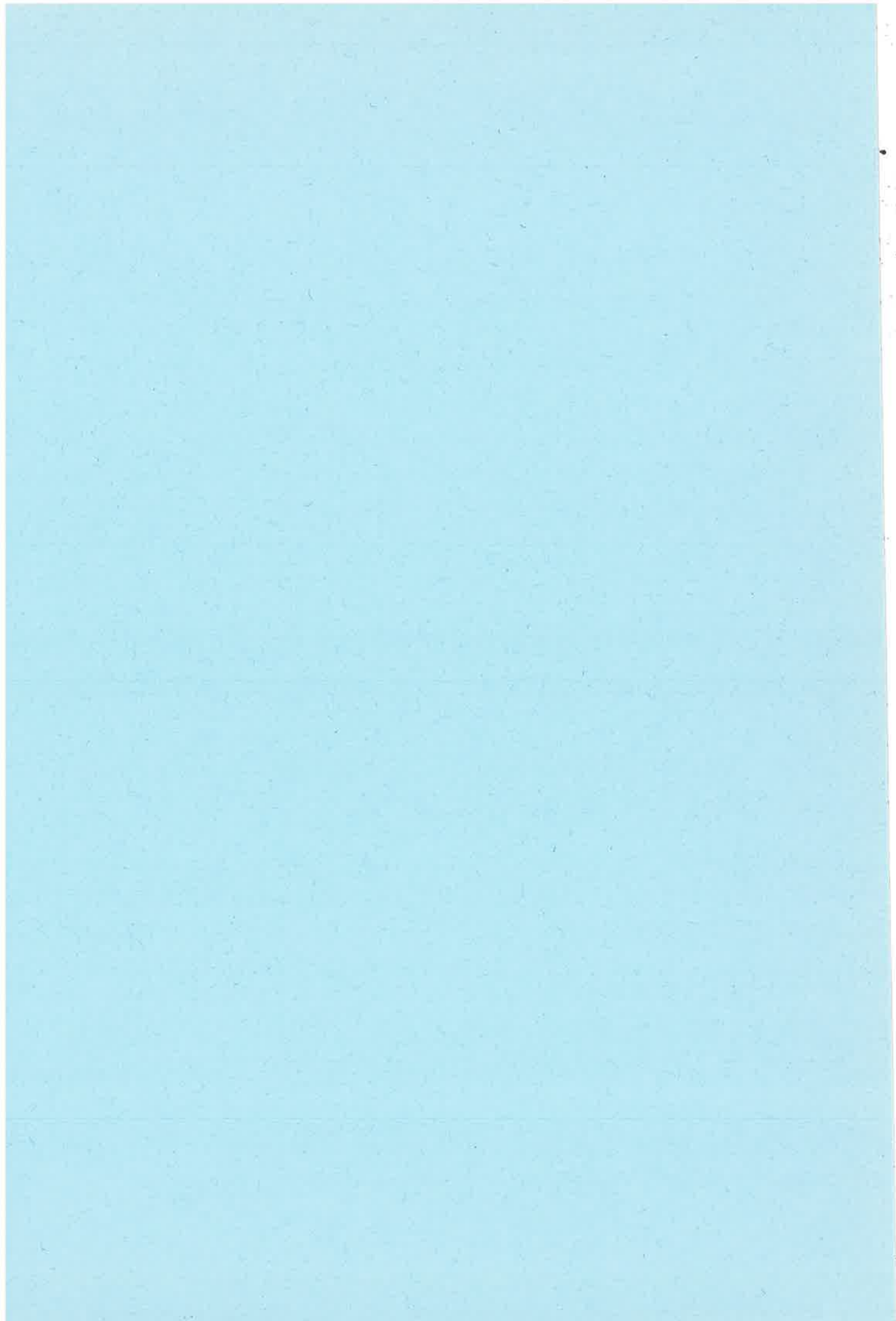
CY2

CY3

PeCH-N

CY34

46, XX



Yu, S., Suthers, G.K. and Mulley, J.C. (1990). A Bc/I RFLP for DXS296 (VK21) near the fragile X. *Nucleic Acids Research*, 18(3), 690.

NOTE: This publication is included in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

<http://dx.doi.org/10.1093/nar/18.3.690-a>



Suthers, G.K., Hyland, V.J., Baker, E., Fernandez, K.E.W., Callen, D.F. and Sutherland, G.R. (1988). TaqI RFLP Identified by probe VK17A (DXS294) at Xq26. *Nucleic Acids Research*, 16(23), 11389.

NOTE: This publication is included in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

<http://dx.doi.org/10.1093/nar/16.23.11389>





cephmap.art  
June 26, 1990

Submitted to Genomics in  
early June, 1990

Genetic mapping of new RFLPs at Xq27-q28.

G.K. Suthers<sup>1</sup>, I. Oberle<sup>2</sup>, J. Nancarrow<sup>1</sup>, J.C. Mulley<sup>1</sup>, V.J. Hyland<sup>1</sup>, P.J. Wilson<sup>3</sup>, J. McCure<sup>1</sup>, C.P. Morris<sup>3</sup>, J.J. Hopwood<sup>3</sup>, J.L. Mandel<sup>2</sup>, G.R. Sutherland<sup>1</sup>.

---

<sup>1</sup>Department of Cytogenetics & Molecular Genetics, and <sup>3</sup>Lysosomal Diseases Research Unit, Department of Chemical Pathology, Adelaide Children's Hospital, North Adelaide, South Australia;  
<sup>2</sup>L.G.M.E./C.N.R.S., INSERM U184, Faculte de Medecine, 11 Rue Humann, 67085 Strasbourg Cedex, France.

---

Corresponding author: Dr GR Sutherland, Department of Cytogenetics and Molecular Genetics, Adelaide Children's Hospital, North Adelaide, SA 5006 Australia.

tel (08)-267 7333

fax (08)-267 7342

Running heading: Linkage map at Xq27-q28

ABSTRACT

The development of the human gene map in the region of the fragile X mutation (FRAXA) at Xq27 has been hampered by a lack of closely linked polymorphic loci. The polymorphic loci, DXS369 (detected by probe RN1), DXS296 (VK21A, VK21C), and DXS304 (U6.2), have recently been mapped to within 5 centimorgans of FRAXA. The order of loci near FRAXA has been defined on the basis of physical mapping studies as cen-F9-DXS105-DXS98-DXS369-DXS297-FRAXA-DXS296-IDS-DXS304-DXS52-qter. The probe VK23B detected HindIII and XmnI restriction fragment length polymorphisms (RFLPs) at DXS297 with heterozygote frequencies of 0.34 and 0.49 respectively. An IDS cDNA probe, pc2S15, detected StuI and TaqI RFLPs at IDS with heterozygote frequencies of 0.50 and 0.08 respectively. Multipoint linkage analysis of these polymorphic loci in normal pedigrees indicated that the locus order was F9-(DXS105,DXS98)-(DXS369,DXS297)-(DXS296,IDS)-DXS304-DXS52. The recombination fractions between adjacent loci were F9-(.058)-DXS105-(.039)-DXS98-(.123)-DXS369-(.00)-DXS297-(.057)-DXS296-(.00)-IDS-(.012)-DXS304-(.120)-DXS52. This genetic map will provide the basis for further linkage studies of both the fragile X syndrome and other disorders mapped to Xq27-q28.

## INTRODUCTION

The seminal paper by Botstein et al., (1980) indicated the power of linkage analysis to localize disease genes and polymorphic loci on the human gene map. This approach has been very successful, as indicated by the increasing volume of linkage data presented at the International Human Gene Mapping Workshops (Human Gene Mapping 10, 1989). However this success has not been uniform throughout the human genome. The fragile X syndrome is the most common cause of familial mental retardation (Sutherland and Hecht, 1985) and consumes significant health care resources in Western societies (Turner et al., 1986). The locus responsible for this syndrome (FRAXA) is located at or very near the rare fragile site at Xq27.3 (Sutherland and Hecht, 1985), but the development of a precise genetic map around FRAXA has been hampered by the lack of closely linked polymorphic loci.

Three polymorphic loci near FRAXA, DXS369 (detected by the probe RN1), DXS296 (VK21A, VK21C), and DXS304 (U6.2), were recently reported (Oostra et al., 1990; Suthers et al., 1989; Dahl et al., 1989). Linkage studies in fragile X pedigrees indicated that these loci lay within 5 centimorgans (cM) of FRAXA (Oostra et al., 1990; Suthers et al., 1989; Vincent et al., 1989). However a combined linkage analysis of these three loci has not been presented.

We have described a panel of cell lines with precisely delineated X chromosome breakpoints which was used to physically map a series of loci in relation to DXS369, DXS296, DXS304, and FRAXA (Suthers *et al.*, 1990). The order of loci near FRAXA was cen - F9 - DXS105 - (DXS98, DXS288) - DXS292 - (DXS369, DXS298) - DXS308 - DXS297 - (DXS293, FRAXA) - DXS296 - DXS295 - IDS - DXS302 - (DXS304, DXS299) - DXS52 - qter.

We have endeavoured to define restriction fragment length polymorphisms (RFLPs) at other loci near FRAXA, and in the present paper describe RFLPs at DXS297 and IDS. A multipoint linkage analysis of both the established RFLPs and these new polymorphisms in a series of normal pedigrees is presented.

## MATERIAL and METHODS

The preparation of genomic DNA, digestion with restriction endonucleases, Southern blotting, radiolabelling of DNA probes, and hybridization conditions were as reported previously (Suthers *et al.*, 1990).

Four new DNA probes were used to search for RFLPs near FRAXA. VK23B was a 4.4 kilobase (kb) EcoRI fragment of VK23 (DXS297) (Hyland *et al.*, 1989) that was subcloned in pUC19. VK16B3 was a 0.9 kb HindIII single-copy DNA fragment of VK16 (DXS293) (Hyland *et al.*, 1989) that was subcloned in pSP64. VK18A was a 1.2 kb SalI-HindIII single-copy DNA fragment of VK18 (DXS295) (Hyland *et al.*, 1989) that was subcloned in pBR328. The method used for subcloning has been described (Greene and Guarente, 1987). The probe pc2S15 was a 1.5 kb IDS cDNA clone (Wilson P.J. *et al.*, manuscript in preparation).

Panels of human genomic DNA containing 13 or more X chromosomes were digested with a variety of restriction endonucleases (New England Biolabs) under the conditions specified by the manufacturer. After Southern blotting the panels were probed with the radiolabelled single-copy probes. With at least 13 X chromosomes being surveyed, the probability of detecting a two-allele RFLP with a rare allele frequency of at least 0.15 was greater than 80% (Aldridge *et al.*, 1984). The linkage

disequilibrium constants between the RFLPs at a locus were estimated in the manner described by Thompson et al. (1988).

Forty normal pedigrees from the Centre d'Etude du Polymorphisme Humain (CEPH) (White and Lalouel, 1988; Dausset et al., 1990) were genotyped for the RFLPs listed in Table 1. Linkage studies utilizing some of the data have been published previously (Oberle et al., 1986, 1987; Arveiler et al., 1988). The data were checked by hand. No double recombinants in the interval F9-DXS52 were observed. All these genotype data have been communicated to the CEPH database and will be included in the CEPH consortium linkage map of the X chromosome.

Linkage analyses were performed using the LINKAGE programs modified for use with the CEPH three-generation pedigrees (Lathrop et al., 1984, 1985, 1986; Lathrop and Lalouel, 1988). As the LINKAGE programs limit analyses incorporating interference to just three loci, interference was not considered in this analysis.

The A-test (Ott 1985; pp. 105-109) (as implemented in the computer program HOMOG2) was used to test for homogeneity in the two-point recombination fractions between F9-IDS and IDS-DXS52. The program determines whether there is sufficient evidence to indicate that the pedigrees can be divided into two groups with different recombination fractions between the loci.



## RESULTS

### Detection of RFLPs

VK23B detected two-allele RFLPs at DXS297 in genomic DNA digested with HindIII (alleles A1,A2) or XmnI (alleles B1,B2) (Fig. 1). The heterozygote frequencies for the two RFLPs were 0.34 and 0.49 respectively. The two RFLPs were in linkage disequilibrium. Among 19 DNA samples from unrelated Caucasian males the haplotype frequencies were A1B1 (0.42), A1B2 (0.21), A2B1(0.37), and A2B2 (0.0). The standardized linkage disequilibrium constant was -2.01 ( $\chi^2_1=4.04$ ;  $p<0.05$ ). VK23B did not detect RFLPs in genomic DNA digested with any of the following enzymes: AvaI, AvaII, BamHI, BclI, BglII, EcoRI, HincII, MspI, PstI, PvuII, Sau3A, StuI, and TaqI.

VK16B3 did not detect an RFLP at DXS293 with the following enzymes: AluI, AvaI, AvaII, BamHI, BanI, BclI, BglI, BglII, BstNI, BstXI, DdeI, DraI, EcoRI, EcoRV, HaeIII, HincII, HindIII, HinfI, MboII, MspI, NsiI, PstI, PvuII, RsaI, SacI, Sau3A, StuI, TaqI, XbaI, and XmnI.

VK18A did not detect an RFLP at DXS295 with the following enzymes: AvaI, AvaII, BamHI, BanI, BanII, BclI, BglI, BglII, BstNI, BstXI, DraI, Eco0109, EcoRI, EcoRV, HaeIII, HincII,

HindIII, HinfI, MboII, PstI, PvuII, RsaI, SacI, StuI, TaqI, XbaI, and XmnI.

The probe pc2S15 detected two-allele RFLPs at IDS in DNA digested with StuI (alleles A1,A2) or TaqI (alleles B1,B2) (Fig. 2). The heterozygote frequencies for the two RFLPs were 0.50 and 0.08 respectively. There was no evidence that the two RFLPs were in linkage disequilibrium. Among 27 X chromosomes from unrelated CEPH males and females the haplotype frequencies were A1B1 (0.60), A1B2 (0.04), A2B1 (0.36), and A2B2 (0.0). The standardized linkage disequilibrium constant was -0.959 ( $\chi^2_1=0.92$ ;  $p>0.1$ ). RFLPs were not detected with the following enzymes: BamHI, BclI, BglII, EcoRI, HincII, HindIII, MspI, PstI, PvuII, and Sau3A.

