

ISTITUTO
CSS-MENDEL
ROMA



Portal for the management of clinical data

Creation of an archive of patients with Joubert syndrome

di

Lorenzo Martucci, Claudia Raponi, Luca Tomaselli

June 29, 2014

This project was developed jointly by the Roma Tre University and the Institute C.S.S. Mendel.
The aim of project is the creation of a web portal server for the management of clinical data on patients with the Syndrome Joubert, in the context of biomedical informatics.

Contents

I	Introduction	3
1	Case study: Joubert Syndrome	4
1.1	About the disease	4
1.2	Contribution of the Institute C.S.S. Mendel	5
1.3	Clinical data	5
2	Main features of the project	9
2.1	Aims	9
2.2	Functions	9
2.3	Tools	10
II	Implementation	11
3	Structure	12
4	Interface	13
5	Model	14
6	Persistence	15
6.1	Setting database	15
6.1.1	Script	17
6.2	Loading data	18
6.2.1	Loading of a single sample	18
6.2.2	Loading of an entire sequencing	20

Part I

Introduction

Chapter 1

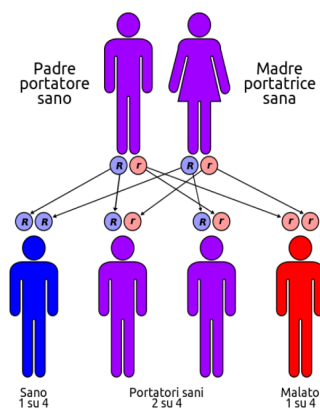
Case study: Joubert Syndrome

1.1 About the disease

Joubert syndrome is a rare genetic disorder that affects the cerebellum, which is the portion of the brain responsible for controlling balance and coordination [1]. The most common symptoms that indicate the presence of this disease are:

- **ataxia**: loss of muscle coordination;
- **hyperpnea**: increase in the depth of breathing compared to the norm;
- **sleep apnea**: absence of external breathing or pauses in breathing of more than 15 seconds;
- abnormal eye and tongue movements;
- **hypotonia**: a reduction in muscle tone of an organ or part of it;

Joubert Syndrome is transmitted through an autosomal recessive trait from both parents carriers.



They are generally referred to as "carriers" for individuals who have a given gene, one healthy and one mutated allele, therefore, with the possibility of transmitting a defective allele to her children.

In autosomal recessive inheritance, 25% of children of two heterozygous parents manifest the character and there is the same probability of transmission to both sons at the end that females.

The prognosis for this disease is very variable. There are cases where patients have minimal disabilities monotorie but who have a good mental development. However, there are other cases in which individuals have a discrete mental retardation and muscle problems more obvious.

Treatment for Joubert syndrome is symptomatic and supportive. In the infant stage it is possible to intervene with stimulation both physical and occupational. Other supportive treatments can affect speech therapy and monitoring of respiration particularly in the neonatal stage. In the absence of a proper diagnosis and appropriate treatment, often leads to death in the second decade of life.

1.2 Contribution of the Institute C.S.S. Mendel

The Institute C.S.S. Mendel is concerned with the study of Joubert syndrome in particular as regards the scope of molecular genetics. Molecular genetics is the field of genetics that focuses on the study of the structure and function of genes at a molecular level. The surveys consist of molecular genetics in the study of DNA and its products, RNA and proteins, whose changes may be related to or responsible for a particular genetic condition.

The study of mutation note on Joubert syndrome by sequencing is one of the main objects of research and analysis of the Institute C.S.S. Mendel since 2003 and in recent years has become a European reference center for the study of these syndromes, receiving case studies from around the world and praising numerous collaborations and partnerships [2].

In particular with the assistance of the University of California San Diego, University of Leeds and the Hospital Necker in Paris, researchers from the Instituto CSS Mendel developed a scientific work inherent in a new gene.

The study shows that the new gene identified stops the activity of cellular structures-like antennae, called "cilia", which lose their ability to transmit and recognize information both for embryonic development and for the smooth functioning of various organs. Joubert syndrome is part of a large group of diseases called "ciliopathies", as the cilia on the cell surface do not follow a proper function is not responding appropriately to the signals.

