Tutorial

1. Introduction

This repository is an open resource provided for International Potato Center (CIP) and PNIA as part of the "Database of the main pathogens detected, identified and characterized in the potato crop in Peru" project.
This web resource harbor virus data (Virome) in Peruvian potato. *Phytophthora infestans* and *Ralstonia solanacearum* pathogens (on potato and other crops in Peru) are also included in this repository.
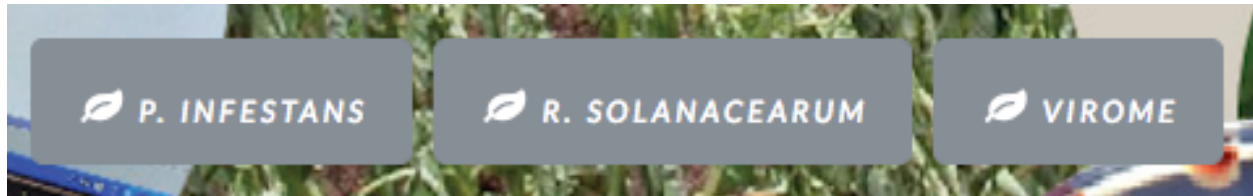
This repository can be accessed following the next link
http://potpathodiv.org

2. Overview

In the main page, you can access to *Phytophthora infestans*, *Ralstonia* or virome repositories.

A brief description of each of these repositories are in the following section on the home page.

Phytophthora infestans and *Ralstonia solanacearum* affect numerous crops of economic importance worldwide. In Peru, these pathogens are the main phytosanitary problems associated with potato. On the other hand, viruses affect potato crops, so the results generated in the detection of these viruses will allow to better understand the diversity of viruses.
Results for *Ralstonia solanacearum*, *Phythopthora infestans* and capacity building

Phytophthora infestans

Late blight (Tizón tardio TT) caused by *P. infestans* is the most destructive potato disease in the world and is consider a very strong plant disease.
See more

Virome

Detection and identification of known or unknown viruses allows us to better understand the diversity and distribution of viruses and their correlation with potato production in Peru.
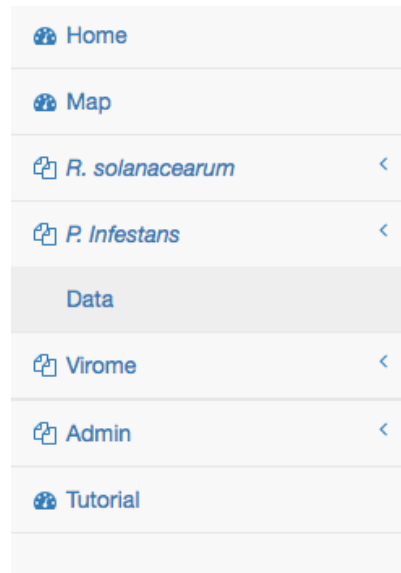See more
See data

Ralstonia solanacearum

Bacterial wilt (marchitez bacteriana MB) caused by *R. solanacearum* produces large losses worldwide due to the variability of the pathogen, wide range of hosts and large geographical distribution.
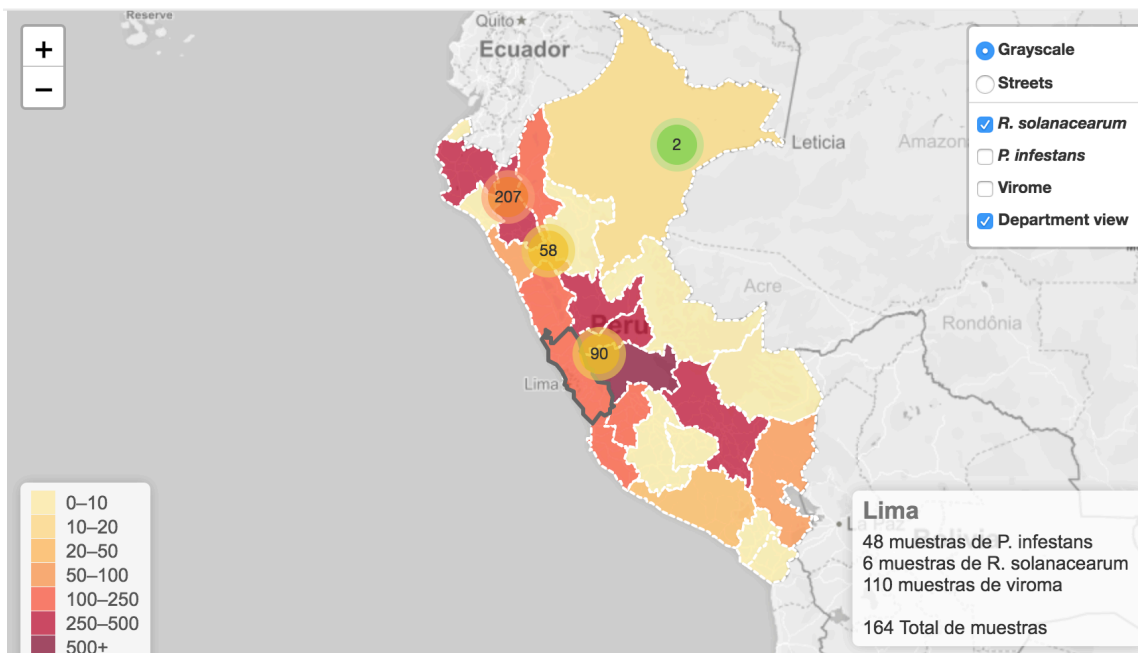See more