### Linkage Analysis

Two-point LOD scores and recombination fractions for all pairwise combinations of loci are summarized in Table 2. The peak LOD scores for the recombination fractions between loci that were adjacent on the basis of physical mapping ranged from 1.19 (DXS297-DXS296) to 18.20 (IDS-DXS304). No recombinants were observed between DXS105-DXS98, DXS98-DXS369, DXS369-DXS297, and DXS296-IDS.

For multipoint linkage analysis, the order of loci determined by physical mapping (Suthers et al., 1990) was assumed, and the relative likelihood of adjacent pairs being inverted was calculated (Fig. 3). In general it was not possible to order adjacent pairs of loci where there was no recombination on two-point linkage analysis. The one exception was DXS98-DXS369 where DXS98 was placed proximal to DXS369. Although no recombination was observed between these two loci, DXS369 demonstrated much closer linkage to distal markers than did DXS98 (Table 2). The odds in favor of placing DXS304 distal to IDS were only 22:1. There was only one recombination event between IDS and DXS304, and this was in a pedigree uninformative at DXS296 and DXS52.

The order of loci suggested by this analysis was

F9-(DXS105,DXS98)-(DXS369,DXS297)-(DXS296,IDS)-DXS304-DXS52. As an aid for ordering other loci in this region, individuals who were recombinant between loci in the interval DXS369-DXS304 are listed in Table 3. Assuming that DXS98 was distal to DXS105 (Keats et al., 1989; Suthers et al., 1990), the recombination fractions between adjacent loci were derived by multipoint linkage analysis and are shown in Fig. 3.

The A-test was used to determine whether the pair-wise LOD scores for the recombination fractions between F9-IDS and IDS-DXS52 were heterogeneous. The locus IDS was chosen because more CEPH pedigrees were informative at this locus than at other nearby

loci. For F9-IDS there were 10 informative pedigrees with 45 phase-known and 27 phase-unknown meioses. There was insufficient evidence to indicate linkage heterogeneity ( $X^2_2=0.078$ ;  $p>0.2$ ). For IDS-DXS52 there were 17 informative pedigrees with 82 phase-known and 39 phase-unknown meioses. Again there was insufficient evidence to indicate linkage heterogeneity ( $X^2_2=2.840$ ;  $p>0.1$ ).

## DISCUSSION

In this study we have described two RFLPs at each of two loci near FRAXA. These RFLPs are informative in a large proportion of women, and they will be of great value in genetic counseling of families with the fragile X syndrome or Hunter syndrome (IDS deficiency). The RFLPs detected by VK23B at DXS297 are in linkage disequilibrium. Although there was no evidence of linkage disequilibrium between the RFLPs at IDS, the power of this assessment to exclude linkage disequilibrium was low (Thompson et al. 1988).

The RFLPs defined at DXS297 and IDS flank FRAXA, and are closer to FRAXA than DXS369 or DXS304 (Fig. 3; Suthers *et al.*, 1990). A study of the linkage relationships of FRAXA and these RFLPs is underway. In the case of Hunter syndrome, a proportion of affected males have deletions at IDS which are detected by pc2S15 (Wilson *et al.*, 1990), and the identification of carriers may be relatively simple. In families with an established mutation where a deletion is not evident, it is now possible to provide carrier risk estimates utilizing RFLPs at IDS and at closely linked flanking loci.

We have also defined the linkage relationships of nine loci at Xq27-q28 in this study. The order of loci determined by linkage analysis was consistent with that determined by physical mapping

studies (Suthers et al., 1990). The physical mapping studies had been based on the assumption that the cell lines used did not have complex rearrangements at Xq27-q28. It is known that such rearrangements can occur, and it is reassuring that the results of this linkage study were compatible with the order based on physical mapping studies.

However the limitations of linkage analysis for ordering loci were evident. Despite genotyping 40 three-generation pedigrees at nine loci there were three pairs of loci that could not be ordered on the basis of linkage analysis, DXS105-DXS98, DXS369-DXS297, and DXS296-IDS. There are three reasons for this. First, the X chromosome has less polymorphic variability than the autosomes (Hofker et al., 1986), and the RFLPs tested at DXS98 and DXS297 were informative in only a few pedigrees. Second, recombination between X-linked loci occurs in only one parent, thus reducing the information that can be derived from a study of X-linked rather than autosomal loci. Third, as more loci are identified in a region it becomes increasingly less likely that recombination will be observed between an adjacent pair of loci. The order of loci shown in Fig. 3 is based on both physical mapping studies (Suthers et al., 1990) and the linkage analysis presented in this paper. There is some independent evidence to support this order. Collated linkage data has placed DXS98 distal to DXS105 (Keats et al., 1989), and IDS has been placed distal to DXS296 in pulsed field gel electrophoresis studies (Hyland et

al., 1990). The location of DXS297 distal to DXS369 (Suthers et al., 1990) has not been independently confirmed.

With the development of a large scale restriction map of the region close to FRAXA (Hyland et al., 1990) it is possible to correlate physical and genetic distances. IDS and DXS304 are within 900 kb of each other, and the recombination fraction between them was 0.012. DXS296 lies approximately 800 kb proximal to IDS (Yu S., personal communication) and no recombinants were observed between these loci.

These linkage studies were performed under the assumption that there was no heterogeneity in the linkage relationships of co-dominant loci. This assumption need not be true. Some evidence has been presented that there is linkage heterogeneity between F9 and DXS52 in fragile X pedigrees (Brown et al., 1987; Risch 1988). Brown et al., (1987) also reported definite evidence of linkage heterogeneity between F9 and FRAXA, but this has been disputed (Clayton et al., 1988). There is no evidence of linkage heterogeneity between FRAXA and DXS52 in fragile X pedigrees (Brown et al., 1987; Risch 1988). Linkage heterogeneity could reflect a sampling fluctuation, or be specific for the fragile X mutation, or also be a feature of normal pedigrees. If linkage heterogeneity is a feature of normal pedigrees, it may be necessary to utilise more than 40 CEPH pedigrees to develop a representative linkage map of the human genome.

It is impossible to test for linkage heterogeneity in the interval F9-FRAXA in normal pedigrees. IDS was chosen as an appropriate locus for this test for two reasons. First, a large number of the CEPH pedigrees were informative at IDS. Second, IDS must be close to FRAXA; IDS lies between FRAXA and DXS304 (Suthers et al., 1990) which are separated by a genetic distance of approximately 3 cM (Vincent et al., 1989). No evidence of linkage heterogeneity was found between F9-IDS and IDS-DXS52 in this study of normal pedigrees.

It is difficult to estimate the power of this study to exclude linkage heterogeneity. In the study of Brown et al., (1987) 24% of fragile X pedigrees showed no recombination between F9-FRAXA while the remainder demonstrated a recombination fraction of 0.37. Ott (1986) and Risch (1988) have tabulated the power of heterogeneity tests for a variety of alternative hypotheses, but they do not consider the combination of values noted by Brown et al. (1987). Cavalli-Sforza and King (1986) suggested that 24 phase-known pedigrees each with four children were sufficient to detect linkage heterogeneity at odds of 10:1 when 20% of the pedigrees were tightly linked and the remainder were unlinked. In this study the pedigrees all had more than four children, thus increasing the power of the study (Ott, 1985; p. 54). On the other hand, under the alternative hypothesis that 80% of the



pedigrees were loosely linked rather than unlinked, the power of the heterogeneity test would be reduced (Risch 1988).

An increasing number of genes are being localized to Xq27-q28 (Mandel et al., 1989). The development of this genetic map will be crucial for the precise localization of disease loci in this region. Such a map also may be the basis for estimating approximate confidence intervals for gene location, and for providing reliable carrier risk estimates to consultands (Suthers and Wilson 1990).

**ACKNOWLEDGEMENTS**

C. Kretz, Helen Kozman, Paul Nelson, and Hilary Barton gave advice and assistance. The following colleagues provided DNA probes: B.A. Oostra (RN1); N. Dahl (U6.2). DNA samples and database software were provided by the Centre d'Etude du Polymorphisme Humain. Dr J. Ott provided the LINKAGE and HOMO2 programs. This research was supported by the National Health and Medical Research Council of Australia, The Adelaide Children's Hospital Research Foundation, the Ministere de la Recherche et de la Technologie (88.C.0178), and C.N.A.M.T.S./I.N.S.E.R.M.

REFERENCES

- 1 ALDRIDGE, J., KUNKEL, L., BRUNS, G., TANTRAVAHU, U., LALANDE, M., BREWSTER, T., MOREAU, E., WILSON, M., BROMLEY, W., RODERICK, T., AND LATT, S. (1984). A strategy to reveal high-frequency RFLPs along the human X chromosome. Am. J. Hum. Genet. 36:546-564.
- 2 ARVEILER, B., OBERLE, I., VINCENT, A., HOFKER, M.H., PEARSON, P.L., AND MANDEL, J.L. (1988). Genetic mapping of the Xq27-28 region: New RFLP markers useful for diagnostic applications in fragile-X and hemophilia-B families. Am. J. Hum. Genet. 42:380-389.
- 3 BOTSTEIN, D., WHITE, R.L., SKOLNICK, M., AND DAVIS, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32:314-331.
- 4 BROWN, W.T., JENKINS, E.C., GROSS, A.C., CHAN, C.B., KRAWCZUN, M.S., DUNCAN, C.J., SKLOWER, S.L., AND FISCH, G.S. (1987). Further evidence for genetic heterogeneity in the fragile X syndrome. Hum. Genet. 75:311-321.

- 5 CAVALLI-SFORZA, L.L., AND KING, M.C. (1986). Detecting linkage for genetically heterogeneous diseases and detecting heterogeneity with linkage data. Am. J. Hum. Genet. 38:599-616.
- 6 CLAYTON, J.F., GOSDEN, C.M., HASTIE, N.D., AND EVANS, H.J. (1988). Linkage heterogeneity and the fragile X. Hum. Genet. 78:338-342.
- 7 DAHL, N., HAMMARSTROM-HEEROMA, K., VAN OMMEN, G.B. AND PETTERSSON, U. (1989). A polymorphic locus at Xq27-28 detected by the probe U6.2 (DXS304). Nucl. Acid. Res. 17:2884-2884.
- 8 DAUSSET, J., CANN, H., COHEN, D., LATHROP, M., LALOUEL, G.M., AND WHITE, R. (1990). Centre d'Etude du Polymorphisme Humain (CEPH): Collaborative genetic mapping of the human genome. Genomics 6:575-577.
- 9 GREENE, J.R., AND GUARENTE, L. (1987). Subcloning. In "Guide to Molecular Cloning Techniques" (S.L. Berger and A.R. Kimmel, Eds.), Methods in Enzymology, Vol. 152, pp. 512-522. Academic Press Inc., Orlando, Florida.
- 10 HOFKER, M.H., SKRAASTAD, M.I., BERGEN, A.A.B., WAPENAAR, M.C., BAKKER, E., MILLINGTON-WARD, A., VAN OMMEN, G.J.B.,

AND PEARSON, P.L. (1986). The X chromosome shows less genetic variation at restriction sites than the autosomes. Am. J. Hum. Genet. 39:438-451.

- 11 HUMAN GENE MAPPING 10 (1989): Tenth International Workshop on Human Gene Mapping. Cytogenet. Cell Genet. 51:1-1148.
- 12 HYLAND, V.J., FERNANDEZ, K.E.W., CALLEN, D.F., MACKINNON, R.N., BAKER, E.G., FRIEND, K., AND SUTHERLAND, G.R. (1989). Assignment of anonymous DNA probes to specific intervals of human chromosomes 16 and X. Hum. Genet. 83:61-66.
- 13 HYLAND, V.J., NANCARROW, J., CALLEN, D.F., SUTHERS, G.K., WILSON, P.J., MORRIS, C.P., HOPWOOD, J.J., DAHL, N.P., AND SUTHERLAND, G.R. (1990). Physical linkage of four loci distal to the fragile X locus (submitted to Genomics, manuscript #D085).
- 14 KEATS, B., OTT, J., AND CONNEALLY, M. (1989). Report of the committee on linkage and gene order. Human Gene Mapping 10 (1989): Tenth International Workshop on Human Gene Mapping. Cytogenet. Cell Genet. 51:459-502.
- 15 KIDD, K.K., BOWCOCK, A.M., SCHMIDTKE, J., TRACK, R.K., RICCIUTI, F., HUTCHINGS, G., BALE, A., PEARSON, P., WILLARD, H.F., GELERNTER, J., GIUFFRA, L., AND KUBZDELA, K. (1989).

Report of the DNA committee and catalogs of cloned and mapped genes and DNA polymorphisms. *Human Gene Mapping 10* (1989): Tenth International Workshop on Human Gene Mapping. *Cytogenet. Cell Genet.* 51:622-947.

- 16 LATHROP, G.M., LALOUEL, J.M., JULIER, C., AND OTT, J. (1984). Strategies for multilocus linkage analysis in humans. *Proc. Natl. Acad. Sci. USA* 81:3443-3446.
- 17 LATHROP, G.M., LALOUEL, J.M., JULIER, C., AND OTT, J. (1985). Multilocus linkage analysis in humans: Detection of linkage and estimation of recombination. *Am. J. Hum. Genet.* 37:482-498.
- 18 LATHROP, G.M., LALOUEL, J.M., AND WHITE, R.L. (1986). Construction of human linkage maps: Likelihood calculations for multilocus linkage analysis. *Genet. Epidemiol.* 3:39-52.
- 19 LATHROP, G.M., AND LALOUEL, J.M. (1988). Efficient computations in multilocus linkage analysis. *Am. J. Hum. Genet.* 42:498-505.
- 20 MANDEL, J.L., WILLARD, H.F., NUSSBAUM, R.L., ROMEO, G., PUCK, J.M., AND DAVIES, K.E. (1989). Report of the committee on the genetic constitution of the X chromosome. *Human Gene*

Mapping 10 (1989): Tenth International Workshop on Human Gene Mapping. Cytogenet. Cell Genet. 51:384-437.

