DEVELOPING A WEB ACCESSIBLE INTEGRATED DATABASE AND

VISUALIZATION TOOL FOR BOVINE QUANTITATIVE TRAIT LOCI

A Thesis

by

PAVANA POLINENI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2004

Major Subject: Computer Science

DEVELOPING A WEB ACCESSIBLE INTEGRATED DATABASE AND

VISUALIZATION TOOL FOR BOVINE QUANTITATIVE TRAIT LOCI

A Thesis

by

PAVANA POLINENI

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Approved as to style and content by:

Richard Furuta	David Adelson
(Co-Chair of Committee)	(Co-Chair of Committee)
Du Li	Valerie E. Taylor
(Member)	(Head of Department)

May 2004

Major Subject: Computer Science

ABSTRACT

Developing a Web Accessible Integrated Database and Visualization Tool for Bovine
Quantitative Trait Loci. (May 2004)

Pavana Polineni, B.Tech., Sri Venkateswara University

Chair of Advisory Committee: Dr. Richard Furuta

A quantitative trait locus (QTL) is the location of a gene that affects a trait that is measured
on a quantitative (linear) scale. Many important agricultural traits such as weight gain, milk
fat content and intramuscular fat in cattle are quantitative traits. There is a need to integrate
genomic sequence data with QTL data and to develop an analytical tool to visualize the data.
Without integration, application of this data to agricultural enterprise productivity will be
slow and inefficient.

My thesis presents a web-accessible tool called the Bovine QTL Viewer developed to solve
this problem. It consists of an integrated database of bovine QTL and the QTL viewer to
view the QTL and their relative chromosomal position. This tool generates dynamic and
interactive images and supports research in the field of genomics. For this tool, the data is
modeled and the QTL viewer is developed based on the requirements and feedback of experts
in the field of bovine genomics.

ACKNOWLEDGEMENTS

I would like to take this opportunity to offer my heartfelt thanks to a number of people without whose help and support this thesis would never have been possible. I would like to thank my advisor, Dr. Richard Furuta. I would like to express great gratitude to Dr. David Adelson for guiding me throughout. I would also like to thank Dr. Du Li.

My parents have provided me with countless opportunities for which I am eternally grateful. Nothing I can say can do justice to how I feel about their support. I must thank my husband without whose encouragement I might never have done this project. Thank you very much, Phani.

I would like to thank Raghu Chaitanya Akkapeddi, Suresh Raja Xavier and Prathyusha Aragonda for their tremendous help and support. I also want to thank all my fellow students and researchers in the department.

I would like to thank Ratmap and USDA for providing the data model and bovine marker data respectively. I would also like to thank Dr. Clare Gill, Ms. Kelli Kay and Dr. Loren Skow and his group for providing their valuable feedback on the tool. Last, but by no means least, it gives me immense pleasure to thank my dearest roommates and friends.

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

A quantitative trait locus (QTL) refers to the location of a gene that affects a trait that is measured on a quantitative (linear) scale. Quantitative traits are typically affected by more than one gene, and also by the environment. Many important agricultural traits such as weight gain, milk fat content and intramuscular fat in cattle are quantitative traits. While, much of the pioneering research into QTL was performed by agricultural geneticists, full identification of the genes involved will require integration of genomic sequence data with QTL data. Without integration, application of these data to agricultural enterprise productivity will be slow and inefficient. In order to facilitate this overall integration, there is a need for an integrated database of Quantitative Trait Loci of cattle and a visualization tool for this database.

This thesis presents the integrated database and visualization tool called Bovine QTL Viewer developed for bovine Quantitative Trait Loci. The data was modeled and the viewer was developed based on the requirements and feedback of experts in the field of bovine genomics. The viewer generates dynamic and interactive images and it is web-accessible.

In this thesis, bovine QTL are introduced and the state of the art is reviewed. The current problem with Bovine QTL data is explained and a solution for the problem, Bovine QTL Viewer, is presented. The design, development and important features of this tool are explained, along with the results of the evaluation done at various stages of the research.

_____

This thesis follows the style and format of *Genome Research*.

2. BOVINE QTL

Every living organism is made up of cells. In every cell, the nucleus has tightly coiled threads of deoxyribonucleic acid (DNA) and associated protein molecules, organized into structures called chromosomes. For each organism, the components of these slender threads encode all the information necessary for building and maintaining life, from simple bacteria to remarkably complex human beings. DNA consists of four different bases: adenine (A), thymine (T), cytosine (C), and guanine (G). The particular order of the bases are arranged in is called the DNA sequence. The sequence specifies the exact genetic instructions required to create a particular organism with its own unique traits. The unit of heredity which contains hereditary information encoded in the form of DNA and is located at a specific position on a chromosome in a cell's nucleus is called a gene. Each individual has a unique sequence of genes, or genetic code.

The genotype of an organism represents the genetic potential or the set of genetic instructions responsible for a particular trait (Wattiaux, M.A. 2003). The genotype can also describe the entire set of genetic instruction inherited by an individual. The phenotype is the value taken by a trait, it is what can be observed or measured. For example, the phenotype may be an individual cow's milk production, the percentage of fat in the milk or a classification score for conformation. The genotype is essentially a fixed characteristic of the organism; it remains constant throughout life and is unchanged by environmental factors. When only one or a few genes are responsible for a trait, the phenotype usually remains unchanged throughout life (e.g., blood type). In this case, the phenotype gives a good indication of the genetic composition of an individual. Such kind of traits are called Qualitative traits. Qualitative traits tend to fall into discrete categories. Usually just one or a few genes have a major effect on qualitative traits and the environment usually has a minor role in influencing the category into which the animal falls. In this case, the phenotype of the animal reflects its genotype. Examples of qualitative traits in dairy cattle are:

• Hair color

• Hereditary defects such as dwarfism

• Presence or absence of horns

• Blood type

However, for some traits, the phenotype changes continually throughout the life of the individual *in response to environmental factors*. In this case, the phenotype is not a reliable indicator of the genotype. This usually occurs when many genes are involved in the expression of a trait such as milk yield. These traits are called Quantitative traits and they differ from qualitative traits in two important ways:

1) They are influenced by many of genes

2) The phenotypic expression is more strongly modulated by the environment than is true for qualitative traits.

Many of the economically important traits in dairy cattle are quantitative traits:

• Milk yield

• Milk composition

• Efficiency of feed conversion

• Disease resistance

The same holds for beef cattle:

• Carcass weight

• Intramuscular fat (marbling)

• Warner-Bratzler shear force (tenderness)

• Feed efficiency

The combined influence of many genes and the effects of the environment on quantitative traits make it much more difficult to determine the genotype accurately compared to qualitative traits. Sometimes the animal's phenotype tells us very little about its genotype. For example, a single lactation record only tells a portion of the information about the genetic merit of a cow for milk yield.

While Quantitative traits result from the interactions of multiple genes of the environment, some of the genes may contribute more to the phenotype than others. These genes are said to have major effects on the phenotype. The detection of these genes of major effect results in the identification of quantitative trait loci (QTL).

A quantitative trait locus (QTL) describes a chromosomal region containing one or more genes involved in the expression of a polygenic trait (Fischer et al. 2003). QTL are identified by the association between a set of genetic markers and the segregation of a significant

portion of the variance of the studied trait. The strength of the association is measured as a linkage score. Significant peak values reveal the most likely position of the QTL, and its limits are set with the help of confidence intervals (Fischer et al. 2003).

3. LITERATURE REVIEW

Visualization tools for genomics ideally should provide universal access to the most current data in an interactive and intuitive graphical user interface. In order to achieve this goal, several graphical software tools have been developed in the last few years. These tools for comparative genome analysis provide us with a powerful view of the differences and similarities between genomes.

One of the first public graphical software tools for comparative genome analysis was published in 1996. It was a series of graphical tools (Dicks 1997), centered on the Oxford Grid (Edwards 1991). The graphics for them were written using the C programming language, and they were integrated with the ACEDB database management system (Durbin and Thierry Mieg 1991), depending directly on data held within the ACEDB database. The ability to link an interactive graphical tool to a data source such as a database and to access that tool within a WWW browser has opened up data analysis of biological information to a wider audience.

Some genome browsers are map based, some are sequence based. Map based browsers can be converted to sequence based browsers by anchoring marker reference ID to a sequence co-ordinate in a sequence database. The Genome Map Viewer (Davenport et al. 1999) is a Java graphical tool recently developed by UK CropNet. The Genome Map Viewer can display elements such as loci and it has been used extensively for crop plant species. The GridMap (Priestly et al. 1999) is a generic grid-drawing Java tool that is particularly useful for the display of comparative genome data. Another recently developed tool for comparing linear maps across different species is the Comparative Genome Mapping Tool (CGMT) (Pecherer et al. 1999) from the National Center for Genome Research (NCGR) in Santa Fe, New Mexico.

Most map viewers were created before genome sequence became available. The National Center for Biotechnology Information (**NCBI**) (National Center for Biotechnology Information 2004) provides an integrated approach to the use of gene and protein sequence

information. NCBI has a map based viewer linked to genome database. The NCBI Map Viewer for cow presents the Meat Animal Research Center (MARC) map, (U.S Meat Animal Research Center 2003) in which each of the markers has been integrated into NCBI's UniSTS resource which provides regular updates of positioning of these markers on sequences available in DDBJ/EMBL/GenBank. It has the biggest, best and most comprehensive dataset. Human genome database, now almost defunct, was the map viewer at one time. Its functions are now present at NCBI.

Annotations i.e knowledge of how that sequence is associated with phenotype or biological phenomena, greatly enhance the biological value of these genome sequences. For this purpose, Genotator (Harris 1997) was developed by the Human Genome Informatics Group. It is a workbench for automated sequence annotation and annotation browsing. Bailey et al. in their paper, 'GAIA: framework annotation of genomic sequence' (Bailey et al. 1998) describe a process of high-throughput, reliable annotation, called framework annotation, which is designed to provide a foundation for initial biologic characterization of previously unexamined sequence. To examine this concept in practice, they have constructed Genome Annotation and Information Analysis (GAIA), a prototype software architecture that implements several elements important for framework annotation. The maps need to be linked to sequence which in turn need to be linked to annotations so that it is clear that identifying a region of the map (genome) tells the user something about the biology, providing sufficient annotation present in the region of interest.

