

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
28 February 2002 (28.02.2002)

PCT

(10) International Publication Number
WO 02/17207 A2

(51) International Patent Classification⁷: **G06F 19/00**

(21) International Application Number: PCT/IB01/01883

(22) International Filing Date: 23 August 2001 (23.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/227,342 23 August 2000 (23.08.2000) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/227,342 (CON)
Filed on Not furnished

(71) Applicant (for all designated States except US): **AREXIS AB** [SE/SE]; Krokslatts Fabriker 30, S-431 37 Molndal (SE).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **LUTHMANN, L.,**

Holger [SE/SE]; Fagerviksvagen 3, S-168 39 Bromma (SE). **ANDERSSON, Leif** [SE/SE]; Bergagatan 30, S-752 39 Uppsala (SE). **WENDEL-HANSEN, Vidar** [SE/SE]; Ringduvevagen 1, S-756 52 Uppsala (SE).

(74) Agent: **ELLINGER, Mark, S.**; Fish & Richardson P.C., 3300 Dain Rauscher Plaza, 60 South Sixth Street, Minneapolis, MN 55402 (US).

(81) Designated States (*national*): CA, JP, US.

(84) Designated States (*regional*): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: GENETIC RESEARCH SYSTEM

(57) Abstract: The invention is directed to a system that provides a flexible genetic information storage and processing structure. This system can be used to facilitate collaboration between genetic researchers within a research group. Multiple genetic research groups can securely and independently access and use the genetic research system.



WO 02/17207 A2

GENETIC RESEARCH SYSTEM

TECHNICAL FIELD

This invention relates to a system for supporting genetic research groups.

BACKGROUND

5 Genetic research is the study of inherited traits, often with the primary goal of identifying DNA mutations that can cause specific health problems. The identification of genetic mutations enables clinicians to predict the likelihood that individuals will develop particular health problems or pass on health risks to their children. As such, efforts to isolate DNA mutations have lead to intense worldwide biomedical research. This genetic
10 research, such as sequencing human genomes, can be exhaustive research and often involves the collaboration between geographically distributed genetic researchers. The research typically requires substantial computing resources and specialized algorithms to analyze and process vast amounts of genetic research data.

SUMMARY

15 In general, the invention is directed to a system that provides a flexible genetic information storage and processing structure in order to facilitate the collaboration of genetic researchers and research groups. Multiple genetic research groups can securely and independently access and use the genetic research system.

According to one aspect, the invention is directed to a web-based genetic research
20 system in which a web server receives genetic research data from a user over a network. A database is coupled to the web server and is configured to store the genetic research data. The database conforms to a database schema that defines a hierarchy of genetic research objects to store the genetic research data. For example, the schema defines a species data object to store species data and a chromosome data object to store
25 chromosome data. Additional objects to store the genetic research data include a gene object, a marker object, an allele object, a genotype object, a phenotype object and a variable object. A number of administration objects provide secure access to the hierarchy of genetic research objects. A project data object allows a system administrator

to configure project views defining access to the genetic research projects according to an access control hierarchy having a plurality of levels.

According to another aspect, the invention is directed to a computer-readable storage medium having data structures thereon to store the genetic research data in hierarchical fashion, thereby providing a flexible storage infrastructure. For example, the data structures include a species data structure to store species data, a chromosome data structure to store chromosome data and a gene data structure to store gene data. Relationships between the data are enforced to ensure data integrity and hierarchical consistency.

According to yet another aspect, the invention is directed to methods and techniques for securely providing access to the genetic research database and the data contained therein. According to the method, an access request is received from a user to access one of a set of genetic research objects within the database that store the genetic research data. A project view object within the database is queried to determine which genetic research objects the user can access. A role object and a privileges object within the database are queried to determine operations that the user is allowed to perform. The user is permitted to access the requested genetic research objects based the accessible objects and the allowed operations.

Various embodiments of the invention are set forth in the accompanying drawings and the description below. Other features and advantages of the invention will become apparent from the description, the drawings, and the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a block diagram illustrating a distributed genetic research environment capable of supporting multiple research groups according to the invention.

Figure 2 is a block diagram illustrating an example of a genetic research system configured according to the invention.

Figure 3 is a block diagram illustrating an example of a schema for a database to store genetic research data.

Figures 4 and 5 illustrate example charts produced by the genetic research system.

Figure 6 is a block diagram illustrating a computer suitable for implementing the various embodiments of the invention.

DETAILED DESCRIPTION

Figure 1 is a block diagram illustrating an environment 2 for facilitating multiple genetic research groups 6. Each genetic research group 6 includes a number of researchers that collaborate toward accomplishing a common goal, such as finding and isolating genes related to metabolic diseases. As described in detail below, genetic research system 8 provides a flexible genetic information storage and processing structure by which individual researchers can securely share research data within research groups 6.

Each researcher uses a corresponding computing device 10 to access genetic research system 8 via network 18. A communication tool, such as a web browser like Internet Explorer™ from Microsoft Corporation of Redmond, Washington, executes in an operating environment on each computing device 10 and allows each researcher to remotely access genetic research system 8. Each computing device 10 represents any general purpose computing device suitable for interacting with network 18 and genetic research system 8. One example of a suitable computing device 10 is a personal computer. However, computing device 10 can be a laptop computer, a handheld computer, a personal digital assistant (PDA), such as a Palm™ organizer from Palm Inc. of Santa Clara, California, or even a network-enabled cellular telephone. Network 18 represents any transmission medium suitable for transmitting digital data. For example network 18 can be a packet-based digital network, such as a private wide area network (WAN) or the Internet, running a network protocol, such as the transmission control protocol / internet protocol (TCP/IP),

Figure 2 illustrates in further detail one embodiment of genetic research system 8. Genetic research system 8 includes one or more web servers 20 coupled to research database 22. Web servers 20 provide an interface for communicating with computing devices 10 and via network 18. In one configuration, web servers 20 execute web server software, such as Internet Information Server™ from Microsoft Corporation, of Redmond, Washington. In another configuration, web servers 20 execute Apache Web

Server™ software within an operating environment provided by the Linux operating system. Web servers 20 provide an environment for interacting with research database 22 according to software modules 24, which can include Lotus scripts, Java scripts, Java Applets, Java servlets, Active Server Pages, web pages written in hypertext markup language (HTML) or dynamic HTML, Active X modules, CGI scripts, and other suitable modules such as stand-alone executables written in C or C++.