After selecting a repository, you can see the main menu.

3. Samples on the map

Select map in the menu to have a view of the samples on the map. Circles show the total number of samples collected in different places, you can **zoom in** or **zoom out** to see samples location in detail.

You can activate and deactivate department view. In the same way, you can filter per disease using radio buttons located in the top right corner.
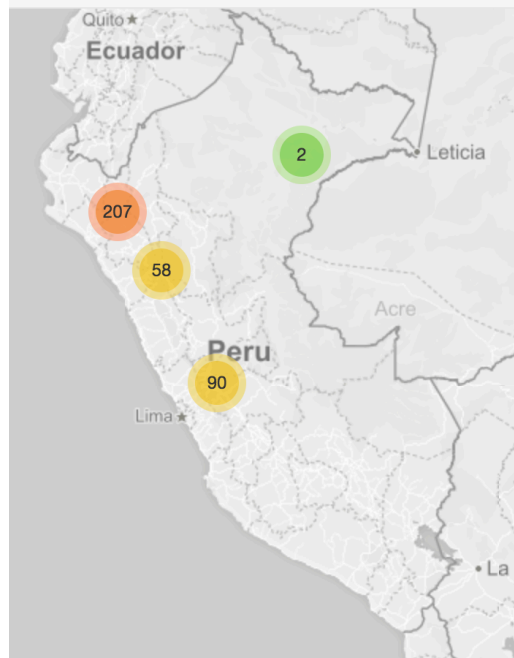
Departments are in color scale, where darkest are departments with more samples and colorless departments with less samples. The department view, activate and deactivate this option. At the bottom right corner in the screen, you can see the number of samples per department.
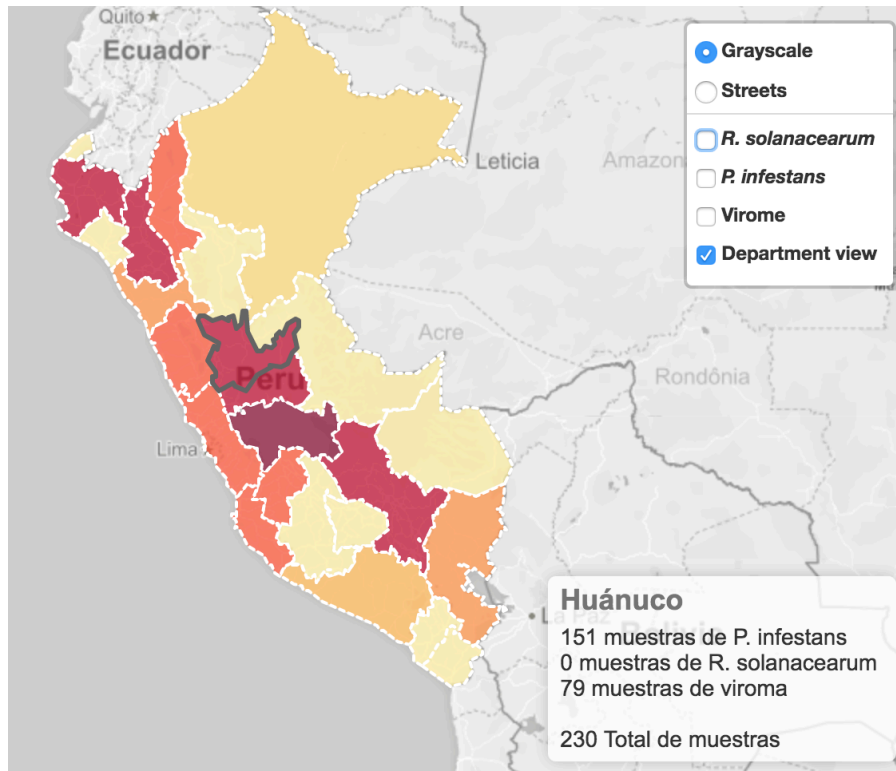
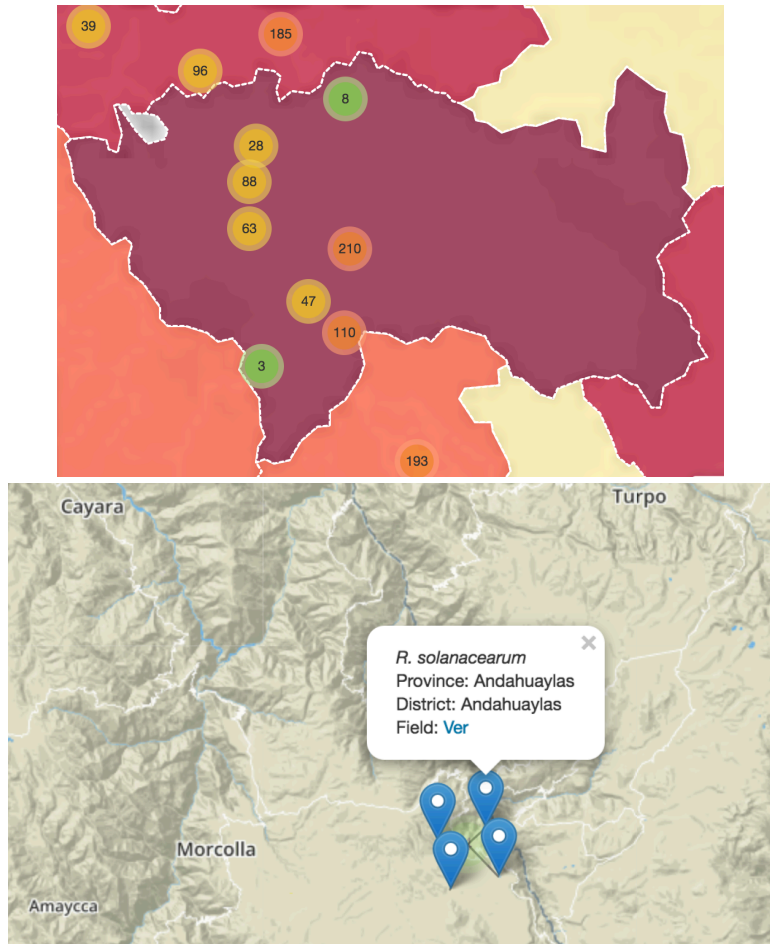Zoom in option allows to see a closer view of locations and their details.