In order to take advantage of available public human genome sequence data and associated annotations, Loraine and Helt (Loraine and Helt 2002) describe visualization techniques for presenting human genomic sequence data and annotations in an interactive, graphical format. These techniques include: one-dimensional, semantic zooming to show sequence data alongside gene structures; color-coding exons to indicate frame of translation; adjustable, moveable tiers to permit easier inspection of a genomic scene; and display of protein annotations alongside gene structures to show how alternative splicing impacts protein structure and function. These techniques are illustrated using examples from two genome browser applications: the Neomorphic GeneViewer annotation tool and ProtAnnot, a prototype viewer which shows protein annotations in the context of genomic sequence.

Similarly a Human Genome Browser (Kent et al. 2002) was developed at UCSC. This browser displays assembly contigs and gaps, mRNA and expressed sequence tag alignments, multiple gene predictions, cross-species homologies, single nucleotide polymorphisms, sequence-tagged sites, radiation hybrid data, transposon repeats, and more as a stack of coregistered tracks. Functional annotations are provided by links to other biological databases such as Online Mendelian Inheritance in Man (OMIM) (National Center for Biotechnology Information 2003) and GeneCards (Crown Human Genome Center and Weizmann Institute of Science 2003)

Helt et al. (Helt et al. 1998) have developed a prototype genome browser applet, BioViews using Java that incorporates a three-level graphical view of genomic data: a physical map, an annotated sequence map, and a DNA sequence display. This genome browser is built on top of extensible, reusable graphic components specifically designed for bioinformatics.

Though the presentation of genomic data in a perspicuous visual format is critical for its rapid interpretation and validation, relatively few public database developers have the resources to implement sophisticated front-end user interfaces themselves. Accordingly, these developers would benefit from a reusable toolkit of user interface and data visualization components. For this purpose, Fischer et al. (Fischer et al. 1999) have designed the bioWidget toolkit as a set of JavaBean components.As a proof of concept, they have used the bioWidget toolkit to create three extensible applications: AnnotView, BlastView and AlignView. Similarly, another project by name The Generic Model Organism System Database Project (GMOD) (Stein et al. 2002), seeks to develop reusable software components for model organism genome databases. In the paper entitled the Generic Genome Browser (GBrowse) (Stein et al. 2002), a Web-based application for displaying genomic annotations and other features is described.

The Rat Genome Database (RatMap) (The RatMap Group 2003), available through the Department for Cell and Molecular Biology, Göteborg University, Sweden, offers information on rat gene nomenclature, chromosomal positions for genes (including predicted positions), gene function, literature references and tool for QTL discovery in rats. Gene records in the database may be recalled with a locus query or by browsing lists, and some

data sets may be downloaded. Based on integrated linkage data, all QTL present in RatMap are aligned in a single map. The Rat QTL-tool makes it possible to survey all QTL and their relative chromosomal position. The appearance of different types of QTL can be selected/deselected according to the users own preferences. By selecting a chromosomal interval, all QTL within that limited region are retrieved. For each individual QTL, detailed information is presented on genes and markers found in the interval, along with flanking markers, peak LOD score etc. The Rat QTL tool is fully integrated with the data content of the RatMap database, ensuring an up-to-date presentation.

Along with the comparisons and annotations, there is also an acute need for reliable up-to-date genomic databases integrating both functional and structural information. Frishman et al. in their paper, 'Functional and structural genomics using PEDANT' (Frishman et al. 2001) describe the PEDANT software system for high-throughput analysis of large biological sequence sets and the genome analysis server associated with it.

4. PROBLEM STATEMENT

Many important agricultural traits such as weight gain, milk fat content and intramuscular fat (marbling) in cattle are quantitative traits. While significant information regarding the mode of inheritance of these traits is available, most of this information is not integrated into a genomic context. As large amounts of genomic sequence data become available (estimated completion time of the bovine genome sequence is July 2005), they require livestock genome researchers to integrate sequence data not only with existing gene maps, but more importantly with QTL and phenotype data. Without integration, application of these data to agricultural enterprise productivity will remain slow and inefficient. In order to facilitate this overall integration, there is a requirement for an integrated database of Quantitative Trait Loci of cattle and an analytical tool for this database such as a visualization component.

In my thesis, I present the creation of a web accessible, integrated database of bovine Quantitative Trait Loci with the possibility of linking it to animal genome data as it becomes available. I also present the development of a QTL viewer to show the contents of the database in a manner this is useful to both novices and experts in genomics.

5. BOVINE QTL VIEWER

Bovine QTL Viewer was developed for the integration of QTL data available for cattle. The tool consists of an integrated database of bovine QTL and the QTL viewer to display QTL and their chromosomal position. Fig. 1 shows the structure of Bovine QTL Viewer.
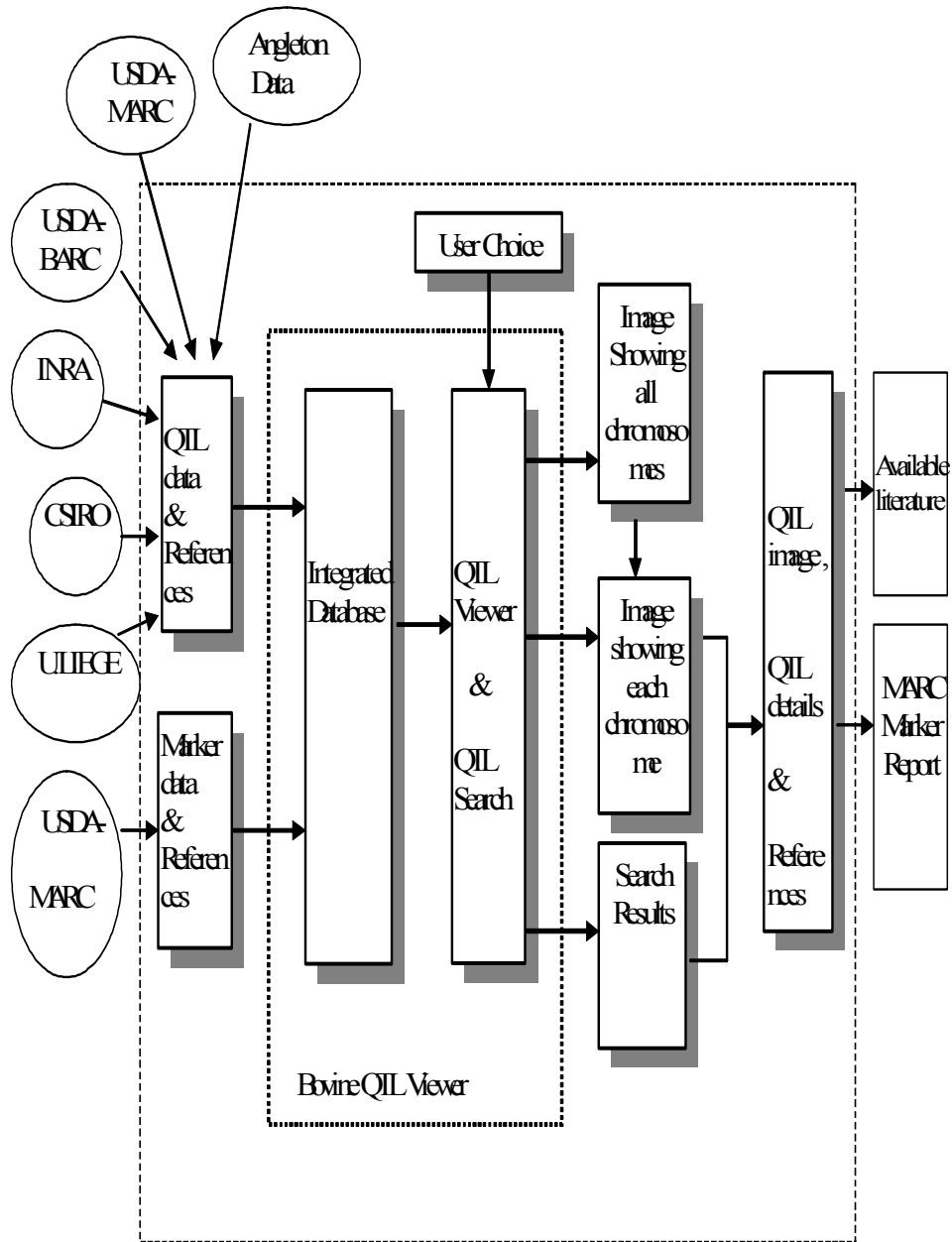
Figure 1: Bovine QTL Viewer

6. INTEGRATED DATABASE

Bovine QTL Viewer was developed for the integration of QTL data available for cattle. The tool consists of an integrated database of bovine QTL and the QTL viewer to display QTL and their chromosomal position.

6.1 DATA MODELING

The QTL data and related information for bovine QTL are gathered from the published work and from existing databases like Angleton, USDA-BARC, INRA, CSIRO, U.LIEGE. An integrated database schema is designed and the database is populated with the gathered data. As shown in Fig. 1, the database mainly consists of QTL data and Marker data.

6.1.1 E-R DIAGRAM

The overall logical structure of the integrated bovine QTL database is represented graphically with the Entity-Relationship diagram shown in the Fig. 2. The diagram consists of these entities:

- QTL information, including all the details of each QTL
- Marker information, including all the details of each marker
- References, including the complete list of references
- Trait information, including the details of the traits of the cattle
- Category information, which gives details about the categories into which traits are divided
- Chromosome information, which has the length of each of chromosome measured in linkage units (Morgans)

The database is modeled such that each trait can fall under multiple categories. In addition, each QTL and marker record can have more than one reference. Finally, the QTL information and Marker information are not linked directly. These data are linked dynamically by the QTL Viewer based on the position of the marker and the position of the QTL on the

chromosome. This will provide flexibility when the database is updated, allowing Marker information and QTL information to be updated independently of each other.