This failure prevents the formation of the embryo proper neural tube, resulting in abnormal development of the central nervous system in children with Joubert syndrome, almost invariably lead to the development of mental retardation of medium to severe. Other anomalies may also affect the retina, the kidneys, the liver and the skeletal system.

1.3 Clinical data

Clinical data that we have been provided by the Institute C.S.S. Mendel relate to an entire sequencing. A sequencing, in the context of molecular biology, is

the process of determining the correct order of nitrogenous bases in a nucleic acid. Each sequencing is constituted by a set of samples and each of these is related to a single patient. A sample has several fields that are divided into an *attributes on the generality of the patient*, an *attributes on specific phenotypic traits of patients* and an *attributes on notes and diagnosis*.

The attributes on the generality of the patient are:

- **FAMILY**: family code;
- **NG**: patient code;
- **YOB**: year of born (that we will not consider in queries);
- **last evaluation** (that we will not consider in queries);
- **sex**: gender of the patient;
- **consang**: presence of consanguineous;

Some partial data relating to the attributes on the generality of the patient are:

2	FAMILY▼	NG▼	YOB▼	last evaluation▼	sex▼	consang▼
87	COR064	1779	1989		f	
88	COR064	1777	1996		m	
89	COR065					
90	COR66	1641	1999		m	n
91	COR67		feto		x	
92	COR68	1690	1996	9	m	x
93	COR069	1692	1988		m	n
94	COR70	1696	2004		m	n
95	COR72	1701	1999		f	y
96	COR74	1830	1990		m	n
97	COR75	1835	2002		m	n
98	COR77	1827	2000		m	n
99	COR79	1838	1990		f	x
100	COR80	1870	2004		m	n
101	COR81	1979	2000		f	y
102	COR82	1884	1988		f	n
103	COR85	1894			f	n
104	COR86	1895	2003		f	x
105	COR87	1898	2004		m	n
106	COR89	1921	2003		m	n
107	COR90	1926	2005		f	n
108	COR 91	1928			m	
109	COR 92	1932	1979		m	n
110	COR 93	1936	1993		m	n
111	COR 95	1943	1999		m	n

Instead the attributes on specific phenotypic traits of patients are divided on the basis of the main traits and specific characters. There are nine main sections and only some of these are analyzed for specific characters. Overall, these are:

- **CNS**: central nervous system;
 - breath;
 - ID: infectious disease(?);
 - hypotonia;
 - ataxia;
 - apraxia;
 - nystagmus;

- **eyes;**
 - leber amaurosis;
 - retinopathy;
 - coloboma;
- **kidneys;**
 - renal failure;
 - NPH: nephronophthisis;
 - cystis;
 - eco/blood alterations;
- **liver;**
 - eco/blood alterations;
 - HF: heart failure(?);
- **polydactyly;**
 - polydactyly postaxial;
 - polydactyly mesa-preaxial;
- **tongue;**
 - cleft lip/palate;
- **heart;**
- **dysmorphic features;**
- **MTI: molar tooth imaging;**
 - E/M cele: encephalo-meningocele;
 - hydroceph;
 - DW: Dandy Walker Malformation;
 - CC hypopl: Corpus Callosum hypoplasia;

Some partial data relating to the attributes on specific phenotypic traits of patients are:

	FAMILY	KIDNEYS	Renal failure	NPH	Cystis	Eco/blood alterations	LIVER	Eco/blood alterations	HF
2									
3	COR000	y	y	n	n	y	y	y	n
4	COR000	y	y	y	n	y	x	x	x
5	COR000	n	n	n	n	n	n	n	n
6	COR000	y	y	y	n	y	n	n	n
7	COR001	y	y	y	n	y	n	n	n
8	COR002	y	y	y	n	y	x	x	x
9	COR002	y	y	y	y	y	n	n	n
10	COR003	y	n	n	y	y	n	n	n
11	COR004	y	n	y	n	y	n	n	n
12	COR004	y	y	y	y	y	n	n	n
13	COR005	n	n	n	n	n	n	n	n
14	COR006	n	n	n	n	n	n	n	n
15	COR007	n	n	n	n	n	x	x	x
16	COR007	n	n	n	n	n	n	n	n
17	COR008	y	n	y	n	y	x	x	x
18	COR009	n	n	n	n	n	y	y	y
19	COR010	n	n	n	n	n	y	y	n
20	COR010	n	n	n	n	n	n	n	n
21	COR011	y	n	n	n	y	n	n	n
22	COR012	n	n	n	n	n	n	n	n

