Proposal

application

Nowadays, medical field became more and more significant in our daily life. This report is talking about the details of the designed app, which related to the medical domain. It can help the researchers found a variant gene group compared with the normal correct gene combination from our tremendous gene pool database Here are some descriptions based on the creativity, challenge, and novelty of the application domain.

creativity
As for creativity, it can be used by researchers and doctors to search for separate gene combination in the field ground on it works on the phone or portable equipment. It can also show the data from main real-time update database. Therefore, the user can check cloud database and view different version data and strain data. Additionally, researchers and doctors can improve or expand the database by their specific ways. When the client uses this app, researchers need visualized the gene sequence of a mutated strain through this app and check gene resistance then make some marks or notes， as well as transiting an appropriate interaction. Therefore, that is also a challenge for UI design.

comparison
Compared with this new app design, another method to innovation is going to be considered which current researcher did. for most lab researchers foster the bacteria in a cultural dish. After several months, it grows up, then researchers experiment several kinds of antibiotics on it in terms of researching the resistance to drugs. This method is more realistic. However, it can waste plenty of time and more difficulty for practical. If this kind of app can be created, the problem will be solved. Doctors can check the gene combination on the phone instead of experiment several times using several months. This method is more efficient.

challenge
Apart from creativity, there are still several challenges during the implementation of this technology. Firstly, the format of the file is FASTA, it is a kind of format that using for gene sequence, we have never used before. Therefore, we need to make an analysis to parse the FASTA file. Additionally, each specific piece of gene combination is diverse. Therefore there are several random possibility forms. This app has to include several kinds of operations which can support multiple gestures. This can make this app more convenient for users and it can also save as much as the time we can. Ultimately, as I mentioned before, this app should be connected with the Azure cloud which can be updated several specific gene combinations at a real-time.

novelty
In addition, this app concludes full of novelty in several fields. It can automatically generate Excel form which imports Jxl packet and create a bitmap, then read and analyze the FASTA documents which is in the format of a bitmap. This is kind of new innovation that did not appear before. There is also a kind of search website which called Genome(https://www.ncbi.nlm.nih.gov/genome/?term=streptococcus+pneumoniae). This website is a gene search website as well as our app. However, it just has the function of

searching. It can not generate anything that can help researchers to understand or study deeply. Our app does have this novel function.

The last thing I should mention is that this kind of app can use in several application domains. It can be used in the research institution, for the research on the resistance of micro-organisms. It can also use in the hospital. The doctor can check patient's gene through this app swiftly.

project requirements

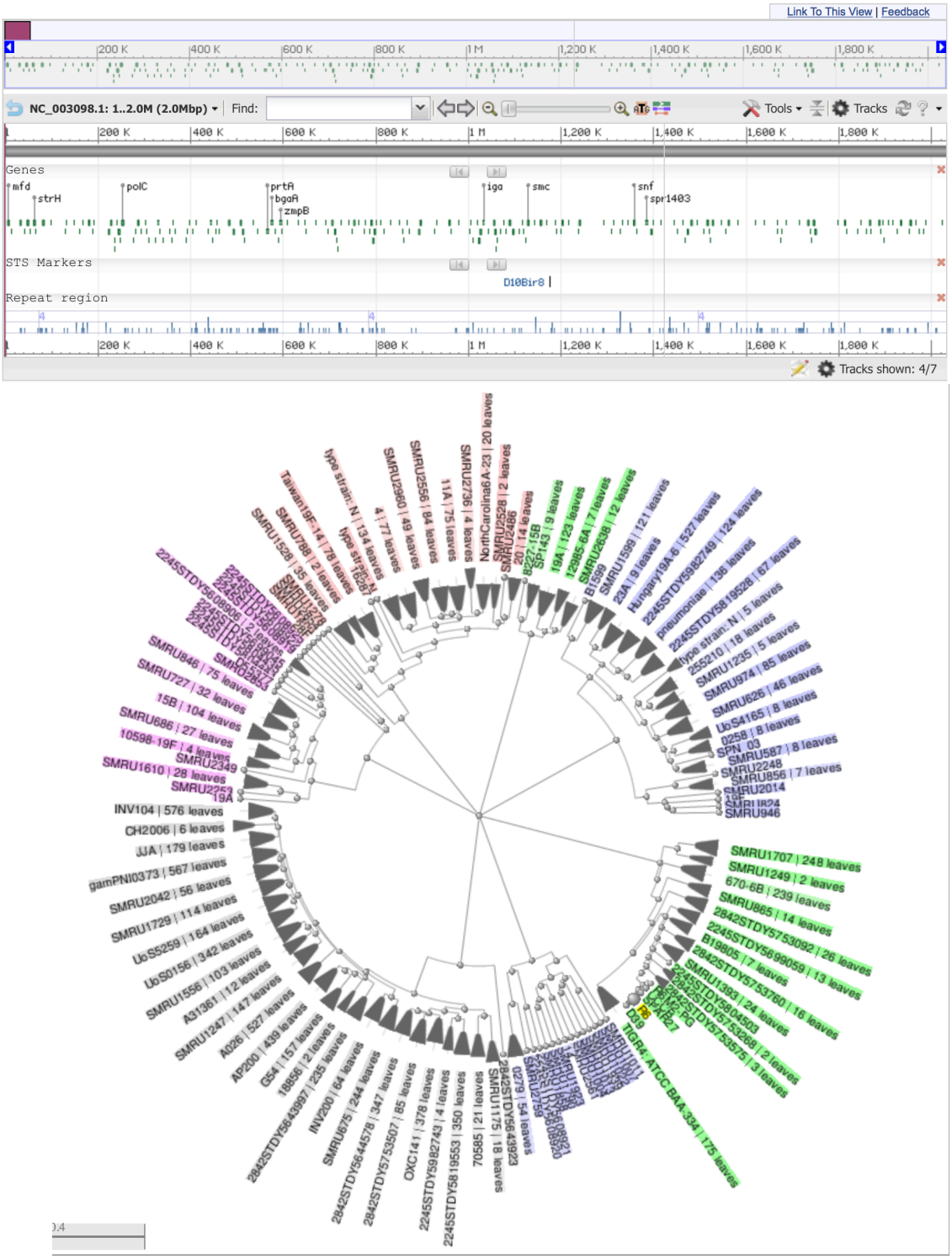
Requirement 1:

The special requirement for input is that the format of the input file is FASTA format, as shown in the figure below, The FASTA file contains the name of the strain and the sequence of genes of a specific location, one FASTA file could contain several different strain. The following example contains two samples SP49, SP52.

```
>sp49_8493_faa
MNKPTILRLIKYLSISFLSLVIAAIVLGGGVFFYYVSKAPSLSESKLVATTSSKIYDNKN
QLIADLGSERRVNAQANDIPTDLVKAIVSIEDHRFFDHRGIDTIRILGAFLRNLQSNSLQ
GGSTLTQQLIKLTYFSTSTSDQTISRKAQEAWLAIQLEQKATKQEILTYYINKVYMSNGN
YGMQTAAQNYYGKDLNNLSLPQLALLAGMPQAPNQYDPYSHPEAAQDRRNLVLSEMKNQG
YISAEQYEKAVNTPITDGLQSLKSASNYPAYMDNYLKEVINQVEEETGYNLLTTGMDVYT
NVDQEAQKHLWDIYNTDEYVAYPDDELQVASTIVDVSNGKVIAQLGARHQSSNVSFGINQ
AVETNRDWGSTMKPITDYAPALEYGIYDSTATIVHDEPYNYPGTNTPVYNWDRGYFGNIT
LQYALQQSRNVPAVETLNKVGLNRAKTFLNGLGIDYPSIHYSNAISSNTTESDKKYGASS
EKMAAAYAAFANGGTYYKPMYIHKVVFSDGSEKEFSNVGTRAMKETTAYMMTDMMKTVLS
YGTGRNAYLAWLPQAGKTGTSNYTDEEIENHIKTSQFVAPDELFAGYTRKYSMAVWTGYS
NRLTPLVGNGLTVAAKVYRSMMTYLSEGSNPEDWNIPEGLYRNGEFVFKNGARSTWNSPA
PQQPPSTESSSSSSDSSTSQSSSTTPSTNNSTTTNPNNNTQQSNTTPDMKLDKLFEKFLS
LFKKETSELEDSDSTILRRSRSDRKKLAQVGPIRKFWRRYHLTKIILILGLSAGLLVGIY
LFAVAKSTNVNDLQNALKTRTLIFDREEKEAGALSGQKGTYVELTDISKNLQNAVIATED
RSFYKNDGINYGRFFLAIVTAGRSGGGSTITQQLAKNAYLSQDQTVERKAKEFFLALELS
KKYSKEQILTMYLNNAYFGNGVWGVEDASKKYFGVSASEVSLDQAATLAGMLKGPELYNP
LNSVEDSTNRRDTVLKNMVAAGYIDKNQETEAAEVDMTSQLHDKYEGKISDYRYPSYFDA
VVNEAVSKYNLTEEEIVNNGYRIYTELDQNYQANMQIVYENTSLFPRAEDGTFAQSGSVA
LEPKTGGVRGVVGQVADNDKTGFRNFNYATQSKRSPGSTIKPLVVYTPAVEAGWALNKQL
DNHTMQYNSYKVDNYAGIKTSREVPMYQALAESLNLPAVATVNDLGVDKAFEAGEKFGLN
MEKVDRVLGVALGSGVETNPLQMAQAYAAFANEGLMPEAHFISRIENASGQVIASHKNSQ
KRVIDKSVADKMTSMMLGTFTNGTGISSSPADYIMAGKTGTTEAVFNPEYTSDQWVIGYT
PDVVISHWLGFPTTDENHYLAGSTSNGAAHVFRNIANTILPYTPGSTFTVENAYKQNGIA
PANTKRQVQTNDNSQTDDNLSDIRGRAQSLVDEASRAISDAKIKEKAQTIWDSIVNLFRM
RKFNSHSIPIRLNLLFSIVILLFMTIIGRLLYMQVLNKDFYEKKLASASQTKITSSSARG
EIYDASGKPLVENTLKQVVSFTRSNKMTATDLKETAKKLLTYVSISSPNLTERQLADYYL
ADPEIYKKIVEALPSEKRLDSDGNRLSESELYNNAVDSVQTSQLNYTEDEKKEIYLFSQL
NAVGNFATGTIATDPLNDSQVAVIASISKEMPGISISTSWDRKVLETSLSSIVGSVSSEK
AGLPAEEAEAYLKKGYSLNDRVGTSYLEKQYEETLQGKRSVKEIHLDKYGNMESVDTIEE
GSKGNNIKLTIDLAFQDSVDALLKSYFNSELENGGAKYSEGVYAVALNPKTGAVLSMSGI
KHDLKTGELTPDSLGTVTNVFVPGSVVKAATISSGWENGVLSGNQTLTDQSIVFQGSAPI