- 21 OBERLE, I., HEILIG, R., MOISAN, J.P., KLOEPFER, C., MATTEI, M.G., MATTEI, J.F., BOUE, J., FROSTER-ISKENIUS, U., JACOBS, P.A., LATHROP, G.M., LALOUEL, J.M., AND MANDEL, J.L. (1986). Genetic analysis of the fragile-X mental retardation syndrome with two flanking polymorphic DNA markers. Proc. Natl. Acad. Sci. USA 83:1016-1020.
- 22 OBERLE, I., CAMERINO, G., WROGEMANN, K., ARVEILER, B., HANAUER, A., RAIMONDI, E., AND MANDEL, J.L. (1987). Multipoint genetic mapping of the Xq26-q28 region in families with fragile X mental retardation and in normal families reveals tight linkage of markers in q26-27. Hum. Genet. 77:60-65.
- 23 OBERLE, I., VINCENT, A., ABBADI, N., ROUSSEAU, F., HUPKES, P.E., HORS-CAYLA, M.C., GILGENKRANTZ, S., OOSTRA, B.A., AND MANDEL, J.L. (1990). A new polymorphism and a new chromosome breakpoint establish the physical and genetic mapping of DXS369 in the DXS98-FRAXA interval. Am. J. Med. Genet. (in press).
- 24 OOSTRA, B.A., HUPKES, P.E., PERDON, L.F., VAN BENNEKOM, C.A., BAKKER, E., HALLEY, D.J.J., SCHMIDT, M., DU SART, D.,

- SMITS, A., WIERINGA, B., AND VAN OOST, B.A. (1990). New polymorphic DNA marker close to the fragile site FRAXA. Genomics 6:129-132.
- 25 OTT, J. (1985). "Analysis of Human Genetic Linkage", Johns Hopkins University Press, Baltimore, Maryland.
- 26 OTT, J. (1986). The number of families required to detect or exclude linkage heterogeneity. Am. J. Hum. Genet. 39:159-165.
- 27 RISCH, N. (1988). A new statistical test for linkage heterogeneity. Am. J. Hum. Genet. 42:353-364.
- 28 ROUSSEAU, F., VINCENT, A., OBERLE, I., AND MANDEL, J.L. (1990). New informative polymorphism at the DXS304 locus, a close distal marker for the fragile X locus. Hum. Genet. 84:263-266.
- 29 SUTHERLAND, G.R., AND HECHT, F. (1985). "Fragile Sites on Human Chromosomes", Oxford University Press, New York.
- 30 SUTHERS, G.K., CALLEN, D.F., HYLAND, V.J., KOZMAN, H.M., BAKER, E., EYRE, H., HARPER, P.S., ROBERTS, S.H., HORS-CAYLA, M.C., DAVIES, K.E., BELL, M.V., AND SUTHERLAND, G.R.



(1989). A new DNA marker tightly linked to the fragile X locus (FRAXA). Science 246:1298-1300.

- 31 SUTHERS, G.K., HYLAND, V.J., CALLEN, D.F., OBERLE, I., ROCCHI, M., THOMAS, N.S., MORRIS, C.P., SCHWARTZ, C.E., SCHMIDT, M., ROPERS, H.H., BAKER, E., OOSTRA, B.A., DAHL, N., WILSON, P.J., HOPWOOD, J.J., AND SUTHERLAND, G.R.(1990). Physical mapping of new DNA probes near the fragile X (FRAXA) with a panel of cell lines. Am. J. Hum. Genet. (in press).
- 32 SUTHERS, G.K., AND WILSON, S.R. (1990). Genetic counseling in rare syndromes: A resampling method for determining an approximate confidence interval for gene location with linkage data from a single pedigree. Am. J. Hum. Genet. (in press).
- 33 THOMPSON, E.A., DEEB, S., WALKER, D., AND MOTULSKY, A.G. (1988). The detection of linkage disequilibrium between closely linked markers: RFLPs at the AI-CIII apolipoprotein genes. Am. J. Hum. Genet. 42:113-124.
- 34 TURNER, G., ROBINSON, H., LAING, S., AND PURVIS-SMITH, S. (1986). Preventive screening for the fragile X syndrome. N. Engl. J. Med. 315:607-609.

- 35 VINCENT, A., DAHL, N., OBERLE, I., HANAUER, A., MANDEL, J.L., MALMGREN, H., AND PETTERSSON, U. (1989). The polymorphic marker DXS304 is within 5 centimorgans of the fragile X locus. Genomics 5:797-801.
- 36 WHITE, R., AND LALOUEL, J.M. (1988). Sets of linked genetic markers for human chromosomes. Annu. Rev. Genet. 22:259-79.
- 37 WILSON, P.J., SUTHERS, G.K., CALLEN, D.F., BAKER, E., NELSON, P.V., COOPER, A., WRAITH, J.E., SUTHERLAND, G.R., MORRIS, C.P., AND HOPWOOD, J.J. (1990). Frequent deletions at Xq28 indicate genetic heterogeneity in Hunter syndrome (submitted).

TABLE 1  
Restriction fragment length polymorphisms used in the linkage study.<sup>a</sup>

Locus	Probe	Enzyme	Alleles	PIC <sup>b</sup>	Location	Reference
<u>F9</u>	pVIII	<u>TaqI</u>	1.8,5/1.3,5 kb	0.33	Xq26.3-q27.1	15
<u>DXS105</u>	cX55.7 55E	<u>TaqI</u> <u>MspI</u>	3.2/4.5 kb 16/10 kb	0.11 0.48	Xq27.1-q27.2	15
<u>DXS98</u>	4D-8	<u>MspI</u>	25/7.8 kb	0.30	Xq27.2	15
<u>DXS369</u>	RN1	<u>XmnI</u> <u>TaqI</u>	1.25/1.1 kb 4.5/4.3 kb	0.48 0.24	Xq27.2-q27.3	24 23
<u>DXS297</u>	VK23B	<u>HindIII</u>	10.5/9.5 kb	0.34	Xq27	this paper
<u>DXS296</u>	VK21A VK21C	<u>TaqI</u> <u>MspI</u>	10.9/9.9 kb 12.7/9.9 kb	0.23 0.31	Xq27.3-q28	30
<u>IDS</u>	pc2S15	<u>StuI</u>	17.8/15,2.8 kb	0.50	Xq28	this paper
<u>DXS304</u>	U6.2 U6.2-20E	<u>TaqI</u> <u>MspI</u> <u>BanI</u>	7.0/3.3 kb 4.5/2.2 kb 9.6/5.8 kb	0.36 0.36 0.49	Xq28	7 28
<u>DXS52</u>	St14-1	<u>TaqI</u>	many	0.80	Xq28	15

a Other RFLPs have been documented at many of the loci. The CEPH pedigrees were not typed for the XmnI RFLP at DXS297 or for the TaqI RFLP at IDS.

b Polymorphism information content; for X-linked loci this is the frequency of females heterozygous at the locus.

TABLE 2

Summary of pair-wise recombination fractions

	<u>F9</u>	<u>DXS105</u>	<u>DXS98</u>	<u>DXS369</u>	<u>DXS297</u>	<u>DXS296</u>	<u>IDS</u>	<u>DXS304</u>	<u>DXS52</u>
<u>F9</u>		.043 14.34	.078 6.38	.221 2.54	.251 1.36	.190 1.28	.193 4.88	.241 2.62	.327 1.46
<u>DXS105</u>	.026 9		0 6.62	.170 6.79	.179 1.38	.142 2.28	.209 4.77	.170 5.50	.279 3.55
<u>DXS98</u>	.044 6	NC 3		0 2.71	.072 2.65	.152 1.19	.125 4.59	.069 5.27	.237 1.81
<u>DXS369</u>	.060 6	.044 9	NC 1		0 4.52	.063 6.08	.054 15.19	.062 12.74	.124 10.83
<u>DXS297</u>	.089 3	.094 2	.072 2	NC 2		.155 1.19	.077 6.85	.042 5.12	.175 2.36
<u>DXS296</u>	.087 4	.078 3	.103 2	.045 5	.105 2		0 15.95	0 7.53	.154 5.06
<u>IDS</u>	.049 10	.049 10	.054 6	.028 10	.045 5	NC 8		.014 18.20	.136 13.93
<u>DXS304</u>	.062 7	.050 8	.049 4	.031 8	.043 3	NC 4	.017 9		.127 11.25
<u>DXS52</u>	.056 12	.048 13	.076 6	.036 11	.074 5	.051 9	.032 17	.036 12	

## Footnote to Table 2

Two-point LOD scores were calculated for each pair of 1 figures above the diagonal are the best estimate recombination fraction (upper figure) and the peak LO (lower figure). The figures below the diagonal are the error of the recombination fraction (upper figure) and th of pedigrees informative for each pair of loci (lower The standard error of the recombination fraction is the root of the variance estimated by the LINKAGE programs. calculated).

TABLE 3

Offspring in the CEPH pedigrees with phase-known recom  
 between loci in the interval DXS369-DXS304.

Closest informative loci	Pedigree number (individual)
<u>DXS369:DXS297</u>	none
: <u>DXS296</u>	13292 (8); 1362 (11).
: <u>IDS</u>	1416 (8); 1421 (12).
: <u>DXS304</u>	884 (14).
 <u>DXS297:DXS296</u>	 1344 (5, 11) <sup>a</sup> .
: <u>IDS</u>	1408 (7).
: <u>DXS304</u>	none
 <u>DXS296:IDS</u>	 none
: <u>DXS304</u>	none
 <u>IDS</u> : <u>DXS304</u>	 1416 (6).

a phase inferred.

## FIGURE LEGENDS

### Figure 1.

a). Mendelian inheritance of the HindIII RFLP detected at DXS297. This is a composite figure made from autoradiographs. DNA size markers indicate the DNA fragment in the two parental lanes on the left of the figure; alleles are indicated in the lanes of the offspring by a the right of the figure. The polymorphic fragments were (A1 allele) and 9.5 kb (A2 allele) long. A faint constant fragment was also detected. Some lane background was not using VK23B, but this did not obscure the polymorphic bands. Among 117 X chromosomes from unrelated CEPH males and females the A1 allele frequency was 0.78.

b). Mendelian inheritance of the XmnI RFLP detected by DXS297. DNA size markers are indicated on the left of the figure and the alleles are indicated on the right. The polymorphic fragments were 10.3 kb (B1 allele) and 6.6 kb (B2 allele) long. No other bands were detected. Among 42 X chromosomes from Caucasian males and females the B1 allele frequency was 0.

**Figure 2.**

a). Mendelian inheritance of the StuI RFLP detected by p IDS. This is a composite figure made from two autoradi. The DNA size markers indicate the DNA fragment sizes in parental lanes on the left of the figure; the two alleles indicated in the lanes of the offspring by arrows to the the figure. The polymorphic fragments were 17.8 kb (A1 and 15.0 and 2.8 kb (A2 allele) long. Constant fragments following sizes were also detected: 2.6, 3.2, 3.4, 3.6, kb. Among 104 X chromosomes from unrelated CEPH males and the A1 allele frequency was 0.55.

b). Mendelian inheritance of the TaqI RFLP detected by p IDS. DNA size markers are indicated on the left of the the alleles are indicated on the right. The polymorphic fragments were 5.1 kb (B1 allele) and 3.8 kb (B2 allele) long. fragments of the following sizes were also detected: 1.1.7, 2.0, 2.8, and 3.3 kb. The polymorphic bands were fainter than the constant bands. Among 27 X chromosomes unrelated CEPH males and females the B1 allele frequency was 0.96.



Figure 3.

Multipoint linkage analysis of nine loci at Xq27-q28. (Up order of loci was derived from physical mapping studies et al., 1990). The odds against inverting adjacent loci calculated using the linkage analysis programs LINK genotypes from the CEPH pedigrees. (Lower) Recombination fractions between the adjacent loci were estimated multipoint analysis of the most likely order.





July 13, 1990

*Submitted in June 1990*

**Genetic mapping of new DNA probes at Xq27  
defines a strategy for DNA studies  
in the Fragile X syndrome.**

G.K. SUTHERS, M.B.B.S., J.C.MULLEY, Ph.D., M.A. VOELCKEL,  
Ph.D., N. DAHL, M.D., Ph.D., M.L. VAISANEN, M.Sc., P.  
STEINBACH, Ph.D., I.A. GLASS, M.B.Ch.B., C.E. SCHWARTZ,  
Ph.D., B.A. van OOST, Ph.D., S.N. THIBODEAU, Ph.D., N.E.  
HAITES, M.B.Ch.B., Ph.D., B.A. OOSTRA, Ph.D., R. GINE, M.D.,  
M. CARBALLO, Ph.D., C.P. MORRIS, Ph.D., J.J. HOPWOOD, Ph.D.,  
G.R. SUTHERLAND, Ph.D., D.Sc.

---

From the Department of Cytogenetics & Molecular Genetics  
(G.K.S, J.C.M., G.R.S.), and the Lysosomal Diseases Research  
Unit, Department of Chemical Pathology (C.P.M., J.J.H.),  
Adelaide Children's Hospital, Adelaide; Centre de Diagnostic  
Prenatal, Hopital d'Enfants de la Timone, Marseille (M.A.V.);  
Department of Medical Genetics, University of Uppsala,  
Uppsala (N.D.); Department of Clinical Genetics, Oulu

University Central Hospital, Oulu (M.L.V.); Department of Clinical Genetics, University of Ulm, Ulm (P.S.); University Department of Medical Genetics, University of Glasgow, Glasgow (I.A.G.); Greenwood Genetics Center, Greenwood (C.E.S.); Department of Human Genetics, University Hospital, Nijmegen (B.A. van O.); Department of Laboratory Medicine, Mayo Clinic, Rochester (S.N.T.); Aberdeen Medical Genetics Group, Department of Molecular and Cell Biology, University of Aberdeen, Aberdeen (N.E.H.); Department of Cell Biology and Clinical Genetics, Erasmus University, Rotterdam (B.A.O.); Institute of Medical Genetics, University of Zurich, Zurich (R.G.); Molecular Genetics, Consejo Superior de Investigaciones Cientificas, Barcelona (M.C.).

Address reprint requests to Dr G.R. Sutherland, Department of Cytogenetics & Molecular Genetics, Adelaide Children's Hospital, North Adelaide, SA 5006 Australia.

