ABSTRACT

The rat Allele Characterization Project undertook to characterize over 4,500 genetic markers in 48 common strains of the laboratory rat. Each marker was genotyped in each strain and the 210,000 allele sizes recorded. The completion of this project has provided a new and interesting datasource with many potential uses for the rat researcher. We describe two novel tools which have been designed to use this dataset to assist in visualizing and analyzing rat genomic structure between the various rat strains.

ACP Haplotyper compares the allele sizes across any or all of the 48 strains with each other and also with an artificial 'ancestor strain' consisting of the most common allele sizes, presumably representing the oldest allele for a given marker. By ordering the alleles according to one of a number of genetic and physical maps, this analysis allows the researcher to visualize the regions of conservation between strains on a broad level. Similarly, this enables interesting inter-strain comparisons, potential evolutionary inferences and other practical benefits such as map placement evaluation.

Genome Scanner also compares the allele sizes though only between two specified strains. Designed as a practical tool to support the day to day work in the laboratory, Genome Scanner simplifies the often tedious task of selecting polymorphic markers spaced at regular intervals across the genome. Markers are arranged according to genetic or radiation hybrid map order and separated into bins of variable size. The user can then simply select a polymorphic marker from each bin and quickly assemble a list of markers spaced across a chromosome, or the entire genome, at the desired resolution. The final list is annotated with the expected allele sizes and marker and assay name information and is suitable for use directly in the lab. Both tools are webbased and utilize Perl/CGI with an Oracle 8i database handling the storage of the allele and marker data, they are available online at http://rgd.mcw.edu/

INTRODUCTION

The Rat Allele Characterization Project (ACP) measured the allele sizes of over 4,500 markers (predominantly simple sequence length polymorphisms) in 48 commonly-used strains of the laboratory rat. Previously, allele size information for newly mapped genetic markers has frequently been limited to a single mapping cross. Consequently, many investigators were forced to characterize all new genetic markers in order determine which markers were informative in a particular cross. The data generated by the ACP provides investigators with a means of quickly selecting informative markers in a vast number of mapping crosses resulting in a significant savings of both time and resources.

Whilst the existence of the ACP data itself was a major advance, the practical application of that data in the laboratory setting was somewhat laborious. For example, to select markers every 10cM across the whole genome for a total genome scan, a printed version of the map was consulted to select markers at an appropriate spacing, these were then checked against the ACP data in a spreadsheet for both strains to see if the marker was polymorphic. If the allele sizes were not polymorphic, another markers had to be selected - this was repeated until the whole genome was covered with polymorphic markers spaced at the required interval.

Following on from our initial work generating the ACP dataset, we now present a pair of web based tools designed to facilitate the use the ACP data in the lab. Genome Scanner uses the data in its conventional setting as a basis for genome scans and provides a far faster and user friendly way to select polymorphic markers for any given chromosome or across the whole genome. ACP Haplotyper extends the potential uses of the ACP data and enables it to be used in a novel context presenting a view of the conservation of regions of the genome between rat strains.

ACP DATA SUMMARY

For each of the 4,500+ markers characterized in the ACP project, if the marker was typed in all 48 strains, there will be 48 allele sizes available. Many of these will be the same as many of these sizes are conserved across some or all of the strains - this is the premise upon which ACP Haplotyper works.

As an example, D19Rat34 is typed in all 48 strains, but there are only 9 different allele sizes: 232 bp, 236, 240, 246, 250, 252, 254,2 56 and 257 (See the XML record in Figure 3). The most frequent size was 250 bp (21 out of 48 strains had this size) and this is refered to as the *Common size* for that marker. The *Common Strain* used in the ACP Haplotyper is a hypothetical strain constructed by taking the common allele size for each marker. One could imagine that the most common size might represent the oldest allele for that particular marker and hence the common strain might represent an archetypal rat strain from which the others have diverged over time.

ACP Haplotyper

Introduction

ACP Haplotyper uses the ACP data and presents it ordered in chromosomal order (based on the selected genetic or radiation hybrid maps) allowing the researcher to observe how blocks of markers are conserved across strains.