The last fields of clinical data are additional information and refer to:

- notes;
- diagnosis;

Some partial data relating to the attributes on notes and diagnosis are:

2	FAMILY ▼	Notes	Diagnosis ▼
27	COR016		
28	COR017		
29	COR018	micropenis	Varadi-Papp syndrome
30	COR019		Joubert+retinal
31	COR020		
32	COR020		
33	COR021		
34	COR022		
35	COR023	sindattilia 2°-3° piedi bilateralmente	classical Joubert
36	COR024	Deafness	
37	COR024		
38	COR025		
39	COR026		
40	COR027		
41	COR027		
42	COR027	macrocefalia	
43	COR028	osti aracnoidea	classical Joubert
44	COR029		classical Joubert
45	COR030	Iperplasia gengivale,atrofia ottica	Joubert+retinal
46	COR030	Iperplasia gengivale,atrofia ottica	
47	COR031		
48	COR032	Rene destro polistico	
49	COR032		

Chapter 2

Main features of the project

2.1 Aims

The aim of the project is to create an archive containing people with Joubert Syndrome. This data file consist of the general characteristics of each patient and their phenotypic traits. The main purpose of this archive is to facilitate the association genotype/phenotype of patients under observation at the Institute C.S.S. Mendel. This report make it easier to identify the pathogenetic genetic variants.

Each patient manifests different phenotypic traits according to the degree of intensity of the syndrome. People affected by a mild form of the syndrome present problems relating to only some physical components, while patients with more intense forms of the syndrome present the other. The association that is created by the definition of the archive identify the physical components (and hence phenotypic) most relevant to recognize the presence of Joubert Syndrome.

The archive is available through a web portal server that allows different interactions with the clinical data of the store.

2.2 Functions

The portal is used by two types of users:

- *admin*: person with the privileges that may adopt measures on the management of the archive;
- *simple user*: person who can only use the portal;

Based on the user's role distinguish different application functionality.

Both types of users can browse the archives through the queries. This function allows users of the portal to view the clinical data on patients who meet the criteria set in the request. In this way, the user sets a query that specifies the parameters that wants to filter. The portal respond to the user by providing the clinical data of individual patients who possess those parameters. The user sees only the general phenotypic characteristics of each patient and then may possibly view traits more specific. In reference to the previous image (mettere numeri alle immagini), the user sees as a possible liver problem, and only then

can see that it is particularly eco/blood alteration problem.

A simple user can only perform simple queries. An admin user not only uses the portal to search clinical data, but also to perform other operations. Among these features, an admin can:

- load clinical data;
 - load an entire sequencing from an Excel file;
 - load a single sample;
- modify existing samples;

The admin user has the skills and privileges to interact with the store in a more specific way. This allows him to intervene in the structural details of the store and make decisions about its organization.

2.3 Tools

The development of the project was supported by the use of different tools:

- **PHP**: server-side scripting language designed for web development but also used as a general-purpose programming language [3];
- **Javascript** (JS): dynamic computer programming language [4];
- **Bootstrap**: front-end framework for developing responsive projects on the web [5];
- **Apache HTTP Server**: web server application [6];
- **MySQL**: open-source relational database management system (RDBMS) [7];

The relevant part of the programming made use of PHP language for writing web interfaces of the portal while JS has been exploited for checking the integrity of the fields of data-entry form or querying. The coding was developed by the **Eclipse IDE**, adapted of the needs of shared work. For these reasons, the editor was supported by suitable plugin that allowed both writing in the languages mentioned above and the management of the project among the various members of the group.

The server side is managed through the use of the Apache HTTP Server, the platform that allows the implementation of the functions of transport and connection information. Web server has permission to provide, by means of dedicated software and user's request, files of any type, such as the web pages of the portal.

The management of persistence was finally entrusted to MySQL. This RDBMS allowed to organize in ordered structures clinical data to manage. He also encouraged interaction with the clinical data, facilitating the handling of the queries.