Supported by the National Health and Medical Research Council of Australia; The Adelaide Children's Hospital Research Foundation; CNAM and Association Francaise; The Swedish Medical Research Council; Deutsche Forschungsgemeinschaft CDFG; Grampian Health Board and Scottish Hospitals Endowments Research Trust.

---

Running heading: DNA studies in the Fragile X syndrome

ABSTRACT

The fragile X syndrome is the most common cause of familial mental retardation, and is characterised by a fragile site at the end of the long arm of the X chromosome. The unusual genetics and cytogenetics of this X-linked condition make genetic counselling difficult. DNA studies were of limited value in genetic counselling because the nearest polymorphic DNA loci had recombination fractions of 12% or more with the fragile X mutation, FRAXA.

Five polymorphic loci have recently been described in this region of the X chromosome. The positions of these loci in relation to FRAXA were defined in a genetic linkage study of 112 affected families. The five loci - DXS369, DXS297, DXS296, IDS, and DXS304 - had recombination fractions of 4% or less with FRAXA. The closest locus, DXS296, was distal to FRAXA and had a recombination fraction of 2%.

The polymorphisms at these loci can be detected in DNA enzymatically digested with a limited number of restriction endonucleases. A strategy for DNA studies which is based on three restriction endonucleases and five probes will detect one or more of these polymorphisms in 95% of women. This strategy greatly increases the utility of DNA studies in providing genetic advice to families with the fragile X

syndrome.

Key Words: Fragile X syndrome, genetic marker, linkage (genetics), chromosome mapping, genetic counseling.



## INTRODUCTION

Mutations of genes on the human X chromosome are a common cause of mental retardation. The incidence of X-linked mental retardation is approximately 1 in 600 male births<sup>1</sup>. The fragile X syndrome is diagnosed in approximately a quarter of these boys, and it is the most common cause of familial mental retardation<sup>1</sup>. The index case within a family is diagnosed on the basis of mental retardation, subtle facial features, and a specific cytogenetic abnormality - a fragile site - on the X chromosome in band Xq27.3<sup>1,2</sup>. Affected boys do not have any other specific clinical or pathological abnormalities. The fragile X syndrome is unique among X-linked disorders in that males may carry the mutation while not expressing the fragile site nor being mentally retarded<sup>3</sup>. Conversely, an unusually large proportion of women who are carriers may be mentally retarded or express the characteristic fragile site<sup>3,4</sup>.

Once a child has been diagnosed as having the fragile X syndrome, other members of the family frequently seek genetic counselling, either to determine their risk of being a carrier or to request prenatal diagnosis<sup>5</sup>. The unusual pattern of inheritance of the syndrome and the lack of an unequivocal marker of those carrying the mutation have made it difficult to provide reliable risk estimates. The fragile

X mutation, FRAXA, has not been isolated, but it is possible to determine an individual's risk of being a carrier by observing the inheritance of DNA polymorphisms located near FRAXA<sup>6</sup>.

In general, the value of a DNA polymorphism for providing genetic counseling in a disorder is determined by two factors<sup>7</sup>. The first is the genetic distance between the polymorphism and the disease gene. FRAXA has been localized to Xq27, at or very close to the fragile site. The polymorphic loci generally used in studies of fragile X families have been F9, DXS105, DXS98, and DXS52<sup>6</sup>. The genetic distances between these loci and FRAXA may be expressed in terms of recombination fractions, and each of them has a recombination fraction of more than 12% with FRAXA<sup>8</sup>. In other words, during meiosis the probability of recombination between each polymorphic locus and FRAXA is more than 12%. Consequently, identification of a fragile X carrier on the basis of the inheritance of one of these polymorphisms would be incorrect in approximately 12% of cases.

In view of the large recombination fractions between the nearest polymorphic loci and FRAXA, an estimate of carrier risk was ideally based on the inheritance of any two polymorphisms which flank FRAXA. In this situation the risk of incorrectly identifying a carrier may be as low the

product of the two recombination fractions (i.e. 1.4%). However, the advantage of using flanking polymorphisms is lost if recombination has occurred in the interval between the two polymorphic loci, making it impossible to determine which of the two polymorphisms was inherited with FRAXA. The probability that such a recombination will occur between two loci is approximately equal to the sum of the recombination fractions (i.e. at least 24%).

The second consideration in choosing a DNA polymorphism is whether the polymorphism is informative in the family being investigated. For X-linked disorders, this is expressed as the probability that a woman is heterozygous at the polymorphic locus. The ideal polymorphic locus for investigating families with the fragile X syndrome would be very close to FRAXA and have a high probability of being informative in the women of the family. In practice it is usually necessary to examine the family at a number of polymorphic loci, and to seek some compromise between testing highly polymorphic loci far from FRAXA and testing closer but relatively uninformative loci. Strategies for DNA studies in fragile X families have been proposed<sup>6,9</sup> which indicate the order in which the polymorphisms should be evaluated to provide the most efficient diagnostic service for the genetic counselor.

Recently, three polymorphic loci, DXS369<sup>10</sup>, DXS296<sup>11</sup>, and DXS304<sup>12</sup>, were reported to have recombination fractions of less than 5% with FRAXA. The locus DXS296 showed no recombination with FRAXA in the first few families studied, and is the closest locus to FRAXA. A combined linkage study of these three loci in fragile X families has not been reported. Recently two more polymorphic loci, DXS297 and IDS, were mapped to Xq27-q28 in normal families (G.K. Suthers et al., submitted); the role of these loci in DNA studies of fragile X families has not been defined.

A collaborative genetic linkage study of FRAXA and these five polymorphic loci was performed using data from 112 families with the fragile X syndrome. The five loci all had recombination fractions of 4% or less with FRAXA. The closest locus was DXS296 which had a recombination fraction of 2% with FRAXA. Over 98% of women were heterozygous at one or more of these loci. On the basis of these results, an efficient strategy for DNA studies in families with the fragile X syndrome is presented.

## **METHODS**

### **Selection of pedigrees**

Data from a total of 153 families were obtained from 13 centers around the world. As affected individuals with no affected relatives could represent new mutations<sup>13</sup>, families were included in the analysis only if at least one family member was mentally retarded and expressed the fragile site at Xq27, and provided at least one further family member expressed the fragile site. Expression of the fragile site was assessed by culturing peripheral lymphocytes under specific conditions<sup>14</sup>. Expression of the fragile site in 1% or more of lymphocytes was regarded as positive.

There was a general selection bias in favor of families having women heterozygous at loci near FRAXA, and a specific bias in favor of families with women heterozygous at DXS296 as this polymorphic locus was the closest to FRAXA. Details regarding the pedigrees are available on request.

### **DNA studies**

Various DNA probes were used to identify restriction fragment length polymorphisms<sup>7</sup> at nine loci near FRAXA. The approximate positions of the loci on the X chromosome are

shown in Figure 1. Details of the polymorphisms are summarized in Table 1. DNA samples were extracted from lymphocytes or lymphoblastoid cell lines of individuals from the families using established methods<sup>21</sup>. DNA samples were enzymatically digested with an appropriate restriction endonuclease, size-fractionated in agarose gels, and transferred to nylon membranes. Probes were radiolabelled with <sup>32</sup>P and hybridized to the membrane-bound DNA samples. Details of each pedigree and the genotypes at each locus were sent to one of us (G.K.S.) for analysis.

### **Linkage analysis**

All the genotype and pedigree data were checked by hand. Pedigrees having a single affected individual or with apparent non-Mendelian inheritance of a polymorphism were excluded. A total of 1368 individuals from 112 pedigrees were included in the analysis. The number of families informative at each locus is shown in Table 2.

Two-point and multipoint linkage analyses were performed using the LINKAGE package of computer programs (Version 5)<sup>22,23</sup>. The genetic parameters relating to FRAXA were as follows: mutant allele frequency (0.0006); mutation rate (0.00024 in males, 0.00048 in females); proportion of individuals with the mutant allele who had mental retardation

or expressed the fragile site (0.80 among males, 0.55 among females)<sup>3,13</sup>. If pedigree data indicated that an apparently normal individual of either sex was an obligate carrier, that individual was coded as affected for the linkage analysis.

For two-point linkage analysis, LOD scores were calculated for recombination fractions between FRAXA and each of the polymorphic loci. The LOD score is a statistical measure of the relative likelihood of a given recombination fraction<sup>24</sup>. The recombination fraction corresponding to the maximum LOD score is the best estimate of the true recombination fraction between the two loci.

Multipoint linkage analysis is a statistical method for localizing a disease gene in relation to polymorphic loci by simultaneous evaluation of all the genotype and pedigree data. For this analysis the genetic location of FRAXA was determined in relation to a known genetic map. The genetic map consisted of the positions of six polymorphic loci, DXS98, DXS369, DXS297, DXS296, IDS, and DXS304. The order of these loci down the X chromosome (Fig. 1) has been determined independently by both physical mapping<sup>25</sup> and genetic linkage studies (G.K. Suthers et al., submitted). The recombination fractions between these six loci have been estimated in a large series of normal pedigrees to be DXS98-(12.3%)-DXS369-(0%)-DXS297-(5.7%)-DXS296-(0%)-IDS-(1.2%)-DXS304 (G.K.

Suthers et al., submitted). The order of these loci and the recombination fractions constituted the genetic map on which FRAXA was localized. Genotype data at one or more of these loci were available from 101 of the fragile X pedigrees.

Multipoint LOD scores are comparable to two-point LOD scores, and are a measure of the relative likelihood of FRAXA being located at a given point on the genetic map<sup>24</sup>. The point on the genetic map corresponding to the highest multipoint LOD score indicates the most likely location of FRAXA.



## RESULTS

### Two-point linkage analysis

The results of two-point linkage analysis of FRAXA and each of the nine polymorphic loci are summarized in Table 2. Recombination was observed between FRAXA and each of the loci. Details of the pedigrees demonstrating recombination and of two-point linkage analysis of all pairs of loci are available.

The locus closest to FRAXA was DXS296, which had a peak LOD score of 33.45 at a recombination fraction of 1.5% with FRAXA. This analysis incorporated the data from the earlier study<sup>11</sup>. Recombination between DXS296 and FRAXA was documented in three affected males from three different families. The adjacent locus IDS had a recombination fraction of 8.9% with FRAXA. The other locus distal to FRAXA, DXS304, had a recombination fraction of 3.1%.

The proximal loci DXS369 and DXS297 had recombination fractions with FRAXA of 6.6% and 4.2% respectively.

### Multipoint linkage analysis

Multipoint LOD scores were calculated for various

positions of FRAXA along the genetic map (Fig. 2). The peak multipoint LOD score was 48.49. The corresponding location of FRAXA was at a recombination fraction of 2.2% proximal to DXS296. The confidence interval<sup>27</sup> for the recombination fraction between FRAXA and DXS296 was 1.0% to 3.7% with FRAXA located proximal to DXS296.

## DISCUSSION

This collaborative linkage study documents the genetic locations of nine polymorphic loci in relation to FRAXA. The relatively large recombination fractions noted between FRAXA and F9, DXS105, and DXS52 (Table 2) are similar to published values<sup>27</sup>. The recombination fraction of 5.8% between FRAXA and DXS98 is similar to the value initially reported<sup>28</sup>, but subsequent pooled studies have indicated that the recombination fraction is more likely to be 15%<sup>8</sup>. The larger value is also more consistent with the relative positions of these polymorphic loci in normal pedigrees (G.K. Suthers et al., submitted).

Two-point linkage analysis of FRAXA with the loci DXS369, DXS297, DXS296, IDS, and DXS304 indicated that the best estimates of the recombination fractions were all less than 10%. The recombination fractions between FRAXA and DXS369 and DXS304 were consistent with published values<sup>10,12</sup>. The recombination fraction between FRAXA and IDS was estimated to be 8.9%. This value seems inconsistent with other data. Physical mapping studies have indicated that IDS lies between DXS296 and DXS304<sup>25</sup>. However, these two loci had recombination fractions with FRAXA of only 1.5% and 3.1% respectively (Table 2). Although the recombination fraction between IDS and FRAXA seemed inappropriately large, IDS was

studied in a small number of families and the confidence interval for the recombination fraction was wide. In normal families there was no recombination between IDS and DXS296 (G.K. Suthers et al., submitted), and the true recombination fraction between FRAXA and IDS is likely to be less than 8.9%.

Multipoint linkage analysis is statistically more efficient than two-point linkage analyses, and provides a more accurate and precise genetic map<sup>22</sup>. Estimates of the recombination fractions between the various loci and FRAXA were derived from Fig. 2 and are summarized in Table 3. On multipoint linkage analysis, the five loci DXS369, DXS297, DXS296, IDS, and DXS304 all had recombination fractions of 4% or less with FRAXA.

This represents a major advance in the development of the genetic map near FRAXA, and has immediate application in genetic counselling. An estimate of genetic risk based on the inheritance of any one of these polymorphic loci would be correct in at least 96% of cases. The inclusion of other pedigree or cytogenetic data in the analysis may reduce the risk even further<sup>6,29</sup>.

Defining a recombination fraction of 2% between a locus and FRAXA could only be achieved with a collaborative study.

Ideally a recombination fraction of this magnitude could be documented by analyzing the DNA of approximately 50 offspring in fragile X pedigrees. However, the presence of unaffected carriers for the fragile X syndrome and the irregular size and structure of human families markedly reduces the amount of information that can be obtained from linkage studies<sup>24</sup>. There has only been one other multipoint linkage study of the fragile X syndrome of this magnitude. Brown et al. described a linkage study of 147 families<sup>30</sup>. The closest polymorphic loci that were localized in that study were F9 and DXS52, each of which have recombination fractions of over 12% with FRAXA.

Less than 50% of women are heterozygous at each of the five polymorphic loci close to FRAXA (Table 1), and at first glance these polymorphisms might appear to be of little added value in studies of fragile X families. However, two factors argue against such a pessimistic conclusion. First, all five loci are close to FRAXA, and an accurate estimate of carrier risk can be made on the basis of the inheritance of just one polymorphism. The probability that a woman would be heterozygous for at least one of the loci is high. Second, a number of the polymorphisms can be detected using the same restriction endonuclease to digest the DNA of family members. The nylon membrane to which the digested DNA is transferred can be reprobbed for a number of different polymorphisms, and

it is possible to rapidly screen the polymorphisms that are close to FRAXA.