6.2 UPDATING QTL DATA

All the QTL data that is available on the web and has been published in research papers has been collected and stored in the database which is updated frequently. Updating is semi-automatic; the administrator is prompted to update the database, and if he/she accepts, the database is updated automatically with the newly entered data.
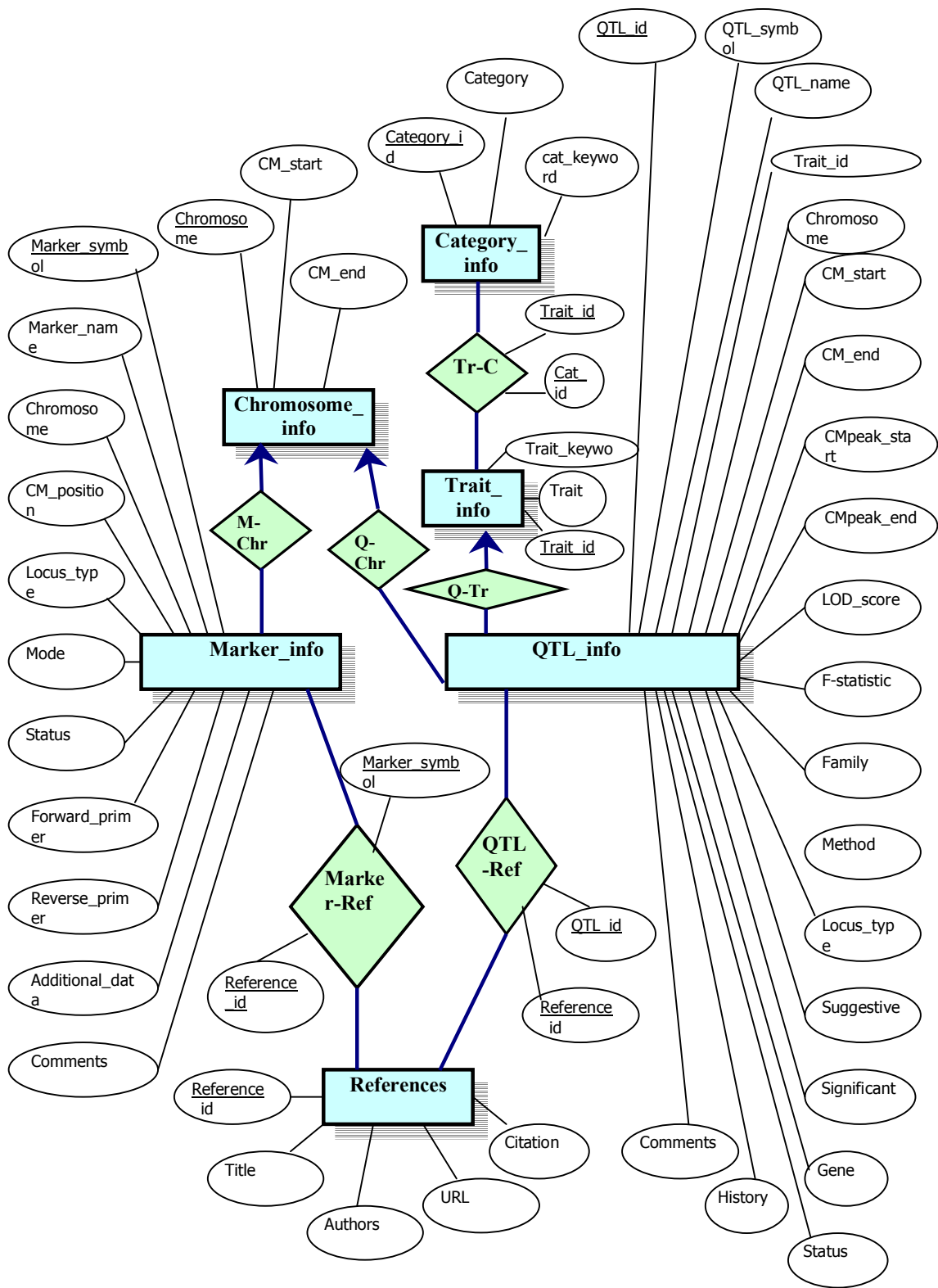
Figure 2: Entity-Relationship diagram

7. VISUALIZATION TOOL

The visualization component, "QTL viewer", displays all QTL and their chromosomal positions. The images displayed are dynamic and interactive.

7.1 DYNAMIC MENUS

7.1.1 BASIC MENU

"Main Menu" is the basic menu that is displayed as soon as the user logs in. It displays the basic menu of trait categories and chromosomes that can be selected, as shown in Fig. 3. The trait categories are retrieved dynamically from the database and displayed. A maximum of 10 categories can be displayed in the image restricting the number of categories that can be selected to 10 (see Color blindness compatibility p.30). If the user attempts to select more than 10 categories, only the first 10 selections are displayed. The QTL search menu is also displayed alongside the basic menu on the same screen. At least one chromosome must be selected for the image to be displayed.

7.1.2 ADVANCED MENU

"Advanced Menu" is a more detailed menu as shown in Fig. 4. As with the "Main Menu", traits present in the database are retrieved dynamically and displayed in the menu. This menu lets the user choose specific traits, LOD score, F-statistic, significance level and chromosomes. Only those QTL which satisfy the given criteria are displayed in the image. A maximum of 10 traits can be selected. If the user attempts to select more than 10 traits, only the first 10 selections are displayed. If no selection of LOD score, F-statistic or significance level is made, then QTL associated with all LOD scores, all F-statistic values and all significance levels are displayed respectively. At least one chromosome must be selected for the image to be displayed.
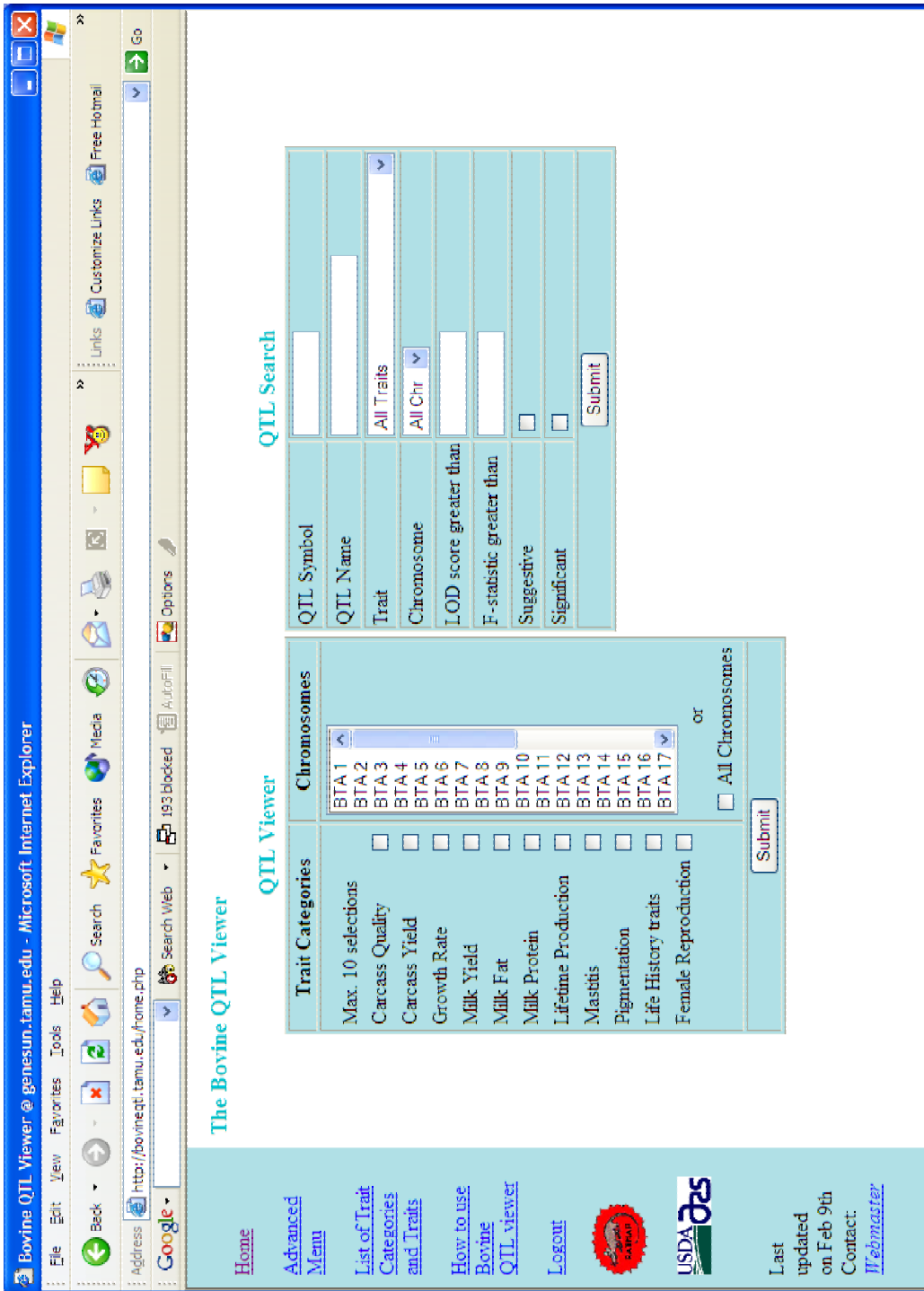
Figure 3: Main Menu

Figure 4: Advanced Menu

7.2 INTERACTIVE IMAGES

7.2.1 IMAGE SHOWING ALL CHROMOSOMES

If all chromosomes are chosen in the basic menu or the advanced menu, an image showing all cattle chromosomes is displayed (Fig. 5 and Fig. 6). In this image, QTL on all chromosomes that satisfy the given criteria are displayed. A separate color is used for each category or trait of QTL. Overlapping QTL are not shown separately, so in order to view them separately, the user needs to go the individual chromosome view by clicking on that particular chromosome.

The image showing all the chromosomes is interactive in two ways. Firstly, any of the chromosomes can be clicked on to go to the image showing that chromosome alone and QTL in detail with overlapping QTL shown separately (Fig. 8). Secondly, any of the labels displayed at the bottom right hand corner of the image can be clicked to get QTL of only that trait or category (Fig. 7). Before clicking on a feature in a zoomed image, the image has to be restored to normal size.

7.2.2 IMAGE SHOWING ONE CHROMOSOME

An image showing a single chromosome is displayed as shown in Fig. 8. This image is shown if a particular chromosome is chosen in the basic menu or the advanced menu or clicked on, in the image displaying all chromosomes. When the comprehensive image is clicked on, to go to the individual chromosome view, all the criteria selected by the user in the previous stage are automatically passed to the next stage (Fig. 9). In this image, the QTL satisfying the given criteria are displayed in detail along with their peak positions. Any overlapping QTL are displayed separately. A separate color is used for each category or trait of QTL and name of the category or trait is displayed beside each QTL.

The image showing each chromosome is interactive in two ways. Firstly, any QTL can be clicked on, to go to the image showing that specific QTL and get all the information about that QTL (Fig. 11). Secondly, any of the labels displayed at bottom of the image can be

clicked on, to get QTL of only that trait or category as shown in Fig. 10. Before clicking on a feature in a zoomed image, the image has to be restored to normal size.

7.2.3 IMAGE SHOWING QTL

This image shows the exact location of the QTL and flanking markers (Fig. 11). Next to the chromosome graphic, the details of the QTL and its markers are displayed. Any of the markers can be selected to display the USDA MARC Marker Report for that particular marker in a separate window as shown in Fig. 12.