Usage:

The opening screen is shown in Figure 1. The application can display the data for all 48 strains or any combination selected from the drop down list. The Hypertensive and Diabetic groups have all 48 strains but are grouped to put known hypertensive or diabetic strains next to each other to facilitate their comparison. A Primary strain is selected, along with a desired color, a secondary strain and color can also be selected. In the resulting haplotype output if a marker in a strain matches the size of the primary strain it will be colored the Primary Strain color. If it is different to the primary but the same as the secondary then it will be colored according to the secondary color. If it matches neither strain's size it will be colored black.

Various other more advanced options are available to alter the final display. The LOD Threshold value can be used to screen out markers placed below a certain LOD threshold. The Basepair range enables one to determine how close two marker sizes have to be to be considered the same. If one set this to 2bp, any marker within +/- 2bp of the primary strain's size would be considered to be the same as the primary strain. One can order the strains in the final output according to one of a number of criteria: Identity, Homology and by strain name (essentially sorted alphabetically). The Identity score for a comparison between Strain A and Strain B is a simple measure of how many markers in Strain A have the same allele size as those in Strain B. The Identity score takes no account of marker order, LOD value of markers, etc. To attempt to get a better representation of interstrain similarity we calculated a Homology percentage which uses a sliding window approach to better account for conserved stretches of markers. The homology calculation also takes into account the LOD score of each marker to ensure that markers placed at a lower confidence level have less of an effect on the final result than those placed at a higher confidence. The window size and slide increment (in cM or cR depending on the map in use) can be set in the final two entry fields.

L/PostScript ACP Haplotyper - Netscap Edit View Go Communicator Help	e	
Parameter	Value	
Map to be used Only v7 genetic maps have LOD data	SHR x BN (v7)	
Chromosome [View Maps]	20 •	
Flanking Marker or Distance #1 Entername/distance to specify the top of the map region		
Flanking Marker or Distance #2 Entername/distance to specify the bottom of the map region		
Strains to Display Multiple selections allowed, select 1 or more strains you wish to view.	All 48 Strains Hypertensive Groups Diabetic Groups ACI AVN	
Primary Strain & color scheme	SHR	Blue •
Secondary Strain & color scheme	BN/SSN •	Red
Advanced Options		
Placement LOD Threshold Use to exclude markers placed at a LOD lower than the threshold value	О	
Basepair Range Allele sizes are considered equal when their values are within the entered basepair range value.	0	
Hide Size Data Hides allele size data, showing a visual haplotype only.	п	
Only show stats (no table) Only displays the analysis statistics, not the visual haplotype data.	п	
Order Strains relative to Primary Strain Orders strains according to % allele size identity or hostology, or alphabetically by strain name	By Percentage: C By Homology: © By Strain Name: C	
Interstrain Homology Calculations		
Calculate Homology Data	₽	
Window Size	10	
Slide Increment	5	
Output Format	PDF •	
	Display Haplotypes	

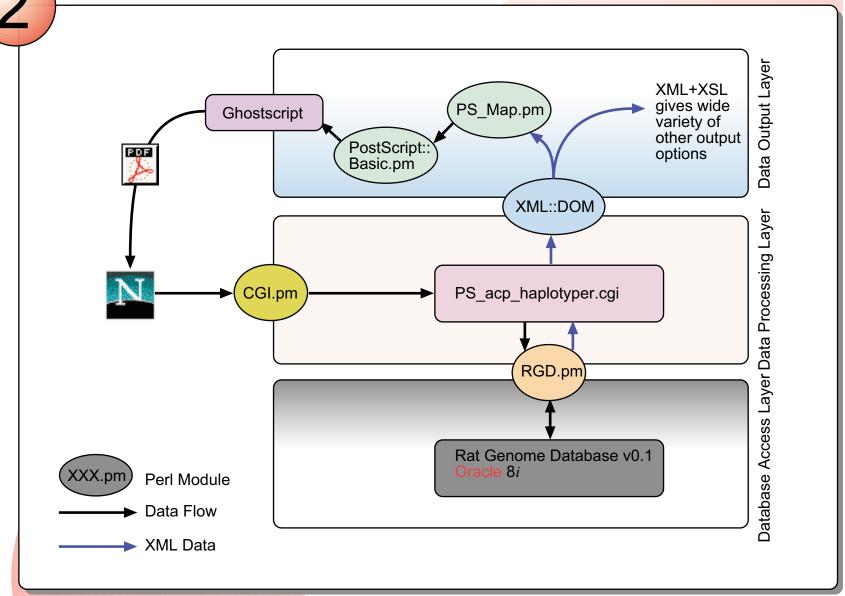
Usage (continued)