An efficient strategy for DNA studies in families with the fragile X syndrome is presented in Table 3. Step 1 involves digesting the DNA samples of family members with three different restriction endonuclease, and using probes which identify polymorphisms at DXS296, IDS, and DXS297. The probability of a woman being heterozygous at one or more of these loci is 80%. In the event that a woman is not informative at these loci, the digested DNA samples may be reprobbed to identify polymorphisms at DXS369 and DXS304 (Step 2). Polymorphisms would be detected in a further <sup>4</sup>15% of women. Using just three enzymes and five probes, <sup>4</sup>98% of women would be heterozygous for at least one of these polymorphisms. The probability that a woman would be heterozygous for two of these polymorphisms which flank FRAXA is <sup>8</sup>68%. Step 3 raises the proportion of women who would be polymorphic at one or more loci to more than 98%.

In presenting this diagnostic strategy, two cautions should be noted. First, careful cytogenetic examination remains crucial to avoid inaccurate diagnosis. A second fragile site has been documented in normal men and women immediately proximal to the fragile site characteristic of the fragile X syndrome<sup>31,32</sup>. If the two fragile sites are not

distinguished, an individual may be incorrectly classified having the fragile X syndrome or being a carrier, and subsequent genetic risk estimates based on DNA studies could be incorrect. Second, the fragile X syndrome is a complex genetic disorder. In all but the simplest of counselling situations, it is advisable to use appropriate computer programs (such as LINKAGE<sup>22</sup>) to integrate the pedigree, cytogenetic, and DNA polymorphism data to provide accurate genetic risk estimates<sup>29</sup>.

It is now possible to correlate physical distances (measured as kilobases of DNA) near FRAXA with genetic distances (measured as recombination fractions). DXS296 and IDS are separated by 800 kilobases (kb) of DNA, and in normal families there was no recombination between them (G.K. Suthers et al., submitted). IDS and DXS304 are no more than 900 kb apart and had a recombination fraction of 1.2%. If this relationship between physical and genetic distances is maintained near FRAXA, FRAXA is approximately two thousand kb proximal to DXS296. The recent cloning of the gene responsible for cystic fibrosis<sup>33</sup> has demonstrated that it is feasible to cover a distance such as this, and so to isolate the fragile X mutation itself.

**ACKNOWLEDGEMENTS**

We would like to thank the many clinicians, genetic counsellors, laboratory colleagues, and forebearing families who made this study possible. We acknowledge the advice and assistance of M.C. Pellisier, I. Boccaccio, U. Pettersson, J.O. Ulmer, and A. Schinzel. The LINKAGE programs were provided by Dr J Ott of Columbia University. Dr J Suthers critically reviewed the manuscript.



**REFERENCES**

- 1 Sutherland GR, Hecht F. Fragile Sites on Human Chromosomes. New York: Oxford University Press, 1985.
- 2 Nussbaum RL, Ledbetter DH. Fragile X syndrome: a unique mutation in man. *Ann Rev Genet.* 1986; 20:109-45.
- 3 Sherman SL, Jacobs PA, Morton NE et al. Further segregation analysis of the fragile X syndrome with special reference to transmitting males. *Hum Genet.* 1985; 69:289-99.
- 4 Turner G, Brookwell R, Daniel A, Selikowitz M, Zilibowitz M. Heterozygous expression of X-linked mental retardation and X-chromosome marker fra(X)(q27). *N Engl J Med.* 1980; 303:662-4.
- 5 Turner G, Robinson H, Laing S, Purvis-Smith S. Preventive screening for the fragile X syndrome. *N Engl J Med.* 1986; 315:607-9.
- 6 Sutherland GR, Mulley JC. Diagnostic molecular genetics of the fragile X. *Clin Genet.* 1990; 37:2-11.
- 7 Gusella JF. DNA polymorphism and human disease. *Ann Rev*

Biochem. 1986; 55:831-54.

- 8 Mandel JL, Willard HF, Nussbaum RL, Romeo G, Puck JM, Davies KE. Report of the committee on the genetic constitution of the X chromosome. Human Gene Mapping 10 (1989): Tenth International Workshop on Human Gene Mapping. Cytogenet Cell Genet. 1989; 51:384-437.
- 9 Heilig R, Oberle I, Arveiler B, Hanauer A, Vidaud M, Mandel JL. Improved DNA markers for efficient analysis of fragile X families. Am J Med Genet. 1988; 3:543-50.
- 10 Oostra BA, Hupkes PE, Perdon LF, et al. New polymorphic DNA marker close to the fragile site FRAXA. Genomics 1990; 6:129-32.
- 11 Suthers GK, Callen DF, Hyland VJ, et al. A new DNA marker tightly linked to the fragile X locus (FRAXA). Science 1989 246:1298-300.
- 12 Vincent A, Dahl N, Oberle I, et al. The polymorphic marker DXS304 is within 5 centimorgans of the fragile X locus. Genomics 1989; 5:797-801.
- 13 Sherman SL, Rogatko A, Turner G. Recurrence risks for relatives in families with an isolated case of the

fragile X syndrome. *Am J Med Genet.* 1988; 31:753-65.

- 14 Sutherland GR. The detection of fragile sites on human chromosomes. In: Adolph KW, ed. *Advanced Techniques in Chromosome Research*. New York: Marcel Dekker (in press).
- 15 Kidd KK, Bowcock AM, Schmidtke J, et al. Report of the DNA committee and catalogs of cloned and mapped genes and DNA polymorphisms. *Human Gene Mapping 10* (1989): Tenth International Workshop on Human Gene Mapping. *Cytogenet Cell Genet.* 1989; 51:622-947.
- 16 Schnur RE, Ledbetter SA, Ledbetter DH, Merry DE, Nussbaum RL. New polymorphisms at the DXS98 locus and confirmation of its location proximal to FRAXA by in situ hybridization. *Am J Hum Genet.* 1989; 44:248-54.
- 17 Oberle I, Vincent A, Abbadi N, et al. A new polymorphism and a new chromosome breakpoint establish the physical and genetic mapping of DXS369 in the DXS98-FRAXA interval. *Am J Med Genet.* (in press).
- 18 Yu S, Suthers GK, Mulley JC. A BclI RFLP for DXS296 (VK21) near the fragile X. *Nucl Acid Res.* 1990; 18:690.
- 19 Dahl N, Hammarstrom-Heeroma K, Goonewardena P, et al.

Isolation of a DNA probe of potential use for diagnosis of the fragile-X syndrome. *Hum Genet.* 1989; 82:216-8.

- 20 Rousseau F, Vincent A, Oberle I, Mandel JL. New informative polymorphism at the DXS304 locus, a close distal marker for the fragile X locus. *Hum Genet.* 1990; 84:263-6.
- 21 Maniatis T, Fritsch EF, Sambrook J. *Molecular Cloning: a Laboratory Manual.* New York: Cold Spring Harbor Laboratory, 1982.
- 22 Lathrop GM, Lalouel JM, Julier C, Ott J. Multilocus linkage analysis in humans: Detection of linkage and estimation of recombination. *Am J Hum Genet.* 1985; 37:482-98.
- 23 Lathrop GM, Lalouel JM. Efficient computations in multilocus linkage analysis. *Am J Hum Genet.* 1988; 42:498-505.
- 24 Ott J. *Analysis of Human Genetic Linkage.* Baltimore: Johns Hopkins University Press, 1985.
- 25 Suthers GK, Hyland VJ, Callen DF, et al. Physical mapping of new DNA probes near the fragile X (FRAXA)

with a panel of cell lines. *Am J Hum Genet.* (in press).

- 26 Conneally PM, Edwards JH, Kidd KK, et al. Report of the committee on methods of linkage analysis and reporting. Eighth International Workshop on Human Gene Mapping. *Cytogenet Cell Genet.* 1985; 40:356-9.
- 27 Keats B, Ott J, Conneally M. Report of the committee on linkage and gene order. *Human Gene Mapping 10* (1989): Tenth International Workshop on Human Gene Mapping. *Cytogenet Cell Genet.* 1989; 51:459-502.
- 28 Brown WT, Wu Y, Gross AC, Chan CB, Dobkin CS, Jenkins EC. RFLP for linkage analysis of fragile X syndrome. *Lancet* 1987; ii:280.
- 29 Mulley JC, Gedeon AK, Thorn KA, Bates LJ, Sutherland GR. Linkage and genetic counselling for the fragile X using DNA probes 52A, F9, DX13, and St14. *Am J Med Genet.* 1987; 27:435-48.
- 30 Brown WT, Gross A, Chan C, et al. Multilocus analysis of the fragile X syndrome. *Hum Genet.* 1988; 78:201-205.
- 31 Ledbetter SA, Ledbetter DH. A common fragile site at Xq27: theoretical and practical implications. *Am J Hum*

Genet. 1988; 42:694-702.

- 32 Sutherland GR, Baker E. The common fragile site in band q27 of the human X chromosome is not coincident with the fragile X. Clin Genet. 1990; 37:167-72.
  
- 33 Rommens JM, Ianuzzi MC, Kerem BS, et al. Identification of the cystic fibrosis gene: chromosome walking and jumping. Science 1989; 245:1059-65.

TABLE 1  
Restriction fragment length polymorphisms used in the linkage study.

Locus	Probe	Enzyme <sup>a</sup>	Heterozygosity <sup>b</sup>	Location	Reference
<u>F9</u>	pVIII	<u>TaqI</u>	0.38	Xq26.3-q27.1	15
<u>DXS105</u> <sup>c</sup>	cX55.7	<u>TaqI</u>	0.11	Xq27.1-q27.2	15
		<u>HindIII</u>	0.48		
		<u>EcoRI</u>	0.40		
		<u>StuI</u>	0.41		
	55E	<u>PstI</u>	0.48		
<u>DXS98</u> <sup>c</sup>	4D-8	<u>MspI</u>	0.30	Xq27.2	15
		<u>XmnI</u>	0.08		
	4D8-B	<u>BglII</u>	0.15	16	
	4D-8IV	<u>XmnI</u>	0.08		
<u>DXS369</u> <sup>c</sup>	RN1	<u>XmnI</u>	0.48	Xq27.2-q27.3	10
		<u>TaqI</u>	0.24		17
<u>DXS297</u> <sup>c</sup>	VK23B	<u>HindIII</u>	0.34	Xq27	*
		<u>XmnI</u>	0.49		
<u>DXS296</u> <sup>c</sup>	VK21A	<u>TaqI</u>	0.23	Xq27.3-q28	11
		<u>BclI</u>	0.23		18
	VK21C	<u>MspI</u>	0.31		11
<u>IDS</u>	pc2S15	<u>StuI</u>	0.50	Xq28	*
		<u>TaqI</u>	0.08		
<u>DXS304</u> <sup>c</sup>	U6.2	many	0.36	Xq28	19
	U6.2-20E	<u>BanI</u>	0.49		20
<u>DXS52</u>	St14-1	<u>TaqI</u>	0.80	Xq28	15

**Footnote to Table 1**

a restriction endonuclease used to digest DNA samples from family members.

b frequency of women heterozygous at the polymor locus.

c women who are not heterozygous for one polymorphis this locus have a reduced probability of b heterozygous for other polymorphisms at the locus.

\* G.K. Suthers et al., submitted



TABLE 2

Summary of two-point linkage analysis of the Fragile X locus, FRAXA, and nearby loci.

	Recombination fractions							LOD <sub>max</sub>	R.F. <sup>a</sup>	C.I. <sup>b</sup>	n <sup>c</sup>
	0.00	0.01	0.05	0.10	0.20	0.30	0.40				
<u>FRAXA</u> vs.											
<u>F9</u>	-26.29	-8.85	1.02	4.88	6.49	5.18	2.66	6.49	.190	(.12-.28)	44
<u>DXS105</u>	-19.68	-5.11	2.55	4.88	5.11	3.38	1.22	5.40	.152	(.08-.24)	37
<u>DXS98</u>	3.65	5.88	6.69	6.47	5.10	3.25	1.36	6.70	.058	(.01-.15)	25
<u>DXS369</u>	5.04	11.31	14.50	14.24	11.06	6.74	2.56	14.62	.066	(.04-.12)	45
<u>DXS297</u>	3.13	4.17	4.46	4.23	3.28	1.99	0.68	4.46	.042	(.005-.18)	12
<u>DXS296</u>	30.75	33.36	32.05	28.58	20.29	11.56	4.08	33.45	.015	(.005-.04)	67
<u>IDS</u>	0.34	2.69	4.01	4.18	3.50	2.32	1.02	4.19	.089	(.02-.23)	16
<u>DXS304</u>	7.57	9.30	9.56	8.74	6.27	3.49	1.09	9.67	.031	(.005-.10)	27
<u>DXS52</u>	-21.42	2.31	16.84	21.07	19.43	12.98	5.43	21.45	.126	(.09-.17)	89

- a recombination fraction corresponding to the maximum LOD score.  
b approximate 90% confidence interval for the recombination fraction<sup>26</sup>.  
c number of families polymorphic at the locus.

Table 3.

Strategy for DNA studies of fragile X families

	Digest DNA with	Probe <sup>a</sup>	DNA with	Recombination fraction with <u>FRAXA</u>
STEP 1				
	<u>TaqI</u> <sup>b</sup>	VK21A	( <u>DXS296</u> ) <sup>c</sup>	2% distal <sup>d</sup>
	<u>StuI</u>	pc2S15	( <u>IDS</u> )	2% distal
	<u>XmnI</u>	VK23B	( <u>DXS297</u> )	4% proximal
STEP 2 (if necessary)				
reprobe	<u>TaqI</u> <sup>b</sup>	U6.2	( <u>DXS304</u> )	3% distal
reprobe	<u>XmnI</u>	RN1	( <u>DXS369</u> )	4% proximal
reprobe	<u>TaqI</u>	RN1	( <u>DXS369</u> )	4% proximal
STEP 3 (if necessary)				
	<u>BanI</u>	U6.2-20E	( <u>DXS304</u> )	3% distal
	<u>MspI</u>	VK21C	( <u>DXS296</u> )	2% distal
reprobe	<u>TaqI</u>	pc2S15	( <u>IDS</u> )	2% distal
	<u>HindIII</u>	VK23B	( <u>DXS297</u> )	4% proximal

- a these probes are available from sources given in the references listed in Table 1.
- b the enzyme/probe combinations of TaqI/VK21A and TaqI/U6.2 could be replaced with MspI/VK21C and MspI/U6.2.
- c the locus detected by each of the probes is indicated.
- d distal/proximal indicates the location of the locus relative to FRAXA.