Software modules 24 are grouped into two general categories: user interface modules 26 and data analysis modules 28. User interface modules 26 include software modules for interacting with the genetic researchers to capture or display genetic research data. Data analysis modules 28 process the genetic data stored within research database 22 to assist and automate the research. For example, in one embodiment data analysis modules 28 include executable software modules that analyze the genetic data stored by research database 22 to locate and map multiple interacting quantitative trait loci (QTL) in a genome.

Research database 22 provides a flexible storage device for storing genetic data received from computing devices 10. In one configuration, research database 22 is implemented using a database engine, such as Oracle™, executing on a database server. In this configuration, the database server is communicatively coupled to web server 20, typically via a packet-based local area network (LAN).

Figure 3 illustrates one example of a schema 30 for research database 22. Database schema 30 illustrates a number of database objects and their inter-relationships that are enforced to ensure data integrity. Generally, the database objects define the data structures for capturing and organizing the genetic research data with database 22. The relationships define the constraints of the database as graphically illustrated by the interconnecting lines and their end points. For example, relationship 68, having one end point indicates a one-to-many relationship between the project object 62 and the role object 32. Thus a single entry in the project object 62 can map to many entries in the role object 32. Conversely, relationship 66 illustrates a many-to-many relationship between privileges object 33 and role object 32.

The objects within database schema 30 fall within two categories: administrative objects 70 and research objects 72. Generally, administrative objects 70 allow research

database 22 to easily support multiple researchers organized into multiple research projects. Research objects 72 store the genetic research data and are hierarchically arranged, from species down to individual, in a manner that facilitates analyzing the genetic research data.

5

Administrative Objects

Administrative objects 70 include four objects: project object 62, role object 32, user object 34 and privileges object 33. User object 34 stores information, such as a name and a password, about each researcher that is authorized to access genetic research system 8. Each entry in user object 34, i.e. each researcher, "relates" to one or more entries in role object 32. In this manner, entries in user object 34 must identify one or more entries in role object 32. When a researcher logs onto genetic research system 8, he or she specifies a project on which to work. Genetic research system 8 presents the researcher with a list of projects for which he or she can access. The researcher can only choose between the projects for which he has a defined role. A role is local to a project and it defines which operations the researchers are allowed to execute. Each role relates to a set of privileges within privileges object 33. Each user can only have one role in a project. As such, role object 32 and privileges object 33 define the access rights for the corresponding researcher, i.e., what functions the researcher can perform on the various research objects 72. In one configuration, the following privileges exist: write access (create, read, update and delete), read access, and no access. In addition, a system administrator has the following non project-specific privileges: create, update and delete projects, create, update and delete users, define access for users in projects.

Project object 62 controls access to the various research objects 72 according to project "views." When creating a new project, an administrator creates a project within project object 62 and defines a view. Each project view defines access to research objects 72 by defining an access hierarchy. Each level of the hierarchy specifies which particular entries can be accessed within a corresponding research object 72. A level within the project view, however, may provide access to no entries. More specifically, the access control hierarchy controlled by the project view is as follows:

- Level 1 – the accessible entries within the sampling unit object 38.
- Level 2 – the accessible entries within individual object 52 that relate to accessible sampling unit specified by level 1.
- 5 ▪ Level 3 – the accessible entries within grouping object 48 that relate to accessible of sampling units specified by level 1.
- Level 4 – the accessible entries within the species object 36.
- Level 5 – the accessible entries within chromosome object 40 that relate to accessible species specified by level 4.
- 10 ▪ Level 6 – the accessible entries within gene object 42 that relate to accessible chromosomes as specified by level 5.
- Level 7 – the accessible entries within marker object 42 that relate to accessible chromosomes as specified by level 5.
- Level 8 – the accessible entries within variable object 58 that relate to accessible species as specified by level 4.

15 For example, a project view can be defined granting full access to a set of species and a set of sampling units. Full access means that the project members will have access to all genetic data within research objects 72, including individuals, groupings, groups, phenotypes, genotypes, chromosomes, markers, etc., that relate to the corresponding sets of species and sampling units for the project. Conversely, a project view can be defined

20 granting limited access such that only one species and one sampling unit can be accessed. In this configuration, the project members will only see one individual in the sampling unit. In particular, no groupings or groups will be accessible. Only one marker is accessible on this chromosome. Only one genotype data is therefore accessible for the marker and individual

25

Research Objects

Research objects 72 are hierarchically arranged to facilitate securely collecting, analyzing and reporting the genetic research data. At the highest level is the species object 36. The species object 36 models information about biological species, such as

30 humans. The chromosome object 40 is used to model a certain chromosome for a certain species. The gene object 42 is used to model information about a specific gene that is

located on a specific chromosome in one species. The marker object 60 stores marker information that identifies a locus on a particular chromosome. The allele object 56 is used to model and store information about a specific allele, i.e. one of the possible variants of a DNA-sequence at a locus on a chromosome. The genotype object 44 is used
 5 to store the observed pair of alleles for an individual at the locus where a specific marker points.

Data file set object 31 stores data file sets, which identify one or more data files. The researcher can generate a data file set by selecting a marker set, a variable set and an appropriate filter. The data file sets are local to a project. Filter object 35 stores one or
 10 more filters, which are logical expressions used for selection of individuals. More specifically, a filter is a boolean expression used for advanced selection of individuals. During the selection process, the expression is evaluated for each individual belonging to the project. The individuals, for which the expression evaluates to true, are said to be selected by the filter. The formal language for writing filter expression is called GQL,
 15 genetic query language. Standard Oracle expressions is a sub set of GQL. A GQL expression can therefore be made up by combinations of parenthesis, logical and numerical operators, standard functions and user defined functions. The only limitation is the length: 2000 characters. In addition, a GQL expression may consist of the following special terms:

- 20 · Individual attributes, e.g. sex or birth date
- Genotype attributes, e.g. allele or raw data for allele
- Phenotype attributes, e.g. value or date
- Set membership, e.g. sampling unit, grouping or group.

Individual attributes are referenced with the prefix: "I." as shown by the following
 25 example: I.SEX. Genotype attributes are referenced with the prefix: "G." as shown by the following example: G.MA001.A1 (allele 1 for marker MA001). Phenotype attributes are referenced with the prefix: "P." as shown by the following example:
 P.EYE_COLOR.VALUE (value of the phenotype eye color). Set membership are referenced with the prefix S. As shown by the following example:
 30 S.HUMAN008.GENERATIONS.F2 (member of group F2 in the grouping
 GENERATIONS in sampling unit HUMAN008). All the above constructs refer to

attributes or membership for the individual under evaluation. It is possible though, to refer to attributes or membership for parents of the individual. This can be made by writing a sequence of M or F after the first prefix, as shown in the following example: P.MM.EYE_COLOR.VALUE (value of eye color for grandmother).

