الفصل الرابع

المحازاة والاشجار العرقية

محازاة التتابع:

فى حالة وجود اثنان من التتابعات يتم عمل الآتى:

- قياس درجة التشابه بينهما.
- تحديد تقابل للأجزاء.
- ملاحظة طرز التحفظ والاختلاف.
- علاقات النشوء.

The dotplot:

وهى صورة بسيطة تعطى رؤية عامة للتشابه بين اثنان من التتابعات. يتكون dotplot من جدول أو مصفوفة. تشير الصفوف الى أجزاء أحد التتابعات بينما تدل الأعمدة على التتابع الآخر. وتظهر بقع فى حالة التماثل، بينما لاتوجد استجابة فى حالة الاختلاف.

مصادر DOTPLOTS على الانترنت (web Resources):

يتيح برنامج E.L. Sonnhammer's Dotter للمستخدم التحكم وطرق الحساب وتغيير مظهر العرض بواسطة ضبط معايير تفاعلية والذى يجب تجهيزه فى الكمبيوتر من على الموقع التالى:

http://www.cgr.ki.se/cgr/groups/sonnhammer/Dotter.html

يقدم الموقع التالى امكانية اجراء dotplotting تفاعلى:

http://www.isrec.isb-sib.ch/java/dotlet/exonintron.html

A PERL program to draw dotplots.

```
#!/usr/bin/perl
#dotplot.pl -- reads two sequences and prints dotplot
# read input
$/ = "";
$_ = <DATA> $_ =~ s/#(.*)\n/\n/g;
$_              =~              /^(.*)\n\s*(\d+)\s+(\d+)\s*\n(.*)\n([A-
Z\n]*)\*\s*\n(.*)\n([A-Z\n]*)\*/;
$title = $1; $nwind = $2; $thresh = $3;
$seqtl = $4; $seq1 = $5; $seqt2 = $6; $seq2 = $7;
$seq1  =~  s/\n//g;  $seq2  =~  s/\n//g;  $n   = length($seq1); $m =
length($seq2);
# postscript header
print  <<   EOF setfont scalefont 20 findfont Helvetica translate 30
setlinewidth  1.75   def load fill f closepath c newpath n rlineto r
moveto m lineto l stroke s %!PS-Adobe->

#print matrix

$dx = 500.0/$n;  $mdx = -$dx; $dy = 500.0/$m;
if  ($dy  <  $dx)  ($dx  =  $dy;}   $dy = $dx; $xmx = $n*$dx; $ymx =
$m*$dx;
print "0 510 m ($title NWIND = $nwind) show\n";
printf  "0  0  m  0  %9.2f  1  %9.2f  %9.2f  1  %9.2f  0   1 c s\n",
$ymx,$xmx,$ymx,$xmx;
for ($k = $nwind - $m + 1; $k < $n - $nwind; $k++) {
    $i = $k; $j = 1; if ($k < 1) ($i = 1; $j = 2 - $k;}
    while ($i <= $n - $nwind && $j <= $m - $nwind) {
        $_    =  (substr($seq1,$i  -1,$nwind)  ^  substr($seq2,$j  -
1,$nwind));
        $mismatch = ($_ =~ s/[^\x0]//g);
        if ($mismatch < $thresh) {
            $xl = ($i - 1)*$dx; $yb = ($m - $j)*$dy;
            printf  "n %9.2f %9.2f m %9.2f 0 r 0 %9.2f r %9.2f 0 r c
f\n",
                    $xl,$yb,$dx,$dy,$mdx;
        }
        $i++; $j++;
    }
}
print "showpage\n";

__END__
ATPases lamprey / dogfish                    #TITLE
15 6                                         #WINDOW, THRESHOLD
Petromyzon marinus mitochondrion             #SEQUENCE 1
ATGACACTAGATATCTTTGACCAATTTACCTCCCCAACA
ATATTTGGGCTTCCACTAGCCTGATTAGCTATACTAGCCCCTAGCTTA
ATATTAGTTTCACAAACACCAAAATTTATCAAATCTCGTTATCACACACTA
CTTACACCCATCTTAACATCTATTGCCAAACAACTCTTTCTTCCAATAAAC
CAACAAGGGCATAAATGAGCCTTAATTTGTATAGCCTCTATAATATTTATC
TTAATAATTAATCTTTTAGGATTATTACCATATACTTATACACCAACTACC
CAATTATCAATAAACATAGGATTAGCAGTGCCACTATGACTAGCTACTGTC
CTCATTGGGTTACAAAAAAAACCAACAGAAGCCCTAGCCCACTTATTACCA
GAAGGTACCCCAGCAGCACTCATTCCCATATTAATTATCATTGAAACTATT
AGTCTTTTTATCCGACCTATCGCCCTAGGAGTCCGACTAACCGCTAATTTA
ACAGCTGGTCACTTACTTATACAACTAGTTTCTATAACAACCTTTGTAATA
ATTCCTGTCATTTCAATTTCAATTATTACCTCACTACTTCTTCTATTA
CTAACAATTCTGGAGTTAGCTGTTGCTGTAATCCAGGCATATGTATTTATT
CTACTTTTAACTCTTTATCTGCAAGAAAACGTTT*
Scyliorhinus canicula mitochondrion          #SEQUENCE 2
ATGATTATAAGCTTTTTTGATCAATTCCTAAGTCCCTCCTTTCTAGGA
ATCCCACTAATTGCCCTAGCTATTTCAATTCCATGATTAATATTTCCAACACCAAC
AATCGTTGACTTAATAATCGATTATTAACTCTTCAAGCATGATTTATTAACCGATT
TATCAACTAATACAACCCATAAATTTAGGAGGACATAAATGAGCTATCTTATTTACA
CTAATATTATTTTTAATTACCATCAATCTTCTAGGTCTCCTTCCATATACTTTTACG
ACAACTCAACTTTCTCTTAATATAGCCTTTGCCCTGCCCTTATGGCTTACAACTGTA
```

```
ATTGGTATATTTAATCAACCAACCATTGCCCTAGGGCACTTATTACCTGAAGGTACCCCA
ACCCCTTTAGTACCAGTACTAATCATTATCGAAACCATCAGTTTATTTATTCGACCATTA
GCCTTAGGAGTCCGATTAACAGCCAACTTAACAGCTGGACATCTCCTTATACAATTAATC
GCAACTGCGGCCTTTGTCCTTTTAACTATAATACCAACCGTGGCCTTACTAACCTCCCTA
GTCCTGTTCCTATTGACTATTTTAGAAGTGGCTGTAGCTATAATTCAAGCATACGTATTT
GTCCTTCTTTTAAGCTTATATCTACAAGAAAACGTATAA*
```