# FIGURE LEGENDS

Figure 1.

The order of polymorphic loci down the X chromosome. The dark band on the ideogram corresponds to the band at Xq27 noted after G-banding of the chromosome. The positions of the loci and of the fragile site in relation to the chromosome bands are approximate.

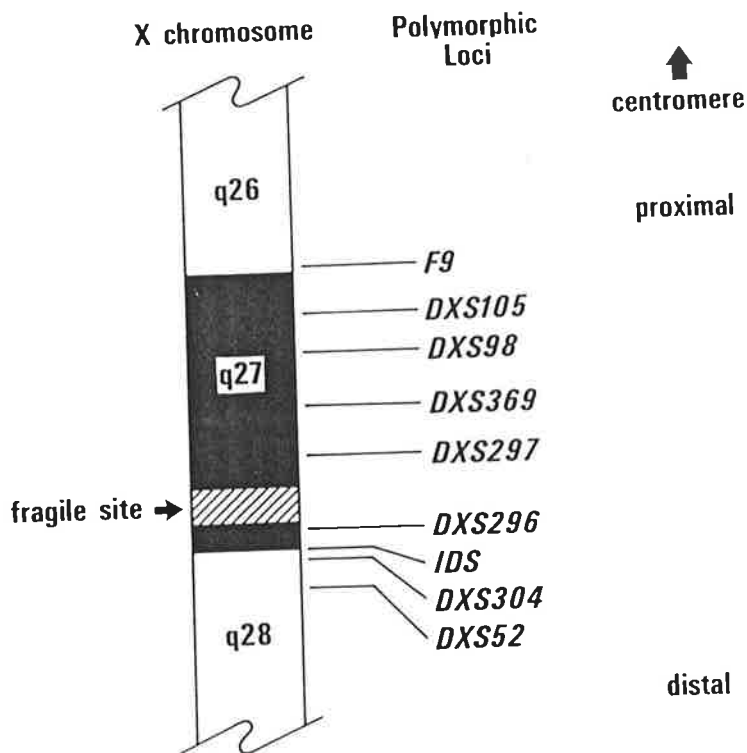
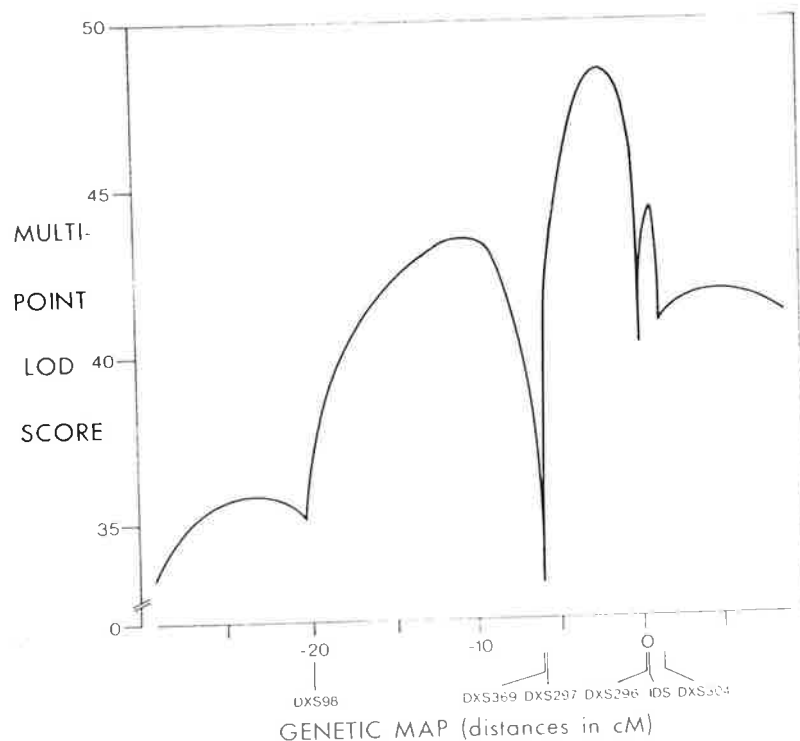


Figure 2.

Multipoint LOD scores for the location of FRAXA are plotted against genetic location along the X chromosome. The background genetic map (X-axis) was derived from a large multipoint linkage study of normal families (Methods). The origin of the map was arbitrarily placed at DXS296. Distances along the map were derived from recombination fractions using Haldane's formula<sup>24</sup>, and are expressed as centimorgans (cM); at recombination fractions of less than 10%, an increase of 1% in the recombination fraction is approximately equal to 1 cM. DXS98 lay 20.2 cM proximal to DXS296, DXS369 and DXS297 lay 6.1 cM proximal to DXS296, IDS was placed coincident with DXS296, and DXS304 lay 1.2 cM distal to DXS296. Multipoint LOD scores for FRAXA location were calculated at 20 points in the interval DXS297-DXS296; calculations were performed at 5 points in each of the remaining intervals. The peak of the multipoint LOD score curve occurred 2.2 cM proximal to DXS296, indicating that this was the most likely location of FRAXA.





Homog.art

July 13, 1990

LINKAGE HOMOGENEITY NEAR THE FRAGILE X LOCUS  
IN NORMAL AND FRAGILE X FAMILIES.

G.K. Suthers<sup>1</sup>, J.C.Mulley<sup>1</sup>, M.A. Voelckel<sup>2</sup>, N. Dahl<sup>3</sup>, M.L. Vaisanen<sup>4</sup>, P. Steinbach<sup>5</sup>, I.A. Glass<sup>6</sup>, C.E. Schwartz<sup>7</sup>, B.A. van Oost<sup>8</sup>, S.N. Thibodeau<sup>9</sup>, N.E. Haites<sup>10</sup>, B.A. Oostra<sup>11</sup>, A. Schinzel<sup>12</sup>, M. Carballo<sup>13</sup>, C.P. Morris<sup>14</sup>, J.J. Hopwood<sup>14</sup>, G.R. Sutherland<sup>1</sup>.

---

<sup>1</sup>Department of Cytogenetics & Molecular Genetics, and <sup>14</sup>Lysosomal Diseases Research Unit, Department of Chemical Pathology, Adelaide Children's Hospital, Adelaide; <sup>2</sup>Centre de Diagnostic Prenatal, Hopital d'Enfants de la Timone, Marseille; <sup>3</sup>Department of Medical Genetics, University of Uppsala, Uppsala; <sup>4</sup>Department of Clinical Genetics, Oulu University Central Hospital, Oulu;

<sup>5</sup>Department of Clinical Genetics, University of Ulm, Ulm;

<sup>6</sup>University Department of Medical Genetics, University of

Glasgow, Glasgow; <sup>7</sup> Greenwood Genetics Center, Greenwood;

<sup>8</sup>Department of Human Genetics, University Hospital, Nijmegen;

<sup>9</sup>Department of Laboratory Medicine, Mayo Clinic, Rochester;

<sup>10</sup>Aberdeen Medical Genetics Group, Department of Molecular and

Cell Biology, University of Aberdeen, Aberdeen; <sup>11</sup>Department of

Cell Biology and Clinical Genetics. Erasmus University,

Rotterdam; <sup>12</sup>Institute of Medical Genetics, University of Zurich,

Zurich; <sup>13</sup>Molecular Genetics, Consejo Superior de Investigaciones

Cientificas, Barcelona.

---

Address reprint requests to Dr G.R. Sutherland, Department of  
Cytogenetics & Molecular Genetics, Adelaide Children's Hospital,  
North Adelaide, SA 5006 Australia.

---

Running heading: Linkage homogeneity at Xq27-q28

SUMMARY

The fragile X syndrome locus, FRAXA, is located at Xq27. Until recently, few polymorphic loci had been genetically mapped close to FRAXA. This has been attributed to an increased frequency of recombination at Xq27, possibly associated with the fragile X mutation. In addition, the frequency of recombination around FRAXA has been reported to vary among fragile X families. These observations suggested that the genetic map at Xq27 was different in normal versus fragile X populations, and that the genetic map also varied within the fragile X population. Such variability would reduce the reliability of carrier risk estimates based on DNA studies in fragile X families.

Five polymorphic loci have now been mapped to within 4 centimorgan of FRAXA - DXS369, DXS297, DXS296, IDS, and DXS304. The frequency of recombination at Xq26-q28 was evaluated using data at these loci and at more distant loci from 112 families with the fragile X syndrome. Two-point and multipoint linkage analyses failed to detect any difference in the recombination fractions in fragile X versus normal families. Two-point and multipoint tests of linkage homogeneity failed to detect any evidence of linkage heterogeneity in the fragile X families.

Therefore, genetic maps derived from large samples of normal families or fragile X families are equally valid as the basis for



*Suthers et al. p. 4*

calculating carrier risk estimates in a particular family.

## INTRODUCTION

The fragile X syndrome is characterised by unusual clinical and cytogenetic features (Sutherland & Hecht 1985; Nussbaum & Ledbetter 1986). The clinical and cytogenetic penetrance of the mutant locus, FRAXA, is incomplete in both males and females, and varies according to the sex and intellectual status of the carrier parent (Sherman et al. 1985). FRAXA is located at - or very close to - the distinctive folate-sensitive fragile site at Xq27.3 which gives the syndrome its name (Sutherland & Hecht 1985). This region of the X chromosome has been the focus of a number of genetic linkage studies in both normal and fragile X families, and it has been suggested that the frequency of recombination in this region may also be unusual.

The proposals regarding frequency of recombination at Xq27 may be summarized as follows. First, *Xq27 is a region of preferential recombination in normal and fragile X families* (Hartley et al. 1984; Szabo et al. 1984; Davies et al. 1985; Oberle et al. 1985, 1987). A 300 kilobase (kb) region of preferential recombination has been documented at Xq26 (Nguyen et al. 1989), but there is no evidence of increased recombination at distal Xq27 (Suthers et al. 1990b) or at Xq28 in normal families (Feil et al. 1990). A two-point linkage study of F9 (at Xq26) and DXS52 (at Xq28) in normal versus fragile X families failed to detect a difference in the recombination fraction (Oberle et al. 1986). The

recombination fraction between these two loci is approximately 0.3, and this comparison has not been repeated using loci at Xq27.

The second proposal is that *the frequency of recombination at Xq27 varies between fragile X families*. In a series of papers, Brown et al. (1985, 1986, 1987, 1988) documented linkage heterogeneity in the two-point recombination fraction between F9 and FRAXA. However, a multipoint analysis failed to corroborate this finding (Clayton et al. 1988). This analysis has not been repeated with polymorphic loci closer to FRAXA than F9.

We have recently defined the locations of a number of polymorphic loci at Xq27-q28 by physical mapping (Suthers et al. 1990a) and by linkage studies in normal families (Suthers et al. 1990b). The genetic map at Xq26-q28 in normal families was determined to be cen-F9-(6.2)-DXS105-(4.1)-DXS98-(14.1)-DXS369-(0.0)-DXS297-(6.1)-DXS296-(0.0)-IDS-(1.2)-DXS304-(13.7)-DXS52-qter (distances in centimorgan [cM] using Haldane's mapping function). The location of FRAXA in relation to this map was documented in a collaborative linkage study of 112 fragile X families (Suthers et al. 1990c). FRAXA was placed 2.2 cM proximal to DXS296. On the basis of this genetic map, the loci DXS369, DXS297, IDS, and DXS304 were all within 4 cM of FRAXA.

The collaborative study provided a rare opportunity to examine both the frequency of recombination and the proposed linkage heterogeneity at Xq27 in a large number of fragile X families.

## MATERIALS AND METHODS

### Linkage analysis.

A total of 1368 individuals from 112 fragile X families were included in the linkage study. Genotypes at the following polymorphic loci were analysed: F9, DXS105, DXS98, DXS369, DXS297, DXS296, IDS, DXS304, and DXS52. Details of family selection, probes used, and DNA methods have been presented elsewhere (Suthers et al. 1990c).

Two-point and multipoint linkage analyses were performed using the programs MLINK, LINKMAP, and ILINK from the LINKAGE package of computer programs (Version 5) (Lathrop et al. 1985; Lathrop & Lalouel 1988). The RFLP allele frequencies have been described (Kidd et al. 1989; Suthers et al. 1990b). The genetic parameters relating to FRAXA were as follows: mutant allele frequency (0.0006); mutation rate (0.00024 in males, 0.00048 in females); penetrance of mental retardation or fragile X expression (0.80 among males, 0.55 among females)(Sherman et al. 1985, 1988). If the pedigree indicated that an apparently normal individual of either sex was an obligate carrier, that individual was coded as affected for the linkage analysis.

### Homogeneity testing.

Homogeneity of two-point recombination fractions was assessed using the HOMOG package of computer programs (Ott 1985). One of

the programs, MTEST, is an implementation of Morton's homogeneity test (Morton 1956). MTEST was used to determine whether there was a significant difference in the recombination fractions between two loci in a normal population versus the fragile X families. The normal population consisted of 40 reference families from Centre d'Etude du Polymorphisme Humain (CEPH) (Dausset et al. 1990).

When testing for heterogeneity of two-point recombination fractions in the fragile X families, one of the implementations of the Admixture test (Ott 1985) was used. The choice of implementation was determined by the hypothesis being tested. As Brown et al. (1987) had suggested that fragile X families could be divided into two groups on the basis of the F9:FRAXA recombination fraction, the program HOMOG2 was used. This program uses maximum likelihood estimation to evaluate three hypotheses. The null hypothesis ( $H_0$ ) is that the two loci being examined are unlinked in all the families ( $\hat{\theta}=0.50$ ). The next hypothesis ( $H_1$ ) is that the two loci are linked at the same recombination fraction in all the families ( $\hat{\theta} < 0.50$ ). The final hypothesis ( $H_2$ ) is that the families comprise two groups, with the loci linked at different recombination fractions ( $\hat{\theta}_1 < \hat{\theta}_2 < 0.50$ ).