NSWYTQAYGSFPITAVQALEYSSNTYMVQTALGLMGQTYQPNMFVGTSNLESAMEKLRST
FGEYGLGTATGIDLPDESTGFVPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANNGVR
VAPRIVEGIYGNNDKGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSGLTTG
RAFSNGALVSISGKTGTAESYVADGQQATNTNAVAYAPSDNPQIAVAVVFPHNTNLTNGV
GPSIARDIINLYQKYHPMNMKWTKRVIRYATKNRKSPAENRRRVGKSLSLLSVFVFAIFL
VNFAVIIGTGTRFGTDLAKEAKKVHQTTRTVPAKRGTIYDRNGVPIAEDATSYNVYAVID
ENYKSATGKILYVEKTQFNKVAEVFHKYLDMEESYVREQLSQPNLKQVSFGAKGNGITYA
NMMSIKKELEAAEVKGIDFTTSPNRSYPNGQFASSFIGLAQLHENEDGSKSLLGTSGMES
SLNSILAGTDGIITYEKDRLGNIVPGTEQVSQRTMDGKDVYTTISSPLQSFMETQMDAFQ
EKVKGKYMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKV
MMLAAAIDNNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRMMTFSQGFAHSSNVGMT
LLEQKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQM
IRAFTAIANDGVMLEPKFISAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTD
PVYGTMYNHSTGKPTVTVPGQNVALKSGTAQIADEKNGGYLVGVTDYIFSAVSMSPAENP
DFILYVTVQQPEHYSGIQLGEFANPILERASAMKDSLNLQTTAKALEQVSQQSPYPMPSV
KDISPGDLAEELRRNLVQPIVVGTGTKIKNSSAEEGKNLAPNQQVLILSDKVEEVPDMYG
WTKETAETLAKWLNIELEFQGSGSTVQKQDVRANTAIKDIKKITLTLGD
>sp52_8529_faa
MNKPTILRLIKYLSISFLSLVIAAIVLGGGVFFYYVSKAPSLSESKLVATTSSKIYDNKN
QLIADLGSERRVNAQANDIPTDLVKAIVSIEDHRFFDHRGIDTIRILGAFLRNLQSNSLQ
GGSTLTQQLIKLTYFSTSTSDQTISRKAQEAWLAIQLEQKATKQEILTYYINKVYMSNGN
YGMQTAAQNYYGKDLNNLSLPQLALLAGMPQAPNQYDPYSHPEAAQDRRNLVLSEMKNQG
YISAEQYEKAINTPITDGLQSLKSASNYPAYMDNYLKEVINQVEEETGYNLLTTGMDVYT
NVDQEAQKHLWDIYNTDEYVAYPDDELQVASTIVDVSNGKVIAQLGARHQSSNVSFGINQ
AVETNRDWGSTMKPITDYAPALEYGVYDSTATIVHDEPYNYPGTNTPVYNWDRGYFGNIT
LQYALQQSRNVPAVETLNKVGLNRAKTFLNGLGIDYPSIHYSNAISSNTTESDKKYGASS
EKMAAAYAAFANGGTYYKPMYIHKVVFSDGSEKEFSNVGTRAMKETTAYMMTEMMKTVLS
YGTGRNAYLAWLPQAGKTGTSNYTDEEIEKHIKNTGYVAPDETFVGYTRKYSMAVWTGYT
NRLTPIMGDGLTVAAKVYRSMMTYLSEGSNPEDWNIPEGLYRNGEFVFKNGARSTWNSPA
PQQPPSTESSSSSSDSSTSQSSSTTPSTNNSTTTNPNNNTQQSNTTPDMKLDKLFEKFLS
LFKKETSELEDSDSTILRRSRSDRKKLAQVGPIRKFWRRYHLTKIILILGLSAGLLVGIY
LFAVAKSTNVNDLQNALKTRTLIFDREEKEAGALSGQKGTYVELTDISKNLQNAVIATED
RSFYKNDGINYGRFFLAIVTAGRSGGGSTITQQLAKNAYLSQDQTVERKAKEFFLALELS
KKYSKEQILTMYLNNAYFGNGVWGVEDASKKYFGVSASEVSLDQAATLAGMLKGPELYNP
LNSVEDSTNRRDTVLQNMVAAGYIDKNQETEAAEVDMASQLHDKYEGKISDYRYPSYFDA
VVNEAVSKYNLTEEEIVNNGYRIYTELDQNYQANMQVVYENTSLFPRAEDGTFAQSGSVA
```