Part II

Implementation

Chapter 3

Structure

- da definire-

Chapter 4

Interface

- da definire-

Chapter 5

Model

- da definire-

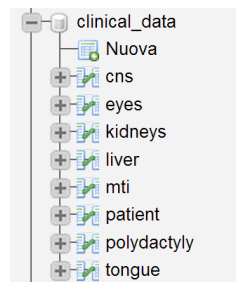
Chapter 6

Persistence

The management of persistence has been supported by the use of the MySQL RDBMS. Its use has facilitated the operations of interactions with clinical data that make up the data base. In fact, all the features that the portal offers require an exchange of information that requires a collaborative work with persistent data. For this reason all the components involved in the persistence have a close correspondence with the operations that a user can take. We analyze below how we have implemented operations involving the storage and management of data through MySQL.

6.1 Setting database

The organization of information through MySQL has been set so as to corresponding clinical data were provided by Institute C.S.S. Mendel. As shown in figure, we have organized the data in nine tables, one of which is related to patient information, one of these for data administrators while the rest have been introduced in order to handle specific aspects of each character of the observed occurrence of the disease.



Patient is the entity described by both attributes on the generality of the patient that the attributes on notes and diagnosis. Each tuple of patient entity is characterized by fields (for each attribute is specified the associated type):

- **ng** - int(11);
- **insertionDate** - timestamp;

- **family** - varchar(10);
- **sex** - char(1);
- **consang** - char(1);
- **cns** - char(1);
- **eyes** - char(1);
- **kidneys** - **char(1)**;
- **liver** - char(1);
- **polydactyly** - char(1);
- **tongue** - char(1);
- **heart** - char(1);
- **dysmorphic** - char(1);
- **mti** - char(1);
- **notes** - varchar(300);
- **diagnosis** - varchar(100);

The primary key for the entity *Patient* consists in ng and insertionDate. The attributes related to the physical characteristics (from the cns to mti) that are taken into account in the analysis of the disease, are described by a single char that can take three values:

- y: yes, the patient presents problems for that character observed;
- n: no, the patient does not present problems for the character observed;
- x or no value: it is not known if the patient develops problems observed for the character;

For some of these traits have been observed also create additional entity to manage the specific traits that are affected. These additional tables are available: *CNS*, *Eyes*, *Kidneys*, *Liver*, *MTI*, *Polydactyl* and *Tongue*. Each of these is described by all of the details observed regarding the character of reference, and attributes ng and insertionDate which constitute the primary key. For example, the entity *Eyes* is described by attributes:

- **ng** - int(11);
- **insertionDate** - timestamp;
- **leber_amaurosis** - char(1);
- **retinopathy** - char(1);
- **coloboma** - char(1);

In the same way before each specific trait may have values: y,n,x (or no value). The relationships between the different entities are managed by referential constraints between the attributes `ng` and `insertionDate` of each tuple.

Finally, the entity *Administrators* allows to manage information relating to people authorized to make changes to the database. In particular the tuples of this entity are characterized by attributes only:

- **name** - `varchar(32)`;
- **password** - `varchar(32)`;

These fields form the primary key for the administrator. The presence of this entity is needed to manage authentication of users when using the portal. Only administrative users that have certain privileges can interact with data in a more specific way. It's therefore necessary to be able to detect the user's identity through identification based on username and password. Then an administrator must already be present in the archive.

6.1.1 Script

The script used in MySQL required to define the different entities is partially shown below. There are no reported all the scripts for creating the tables for the characters in which the disease manifests itself; it has been reported that only the entity *Eyes*. The entities missing are generated in a similar way, by modifying the various attributes.

Table structure *Administrator*

```
CREATE TABLE IF NOT EXISTS 'administrators' (
  'name' varchar(32) NOT NULL,
  'password' varchar(32) NOT NULL,
  PRIMARY KEY ('name','password')
) ENGINE=InnoDB DEFAULT CHARSET=latin1;
```

This entity must necessarily be populated by administrative users. The dump data to the table *Administrator* is:

```
INSERT INTO 'administrators' ('name', 'password') VALUES
('claudia', 'claudia'),
('lorenzo', 'lorenzo'),
('luca', 'luca'),
('tommaso', 'tommaso');
```