The output is currently returned as an Adobe Acrobat PDF document which can be opened by the freely available Acrobat Reader software or directly in a browser equiped with the acrobat reader plugin. The architecture of the application and how the PostScript and PDF documents are created are shown in **Figure 2**. At the present time only PDF output is available though through the use of the eXtensible Markup Language (XML) the data can be easily converted to other formats as required. A sample XML record is shown in **Figure 3**. For more information on XML see the URLs in the References section.

The PDF document format is very useful as it allows the creation of reasonably small multipage documents which can be printed out at high resolution. PDF also allows advanced features such as text searches so one can quickly find a marker on a map for example. The first page of the output is not shown and consists of a summary of the haplotyper's parameters. The second and third pages are matrices of inter-strain identity and homology percentages, respectively. These display how closely related two strains are over the chromosome or region of the chromosome selected. Two versions of the matrix are created, one containing the numerical values and another color coded to create a chromogram of the relationships - a homology chromogram is shown in **Figure 4**.

The Haplotype data for Chromosome 20 is shown in **Figure 5** and represents the main output of the application. On the left hand side the markers are listed in map order along with their LOD scores and map distances. On the right is the haplotype data for the 48 strains. In this example the primary strain was SHR, the secondary strain BN/SSN. Strains with allele sizes matching those of the SHR are colored blue, those matching the BN/SSN are shown in red. If the SHR and BN sizes are the same, they are colored blue. Black squares represent allele sizes different from both SHR and BN/SSN, white squares indicate that there is no data for that marker in that particular strain. The strains are ordered according to their homology to the Primary strain, SHR. One would expect to see more blue on the left where the strains are more similar and less blue and more red to the right as they differ more.

One can quickly see that there are blocks of conserved alleles scattered across the chromosome and areas where SHR alleles are quite rare in the general population and areas where they are common. This data combined with QTL information could be useful in narrowing down potential regions of a chromosome where a candidate gene might lie. If the SHR was the affected strain in a SHRxBN/SSN cross one might expect to find a QTL in areas where the SHR and BN/SSN alleles differed.



3

Sample XML output

from ACP Haplotyper.

A full XML record would contain all

the markers on chromosome 19 for

thes selected map.

<rgd_map_raw_data chromosome="19"> <marker>

> <name>d19rat34</name> <abs_distance>0.0</abs_distance> <f_or_p>P</f_or_p>

<lod>.2</lod>

<allele_size strain="ACI">256</allele_size>
<allele_size strain="AVN">257</allele_size>
<allele_size strain="BB(DP)">252</allele_size>

<allele_size strain="BB(DR)">252</allele_size>
<allele_size strain="BC/CPBU">252</allele_size>

<allele size strain="BDIX">240</allele size>

<allele_size strain="BDVII">252</allele_size>

<allele_size strain="BN/CUBLX">252</allele_size>

<allele_size strain="BN/SSN">252</allele_size>
<allele_size strain="BP">252</allele_size>

<allele_size strain="BUF">252 \frainele_size>
<allele_size strain="BUF">254 \frac{1}{2} \f

<allele size strain="COP">250</allele size>

<allele_size strain="DA">256</allele_size>

<allele_size strain="DRY">250</allele_size>

<allele_size strain="F344">250</allele_size>
<allele_size strain="FHH">250</allele_size>