The transformed values of the three likelihoods ( $-2\ln(\text{likelihood})$ ) may be compared using the  $\chi^2$  distribution with one ( $H_1$  vs  $H_0$ ) or two ( $H_2$  vs.  $H_1$ ) degrees of freedom. Under these

conditions, the tests are generally conservative. Tests of homogeneity were not performed for all two-point analyses in the fragile X families. These tests have limited power (Ott 1986; Risch 1988), and there is little point in performing the test with limited data. Moreover, performing multiple comparisons would have reduced the power of the tests even further.

The programs HOMOG2, HOMOG3, and HOMOG4 were used to perform multipoint tests of linkage homogeneity. In the multipoint linkage analysis of fragile X families (Suthers et al. 1990c), the likelihood (expressed as a multipoint LOD score) of FRAXA being located at specific points along a predefined genetic map had been calculated for each family using the program LINKMAP. The background genetic map used for this analysis had been determined independently in the CEPH families (Suthers et al. 1990b), and consisted of the loci DXS98, DXS369, DXS297, DXS296, IDS, and DXS304. For homogeneity testing, multipoint LOD scores were calculated at 10 locations in the interval DXS297:DXS296, and at 5 locations in each of the other intervals.

In analyzing these data with a multipoint linkage homogeneity test, the null hypothesis ( $H_0$ ) was that FRAXA was not located on the genetic map;  $H_1$  was that there was a single location for FRAXA on the map;  $H_2$  was that the families could be divided into two, three, or four groups (depending on the program used) with different FRAXA locations. When performing a multipoint test of

linkage homogeneity, the differences in the likelihoods do not have a defined distribution (Clayton et al. 1988; J. Ott, personal communication), and significance values cannot be assigned to differences in the likelihoods of the three hypotheses.



## RESULTS

### Frequency of recombination in normal vs. fragile X families.

Two-point linkage analyses were performed using the program ILINK. The peak LOD scores and best estimates of the recombination fractions ( $\hat{\theta}$ ) for all pairs of loci in the 112 fragile X families are summarized in Table 1. For comparison, the results of two-point linkage analyses of the same loci in the 40 CEPH families are also presented. There were 16 pairs of loci where the estimated recombination fractions in the two populations differed by 0.05 or more. However, the peak LOD scores were generally low in one or both of the populations, and little significance could be attached to the differences in recombination fractions.

One pair of loci, F9:DXS105, differed by 0.06 in the recombination fractions in the two populations, and the peak LOD scores in the two populations were both above 10.0. The program MTEST was used to determine whether this difference in recombination fraction was significant; the two recombination fractions were not significantly different ( $\chi^2_1=2.98$ ;  $p=0.08$ ).

Forty nine of the fragile X families were informative at two or more of the loci that were within 4 cM of FRAXA. The recombination fractions between these five loci were estimated

simultaneously in the fragile families using the program ILINK.

The recombination fractions were as follows:

DXS369-(0.000)-DXS297-(0.069)-DXS296-(0.000)-IDS-(0.018)-DXS304.

In the CEPH families these recombination fractions were

DXS369-(0.000)-DXS297-(0.057)-DXS296-(0.000)-IDS-(0.012)-DXS304.

In the fragile X families, the difference in the transformed likelihoods ( $-2\ln(\text{likelihood})$ ) of these two sets of recombination fractions was not significant ( $\chi^2_4=0.35$ ;  $p>0.5$ ).

#### Tests of linkage homogeneity in the fragile X families.

Four loci were informative in more than 40 of the fragile X families, F9, DXS369, DXS296, and DXS52. Two-point homogeneity tests were performed for the recombination fractions between each of these loci and FRAXA. The data for each test were two-point LOD scores at 10 values of  $\theta$  between 0.00 and 0.45. As the DXS296:FRAXA LOD score curve had a sharp peak at  $\theta=0.015$ , LOD scores at  $\theta=0.01$ , 0.02, 0.03, and 0.04 were also included.

The results of the homogeneity tests are summarized in Table 2. For each of the four loci, the recombination fraction between the locus and FRAXA was significantly less than 0.50, i.e.  $H_0$  was rejected in favor of  $H_1$ . However, for none of the loci was there evidence of linkage heterogeneity, i.e. there was insufficient evidence to reject  $H_1$  in favor of  $H_2$ . In Table 2 the estimates of  $\theta$  between each locus and FRAXA under the hypothesis of linkage homogeneity ( $H_1$ ) are slightly different from those listed in

Table 1. The reason for this is that the data entry for the HOMOG programs consists of LOD scores at specified recombination fractions (calculated with the program MLINK). The estimate of  $\hat{\theta}$  determined by HOMOG2 is limited to one of the specified  $\theta$  values. On the other hand, the values in Table 1 were determined iteratively using the program ILINK, and there is no restriction on the estimated value of  $\hat{\theta}$ .

Multipoint homogeneity testing of FRAXA location was performed with multipoint LOD scores from 101 of the fragile X families. Each of these families was informative at one or more of the loci DXS98, DXS369, DXS297, DXS296, IDS, and DXS304. Under the hypothesis of a single location for FRAXA ( $H_1$ ), the most likely location of FRAXA was 2.2 cM proximal to DXS296. The odds in favor of  $H_1$  vs.  $H_0$  were  $10^{48}:1$ . Under the hypothesis ( $H_2$ ) of two locations for FRAXA, FRAXA was located at DXS98 in 15% of the families and 1.6 cM proximal to DXS296 in the remainder. The odds in favor of  $H_2$  vs.  $H_1$  were 2:1. The odds in favor of there being three or four locations for FRAXA rather than one were also 2:1.

## **DISCUSSION**

The fragile X syndrome is the only clinical disorder known to be associated with expression of a fragile site (Sutherland & Hecht 1985). It is also unique in being the only X-linked disorder with incomplete penetrance in males. It was therefore intriguing when linkage analysis suggested that the linkage relationships around FRAXA might be unusual.

### Recombination fractions in normal vs. fragile X families

Investigation of the linkage relationships around FRAXA is dependent on the availability of DNA probes which detect polymorphisms in the region. Clusters of polymorphic loci were identified at Xq26 and Xq28 (Szabo et al. 1984; Oberle et al. 1985, 1987), but until recently there had been few loci mapped close to FRAXA by linkage analysis. This lack of polymorphic loci was attributed to either a high frequency of recombination in the region of FRAXA, or to a selection bias in isolating probes from Xq27 (Oberle et al. 1987). There is no evidence of increased recombination at Xq27 or Xq28 in normal families (Suthers et al. 1990b; Feil et al. 1990), and it has been suggested that the region around FRAXA could contain repeated sequences and hence be under-represented when screening for unique DNA probes from genomic libraries (Hyland et al. 1989).

In the present study, a comparison of two-point and multipoint linkage relationships of loci close to FRAXA in 40 normal families and over 40 fragile X families failed to detect any difference in the recombination fractions in normal versus fragile X families. Thus there is little evidence to support the contention that Xq27 is a region of preferential recombination in either normal families or in fragile X families.

Linkage homogeneity in fragile X families

The second proposal considered in this study was whether the frequency of recombination around FRAXA varies among fragile X families. The first study of the recombination fraction between FRAXA and F9 estimated the recombination fraction to be zero (Camerino et al. 1983), but later investigations indicated that the recombination fraction was much higher (Choo et al. 1984; Warren et al. 1985). Subsequently it was suggested that the frequency of recombination in this interval may vary among fragile X families (Brown et al. 1985, 1986, 1987, 1988; Oberle et al. 1986). In an analysis of 106 families, 20% had no recombination between F9 and FRAXA, while the remainder had a recombination fraction of 0.35 (Brown et al. 1988). There was no evidence of linkage heterogeneity between FRAXA and the distal locus DXS52. These studies were limited to analyzing loci that have recombination fractions of more than 0.12 with FRAXA.

In the present study, tests of two-point linkage homogeneity between FRAXA and F9 or DXS52 failed to document linkage heterogeneity (Table 2). Tests of two-point linkage homogeneity using the closely linked loci DXS369 and DXS296 also failed to document heterogeneity in the fragile X families.

In the absence of a clear clinical or cytogenetic distinction among fragile X families, there is no *a priori* reason why there should be just two groups of fragile X families rather than many groups. In any linkage study, the peak LOD score will occur at a different recombination fraction in each family studied. This reflects the different pedigree structures, numbers of informative women in each family, and the stochastic nature of recombination. However, in a multipoint test of linkage homogeneity based on six loci close to FRAXA and 101 families, the odds in favor of their being two, three, or four locations for FRAXA rather than one were only 2:1.

The only other multipoint test of linkage homogeneity in the fragile X syndrome also failed to detect heterogeneity. Clayton et al. (1988) took essentially the same data set as Brown et al. (1987), and performed a multipoint test of homogeneity similar to that described for the present study. The odds in favor of there being two loci for FRAXA were only 2:1. In discussing the initial conflicting reports of the F9:FRAXA recombination fraction, the authors considered the possibility that there is a familial

predisposition to recombination at Xq27. Such a predisposition need not be specific to the fragile X syndrome, but may be a feature of the normal population. However, two-point tests of linkage homogeneity in the CEPH families found no evidence of heterogeneity at Xq27 (Suthers et al. 1990b).

It is not clear why the conclusions of the present study and of Clayton et al. (1988) differ from that of Brown et al. (1985, 1986, 1987, 1988). In assembling data for this analysis, there was a selective bias in favor of families informative at loci close to FRAXA. There was no bias in terms of the clinical or cytogenetic characteristics of affected males, or of pedigree structure (other than ensuring that at least two individuals in each pedigree expressed the fragile X)(Suthers et al. 1990c). There are three possible explanations for the difference. The first is that these results represent some statistical fluctuation, and will not be discussed further. The second possibility is that this study lacked sufficient power to detect linkage heterogeneity. The third possibility is that Brown et al. obtained an incorrect non-conservative result with the Admixture test.

The power of a two-point homogeneity test varies according to the hypothesis being tested. Tables have been published giving the power of various tests for different hypotheses (Cavalli-Sforza & King 1986; Ott 1986; Risch 1988), but the specific hypothesis

proposed by Brown (1988) is not listed. Moreover, the distribution of likelihood ratios in multipoint tests of linkage homogeneity is not known, and the power of this analysis cannot be estimated. On the other hand, this analysis was based on a large number of families. Two-point linkage heterogeneity between F9 and FRAXA has been documented in as few as six large families (Oberle et al. 1986), and the study of Brown et al. (1987) included just 32 families.

Although homogeneity tests are generally conservative (Risch 1988), the test results can be inaccurate if an insufficient number of LOD scores are presented as data. In particular, if the data does not include LOD scores at  $\hat{\theta}$  (estimated iteratively), the likelihood of  $H_1$  may be incorrectly low with the result that  $H_1$  is incorrectly rejected in favor of  $H_2$ . As an example, a two-point test of linkage homogeneity between FRAXA and DXS296 rejected  $H_1$  in favor of  $H_2$  ( $X^2_2=5.9$ ;  $p=0.03$ ) if LOD scores were entered for just 10 values of  $\theta$  at intervals of 0.05. If LOD scores at values of  $\theta$  close to 0.015 were included, there was insufficient evidence to reject  $H_1$  in favor of  $H_2$  (Table 2).

It is unlikely that the linkage heterogeneity reported between F9 and FRAXA by Brown et al. in 1987 was due to a non-conservative result of the Admixture test as the value of  $\hat{\theta}$  estimated under  $H_1$  by the HOMOG program was the same as that determined by two-point linkage analysis ( $\theta=0.21$ ). The estimate of  $\hat{\theta}$  under  $H_1$  was not



presented in the paper by Brown et al. (1988), and no conclusion can be drawn regarding a possible non-conservative result. The Admixture test was used by Brown et al. in 1985, but significant heterogeneity was not detected with this test. However, heterogeneity was detected using Morton's test in a subset of large families.

### Conclusion

An accurate genetic map is a prerequisite for calculating reliable genetic risk estimates in any disorder. For those involved in genetic counselling of families with the fragile X syndrome, the conclusions of this analysis are encouraging. There is no evidence that the genetic map at Xq27 is different in fragile X families versus the normal population, nor is there evidence of linkage heterogeneity among fragile X families. Therefore, genetic distances that have been estimated in normal families or fragile X families can be incorporated into genetic risk analyses with confidence.

**ACKNOWLEDGEMENTS**

We acknowledge the advice and assistance of M.C. Pellisier, I. Boccaccio, U. Pettersson, J.O. Ulmer, A Smits, JCFM Dreesen, and R. Gine. The LINKAGE and HOMOG programs were provided by Dr J Ott of Columbia University. This project was supported by the National Health and Medical Research Council of Australia; The Adelaide Children's Hospital Research Foundation; CNAM and Association Francaise; The Swedish Medical Research Council; Sigrid Juselius Foundation; Deutsche Forschungs-gemeinschaft CDFG; Grampian Health Board and Scottish Hospitals Endowments Research Trust.

REFERENCES

Brown WT, Gross AC, Chan CB, Jenkins EC (1985). Genetic linkage heterogeneity in the fragile X syndrome. *Hum Genet* 71:11-18.

Brown WT, Gross AC, Chan CB, Jenkins EC (1986). DNA linkage studies in the fragile X syndrome suggest genetic heterogeneity. *Am J Med Genet* 23:643-664.

Brown WT, Jenkins EC, Gross AC, et al. (1987). Further evidence for genetic heterogeneity in the fragile X syndrome. *Hum Genet* 75:311-321.

Brown, WT, Gross A, Chan C, et al. (1988). Multilocus analysis of the fragile X syndrome. *Hum Genet* 78:201-205.

Camerino G, Mattei MG, Mattei JF, Jaye M, Mandel JL (1983). Close linkage of fragile X-mental retardation syndrome to hemophilia B and transmission through a normal male. *Nature* 306:701-704.

Cavalli-Sforza LL, King MC (1986). Detecting linkage for genetically heterogeneous diseases and detecting heterogeneity with linkage data. *Am J Hum Genet* 38:599-616.

Choo KH, George D, Filby G, et al. (1984). Linkage analysis of

X-linked mental retardation with and without fragile-X using factor IX gene probe. *Lancet* ii:349.

Clayton JF, Gosden CM, Hastie ND, Evans HJ (1988). Linkage heterogeneity and the fragile X. *Hum Genet* 78:338-342.