Table structure *Patient*

```
CREATE TABLE IF NOT EXISTS 'patient' (
  'ng' int(11) NOT NULL,
  'insertion_date' timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP
    ON UPDATE CURRENT_TIMESTAMP,
  'family' varchar(10) NOT NULL,
  'sex' char(1) NOT NULL,
  'consang' char(1) NOT NULL,
```

```

'cns' char(1) NOT NULL,
'eyes' char(1) NOT NULL,
'kidneys' char(1) NOT NULL,
'liver' char(1) NOT NULL,
'polydactyly' char(1) NOT NULL,
'tongue' char(1) NOT NULL,
'heart' char(1) NOT NULL,
'dysmorphic' char(1) NOT NULL,
'mti' char(1) NOT NULL,
'notes' varchar(300) NOT NULL,
'diagnosis' varchar(100) NOT NULL,
PRIMARY KEY ('ng','insertion_date')
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

```

Table structure *Eyes*

```

CREATE TABLE IF NOT EXISTS 'eyes' (
'ng_eyes' int(11) NOT NULL,
'insertion_date_eyes' timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP
ON UPDATE CURRENT_TIMESTAMP,
'leber_amaurosis' char(1) NOT NULL,
'retinopathy' char(1) NOT NULL,
'coloboma' char(1) NOT NULL,
PRIMARY KEY ('ng_eyes','insertion_date_eyes')
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

```

6.2 Loading data

Loading data into the database follows the two modes provided by the functions that the portal offers the user: load a single sample or load an entire sequencing.

6.2.1 Loading of a single sample

An admin user can upload data is related to a whole load sequencing both the individual samples. Analyze below the two different loading modes.

It was initially defined a class *LoadRecord* inside of which have been implemented methods that allow the loading of a single sample associated with a single patient. In particular the *Load* function is responsible for defining the SQL strings that allow the rescue of the various fields of each table.

This function has allowed to define the set of attributes, that must manage to analyze the syndrome, in a data structure:

```

public static function Load( $record, $row=1) {

    $lr = new LoadRecord();
    $record = $lr->insertX($record);

    $log = new Log();
    list($FAMILY, $NG, $YOB, $last_evaluation, $sex, $consang,
         $CNS, $breath, $id, $hypotonia, $ataxia, $apraxia, $nystagmus,
         $EYES, $leber_amaurosis, $retinopathy, $coloboma,
         $KIDNEYS, $renal_failure, $nph, $Cystis, $eco_blood_alterations,

```

```

$LIVER, $eco_blood_alterations_liver, $hf,
$POLYDACTYLY, $postaxial, $mesa_preaxial,
$TONGUE, $cleft_lip_palate, $HEART, $DYSMORPHIC_FEATURES,
$MTI, $em_cele, $hydroceph, $dw, $cc_hypopl, $Notes, $Diagnosis) = $record;

$success = 0;

if($NG != null && is_numeric($NG)){
    $mysqli = new mysqli('localhost', 'root', '', 'clinical_data');
    $mysqli->autocommit(true);

    date_default_timezone_set("Europe/Rome");
    //$insertion_date = date("Y-m-d H:i:s");
    $insertion_date = date("Y-m-d");

    //riempie tabella patient

    //ATTENZIONE OVVIAMENTE NON VA BENE COSI', HO RISOLTO IL PROBLEMA TOGLIENDO LE NOTE
    if($NG == 1936)
        $Notes = "";

```

In the same function have also been defined SQL statements for the composition of the database.

In this way it was implemented the inclusion of information relating to patients:

```

$query= sprintf("INSERT INTO patient (ng, insertion_date,
    family, sex, consang, cns, eyes, kidneys, liver, polydactyly,
    tongue, heart, dysmorphic, mti, notes, diagnosis) VALUES
    (' %s', ' %s', ' %s', ' %s', ' %s', ' %s', ' %s', ' %s', ' %s', ' %s',
    ' %s', ' %s', ' %s', ' %s', ' %s')", $NG, "ciao", $insertion_date, $FAMILY,
    $sex, $consang, $CNS, $EYES, $KIDNEYS, $LIVER, $POLYDACTYLY,
    $TONGUE, $HEART, $DYSMORPHIC_FEATURES, $MTI, $Notes, $Diagnosis);

$dati=$mysqli->query($query);
if($dati) {
    $success++;
}
else {
    $log->write( "fallimentoPatient riga: $row" . "<br />");
}