<allele_size strain="FHH">250</allele_size>

<allele_size strain="GK">240</allele_size>

<allele_size strain="IS/KYO">240</allele_size>

<allele_size strain="LE">246</allele_size>

<allele_size strain="LEW">250</allele_size>

<allele_size strain="LH">252</allele_size>

<allele_size strain="LN">250</allele_size>

<allele_size strain="LOU/C">240</allele_size>

<allele_size strain="M520">250</allele_size>

<allele_size strain="MHS">250</allele_size>

<allele_size strain="MNR">240</allele_size>

<allele_size strain="MNRA">250</allele_size>
<allele_size strain="MNS">250</allele_size>

callele size strain "MD"> 250 validio size

<allele_size strain="MR">250</allele_size>

<allele_size strain="NEDH">250</allele_size>

<allele_size strain="NP">250</allele_size>

<allele_size strain="ODU">240</allele_size>

<allele_size strain="OKA">240</allele_size>

<allele_size strain="OM">232</allele_size>

<allele size strain="P">250</allele size>

<allele size strain="PVG">240</allele size>

<allele size strain="SD">250</allele size>

<allele_size strain="SHR">240</allele_size>

<allele_size strain="SHRSP">240</allele_size>

<allele_size strain="SR/JR">250</allele_size>

<allele size strain="SS/JR">250</allele size>

<allele size strain="WAG">250</allele size>

<allele_size strain="WF">250</allele_size>

<allele size strain="WIST">236</allele size>

<allele size strain="WKY">240</allele size>

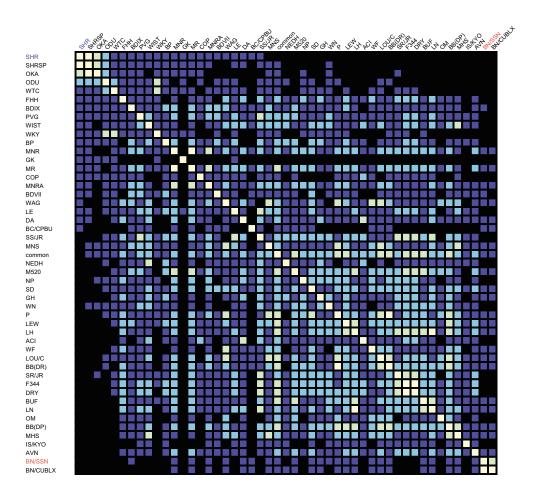
<allele size strain="WN">250</allele size>

<allele_size strain="WTC">240</allele_size>

<common_allele_data size="250" frequency="21"
 total allele num="48" num unique_alleles="9"/>

<allele size strain="common">250</allele size></marker>

Homology chromogram over displayed region [from 0.0 to the end]



ACP Haplotyper analysis of Chr: 20 at Thu Oct 14 14:23:45 1999

Primary Strain: SHR Secondary Strain: BN/SSN

LOD Threshold: 0, BP Range: 0

Strain ID % Scale:



Discussion

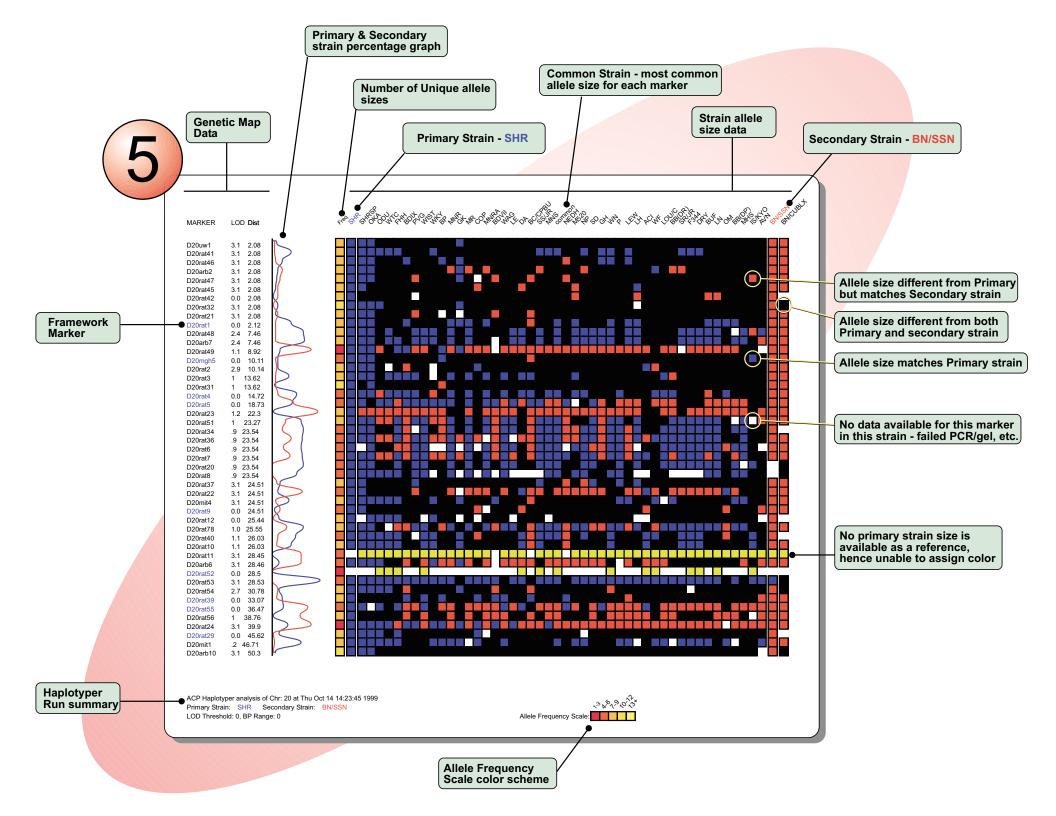
The Haplotype data itself is proving to be extremely useful when combined with experimentally determined QTL locations. Whilst it is unlikely that one will be able to predict the potential location of a candidate gene simply by looking at how two strains differ in regions across the genome, it should certainly be possible to use this data to narrow down target regions for futher study. We are currently evaluating this approach in the lab using QTLs we ourselves are working on and also by examining published QTL and gene data to more fully determine the usefulness of this tool.

Informatics Discussion

The application is written using Perl and runs on a Solaris server which also houses our Oracle 8i database. A number of things are notable about this application, the first being its use of XML, the second the generation of PostScript/PDF documents on the fly, an approach that has not previously been exploited.

XML (eXtensible Markup Language) is touted as the next greatest thing to hit the web, separating web content from how this content is displayed. By marking up your data with XML you accurately define the information present in the document thereby making it easy for other applications to use your data and providing a way to escape the multitude of format incompatibilities that currently plague computing. XML is being used in ACP Haplotyper as a test to see how easy it is to use and how one might be able to use XML more comprehensively in the future. At the present time there are some technological issues that have arisen - the XML::DOM perl module that allows perl scripts access to the structure of the XML document is very slow creating a large time delay from the submission of the web form to the appearance of the PDF file. It is hoped that other approaches may circumvent this problem (using streaming XML rather than the DOM) and as the XML::* modules continue to develop they are sure to become more efficient.

The generation of PDF documents on the fly is made possible through the use of the PostScript::Basic module which was written by the presenting author as a replacement for the GD.pm module used to create GIF images via perl. PostScript is infinitely better for generating high quality images and when combined with PDF provides a way to generate high quality text and images which are accessable to virtually all users. Through the use of the PostScript::Basic module, PostScript code is easily generated (without requiring any knowledge of the PostScript language itself) and is then passed to Ghostscript which converts the PostScript to PDF format for on-line presentation. The documentation for PostScript::Basic.pm is available on request and can be found along with a downloadable archive of the module at the URL listed in the reference section.



Genome Scanner

Introduction

Genome Scanner displays polymorphic markers for a cross between two selected rat strains on a particular chromosome. Its primary use is to enable researchers to select polymorphic markers for a chromosome or genome-wide scan.