7.3 QTL SEARCH

The search menu is displayed next to basic menu, as shown in Fig. 3. If the selection criteria are submitted, a list of QTL satisfying the given criteria are displayed in a table. Each of these entries can be clicked on to go to the image showing that QTL.
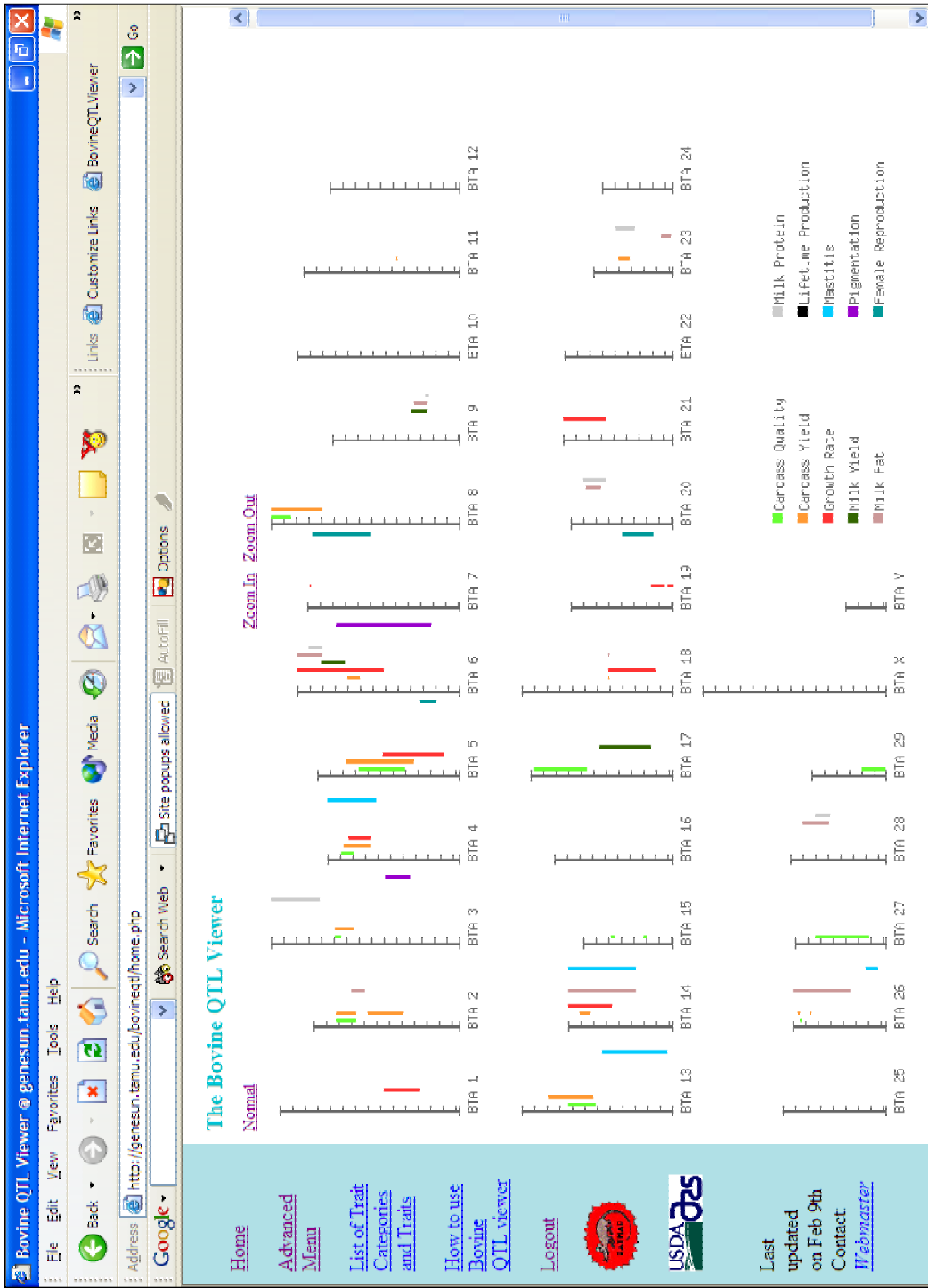
Figure 5: Image showing all the chromosomes and the QTL of selected trait categories