Dausset J, Cann H, Cohen D, Lathrop M, Lalouel JM, White R (1990). Centre d'Etude Polymorphisme Humain (CEPH): Collaborative genetic mapping of the human genome. *Genomics* 6:575-577.

Davies KE, Mattei MG, Mattei JF, et al. (1985). Linkage studies of X-linked mental retardation: High frequency of recombination in the telomeric region of the human X chromosome (fragile site/linkage/recombination/X chromosome). *Hum Genet* 70:249-255.

Feil R, Palmieri G, d'Urso M, Heilig R, Oberle I, Mandel JL (1990). Physical and genetic mapping of polymorphic loci in Xq28 (DXS15, DXS52, and DXS134): Analysis of a cosmid clone and a yeast artificial chromosome. *Am J Hum Genet* 46:720-728.

Hartley DA, Davies KE, Drayna D, White RL, Williamson R (1984). A cytological map of the human X chromosome - evidence for non-random recombination. *Nucl Acid Res* 12:5277-5285.

Hyland VJ, Fernandez KEW, Callen DF, et al. (1989). Assignment of

anonymous DNA probes to specific intervals of human chromosomes 16 and X. *Hum Genet* 83:61-66.

Kidd KK, Bowcock AM, Schmidtke J, et al (1989). Report of the DNA committee and catalogs of cloned and mapped genes and DNA polymorphisms. *Human gene mapping 10* (1989): Tenth international workshop on human gene mapping. *Cytogenet Cell Genet* 51:622-947.

Lathrop GM, Lalouel JM, Julier C, Ott J (1985). Multilocus linkage analysis in humans: Detection of linkage and estimation of recombination. *Am J Hum Genet* 37:482-498.

Lathrop GM, Lalouel JM (1988). Efficient computations in multilocus linkage analysis. *Am J Hum Genet* 42:498-505.

Morton NE (1956). The detection and estimation of linkage between the genes for elliptocytosis and Rh blood type. *Am J Hum Genet* 8:80-96.

Nguyen C, Poustka AM, Djabali M, et al. (1989). Large scale mapping and chromosome jumping in the q27 region of the human X chromosome. *Genomics* 5:298-303.

Nussbaum RL, Ledbetter DH (1986). Fragile X syndrome: A unique mutation in man. *Ann Rev Genet* 20:109-145.

Oberle I, Drayna D, Camerino G, White R, Mandel JL (1985). The telomeric region of the human X chromosome long arm: Presence of a highly polymorphic DNA marker and analysis of recombination frequency. Proc Natl Acad Sci USA 82:2824-2828.

Oberle I, Heilig R, Moisan JP, et al. (1986). Genetic analysis of the fragile-X mental retardation syndrome with two flanking polymorphic DNA markers. Proc Natl Acad Sci USA 83:1016-1020.

Oberle I, Camerino G, Wrogemann K, Arveiler B, Hanauer A, Raimondi E, Mandel JL (1987). Multipoint genetic mapping of the Xq26-q28 region in families with fragile X mental retardation and in normal families reveals tight linkage of markers in q26-27. Hum Genet 77:60-65.

Ott J (1985). The Analysis of Human Genetic Linkage. Baltimore: Johns Hopkins University Press.

Ott J (1986). The number of families required to detect or exclude linkage heterogeneity. Am J Hum Genet 39:159-165.

Risch N (1988). A new statistical test for linkage heterogeneity. Am J Hum Genet 42:353-364.

Sherman SL, Jacobs PA, Morton NE, et al. (1985). Further

segregation analysis of the fragile X syndrome with special reference to transmitting males. *Hum Genet* 69:289-299.

Sherman SL, Rogatko A, Turner G (1988). Recurrence risks for relatives in families with an isolated case of the fragile X syndrome. *Am J Med Genet* 31:753-765.

Sutherland GR, Hecht F (1985). *Fragile Sites on Human Chromosomes*. New York: Oxford University Press.

Suthers GK, Hyland VJ, Callen DF, et al. (1990a). Physical mapping of new DNA probes near the fragile X (FRAXA) with a panel of cell lines. *Am J Hum Genet* (in press).

Suthers GK, Oberle I, Nancarrow J, et al. (1990b). Genetic mapping of new RFLPs at Xq27-q28. (submitted)

Suthers GK, Mulley JC, Voelckel MA, et al. (1990c). Genetic mapping of new DNA probes at Xq27 defines a strategy for DNA studies in the fragile X syndrome. (submitted).

Szabo P, Purrello M, Rocchi M, et al. (1984). Cytological mapping of the human glucose-6-phosphate dehydrogenase gene distal to the fragile-X site suggests a high rate of meiotic recombination across this site. *Proc Natl Acad Sci USA* 81:7855-7859.

*Suthers et al. p. 27*

Warren ST, Glover TW, Davidson RL, Jagadeeswaran P (1985).  
Linkage and recombination between fragile X-linked mental  
retardation and the factor IX gene. Hum Genet 69:44-46.



Table 1

Summary of two-point linkage analysis of loci at Xq27-q28 in fragile X and CEPH families.

	<u>F9</u>	<u>DXS105</u>	<u>DXS98</u>	<u>DXS369</u>	<u>DXS297</u>	<u>FRAXA</u>	<u>DXS296</u>	<u>IDS</u>	<u>DXS304</u>	<u>DXS52</u>
<u>F9</u>		.104 .043	.041 .078	.211 .221	ND .251	.190 ND	.300 .190	.209 .193	.099 .241	.333 .327
<u>DXS105</u>	10.63 14.34		.045 .000	.211 .170	.155 .179	.152 ND	.350 .142	.120 .209	.306 .170	.278 .279
<u>DXS98</u>	5.01 6.38	3.56 6.62		.090 .000	.000 .072	.058 ND	.123 .152	.062 .125	.274 .069	.194 .237
<u>DXS369</u>	1.46 2.54	2.03 6.79	2.28 2.71		ND .000	.066 ND	.062 .063	.000 .054	.093 .062	.184 .124
<u>DXS297</u>	ND 1.36	1.34 1.38	3.38 2.65	ND 4.52		.042 ND	.033 .155	.066 .077	.000 .042	.500 .175
<u>FRAXA</u>	6.49 ND	5.40 ND	6.70 ND	14.62 ND	4.46 ND		.015 ND	.089 ND	.031 ND	.126 ND
<u>DXS296</u>	0.26 1.28	0.59 2.28	2.61 1.19	9.54 6.08	5.76 1.19	33.45 ND		.000 .000	.019 .000	.092 .154
<u>IDS</u>	1.86 4.88	1.65 4.77	2.52 4.59	0.90 15.19	3.18 6.85	4.19 ND	4.10 15.95		.000 .014	.076 .136
<u>DXS304</u>	0.86 2.62	0.45 5.50	0.32 5.27	3.91 12.74	0.60 5.12	9.67 ND	11.85 7.53	0.92 18.20		.161 .127
<u>DXS52</u>	2.62 1.46	3.16 3.55	4.40 1.81	7.00 10.83	0.00 2.36	21.45 ND	23.11 5.06	7.79 13.93	5.10 11.25	

Note: The figures above the diagonal are the best estimates of the recombination fractions between the specified loci. The figures below the diagonal are the peak LOD scores. In each case, the upper figure is the result of analysis in the fragile X families (present paper), and the lower figure is from analysis of the CEPH families (Suthers et al. 1990b). (ND no data)

Table 2

Homogeneity tests of two-point recombination fractions in the fragile X families.

<u>F9:FRAXA</u>	(44 families)	ln(relative likelihood)	
H <sub>0</sub> :	loci unlinked in all families	0.00	\ / $\chi^2_1=29.9$ (p<0.0001)
H <sub>1</sub> :	loci linked at $\theta=0.20$ in all families	14.96	
H <sub>2</sub> :	loci linked at $\theta=0.00$ in 5% of families, at $\theta=0.20$ in 95% of families.	15.04	/ \ $\chi^2_2=0.16$ (p>0.4)
 <u>DXS369:FRAXA</u> (45 families)			
H <sub>0</sub> :	loci unlinked in all families	0.00	\ / $\chi^2_1=66.8$ (p<0.0001)
H <sub>1</sub> :	loci linked at $\theta=0.05$ in all families	33.39	
H <sub>2</sub> :	loci linked at $\theta=0.00$ in 35% of families, at $\theta=0.10$ in 65% of families.	33.77	/ \ $\chi^2_2=0.76$ (p>0.3)
 <u>DXS296:FRAXA</u> (67 families)			
H <sub>0</sub> :	loci unlinked in all families	0.00	\ / $\chi^2_1=153.8$ (p<0.0001)
H <sub>1</sub> :	loci linked at $\theta=0.02$ in all families	76.89	
H <sub>2</sub> :	loci linked at $\theta=0.01$ in 48% of families, at $\theta=0.02$ in 52% of families.	77.00	/ \ $\chi^2_2=0.22$ (p>0.4)
 <u>DXS52:FRAXA</u> (89 families)			
H <sub>0</sub> :	loci unlinked in all families	0.00	\ / $\chi^2_1=97.7$ (p<0.0001)
H <sub>1</sub> :	loci linked at $\theta=0.15$ in all families	48.83	
H <sub>2</sub> :	loci linked at $\theta=0.05$ in 50% of families, at $\theta=0.20$ in 50% of families.	49.96	/ \ $\chi^2_2=2.27$ (p>0.1)

Note to Table 2: The best estimates of  $\theta$  and of the proportion of the families in each group were obtained under each hypothesis, and are listed in the Table. The significance of differences in the relative likelihoods of the hypotheses are indicated to the right of the Table.



Suthers, G.K., Turner, G. and Mulley, J.C. (1988). Case report: Fragile X syndrome and nephrogenic diabetes insipidus. *American Journal of Medical Genetics*, 30(1/2), 231–236.

NOTE: This publication is included in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

<http://dx.doi.org/10.1002/ajmg.1320300123>

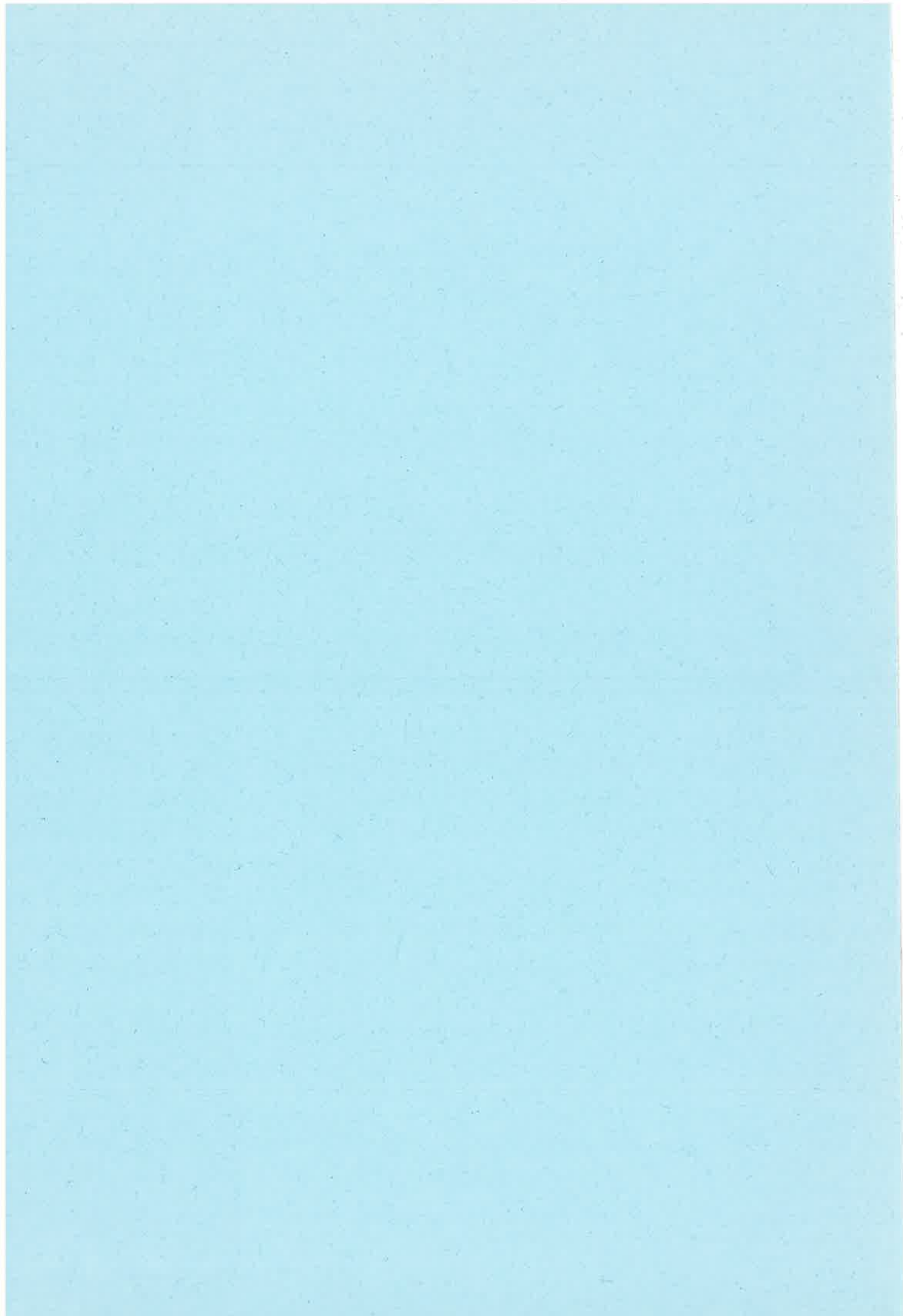


Suthers, G.K., Davies, K.E., Baker, E. and Sutherland, G.R. (1989). TaqI RFLP identified by probe 1A1 [DXS374] at Xq28. *Nucleic Acids Research*, 17(21), 8901.

NOTE: This publication is included in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

<http://dx.doi.org/10.1093/nar/17.21.8901>





Suthers, G.K. and Sutherland, G.R. (1990). Recombination and the fragile X. *Human Genetics*, 85(1), 141–142.

NOTE: This publication is included in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

<http://dx.doi.org/10.1007/BF00276345>