- 5 Sampling unit object 38 models specific sampling units within a species. As such, sampling units are sets of individuals of a species that are collected and processed together. In addition, the sampling unit object 38 interrelates individuals, groupings and groups. The individual object 52 stores general information about an individual belonging to a sampling unit including a unique identity, an alias, father, mother, sex, and
- 10 birth date. Grouping object 48 is used to model a set of groups belonging to a genetically interesting grouping within a sampling unit. Examples of such groupings are families and generations. Group object 50 is used to model a set of individuals belonging to some kind of genetically interesting group, such as a particular family. The sample object 54 models and stores information about the specific samples taken from individuals. The
- 15 phenotype object 46 is used to model and store an observed value for an individual and a specific variable, as stored in the variable object 58. Variables must be defined in order to store phenotypes in the database. Marker set object 61 stores sets of markers. Similarly, variable set object 59 stores sets of variables.

- The following are examples of each object within the schema 30 of the genetic research
- 20 database 22 including the attributes that are stored for each object.

Allele

Name	Type	Example	Description
Name	Integer	3	Order number within marker
Comment	Text[50]	"Micro satellite allele size = 24"	Free text

Chromosome

Name	Type	Example	Description
Name	Text[12]	"45"	Name of chromosome
Alias	Text[12]	"X"	Alias for chromosome
Comment	Text[50]	-	Free text

Data File Set

Name	Type	Example	Description
Name	Text[20]	"DFS2001A"	Name of data file set
Type	Text[20]	"LINKAGE"	Type of data file set
Comment	Text[256]	-	Free text

Filter

Name	Type	Example	Description
Name	Text[20]	"MALE"	Name of filter
Type	Text[2000]	"I.SEX = 'M'"	Logical expression (GQL)
Comment	Text[256]	-	Free text

5 Gene

Name	Type	Example	Description
Name	Text[12]	"GEN4241"	Name of gene
Alias	Text[12]	-	Alias for gene
Comment	Text[50]	-	Free text

Genotype

Name	Type	Example	Description
Raw data 1	Text[10]	"132,7"	Raw data value for allele 1
Raw data 2	Text[10]	"135,9"	Raw data value for allele 2
Reference	Text[20]	"GEL3456"	Reference to raw data, e.g. microfilm or gel
Comment	Text[50]	-	Free text

Group

Name	Type	Example	Description
Name	Text[12]	"FAM023"	Name of group
Comment	Text[50]	-	Free text

10

Grouping

Name	Type	Example	Description
Name	Text[12]	"ICELAND1"	Name of grouping
Comment	Text[50]	-	Free text

Individual

Name	Type	Example	Description
Identity	Text[12]	"610710-6913"	Unique identity within sampling unit
Alias	Text[12]	"A. Nyberg"	Alias for individual
Father	Reference	-	DB reference to father, can be NULL
Mother	Reference	-	DB reference to mother, can be NULL
Sex	M or F	"M"	Male or Female
Birth_date	Date	1961-07-10	Date of birth
Comment	Text[50]	-	Comment

Marker

Name	Type	Example	Description
Name	Text	"MA4001"	Name of marker
Alias	Text	"Marker 1"	Alias for marker
Comment	Text	-	Free Text

5

Marker Set

Name	Type	Example	Description
Name	Text[20]	"MS Chr 12"	Name of marker set
Comment	Text[256]	-	Free Text

Phenotype

Name	Type	Example	Description
Value	Text[20]	"145,35"	Observed value
Date	Date	2000-01-30	Date of observation
Reference	Text[20]	"LMX037.DAT"	Reference to raw data, e.g. blood sample
Comment	Text[50]	-	Free text

10 Privileges

Name	Type	Example	Description
Name	Text[8]	"UPD_IND"	Privilege
Comment	Text[256]	-	Free text

Project

Name	Type	Example	Description
Name	Text[12]	"PROJ 452"	Name of project
Views		-	Defines the project views
Comment	Text[50]	-	Free text

Role

Name	Type	Example	Description
Name	Text[20]	"Admin"	Name of role
Comment	Text[256]	-	Free text

5

Sample

Name	Type	Example	Description
Name	Text[12]	"BLOOD"	Name of sample
Tissue	Text[20]	-	Tissue type
Date	Date	-	Date when the sample was taken
Treatment	Text[20]	-	
Experimenter	Text[20]	-	
Storage	Text[20]	-	
Comment	Text[50]	-	Free text

Sampling Unit

Name	Type	Example	Description
Name	Text[12]	"HUMAN"	Name of species
Comment	Text[50]	-	Free text

10 Species

Name	Type	Example	Description
Name	Text[12]	"HUMAN"	Name of species
Comment	Text[50]	-	Free text

User Object

Name	Type	Example	Description
Identity	Text[4]	"ANNY"	User identity of user
Password	Text[6]	"ABC123"	Password for user
Name	Text[20]	"Anders"	Full name of user
Comment	Text[50]	-	Free text

Variable

Name	Type	Example	Description
Name	Text[20]	"WEIGHT8W"	Name of phenotype variable
Comment	Text[50]	-	Free text

5 Variable Set

Name	Type	Example	Description
Name	Text[20]	"VS 32"	Name of variable set
Comment	Text[256]	-	Free text

Relations

As described above, schema 30 of defines the relationships between the various objects of genetic research database 22. These relationships define the constraints of database 22 and are graphically illustrated by the interconnecting lines and their end points. The following section describes some of the relations in the information structure in more detail.

Project - Species

A project can work with several species, and several projects can work with one species.

Project – Sampling Unit

A project can be defined to include one or more sampling units. A sampling unit can, on the other side, belong to one or more projects.

Project – Filter

A filter can only belong to one project.

Project – Role

Each project has a set of roles. -

5 Species – Sampling Unit

A sampling unit relates to only one species, i.e. all of its members (individuals) must belong to the same species.

Species – Filter

A filter is defined for one species.

10 Sampling Unit - Individual

A sampling unit consists of a set of individuals, but an individual can only belong to one sampling unit.

Sampling Unit - Grouping

15 A sampling unit can have several groupings, but a grouping can only belong to one sampling unit.

Grouping - Group

A grouping consists of a set of groups, but a group can only belong to one grouping.

Group - Individual

20 A group consists of a set of individuals, and an individual can belong to several groups.

Species - Variable

Variables are defined for species.

Allele – Genotype (two relations) ,

The relations corresponds to the two genotype values (alleles) that has been observed for a marker and an individual.