Figure 6: Image showing all the chromosomes and the QTL of selected traits, with a suggestive level of confidence and F-statistic greater than 2.3
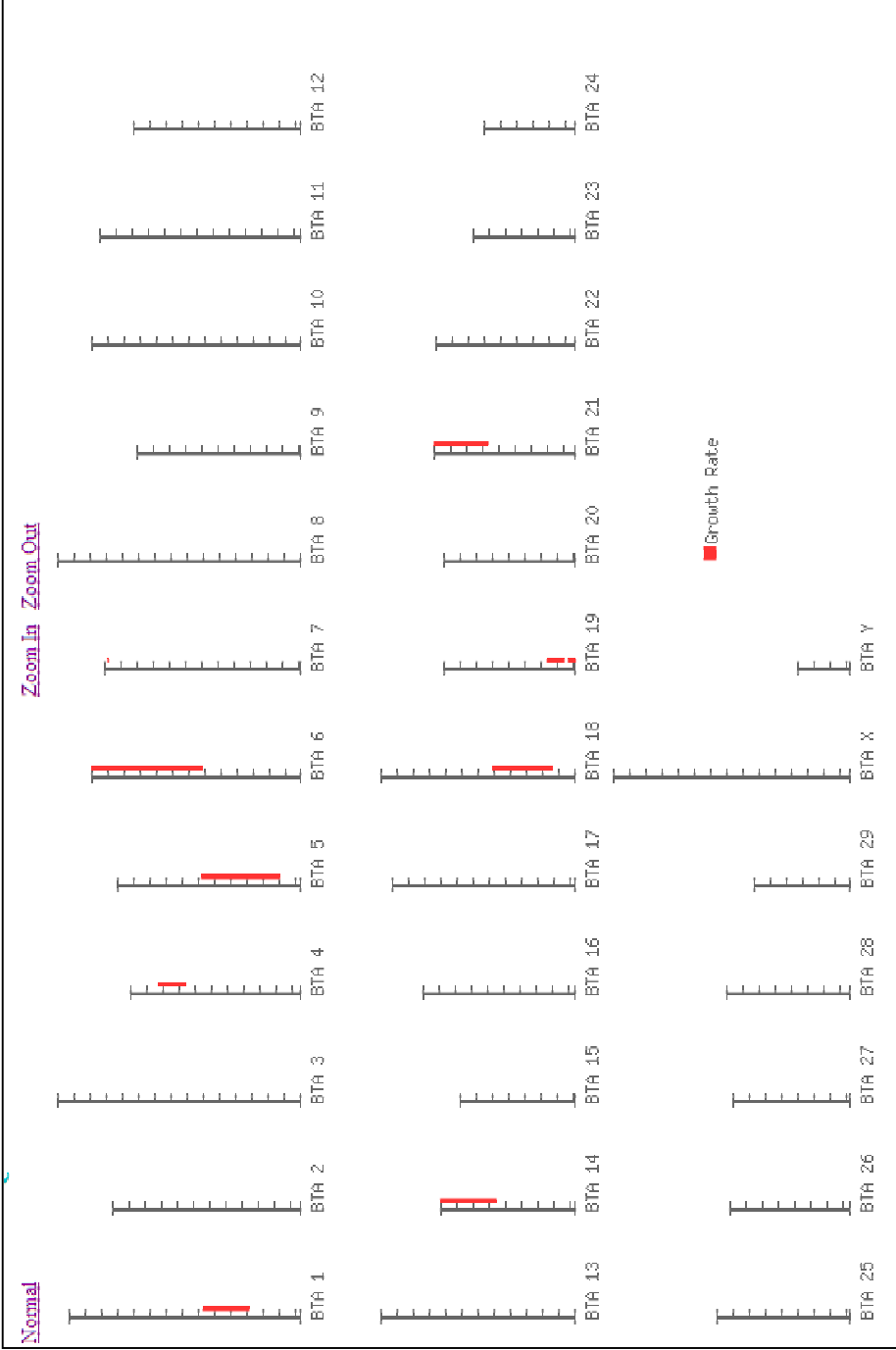
Figure 7: Image showing all the chromosomes and the QTL of Growth Rate, when Growth Rate is clicked on among the labels of Figure 5
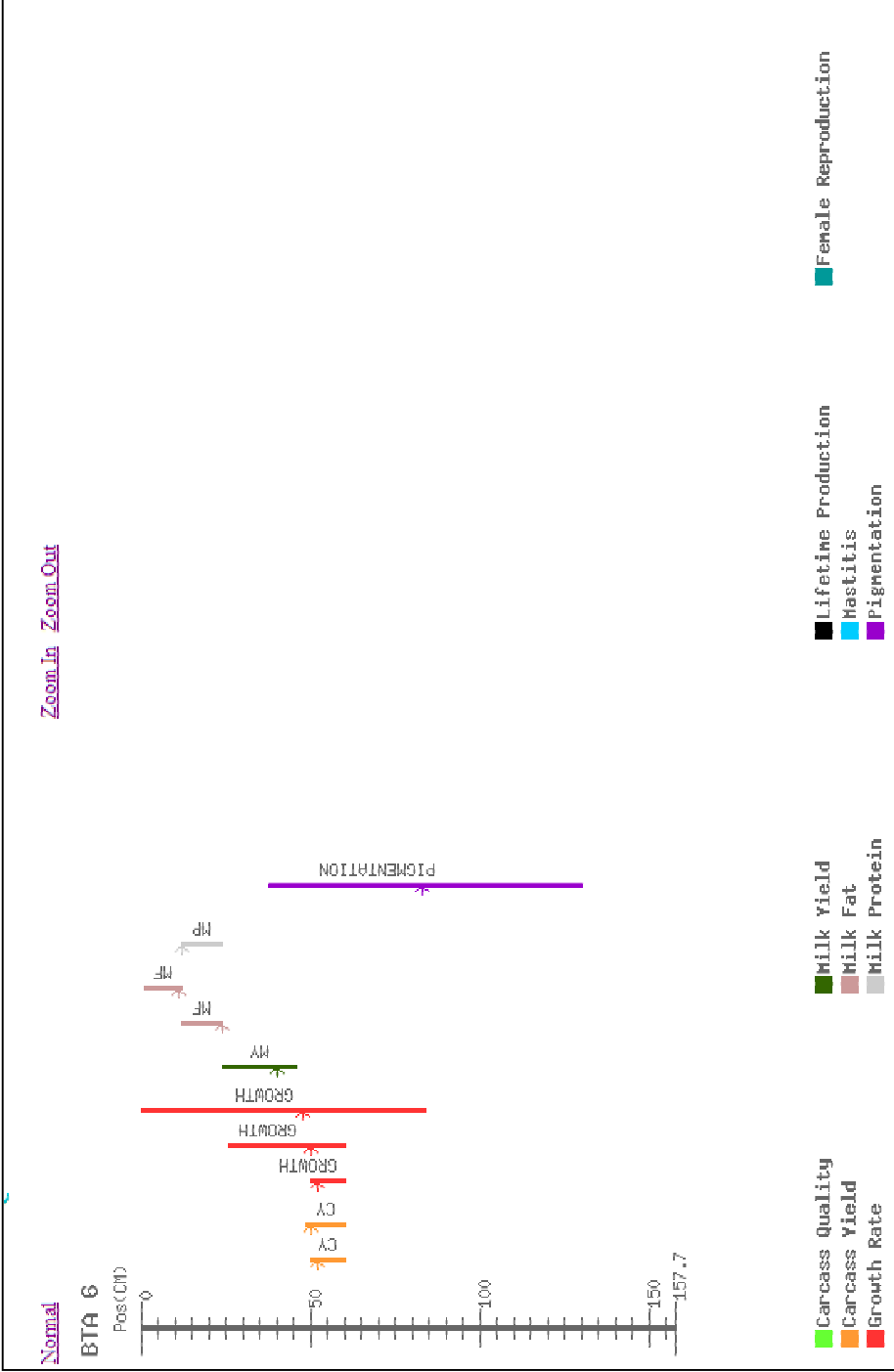
Figure 8: Image showing the chromosome BTA 6 and the QTL for selected trait categories, obtained when BTA 6 is clicked on in Figure 5
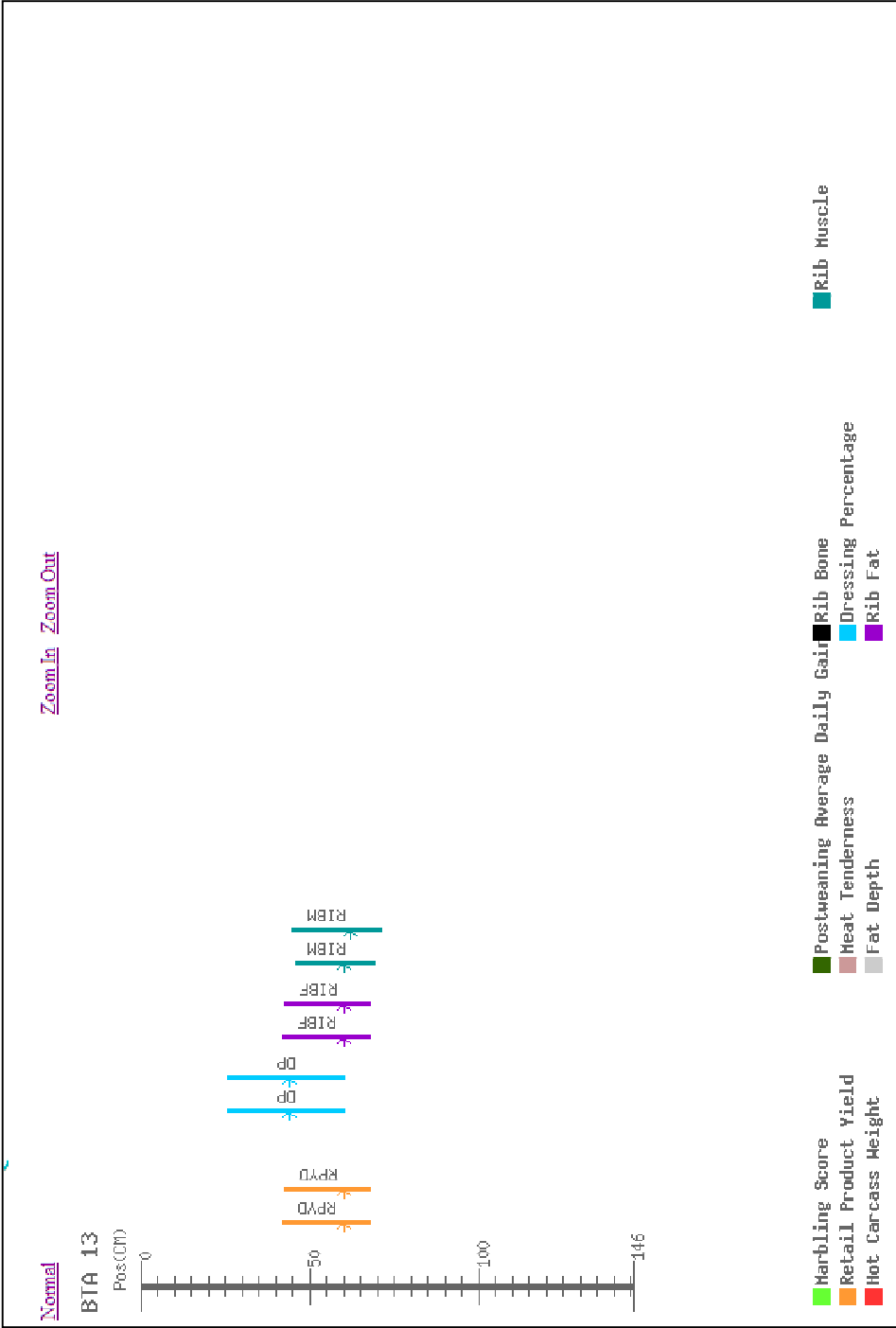
Figure 9: Image showing the chromosome, BTA 13 and the QTL for selected traits, with a suggestive level of confidence and F-statistic greater than 2.3, obtained by clicking on BTA 13 in the image in Figure 6
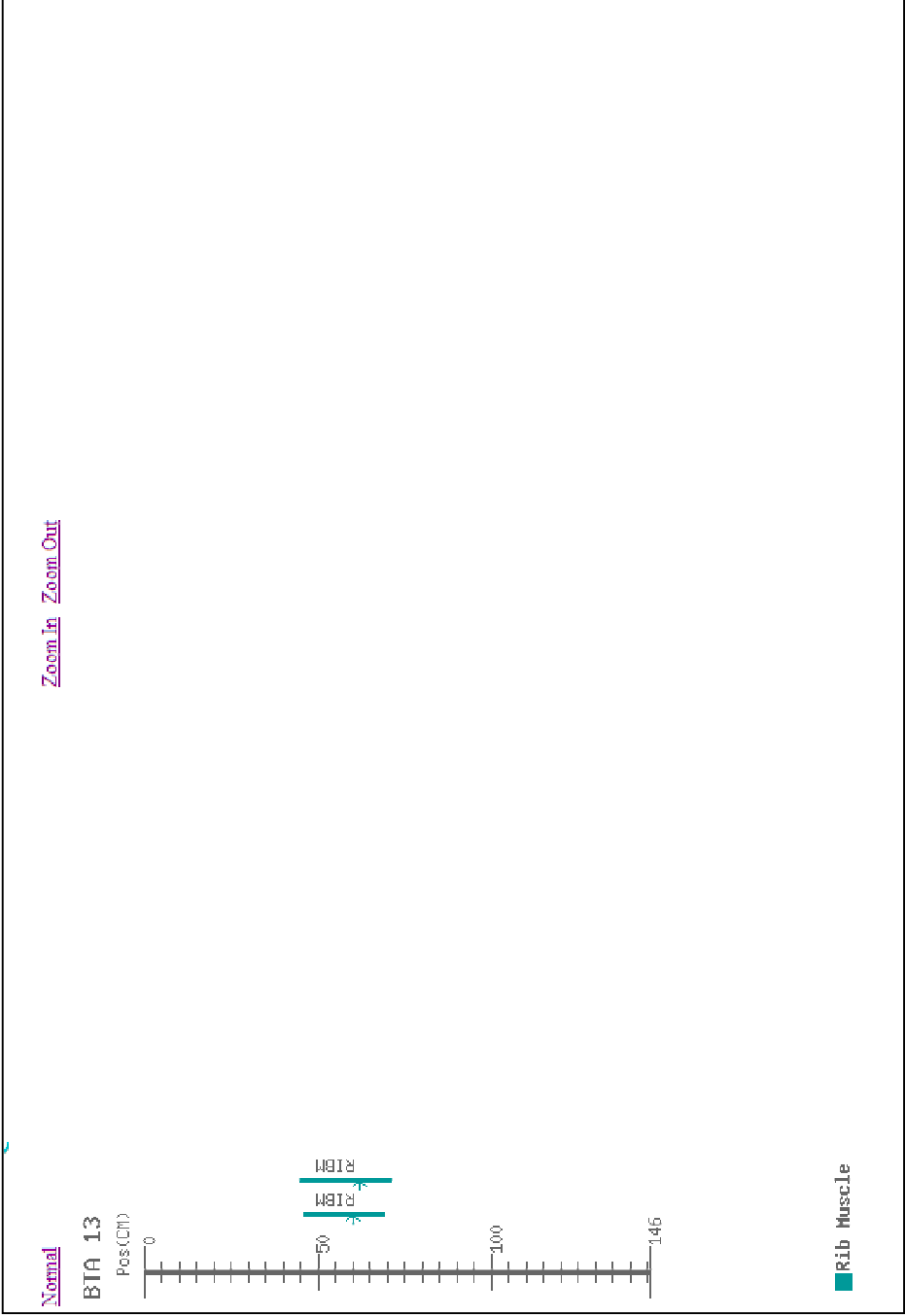
Figure 10: Image showing the chromosome, BTA 13 and the QTL of Rib Muscle, with a suggestive level of confidence and F-statistic greater than 2.3, obtained by clicking on the label 'Rib Muscle' in the image in Figure 9

Normal   Zoom In Zoom Out

QTL for RIBM on BTA 13

Pos(CM)

40   4545   50   55   PEAK 60   65   70 71   75

CSSM034

BMS485

QTL Id	61
Trait	Rib Muscle
Chromosome	13
F-statistic	15.5
Family	Progeny from a Bos indicus X Bos taurus sire mated to Bos taurus cows
Method	Regression suggested by Knott et al (1996)
Significance Level (Suggestive/Significant)	Suggestive
Status	Public
References	List of references
Markers	Position

(Click on the marker name to view MARC Marker Report)