Requirement 2:

The second challenge is the visual request. According to the examples, as shown in the figure. A visualization method is printed on a table, and the other method prints out for a graphical method. The main propose of this application is to display an effetely Interactive visualization tools.



Requirement 3:

The client of this application not only required to view the gene sequence but also be able to operate on the database, for example, once a new resistance gene been classified as a resistance gene then the application should add to the database by the client.

Project solutions

1, **FASTA file**

The file should be stored in the storage of the mobile device, the FASTA file can be an email attachment, it can be stored in a specific location. So the client should be able to load in the
SD card directory and load the specified FASTA file. By reading and writing files, find specific characters of the file, parse the file and write to a special entity.

2, **Data Analysis solution**:

The subject side provides a version of the resistance gene category, as shown in the figure.

Spec ID	Year	ST	MIC (mg/L)	pbp1a (1-707bp)			pbp2a (708 - 1438)		pbp2b (1439 -2119)					pbp2x (2120 - 2870)				
Amino acid position				370-373	428-432	574-575	1169	1294	1825-1828	1865	1866	1882-1885	2054-2058	2456-2459	2513-2516	2519-2520	2665-2668	
Susceptible amino acid code				STMK	SRNVP	TS	S	V	SVVK	T	Q	SSNT	KTGTA	STMK	HSSN	MT	LKSG	

According to this category, the gene sequence in the FASTA file can compare which genes are mutated. As the fast that the capacity of computing ability of the mobile phone is slower than the pc.

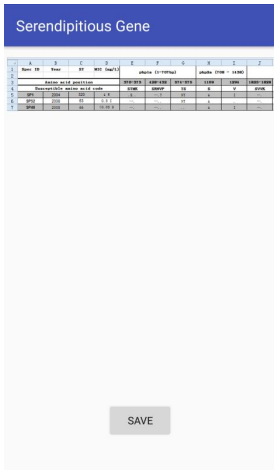
If the file load on the mobile side of the memory and computing occupation will extremely large. Our solution is that during the first time that the FASTA file be loaded to the mobile, the contents of the file should be resolved and saved to the database, each time you select the file from the database you can view the gene sequence information without load the entire file.

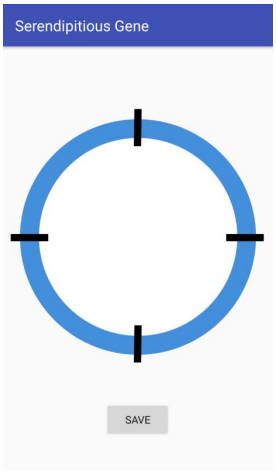
Another reason for using this method is that because of the same name of mutated strain like SP1, its gene sequence of one specific strain is constant, and if there are two SP1 strain in both FASTA files, then the unnecessary input and output times are saved.