As described above, genetic research system 8 provides a flexible genetic information storage and processing structure by which individual researchers can securely share research data within research groups 6. Researchers interact with genetic research system 8 and invoke data analysis modules 28 to process the genetic data stored within research database 22. Figures 4 and 5 illustrate two example output charts produced by genetic research system 8 upon processing the stored genetic data. Genetic research system 8 communicates the charts to research computer 10 for display to the user. Figure 4 is an example of a genetic map that illustrates the distance and relative order between a set of markers within marker object 60 and a related chromosome within chromosome object 40. Figure 4 also shows confidence intervals for the variables over a genetic map. Figure 5 shows linkage values (lod scores) for variables within variable object 58 over the relative markers. These charts are but two examples of output generated by genetic research systems. Other charts, and output generally, is readily produced by other specialized data analysis modules 28.

Figure 6 illustrates a programmable computing system (system) 100 that provides an operating environment suitable for use as a research computer 10 or as a server within genetic research system 8. The system 100 includes a processor 112 that in one embodiment belongs to the PENTIUM[®] family of microprocessors manufactured by the Intel Corporation of Santa Clara, California. However, the invention can be implemented on computers based upon other microprocessors, such as the MIPS[®] family of microprocessors from the Silicon Graphics Corporation, the POWERPC[®] family of microprocessors from both the Motorola Corporation and the IBM Corporation, the PRECISION ARCHITECTURE[®] family of microprocessors from the Hewlett-Packard Company, the SPARC[®] family of microprocessors from the Sun Microsystems Corporation, or the ALPHA[®] family of microprocessors from the Compaq Computer Corporation. In various configurations, system 100 represents any server, personal computer, laptop or even a battery-powered, pocket-sized, mobile computer known as a hand-held PC or personal digital assistant (PDA).

System 100 includes system memory 113, including read only memory (ROM) 114 and random access memory (RAM) 115, which is connected to the processor 112 by a system data/address bus 116. ROM 114 represents any device that is primarily read-only including electrically erasable programmable read-only memory (EEPROM), flash memory, etc. RAM 115 represents any random access memory such as Synchronous Dynamic Random Access Memory.

Within the system 100, input/output bus 118 is connected to the data/address bus 116 via bus controller 119. In one embodiment, input/output bus 118 is implemented as a standard Peripheral Component Interconnect (PCI) bus. The bus controller 119 examines all signals from the processor 112 to route the signals to the appropriate bus. Signals between the processor 112 and the system memory 113 are merely passed through the bus controller 119. However, signals from the processor 112 intended for devices other than system memory 113 are routed onto the input/output bus 118.

Various devices are connected to the input/output bus 118 including hard disk drive 120, floppy drive 121 that is used to read floppy disk 151, and optical drive 122, such as a CD-ROM drive that is used to read an optical disk 152. The video display 124 or other kind of display device is connected to the input/output bus 118 via a video adapter 125.

Users enter commands and information into the system 100 by using a keyboard 140 and/or pointing device, such as a mouse 142, which are connected to bus 118 via input/output ports 128. Other types of pointing devices (not shown) include track pads, track balls, joysticks, data gloves, head trackers, and other devices suitable for positioning a cursor on the video display 124.

System 100 also includes a modem 129 that may be internal or external to the system 100. The modem 129 is typically used to communicate over wide area networks (not shown), such as the global Internet using either a wired or wireless connection.

Software applications 136 and data are typically stored via one of the memory storage devices, which may include the hard disk 120, floppy disk 151, CD-ROM 152 and are copied to RAM 115 for execution. In one embodiment, however, software applications 136 are stored in ROM 114 and are copied to RAM 115 for execution or are executed directly from ROM 114.

In general, the operating system 135 executes software applications 136 and carries out instructions issued by the user. For example, when the user wants to load a software application 136, the operating system 135 interprets the instruction and causes the processor 112 to load software application 136 into RAM 115 from either the hard disk 120 or the optical disk 152. Once one of the software applications 136 is loaded into the RAM 115, it can be executed by the processor 112. In case of large software applications 136, processor 112 loads various portions of program modules into RAM 115 as needed.

The Basic Input/Output System (BIOS) 117 for the system 100 is a set of basic executable routines that have conventionally helped to transfer information between the computing resources within the system 100. Operating system 135 or other software applications 136 use these low-level service routines. In one embodiment system 100 includes a registry (not shown) that is a system database that holds configuration information for system 100. For example, the Windows[®] operating system by Microsoft Corporation of Redmond, Washington, maintains the registry in two hidden files, called USER.DAT and SYSTEM.DAT, located on a permanent storage device such as an internal disk.

The invention has been described in terms of particular embodiments. These and other embodiments are within the scope of the following claims.

What is claimed is:

1. A computer-readable storage medium having data structures stored thereon comprising:
 - a species data structure to store species data;
 - 5 a chromosome data structure to store chromosome data, wherein the chromosome data relates to one of the species; and
 - a gene data structure to store gene data, wherein the gene data relates to at least one entry in the chromosome data structure and at least one entry in the species data structure.
- 10 2. The computer-readable storage medium of claim 1 further comprising a marker data structure to store locus information, wherein the marker information relates to an entry in the chromosome data structure and an entry in the gene data structure, and further wherein the marker information identifies a locus relating to a
- 15 particular chromosome.
3. The computer-readable storage medium of claim 2 further comprising an allele data structure to store allele information identifying variants, wherein the variant relates to at least one entry in the marker data structure that identifies a particular
- 20 locus on one of the chromosomes.
4. The computer-readable storage medium of claim 3 further comprising a genotype data structure to store observed data values for a pair of alleles for one of the loci identified by one of the markers in the marker data structure.
- 25 5. The computer-readable storage medium of claim 1 further comprising a sampling unit data structure to store an identification of one of the species within the species data structure.