CSSM034	45.100
RM185	45.100
BL41	45.200
BMS1353	45.300
BMS2361	45.500
INRA134	45.500
BMS1148	45.800

Figure 11: Image showing the QTL for the trait, Rib Muscle on BTA 13, obtained by clicking on the QTL for Rib Muscle on BTA 13 in the image shown in Figure 10
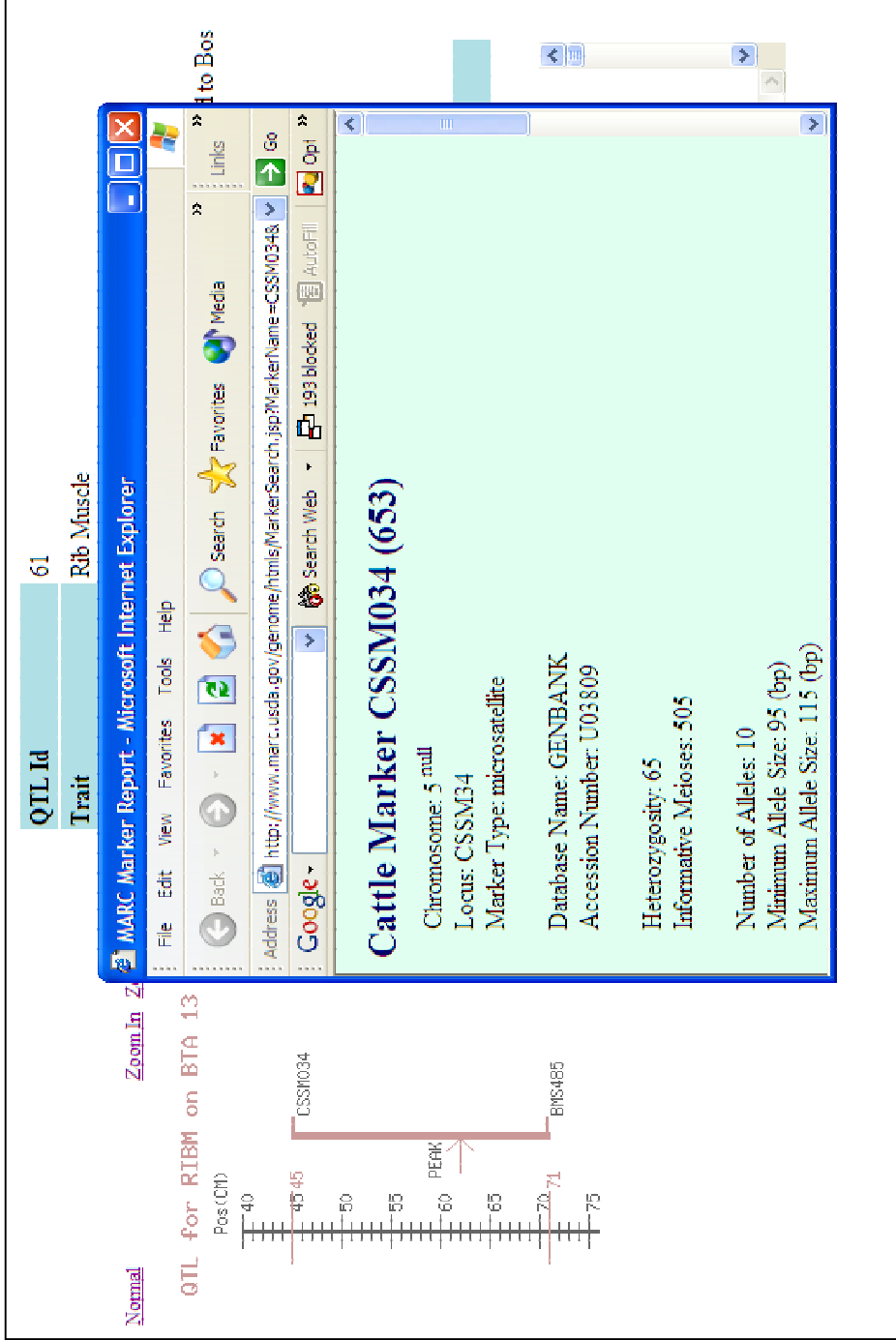
Figure 12: Image showing the MARC Marker Report for the marker CSSM034 in a separate window, obtained by clicking on the marker, CSSM034 in the screen shown in Figure 11

7.4 ADDITIONAL FEATURES

7.4.1 ZOOM FEATURE

All the images displayed can be zoomed in or out as shown in Fig. 14 and Fig. 15. At the top of each image, there are links, 'Zoom in' and 'Zoom out' provided for this purpose. The degree of zoom depends on the number of times the links are clicked on. The link 'Normal' is provided to bring the image to its original size.

7.4.2 COLOR BLINDNESS COMPATIBILITY

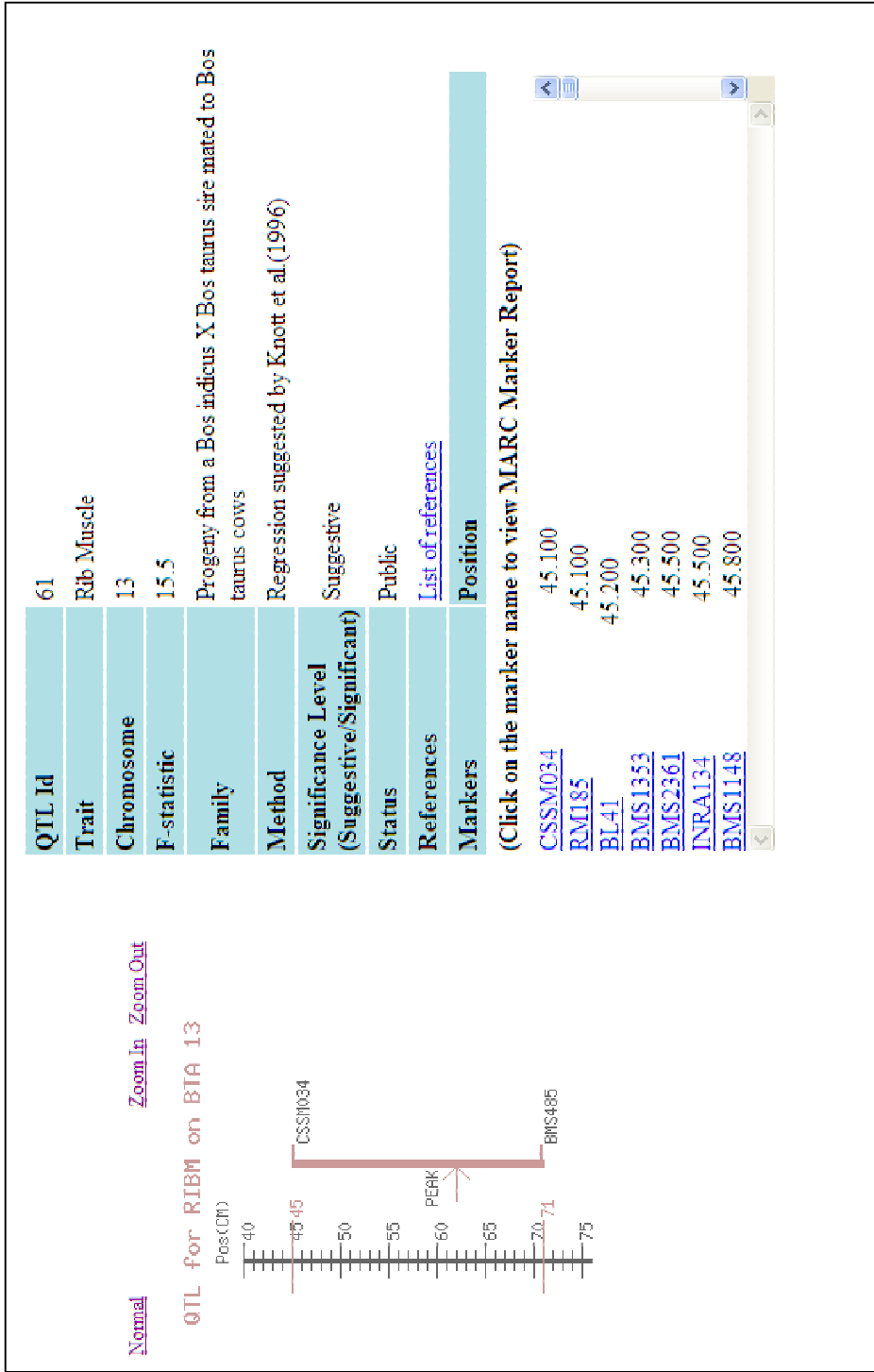All images are displayed compatible with the specifications for graphics for color bind people. I use only those colors recommended by BT Exact Technologies as Safe Web Colours for Colour Deficient Vision (Rigden 1999) (BT Exact Technologies 2003). Ten distinguishable colors for color blind people are used to represent the QTL of each trait or category. Color Blind accommodation is in accordance with: CBC: Color Blindness Check (Q42 2004).

7.4.3 WEB-ACCESSIBILITY AND DIFFERENT LEVELS OF ACCESS

The Bovine QTL Viewer is web-accessible. Users are given privileges of different levels for viewing or editing data. These levels correspond to a 2x2 matrix of viewing and editing, by public and private domain data (Fig. 13). For example, for editing private domain data, the user needs login name, password, privilege to edit QTL data and privilege to edit private domain data. The data sent to the server is encrypted for security purposes.