R. solanacearum
Province: Andahuaylas
District: Andahuaylas
Field: Ver

4. Field view

Each repository contains a data option where you can find the field and sample list. For example, select *R. solanacearum* and data from the menu to get the list of fields and samples evaluated.
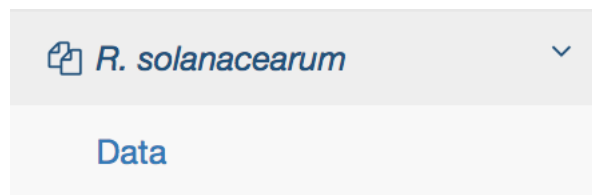
The field list shows geo-referenced data for each field. It is possible to search field names using the box located in the right top corner on the table. To sort the data, you can click the arrow in the column. In the bottom part, you change the number of page that you want to see.

Ralstonia solanacearum data

Field list   Samples

Show 10 ▼ entries   Search: [   ]

Country	Field ID	Department	Province	District	Latitude	Longitude	Altitude
Peru	Ama	Amazonas	Chachapoyas	Chachapoyas	-6.38389	-78.45222	2155
Peru	Ama	Amazonas	Luya	Longuita	-6.39944	-77.98153	2718
Peru	Ama	Amazonas	Luya	María	-6.42978	-77.96864	2672
Peru	Anc	Ancash	Carhuáz	Acopampa	-9.46972	-78.08389	2800
Peru	Anc	Ancash	Carhuáz	Carhuáz	-9.20119	-77.61925	3406
Peru	Anc	Ancash	Huaylas	Caraz	-9.08833	-78.35111	NA
Peru	Anc	Ancash	Santa	Caceres	-9.02750	-78.24056	NA
Peru	Anc	Ancash	Yungay	Yungay	-9.24000	-78.23972	NA
Peru	Apu	Apurimac	Andahuaylas	Andahuaylas	-14.10472	-73.63556	NA
Peru	Caj	Cajamarca	Cajabamba	Cajabamba	-7.58331	-78.07756	2781

Showing 1 to 10 of 56 entries   Previous 1 2 3 4 5 6 Next

Field ID column have a link to the Field page, where you can see more detailed information. The Field page show each field on the map and list all the samples belonged to that specific field.