Usage:

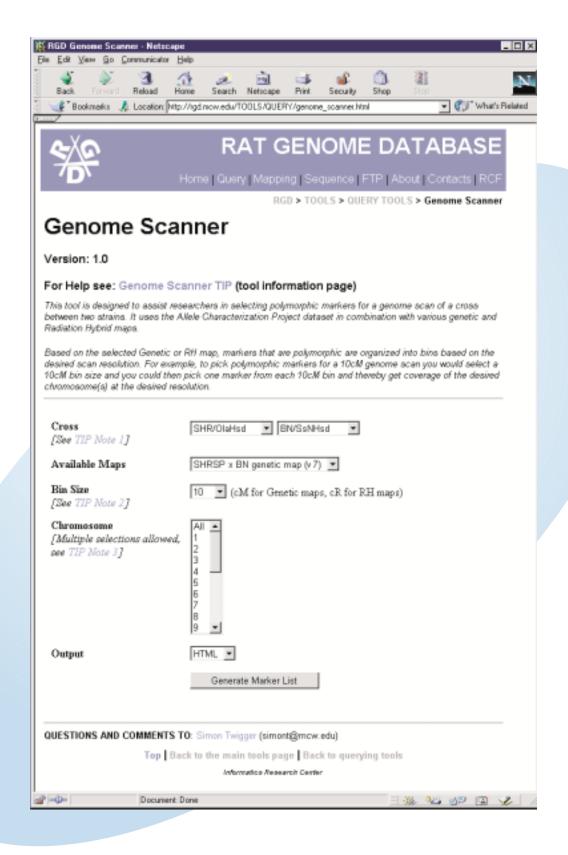
The researcher selects (see Figure 1):

- The two strains to be crossed out of the 48 available strains in the ACP dataset
- The genetic or radiation hybrid map to be used to order the markers
- The chromosome(s) to be studied
- The desired Bin size (10cM is normal when using genetic maps,100-150cR would be approximately equivalent when using one of the RH maps)
- The output format (normally this will be HTML, csv is an alternative option)

Genome scanner will return a (potentially large) HTML table listing all the polymorphic markers between the two selected strains on the selected chromosome, **Figure 2**. In this case the cross was between SHR x BUF on chromosome 20, using the SHRSPxBN genetic map as a reference. Marker names and assay names are shown along with genetic map location and LOD scores. The assay names are color coded according to the LOD score of the marker on the map. Allele sizes for both selected strains are shown along with the difference in size between the two strains. A checkbox is provided for each marker so the researcher can assemble a run list of markers from the complete list of polymorphic markers.

Following marker selection, clicking the Generate Runlist button at the bottom of the page will bring up a Runlist suitable for printing out. The runlist includes a summary of the strains selected and other parameters plus the selected markers which optionally can be sorted according to Assay Name to make primer retrieval a simpler task, **Figure 3**.

Available Maps: Current and past WI/MIT Genetic Maps (SHRxBN, FHHxACI) plus Radiation Hybrid framework and Placement maps (Steen, etal 1999)







Siny Mitting Sepanda FTP About Contacts

RGB > TOOLS > OILTRY TOOLS > Gaverne Scarner

Genome Scanner

Note: Markers with an allele size difference between 4:25 bp are colored with a green background to facilitate picking markers which will have convenient allele separation when run on polyacrylamide gels.

To generate a run list, check each marker you wish to add to the run list, enter in any desired comments in the fields provided at the bottom of this page and click the "Generate Run List" button to produce a run list of only those markers you selected. This list can be ordered by the Assay Name to facilitate primer selection.