3, **visualisation solution**

In terms of visualization, we consider the form like excel and graphics have their own advantages, so these two visual methods will be provided. The advantages of the form can show multiple strain, so we recommend using the form when the user views a FASTA file, as shown in the figure.

The advantage of graphics is good interactivity, and when we need to see a specific strain, we believe that the graphics are more convenient. We will provide simple gesture operation, by clicking to view the gene sequence information of a specific location, through different color to distinguish between mutation and no mutation, by scaling the picture to display and hide the genetic information. Realise visualization through customise View and Draw methods.





4, **Database solution**

For the subject of the operation of the database requirements. We give such a design, we believe that each client should have the right to retain the private database, the client can also view the server to provide different versions of the database.

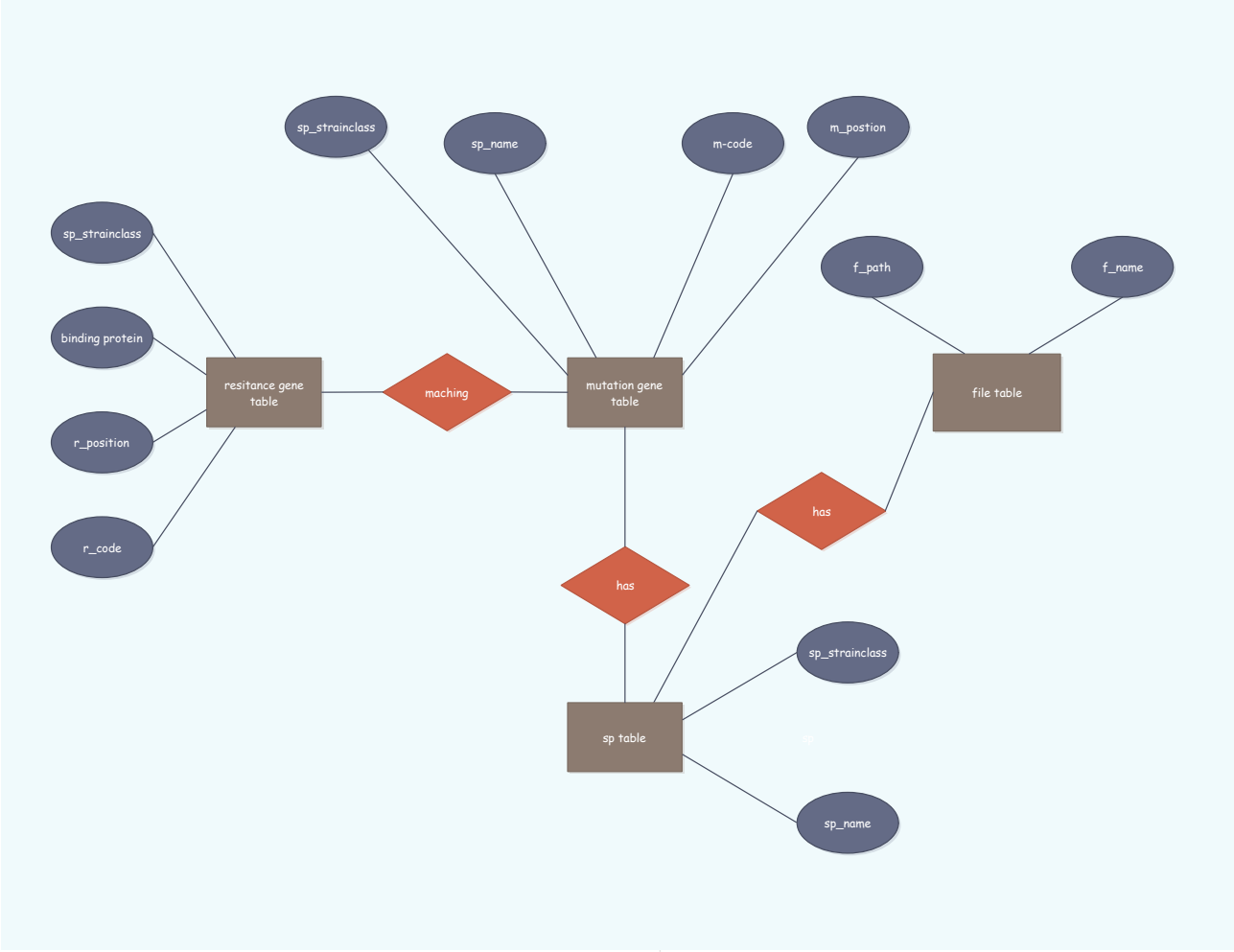
The server should store all the history data of the database version, the client upload their own private database, user can also request the server to check the database version and download an XML format file stored in a specified folder, the customer's local database should
also Save as an XML file in a specific folder, user can view the cloud by comparing the database to choose whether to update the database. This will not only meet the requirements of customers to modify the database also satisfy the cloud database to keep up to date.

```
<?xml version = '1.0' encoding = 'UTF-8' ? >
        <row>
                <binding protein> pbp 1a</binding protein>
                <position>370</position>
                <code>S</code>
        </row>
        <row>
                <binding protein> pbp 1a</binding protein>
                <position>371</position>
                <code>T</code>
        </row>
        <row>
                <binding protein> pbp 1a</binding protein>
                <position>372</position>
                <code>M</code>
        </row>
        <row>
                <binding protein> pbp 1a</binding protein>
                <position>373</position>
                <code>K</code>
        </row>
        <introduction> current version update:…</introduction>
```