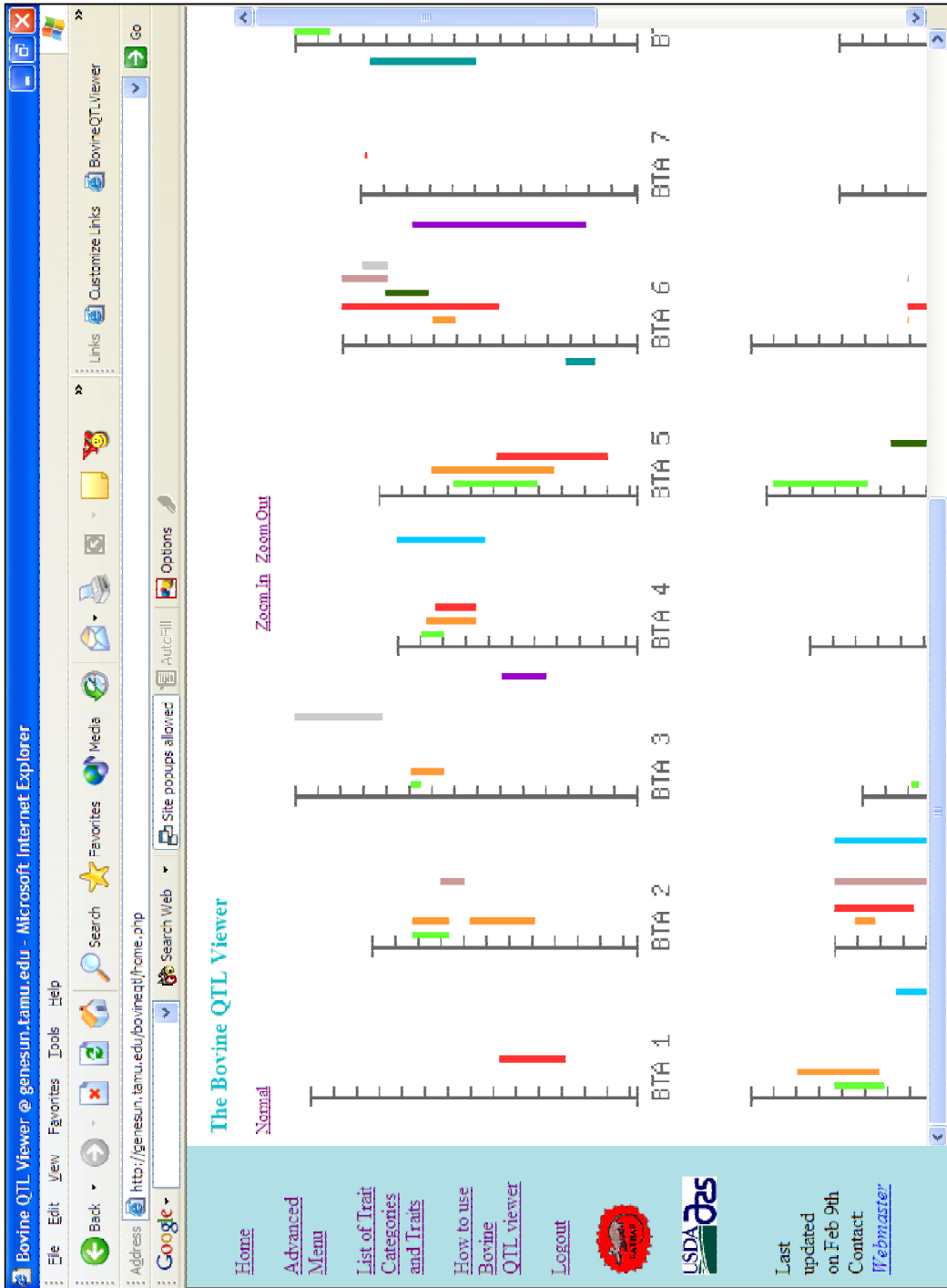| *Type of Data* → *Privilege* ↓	**Public Domain Data**	**Private Domain Data**
Viewing Data	Guest	User
Editing Data	Authorized User	Authorized User

Figure 13: Table showing different levels of access
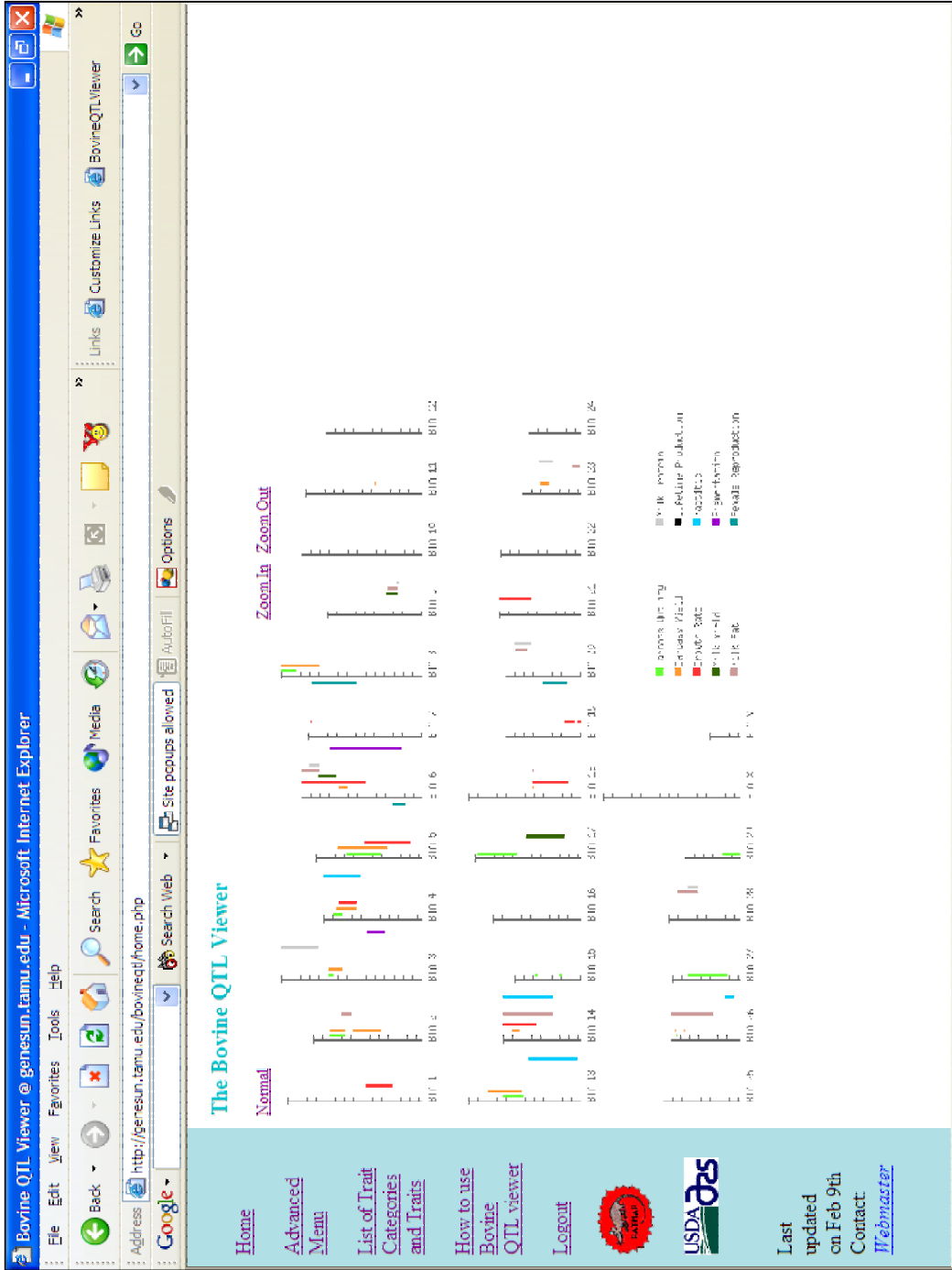
Figure 14: Zoomed portion of image shown in Figure 5

Figure 15: Zoomed out version of the image shown in Figure 5

8. HEURISTIC EVALUATION

To ensure the effectiveness of the proposed system, the Bovine QTL Viewer was evaluated in three stages during the design and development of the tool. The First stage of evaluation was immediately after the design was completed. Then the second stage was after the tool was developed. The third stage of evaluation was after changes were made according to the feed back got in the second stage. The evaluations were performed by the professors, researchers and students of department of Animal Science and department Veterinary medicine of Texas A&M University. A total of eleven evaluators evaluated the tool in various aspects like the information presented, ease of navigation, menus given, search criteria, flexibility of the viewer. In addition to these, all the evaluators gave their opinions and suggestions on the tool.

8.1 FIRST STAGE OF EVALUATION

The first stage of evaluation was conducted after the proposed system was designed. The data model and the pictures of the screen shots of the tool depicting its usage and navigation were presented to the evaluators (Fig. 16, Fig. 17, Fig. 18, Fig. 19, Fig. 20, Fig. 21). The results of the first stage were immensely useful in making modifications in the design according to the needs of the researchers. In this stage, one of the evaluators suggested that there is a need of hierarchy of menu instead of just one type of menu which led to inclusion of an advanced menu and presenting only categories in the main menu. It was expressed that the traits need to be categorized to have the hierarchy. Another evaluator wanted the search criteria to include statistical scores and confidence levels which are later included in advanced menu. Another evaluator suggested dynamic menus instead of static menus. Another suggestion was highlighting the required QTL in the image. Another evaluator expressed the importance of presenting the QTL and marker details. There was also a suggestion to include the breed information in the search criteria.

Figure 16: Picture of Main Menu used in first stage evaluation

THE BOVINE QTL VIEWER

ADVANCED SELECTION FOR QTL VIEWER

LIST OF TRAITS
Select any 10 traits

Birth Traits	Marbling Score	Rib Fat	Hot Carcass Wt	Yearling
Birth Weight	Fat depth	.	.	Weight
Gestation	Retail Product	.	.	.
Length	Yield	.	.	.
Calving ease	Rib Muscle			.
.	.			
.	.			
.	.			

Workability
Milking Speed
Temperament
.

LOD score greater than [____]   Suggestive ☐   Chromosome [____]   **submit**
F-statistic greater than [____]   Significant ☐   All Chromosomes ☐