Marker Name Color Scheme:

- Framework marker on genetic map (LOD > 3)
 Unique placement marker on genetic map (LOD > 2)
 Non-unique placement marker on genetic map (LOD < 2)

Chromosome: 20 |

Marker	Assay Name	Distance	LOD	SHRIOLAHSD Size	BN/SSNHSD Size	Difference	Add to Runlist
Chromoso	me 20 (Bin	1)(=)					
D20Rat46	R0240-B10	2.08	3.1	150	144	.0	E:
D20Rat47	R0234-A10	2.08	3.1	133	118	15	
D20Uw1	R1103-C12	2.08	3.1	130	144	14	Г
D20Arb2	R0270-D07	2.08	3.1	180	160	20	Ε:
D20Rat41	R0058-D07	2.08	3.1	196	206	10	Г.
D20Rat42	R0079-B08	2.08	0.0	132	144	12	C:
D20Rat45	R0233-B06	2.08	3.1	122	114	8	г
D20Rat21	R0057-A12	2.08	3.1	182	162	20	г
D20Rat32	R0071-B03	2.08	3.1	157	171	14	C.
D20Rat1	R0013-C05	2.12	0.0	228	237	9	г
D20Arb7	R0270-D11	7.4699	24	165	169	4	D:

Chromoso	ome 20 (Bin 4)					
D20Rat54	R0056-D04 30.7899	27	198	208	10	г
D20Rat39	R0078-B06 33 0799	0.0	220	222	2	E
D20Rat55	R0089-D11 36 48	0.0	220	195	25	г
D20Rat56	R0236-B03 38 7699	1	141	143	2	Г
D20R#t24	R0081-B08 39.91	3.1	122	124	2	C
Chromoso	ome 20 (Bin 5)					
D20Rat29	R0062-B10 45.62	0.0	231	227	4	Г
D20Mit1	R1103-D05-46.71	2	125	133	8	г
Chromoso	ome 20 (Bin 6)					
D20Arb10	R0270-D06-50:3099	3.1	281	269	12	II

Selected Map SHRSP(RIV x BN genetic maps (v 7) SHR/OLAHSO » BN/SSNHSD polymorphic markets, 10 cM/cl Run List Comments:

Wet Apr 12 11 56 23 2000

Sort Run List by Assay Name

Run List Date:

Generate Runlist

Return to Genume Scarner Questions and Comments to Simon Teigger

Dogwest Days









RAT GENOME DATABASE

Home | Query | Mapping | Sequence | FTP | About | Contacts | RCF

RGD > TOOLS > QUERY TOOLS > RGD Genome Scanner

RGD Genome Scanner

Version: 1

For Help see: Main Help Page

Selected Map: SHRSP/RIV x BN genetic maps (v 7)
Cross: SHR/OLAHSD x BN/SSNHSD

Date: Wed Apr 12 11:56:23 2000

Comments: SHR/OLAHSD x BN/SSNHSD polymorphic markers, 10 cM/cR intervals, Chr's: 20

Chromosome	Marker Name	Assay Name	Distance	LOD	SHR/OLAHSD size	BN/SSNHSD size	Difference	Run List Comments
20	D20Arb2	R0270-D07	2.08	3.1	180	160	20	
20	D20Rat32	R0071-B03	2.08	3.1	157	171	14	
20	D20Rat54	R0056-D04	30.7899	2.7	198	208	10	
20	D20Mit1	R1103-D05	46.71	.2	125	133	8	
20	D20Arb10	R0270-D06	50.3099	3.1	281	269	12	

There were 5 markers selected for this runlist.

Return to Genome Scanner Author: Simon Twigger



Document: Done



References & URLs

Software: contact Simon Twigger - simont@mcw.edu

Genome Scanner: http://rgd.mcw.edu/TOOLS/QUERY/genome_scanner.html
ACP Haplotyper: contact author (simont@mcw.edu) for current working URL

PostScript::Basic.pm http://www.lgr.mcw.edu/LGR/research/tools/

Projects and related information

Genetic & RH Maps: http://www.lgr.mcw.edu/LGR/research/rhp/steen 1999.html

Steen etal. Genome Research, May 1999

Allele Characterization Project: http://www.lgr.mcw.edu/LGR/research/lgr_acp.html

Rat Genome Database: http://rgd.mcw.edu/

XML

Perl & general references http://www.xml.com/pub/Perl XML & BioPerl http://bio.perl.org/Projects/XML/

Bio-XDK Links http://www.vsms.nottingham.ac.uk/biodom/bio-xdk/docs/links.html