5, **cloud server**

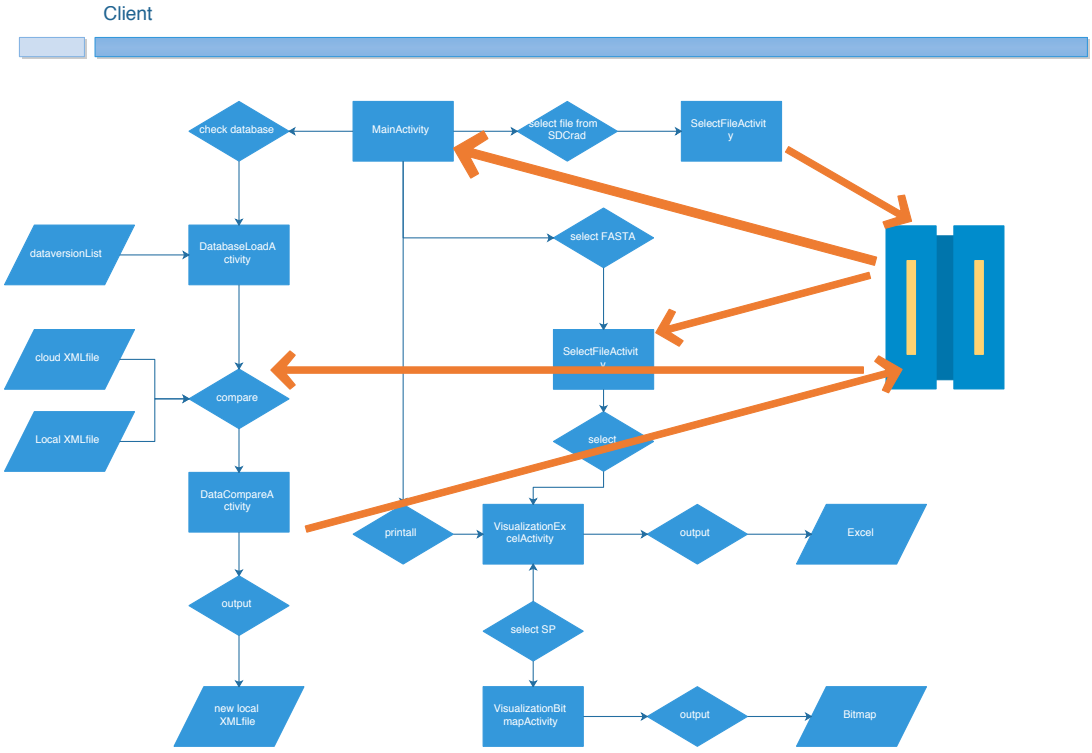
Cloud server function As above, because the cloud server architecture is relatively simple, we intend to build a TCP server through its HTTP protocol for data transmission, different versions of the resistance gene data version stored in the XML file. Specifically, We did not decide which cloud platform to adopt.

Client database design




```
public static final String DATABASE_NAME="SerendipitiousGene.db";

public static final String RESISTANCEGENE_TABLE="resistancegenetable";
public static String RESISTANCEGENE_CREATE="create table resistancegenetable ( strainclass varchar(20) , bindingprotein varchar(10) , position integer , orginalcode varchar(1) )";

public static final String MUTATIONGENE_TABLE = "mutationgenetable";
public static String MUTATIONGENE_CREATE ="create table mutationgenetable ( strainclass varchar(20) , sp varchar(10) , position integer , mutationcode varchar(1)) ";

public static final String FILE_Table = "filetable";
public static String FILE_CREATE= "create table filetable ( path varchar(70) , name varchar(70))";

public static final String SP_TABLE = "sptable";
public static String SP_CREATE = "create table sptable ( strainclass varchar(20), sp varchar(10) )";
```

Client architecture



In term of the client architecture, because of Android application base on the MVC, the application can generally divers to 6 packets.