محازاة التتابع بواسطة Dotplots:

لا يقدم **Dotplots** فقط صورة شاملة لتشابه اثنان من التتابعات ولكن أيضا يعطى مجموعة متكاملة ذات جودة نسبية للمحازاة المحتملة وذلك بالمرور على الصورة من الجانب الأيسر العلوى الى الجانب الأيمن السفلى.

محازاة التتابع المتعدد Multiple sequence alignment :

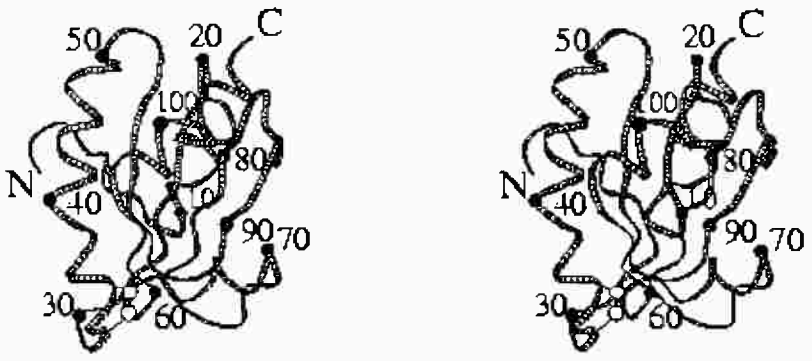
تكون نتائج الكشف عن العلاقات البعيدة بين التتابعات أكثر واقعية فى حالة تعدد التتابعات. وأيضا تعطى أدوات التنبؤ بالتركيب نتائج أفضل عندما تعتمد على محازاة تتابعات متعددة وذلك بالمقارنة بالتتابعات المنفردة. ولذلك يجب أن تحتوى محازاة التتابع المتعدد على توزيع من التتابعات قريبة وبعيدة الصلة.