Figure 17: Picture of Advanced Menu used in first stage evaluation

THE BOVINE QTL VIEWER

Click on an individual chromosome or individual QTL for more information

BTA 1   BTA 2   BTA 3   BTA 4   BTA 5   BTA 6   BTA 7   BTA 8   BTA 9   BTA   BTA

BTA   BTA   BTA   BTA   BTA   BTA   BTA   BTA   BTA   BTA   BTA

BTA 23   BTA 24   BTA 25   BTA 26   BTA 27   BTA 28   BTA 29   BTA X   BTA Y

◆Birth Weight
◆Marbling Score
◆Fat Depth
◆Rib Muscle
◆Rib Fat

Figure 18: Picture of comprehensive image used in first stage evaluation
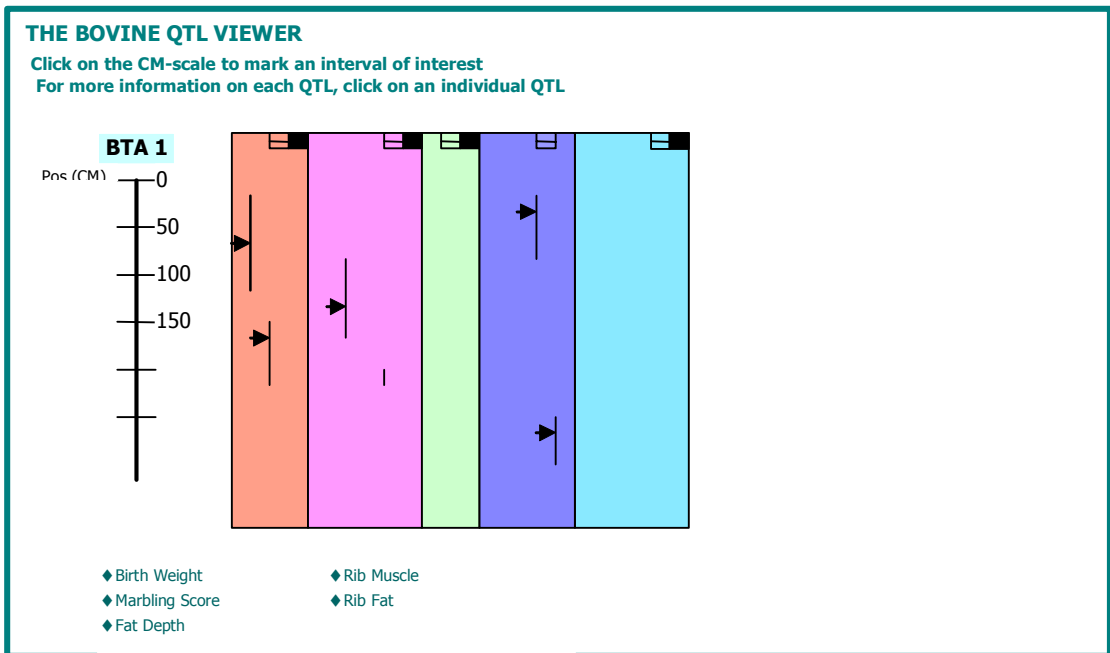
Figure 19: Picture of a single chromosome used in first stage evaluation
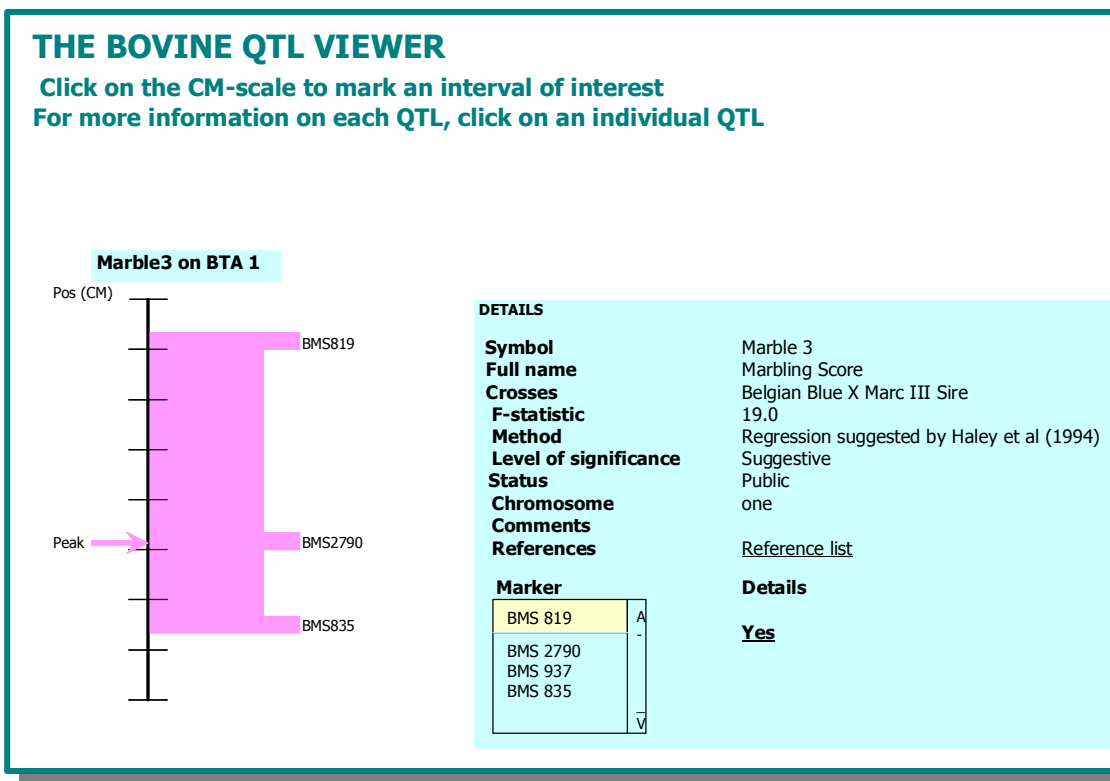
Figure 20: Picture of an individual QTL used in first stage evaluation

Figure 21: Picture showing the navigation of the tool used in first stage evaluation

8.2 SECOND STAGE OF EVALUATION

The second stage of evaluation was conducted after the proposed system was developed according to the design. The tool was made web-accessible, a guest login was created and the evaluators were given access to the tool. The evaluators tried the tool and came up with some important suggestions. They were useful in making the tool more effective. One of the evaluators suggested that it was better to click instead of mouse over for zoom in and zoom out features. The mouse over confuses the users. Another suggestion was not to take the users directly to the QTL details when they click on a QTL in the comprehensive image. As the overlapping QTL cannot be seen in this image, this operation can be misleading. Another evaluator suggested that it was better to link the tool directly to the MARC report for marker details.

8.3 THIRD STAGE OF EVALUATION

Minor modifications were made to the tool according to the suggestions obtained from the second stage of evaluation and the third stage of evaluation was conducted. In the third stage of evaluation, most of the responses I got were positive. The evaluators were satisfied with the menus and the flexibility of the viewer. They commented that the organization of the tool is clear and the system is easy to navigate. In this stage, there were concerns about the curation of the database. Also one of evaluators suggested that it would be nice to have an option to search the database by author. This would be useful to double-check that studies are included in the database and also useful when other people refer to a past study. It was also suggested to include the private domain data in the database.

8.4 DISCUSSION

The feedback got from the evaluations conducted was valuable in making the system more efficient and effective. The evaluation results provided both low-level fixes high-level issues for the data model, the viewer and the representation as a whole. Based on the suggestions obtained from evaluators, a detailed menu was provided, statistical scores and confidence levels were included in the search criteria, menus were made dynamic, user is allowed to select a trait directly from the image and only QTL of selected trait are shown in the image, the tool is directly linked to MARC marker report and the navigation and zooming features of the tool were improved. While most of the suggestions were implemented, there are a few suggestions, like including the breed information in the search criteria and searching the database by author, which have to be given thought.

9. CONCLUSION

The Bovine QTL Viewer described in this thesis provides a comprehensive view of the QTL detected in the cattle genome in support of research into cattle QTL. This tool is useful for animal scientists, geneticists, students studying animal genomics or anyone who is interested in finding out what is known about genotype/phenotype associations in cattle.

REFERENCES

Bailey, L.C., Jr., Fischer, S., Schug, J., Crabtree, J., Gibson, M. and Overton, G.C. 1998. GAIA: Framework annotation of genomic sequence. *Genome Res*. **8:** 234-250

BT Exact Technologies. 2003. Safe Web Colours for colour-deficient vision. http://more.btexact.com/people/rigdence/colours/.

Crown Human Genome Center and Weizmann Institute of Science. 2003. Gene Cards. http://bioinformatics.weizmann.ac.il/cards/.

Davenport, G.F., O'Malia, A.D. and Dicks, J.L. 1999. The Genome Map Viewer. http://synteny.nott.ac.uk/software.html.

Dicks, J. 1997. The display of comparative mapping data using ACEDB. *Genetic Mapping of Disease Genes* (eds. Pawlowitski IH and Edwards JH). Academic Press. (In Press): 221-235.

Durbin, R. and Thierry Mieg, J. 1991. ACEDB. http://www. acedb.org/.

Edwards, J.H. 1991. The Oxford Grid. *Ann Hum Genet* **55:** 17-31

Fischer, G., Ibrahim, S.M., Brockmann, G.A., Pahnke, J., Bartocci, E., Thiesen, H.J., Serrano-Fernández, P. and Möller, S. 2003**.** Expressionview: visualization of quantitative trait loci and gene-expression data in Ensembl. *Genome Biology* **4:**R77

Fischer, S., Crabtree, J., Brunk, B., Gibson, M. and Overton, G.C. 1999. bioWidgets: Data interaction components for genomics. *Bioinformatics* **15:** 837-846

Frishman, D., Albermann, K., Hani, J., Heumann, K., Metanomski, A., Zollner, A. and Mewes, H.W. 2001. Functional and structural genomics using PEDANT. *Bioinformatics*, **17:** 44-57

Harris, N.L. 1997. Genotator: A workbench for sequence annotation. *Genome Res.* **7:** 754-762

Helt, G.A., Lewis, S., Loraine, A.E. and Rubin, G.M. 1998. BioViews: Java-based tools for genomic data visualization. *Genome Res.* **8:** 291-305

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. 2002. The human genome browser at UCSC. *Genome Res*. **12:** 996-1006.

Loraine, A.E. and Helt, G.A. 2002. Visualizing the genome: techniques for presenting human genome data and annotations. *BMC Bioinformatics* **3:**19.

National Center for Biotechnology Information. 2003. OMIM-Online Mendelian Inheritance in Man. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM.

National Center for Biotechnology Information. 2004. NCBI Map Viewer for Bos taurus. http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9913.

Pecherer, R.M., Polacco, M., Zhang, H.B., Beavis, W.D., Cartner, T. and Shroeder, S. 1999. Comparative Genome Mapping Tool. http://www.ncgr.org/research/cgmt/.

Priestly, M.H., Dickson, J.S. and Dicks, J.L. 1999. GridMap III. http://synteny.nott.ac.uk/software.html.

Q42. 2004. CBC: Color Blindness Check. http://www.q42.nl/demos/colorblindnesssimulator/.

Rigden C, 1999. 'The eye of the beholder'- Designing for color-blind users. *British Telecommunications Engineering Journal* **17:** 2-6. http://more.btexact.com/people/rigdence/colours/colours.pdf.

Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J.E., Harris, T.W., Arva, A. *et al*. 2002. The generic genome browser: a building block for a model organism system database. *Genome Res*. **12**: 1599–1610.

The RatMap Group. 2003. Rat Map: QTL Viewer. http://ratmap.gen.gu.se/.

U.S Meat Animal Research Center. 2003. MARC Map for Bos Taurus. http://www.marc.usda.gov/.

Wattiaux, M.A. 2003. Dairy Essentials- Reproduction and Genetic Selection: Basic Genetic Concepts. http://babcock.cals.wisc.edu/de/pdf/14_e.pdf.

VITA

Name: Pavana Polineni

Permanent Address: C/O P. Ramakrishna Chowdary
                               15-40/3, Padmavathi Nagar,
                               Tirupati – 517502.
                               Andhra Pradesh. INDIA.

Education: M.S. Computer Science, Texas A&M University, 2003
                               B.Tech. Computer Science and Engineering, Sri Venkateswara
                               University, Tirupati, Andhra Pradesh, India, 2001.