Firstly, a **Activity packet** to control in which all class extends Activity, to ensure operation and connection between view and control part.

Secondly, **Entity packet** used to customise entity. The entity implements Serializable to allow the entity can use for database using. For example, SusceptibleAminoAcidCode is a entity which use to record the total gene sequence, and SPname …, the attribute also same as the database and cloud database attributes.

Thirdly, **SDCradUtil** packet used to connect local environment, the user mobile device storage.Allow user to write XML file, output bitmap, generate Excel. Input and parse XML file.

Fourthly,**View** packet, this packet is used to implement the visualisation. by extends View, our application is able to build our own visualization function, in addition, it also allow the application can listen to all Event in the view.

Firthly,**DatabaseUtil** packet is used to control the local database the class will extends android.SQLiteOpenHelper, create local database table, user can modify private database, and load tables from XML file.

Sixthly,**HttpUtil packet** is used to all function connect to the cloud server. the internet function possible need System Server to support the internet function and Broadcast to transmit
message to the destination.

technologic approach

STORE (ANDROID.SQLITEOPENHELPER)
1. Update the database
2. Read the local location
3. Create new Excel (import Jxl)
4. Create a new bitmap
5. Read and parse the FASTA file (file Inputstream)

VISUALIZE (ANDROID.VIEW)
1. Define view
2. put (view changes)
3. click (onTouchEvent)
4. pan (onTouchEvent)
5. visual excel table (import Jxl)
6. zoom (onTouchEvent)
7. translation (onTouchEvent)

UI (GITHUB)
1. GridView control the last item + button
2. GridView control can choose
3. gradually increase

NETWORK (JAVA.NET.HTTPURLCONNECTION)
1. Upload the database
2. Download the database
3. Retrieve the version number

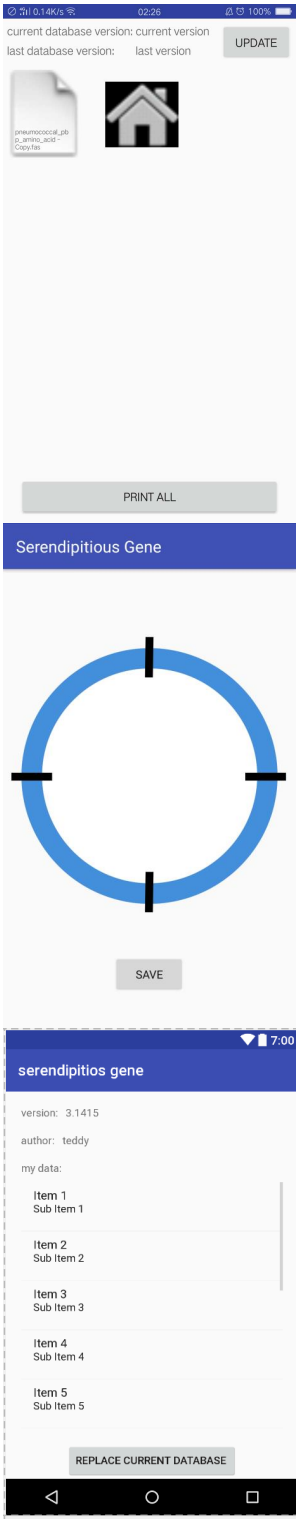
MAIN FUNCTION
1. Resistance to genes to be reported