الاستدلال التركيبى بواسطة محازاة التتابع المتعدد:

يمكن التعرف على الملامح التركيبية والوظيفية بواسطة المحازاة للتتابعات المتعددة من خلال موقع الخادم التالى:

http://www.bio.cam.ac.uk/cgi-bin/seqlogo/logo.cgi

ومثال ذلك التعرف على الملامح التركيبية لعدد ١٦ ثيوريدوكسينات (انزيمات موجودة فى كل الخلايا وتساهم فى مدى واسع من العمليات البيولوجية):



The structure of *E. coli* thioredoxin [2TRX] contains a central five-stranded β-sheet flanked on either side by α-helices. Residue numbers correspond to those in the multiple sequence alignment table. The N- and C-termini are also marked. Spheres indicate positions of the Cα atoms of every tenth residue. The reactive disulphide bridge between Cys32 and Cys35 appears between the numberals 30 and 60 .

Hiden Markov Models (HMMS) :

وهى نماذج تركيب حسابية لوصف طرز عائلات التتابعات المتماثلة. كما أنها تعتبر من الأدوات القوية للكشف عن القرابة البعيدة والتنبؤ بطرز طى البروتين.

مصادر على الانترنت Web Resources :

http://cse.ucsc.edu/research/compbio/sam.html

http://cse.ucsc.edu/research/compbio/HMM-apps/HMM-applications.html

http://www.sanger.ac.uk/Software/Pfam.

http://stash.mre-lmb.cam.ac.uk/SUPERFAMILY/

النشوء النوعى Phylogeny :

هناك العديد من الأمثلة للتطور فى البروتينات والجينوم. ويهدف هذا المجال العمل على التوصل الى العلاقات بين الأنواع والعشائر والأفراد والجينات وذلك باستخدام الخصائص الجزيئية.

أشجار القرابة الوراثية:

وتمثل الشجرة فى علم الحاسبات شكلا خاصا يتكون من نقط nodes يتـــم توصيلها ببعض عن طريق خطوط edges. والشجرة قد تكـون rooted أو unrooted. والشكل الأخير لا يظهر النسب بينما فى الشكل الأول ينحـــدر من كل نقطة أصلان ويسمى binary tree ويستخدم برنامج PERL لرســم الشجرة كما يلى:

```
 drawtree.prl - PERL

program to draw binary trees.

#!/usr/bin/perl
#drawtree.pl -- draws binary trees (root at top)
#usage:  echo '(A((BC)D)(EF))' | drawtree.pl > output.ps

print <; chop($tree); $_ = reverse($tree); s/[()]//g;


$x = 0; $y = 0;
while ($nd = chop()) {
    print "$x $y m ($nd) stringwidth pop -0.5 mul 0 rm ($nd)
show\n";
    $xx{$nd} = $x; $x+=20; $yy{$nd} = 10;
}


while ($tree =~ s/(?([A-Z])([A-Z]))?/1/) {
    print "n $xx{$1} $yy{$1} m\n";
    ($yy{$1} > $yy{$2}) || ($yy{$1} = $yy{$2}); $yy{$1} += 20;
    print "$xx{$1} $yy{$1} l $xx{$2} $yy{$1} l $xx{$2} $yy{$2} l
s\n";
    $xx{$1} = 0.5*($xx{$1} + $xx{$2});
}print "n $xx{$tree} $yy{$tree} m 0 20 rl s showpage\n";


$rx = 2*$x + 30; $yt = 2*$yy{$tree} + 146;
print "%%BoundingBox: 40 95 $rx $yt\n";
```

مصادرعلى الانترنت لأدوات الوراثة العرقية :

http://evolution.genetics.washington.edu/phylip/software.html