6. The computer-readable storage medium of claim 5 further comprising an individual data structure to store individual data including at least one of a unique identity, an alias, a father, a mother, a gender, a birth date, and further wherein the individual data relates to an entry within the sampling unit data structure.
- 5
7. The computer-readable storage medium of claim 6 further comprising a grouping data structure to store a name of a grouping, wherein the grouping is a family or a generation.
- 10
8. The computer-readable storage medium of claim 7 further comprising a group data structure to store data entries that identify a set of individuals within the individual data structure, wherein an entry within the group data structure relates to an entry within the grouping data structure.
- 15
9. The computer-readable storage medium of claim 6 further comprising a sample data structure to store information identifying a biological sample taken from an individual including tissue type, date, treatment, experiment, storage, wherein an entry within the sample data structure relates to an entry within the individual data structure.
- 20
10. The computer-readable storage medium of claim 6 further comprising a variable data structure to store identifiers for phenotype variables, wherein each entry within the variable data structure relates to an entry within the species data structure.
- 25
11. The computer-readable storage medium of claim 10 further comprising a phenotype data structure to store an observed value, wherein an entry within the phenotype data structure relates to an entry within the individual data structure and an entry within the variable data structure.
- 30

12. The computer-readable storage medium of claim 5 further comprising a project data structure to store information of one or more research projects, wherein the project data structure has a many to many relationship with the species data structure and a many-to-many relationship with the sampling unit data structure.
- 5
13. The computer-readable storage medium of claim 12 further comprising a role data structure and a privileges data structure to store project access information, wherein the role data structure has a many-to-many relationship with the project data structure.
- 10
14. The computer-readable storage medium of claim 13 further comprising a user data structure to store information of a genetic researcher including at least one of a unique identity, a password and a full name, wherein an entry within the user data structure relates to an entry within the access data structure.
- 15
15. The computer-readable storage medium of claim 12, wherein the project data structure stores information defining one or more project views, wherein each view defines access to one or more of the other data structures according to an access control hierarchy.
- 20
16. The computer-readable storage medium of claim 15, wherein a first level of the access control hierarchy defines whether a user can access to a sampling unit data structure, a second level of the hierarchy defines access to an individual data structure, wherein the third level defines access to a groupings data structure, wherein the fourth level defines access to a species data structure, wherein the fifth level defines access to the chromosomes data structure, wherein the sixth level defines access to the genes data structure, wherein the seventh level defines access to a markers data structure, wherein the eighth level defines access to a variables data structure.
- 25
- 30

17. The computer-readable storage medium of claim 13, wherein the privileges data structure stores access rights for a corresponding entry in the role data structure, wherein the access types include write access, read access and no access, and further wherein the access rights define access to entries within the other data structures such that the access rights and project views collectively define whether a user can manipulate a particular entry within the data structures.
18. A system comprising:
- a web server to receive genetic research data from a user;
 - a database communicatively coupled to the web server and configured to store the genetic research data, wherein the database conforms to a database schema defining a plurality of database objects including:
 - a species data object to store species data;
 - a chromosome data object to store chromosome data, wherein the chromosome data relates to an entry within the species data object; and
 - a project data object to store information of one or more research projects, wherein each entry in the project data object relates to one or more entries in the species data structure.
19. The system of claim 18 further comprising a role data object and a privileges data object to store project access information, wherein each entry in the role data object relates to at least one entry in the project data object.
20. The system of claim 19 further comprising a user data object to store information of a genetic researcher including at least one of a unique identity, a password and a full name, wherein an entry within the user data object relates to an entry within the role data object.

21. The system of claim 18, wherein the project data object stores information defining one or more project views, wherein each view defines access to each of the other data objects according to an access control hierarchy having a plurality of levels.
- 5
22. The system of claim 21, wherein the access control hierarchy includes:
- a first level defining access to a sampling unit data object,
 - a second level defining access to an individual data object as a function of the first level,
 - 10 a the third level defining access to a groupings data object as a function of the first level.
23. The system of claim 21, wherein the access control hierarchy includes:
- a first level defining access to the species data object,
 - 15 a second level defining access to the chromosomes data object as a function of the first level,
 - a third level defining access to the genes data object as a function of the second level, and
 - a fourth level defining access to a markers data object as a function of the second level.
- 20
24. The system of claim 20, wherein the role data object and the privileges data object stores access rights for a corresponding entry in the user data object, wherein the access types include write access, read access and no access, and further wherein
- 25 the access rights define whether the corresponding user can access particular entries within the other data objects such that the access rights and project views collectively define whether the user can manipulate a particular entry within the data objects.

25. The system of claim 20, wherein the database scheme defines a gene data object to store gene data, wherein the gene data relates to at least one entry in the chromosome data object and at least one entry in the species data object.
- 5 26. The system of claim 20, wherein the database scheme defines a marker data object to store locus information, wherein the marker information relates to an entry in the chromosome data object and an entry in the gene data object, and further wherein the marker information identifies a locus relating to a particular chromosome.
- 10 27. The system of claim 20, wherein the database scheme defines an allele data object to store allele information identifying variants, wherein the variant relates to at least one entry in the marker data object that identifies a particular locus on one of the chromosomes.
- 15 28. The system of claim 20, wherein the database scheme defines a genotype data object to store observed data values for a pair of alleles for one of the loci identified by one of the markers in the marker data object.
- 20 29. The system of claim 20, wherein the database scheme defines a sampling unit data object to store an identification of one of the species within the species data object.
- 25 30. The system of claim 29, wherein the database scheme defines an individual data object to store individual data including at least one of a unique identity, an alias, a father, a mother, a gender, a birth date, and further wherein the individual data relates to an entry within the sampling unit data object.

31. The system of claim 29, wherein the database scheme defines a grouping data object to store a name of a grouping, wherein an entry within the grouping identifies a family or a generation and relates to an entry within the sampling unit data object.
- 5
32. The system of claim 31, wherein the database scheme defines a group data object to store data entries that identify a set of individuals within the individual data object, wherein an entry within the group data object relates to an entry within the grouping data object.
- 10
33. The system of claim 32, wherein the database scheme defines a sample data object to store information identifying a biological sample taken from an individual including tissue type, date, treatment, experiment, storage, wherein an entry within the sample data object relates to an entry within the individual data object.
- 15
34. The system of claim 20, wherein the database scheme defines a variable data object to store identifiers for phenotype variables, wherein each entry within the variable data object relates to an entry within the species data object.
- 20
35. The system of claim 33, wherein the database scheme defines a phenotype data object to store an observed value, wherein an entry within the phenotype data object relates to an entry within the individual data object and an entry within the variable data object.
- 25
36. A method for providing access to a genetic research database comprising:
receiving a request from a user to access one of a set of genetic research objects within the database, wherein the set of genetic research objects store genetic research data;
querying a project view object within the database to determine which
30 entries within the genetic research objects the user can access;

querying a role data object and a privileges data object within the database to determine a set of operations that the user is allowed to perform; and permitting the requested access based the accessible objects and the allowed operations.

5

37. The method of claim 36, wherein the project view object hierarchically relates the genetic research objects, and further wherein querying the project view object includes determining the accessible genetic research objects at each level of the hierarchy.