```

In a similar way we have defined the inclusion of detailed clinical data relating to the main aspects of the disease:

```

if($success > 0) {
    $query= sprintf("INSERT INTO cns(ng_cns, insertion_date_cns,
    breath, id, hypotonia, ataxia, apraxia, nystagmus) VALUES
    (' %s', ' %s', ' %s', ' %s', ' %s', ' %s', ' %s', ' %s')", $NG, $insertion_date,
    $breath, $id, $hypotonia, $ataxia, $apraxia, $nystagmus);

    $dati=$mysqli->query($query);
    if($dati) {
        $success++;
    }
}

```

```

        else {
            $log->write( "fallimentoCNS riga: $row" . "<br />");
            [...]
        }
    }
    $mysqli->close();

```

In the [...] we proceeded in the same way to the management of data entry for the other generic aspects.

6.2.2 Loading of an entire sequencing

The portal also offers the ability to upload a sequencing that allows to insert into the database multiple samples simultaneously. These samples need to be collected in an Excel file according to a specific pattern of the composition of attributes. The interaction between the portal and the management of data via Excel spreadsheet was supported by **PHPExcel**.

PHPExcel is a "*Project providing a set of classes for the PHP programming language, which allow you to write to and read from different spreadsheet file formats, like Excel (BIFF) .xls, Excel 2007 (OfficeOpenXML) .xlsx, CSV, Libre/OpenOffice Calc .ods, Gnumeric, PDF, HTML, ... This project is built around Microsoft's OpenXML standard and PHP*"[8]. It is a library that has facilitated the reading of the Excel input file and allowed to extrapolate from these clinical and general information for the definition of the database. His contribution has been exploited in the class *Read Excel* in which the *Read* method was responsible for the loading the file:

```

public static function Read($path) {
    try {
        $inputFileType = PHPExcel_IOFactory::identify($path);
        $objReader = PHPExcel_IOFactory::createReader($inputFileType);
        $objPHPExcel = $objReader -> load($path);
    } catch (Exception $e) {
        die('Error loading file "' . pathinfo($path,
            PATHINFO_BASENAME) . '":' . $e -> getMessage());
    }

    // Get worksheet dimensions
    $sheet = $objPHPExcel -> getSheet(0);
    $highestRow = $sheet -> getHighestRow();
    $highestColumn = $sheet -> getHighestColumn();

    $rowAdd = 0;
    for ($row = 2; $row <= $highestRow; $row++) {
        // Read a row of data into an array
        $rowData = $sheet -> rangeToArray('A' . $row . ':' .
            $highestColumn . $row, NULL, TRUE, FALSE);

        $loadRecord = new LoadRecord();
        $rowAdd += $loadRecord->load($rowData[0], $row);
    }

    return $rowAdd;
}

```

Within this function has been taken over the previously described method. In this way, the loading of a entire sequencing invokes the method *LoadRecord* for loading a single sample.

The function *Read* takes as a parameter the path of traceability of the excel file to be loaded. An application of this function is shown in the class *FromExcelToDB* in which is set the loading data:

```
$log = new Log();
$log->emptyAll();

$path = 'db_clinico.xlsx';
$readExcel = new ReadExcel();
$recordAdd = $readExcel->read($path);

echo ("Sono stati aggiunti $recordAdd record" . "<br />");

echo ($log->read());
```

In this case, *'db_clinico.xlsx'* represents our input excel file to be loaded into the database.

Bibliography

- [1] http://en.wikipedia.org/wiki/Joubert_syndrome;
- [2] [http://www.css-mendel.it/ricerca/sindrome di joubert /](http://www.css-mendel.it/ricerca/sindrome_di_joubert/)
- [3] <http://en.wikipedia.org/wiki/PHP>;
- [4] <http://en.wikipedia.org/wiki/Javascript>;
- [5] <http://getbootstrap.com/>;
- [6] [http://en.wikipedia.org/wiki/Apache HTTP Server](http://en.wikipedia.org/wiki/Apache_HTTP_Server);
- [7] <http://en.wikipedia.org/wiki/MySQL>;
- [8] <https://phpexcel.codeplex.com/>;