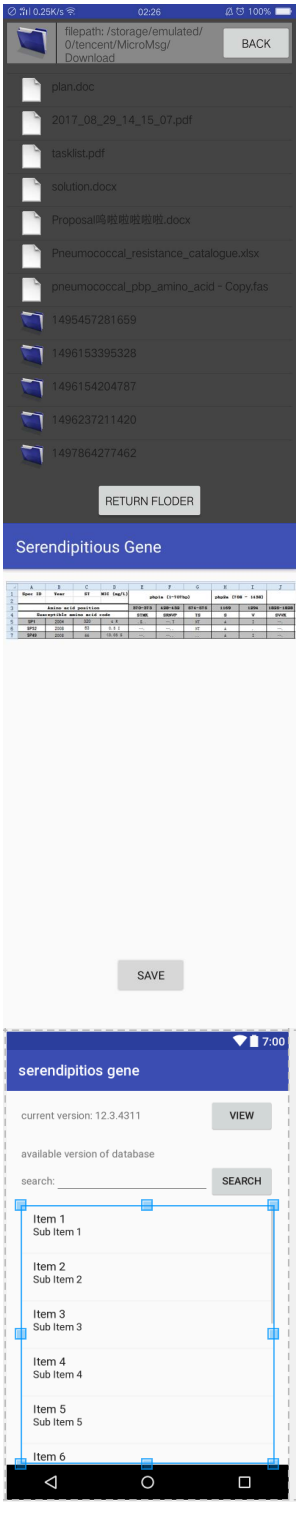
CLOUD

DATA STORAGE
1. Update the database
2. Retrieve the database version

NETWORK
1. get the request, return the database version number
2. get the request, get the database
3. get the request to return to a version of the database

UI design


current database version: current version
last database version: last version
UPDATE
PRINT ALL
Serendipitious Gene
SAVE
7:00
serendipitios gene
version: 3.1415
author: teddy
my data:
Item 1
Sub Item 1
Item 2
Sub Item 2
Item 3
Sub Item 3
Item 4
Sub Item 4
Item 5
Sub Item 5
REPLACE CURRENT DATABASE


filepath: /storage/emulated/0/tencent/MicroMsg/Download
BACK
plan.doc
2017_08_29_14_15_07.pdf
tasklist.pdf
solution.docx
Proposal嗚啦啦啦啦.docx
Pneumococcal_resistance_catalogue.xlsx
pneumococcal_pbp_amino_acid - Copy.fas
1495457281659
1496153395328
1496154204787
1496237211420
1497864277462
RETURN FLODER
Serendipitious Gene
SAVE
7:00
serendipitios gene
current version: 12.3.4311
VIEW
available version of database
search:
SEARCH
Item 1
Sub Item 1
Item 2
Sub Item 2
Item 3
Sub Item 3
Item 4
Sub Item 4
Item 5
Sub Item 5
Item 6

plan

week5

1. Analysis the user demands of this project and divide the main two part of basic function into server layer and application layer, build frame for application in order to achieve a large number of functions(all member)
2. Load file which provided by users from gene pool such as FASTA fileParse file which provided by users , and compare the load file with database, update the local database and Draw E-R graph and build database(Shengyuan Sun)
3. Design the application of construction and a large number of details for example Client architecture and Client database design(Yanlong Guan and Kai Liu)

week6

1. Using Android Studio to build data function for instance Excel display function in order to display the gene list information(import some jar packages), it also can provide save data function to save excel data to local device and server(Yanlong Guan)
2. Analysis the logic connection relationship between the different pages in this application and make connection test for application. Build bitmap function to save local data.(Kai Liu)
Create the XML documents to deploy the configuration of the application
3. Build service platform include client, database and service, create the new database in order to save a large number of data from the gene pool and this cloud database also can achieve the compare the data from client device.(Shengyuan Sun)

week7

1. Connect this application to cloud service and test the basic function(Shengyuan Sun)
2. Check database vision function and depend on case to update local database to server or upload local database to server(Yanlong Guan)
3. Find new challenge, use the visualized graph to operate the data and combine the function of code and graph(Kai Liu)

week8

1. Redesign the User Interface (UI) of application and complete details. Collect some picture to prepare layout file and background picture(Kai Liu)
2. Add some detail function by discussion for example add more detail information when user upload files or update files(Shengyuan Sun)
3. Debug some bugs(Yanlong Guan)

week9

1. Modify the User Interface(UI) of application and consummate the details of interface in order to make the application become more friendly and easy to use(Kai Liu)
2.Test the application in different digital device in order to improve the compatibility of the application(Shengyuan Sun)
3. Test the application in different digital system and environment in order to improve the compatibility of the application(Yanlong Guan)

week10

1. Review application include some functions and UI(Shengyuan Sun)
2.Prepare presentation and make PowerPoint(Yanlong Guan)
3. Zip all the files and documents and upload on eLearning broad(Kai Liu)

Reflection

During designing this application, we search for a large amount of knowledge and useful information about the gene from the website such as https://www.ncbi.nlm.nih.gov/genome/?term=streptococcus+pneumoniae. Comparing with this website, we found some viewpoint and innovations for creating into project application. In addition, in order to develop this app, we try our best to focus on the users' demand and add some humanized and gestures so as to make this app more modern and useful. Apart from this, we also pay attention to the visualization so that achieve in this app. It is easy to be visualized but difficult for users to adapt the operation interface and the Excel form it produces. There is also a massive challenge for our group to overcome.