10

1/5

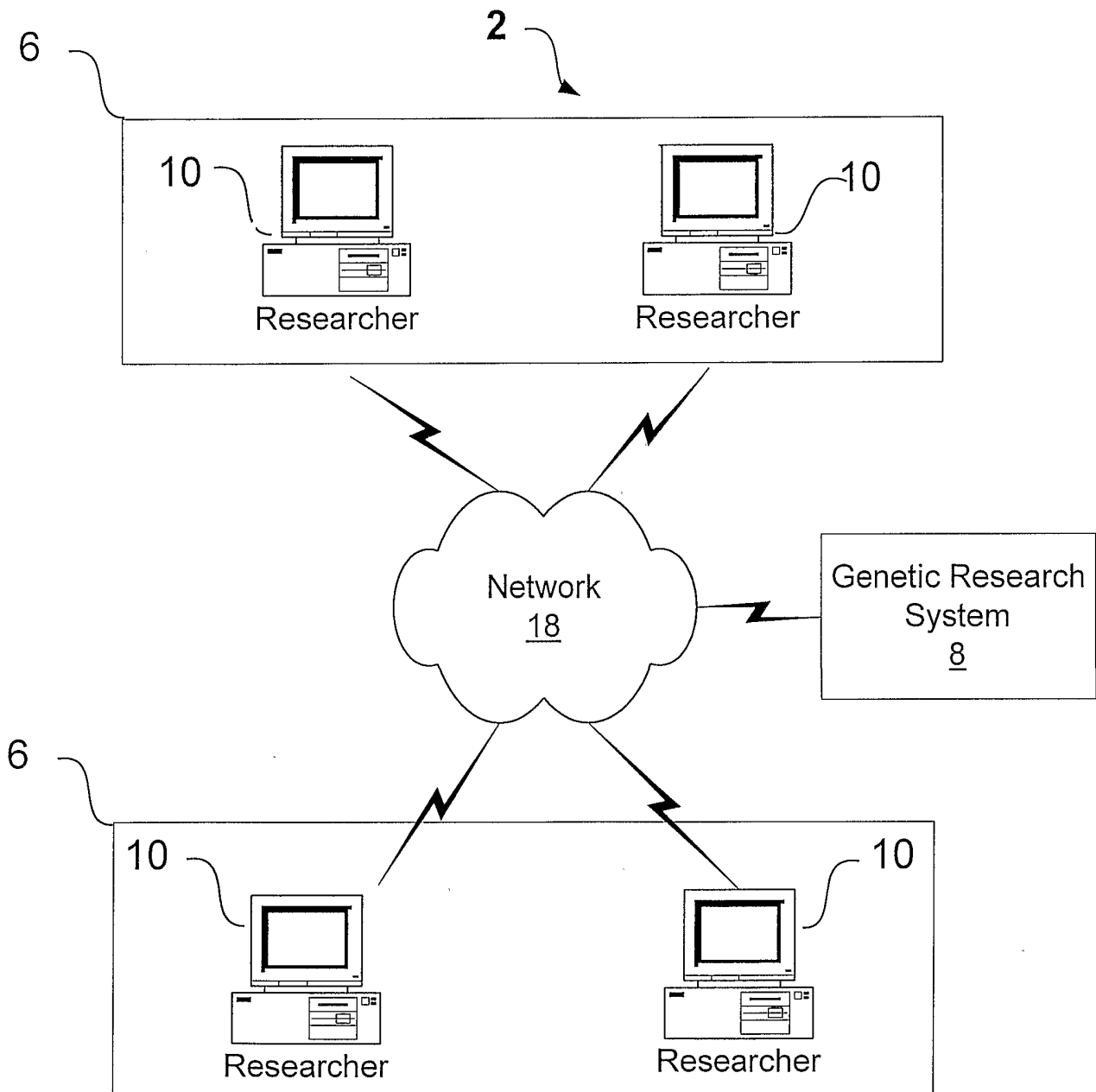


FIG. 1

2/5

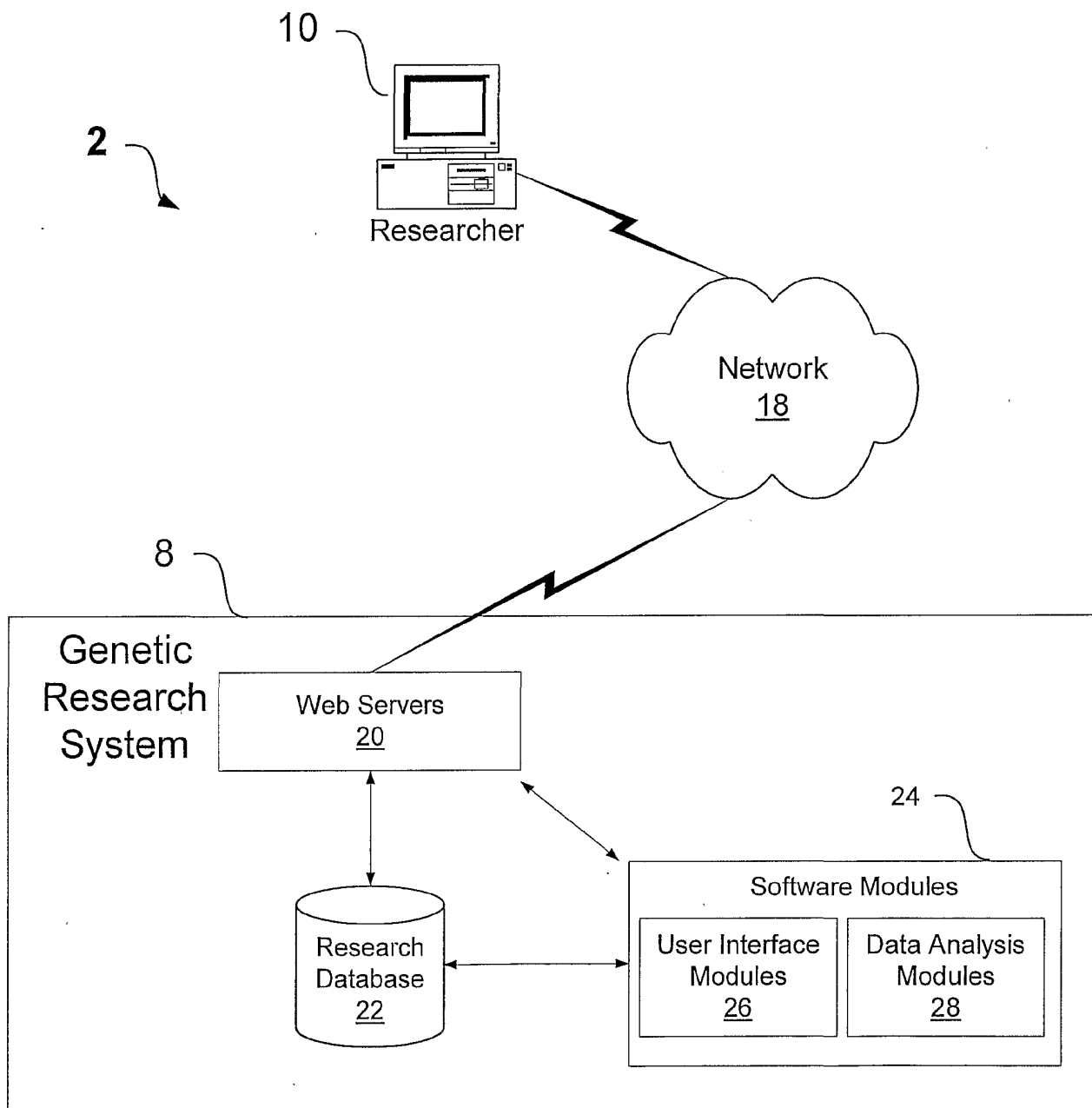


FIG. 2

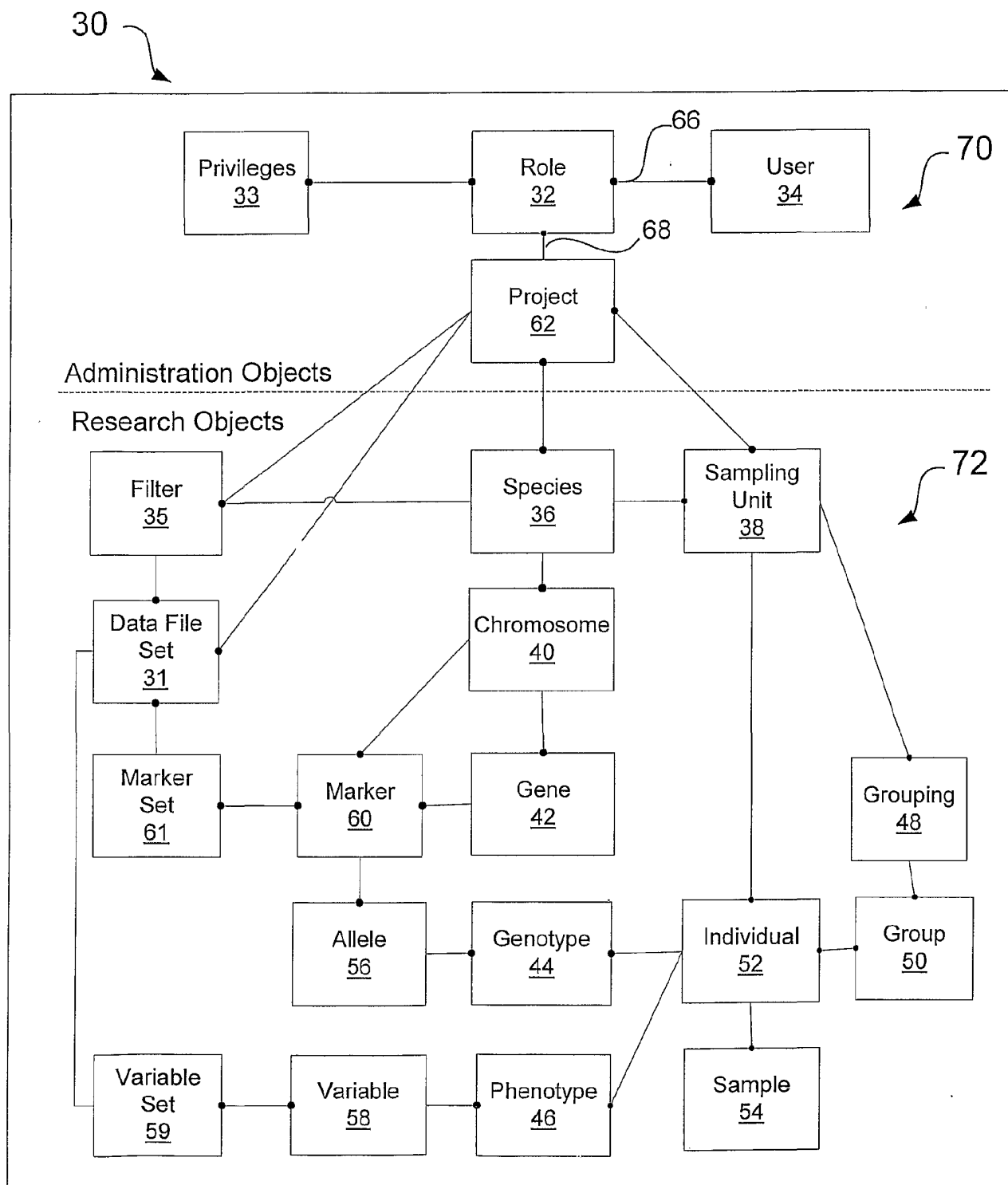


FIG. 3

4/5

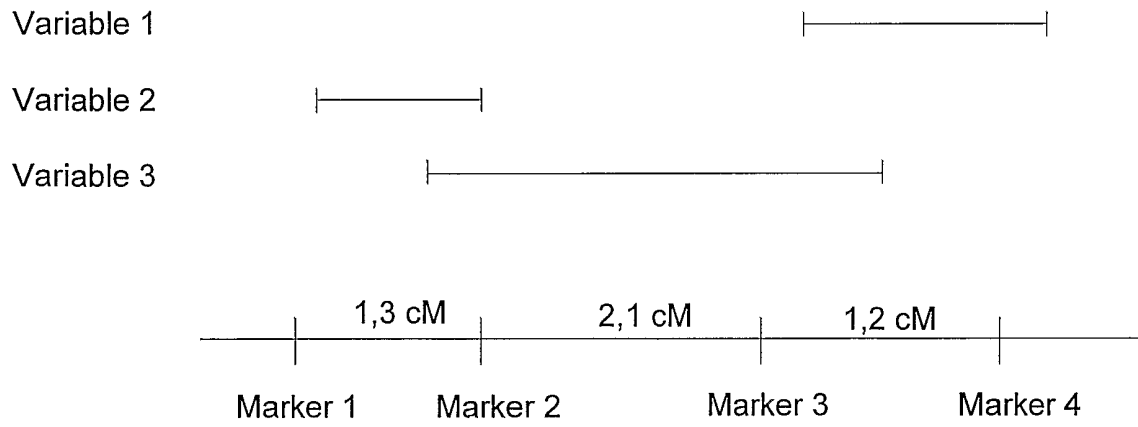


FIG. 4

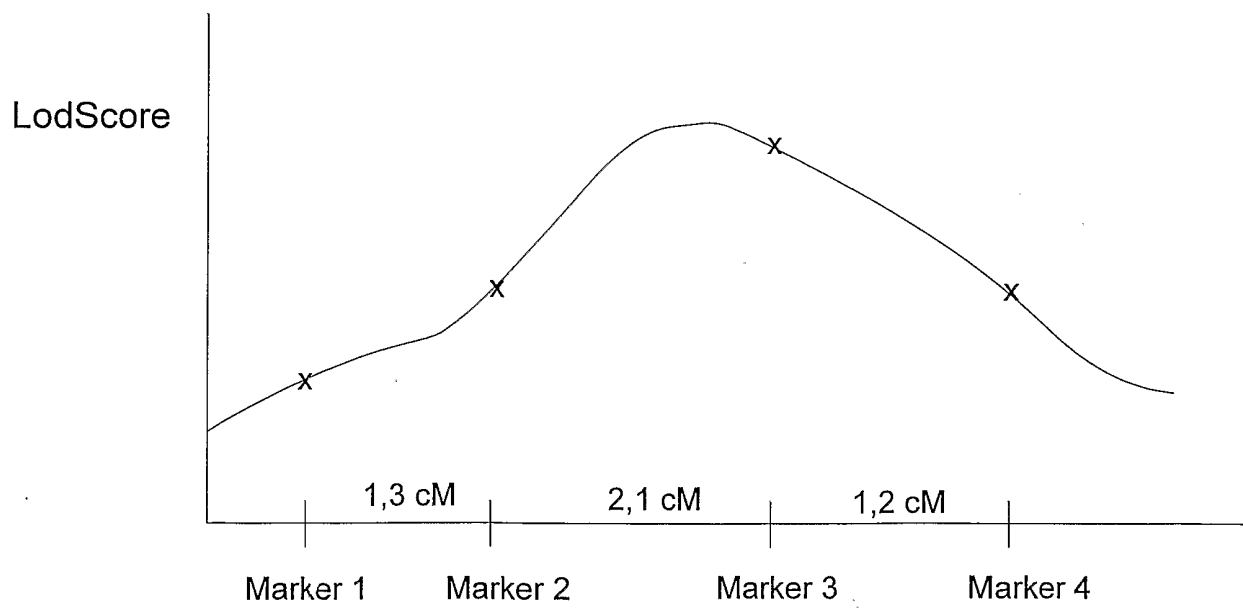


FIG. 5

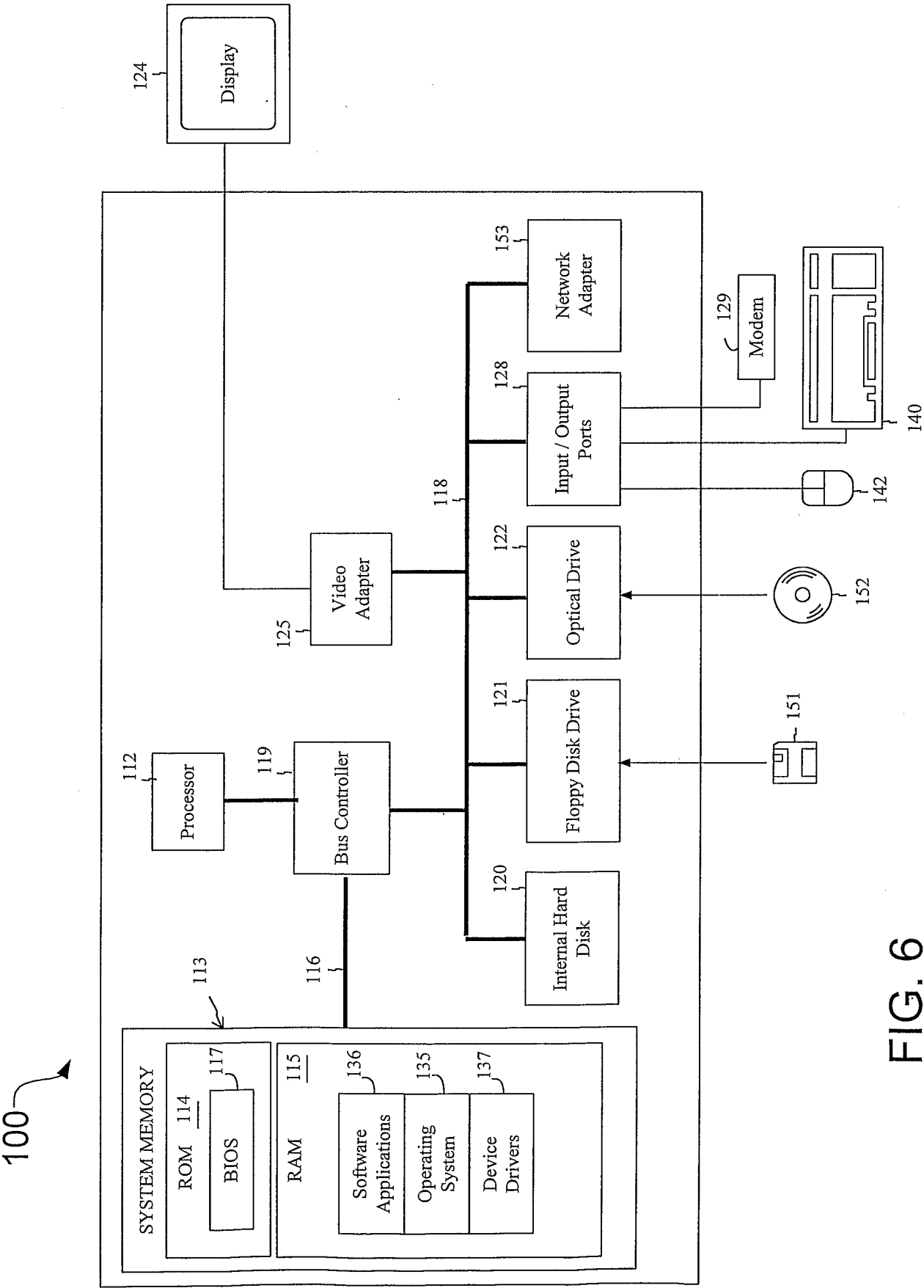


FIG. 6