Field (Ama) information

Field ID:	Ama
Country:	Peru
Departament:	Amazonas
Province:	Luya
District:	Lamud
Latitude:	-6.10075°
Longitude:	-77.9497777777778°
Altitude:	2330 m

Field Map | Field Image

Samples in field Ama

Show 10 entries   Search: [   ]

Sample	Pathogen Host	Date	Race	Genotypic method	Clonal lineage
PAMA017	S. tuberosum Group Tuberosum	2016	1,7	SSR	EC-1
PAMA018	S. tuberosum Group Tuberosum	2016	1,7	SSR	EC-1
PAMA019	S. tuberosum Group Tuberosum	2016	1,11	SSR	PE-3
PAMA020	S. tuberosum Group Tuberosum	2016	NA	SSR	PE-3
PAMA030-FTA	S. tuberosum Group Tuberosum	2016	NA	SSR	EC-1
PAMA031-FTA	S. tuberosum Group Tuberosum	2016	NA	SSR	EC-1

5. Samples view

Next to the field tab, you can see the samples tab, where according to the pathogen of interest, evaluations and data are listed.

Show 10 ⬍ entries

Search: [   ]

CIP code	Field	Host	Year	Biovar	Phylotype	Endoglucanase (egl) sequevar	NCBI accession number for partial sequence of egl gene
CIP-003	Piu	Solanum tuberosum	1974	1	IIB	1	MF461787
CIP-004	Lor	Musa paradisiaca	1966	1	IIB	4	MF461808
CIP-010	Jun	Solanum tuberosum	1979	2T	IIB	25	MF461731
CIP-012	Caj	Solanum tuberosum	1973	2A	IIB	1	MF461735
CIP-013	Caj	Solanum tuberosum	1974	2A	IIB	1	MF461739
CIP-014	Caj	Solanum tuberosum	1974	2A	IIB	1	MF461742
CIP-015	Hua	Solanum tuberosum	1973	1	IIA	41	MF461746
CIP-016	Caj	Solanum tuberosum	1974	2A	IIB	1	MF461747
CIP-017	Caj	Solanum tuberosum	1974	2A	IIB	1	MF461750
CIP-018	Lor	Solanum tuberosum	1975	1	IIA	50	MF461756

Showing 1 to 10 of 358 entries

Previous | 1 | 2 | 3 | 4 | 5 | … | 36 | Next

6. Images from fields, plants, and leaves

Virome repository has images from the fields, plants, and leaves collected from each sample. Plants and leaves show symptoms, if any, induced for the virus infection.

7. Sequences analysis result

Virome repository has sequences analysis results for each sample. To access to this page, click in any sample from the sample view tab. This page has 4 sections: sample cleaning, graph for length distribution, blastn and blastx for virus detection, and fastQC report. It is possible to download clean reads from the sample cleaning section.

Sample Cleaning   ⌄

Download Clean Reads

sample ID	Total Reads	Reads to be cleaned	% to be cleaned	Final Clean Reads	% Final Clean Reads
Caj-001	3693048	1846855	50.01	1846193	49.99

Length Distribution   ⌄

Virus Identified by nucleotide similarity (BLASTN)   ⌄

Virus Identified by translated protein similarity (BLASTX)   ⌄

FastQC   ❯

8. Upload data (Needs admin access)

To upload data, you need an admin account. Please contact Segundo Fuentes (S.FUENTES@CGIAR.ORG) to get an account.

Go to the menu and select Admin and choose Login. Login with your user and password provided.

Login

Username: |

Password:

Login

Go to the menu and choose Admin, and choose Upload data.

Download the excel template from this page. A specific example is provided as well.

Be careful with some common errors, you can find a file with instructions in the error list link.

1. Select the corresponding database:

2. Select an option. **Upload new data** option is the common case and correspond to upload new registries in the database. If you want to modify an existing sample that is currently in the database, select **Update data** and upload a new excel file just with the correct information for each sample to be changed, it will replace the existing information. If you need do delete a registry, select **Delete data** and in the next section upload the excel file with the sample ID values in the ID column to be deleted.

Select an option.
○ Upload new data   ○ Update data   ○ Delete data

3. Choose the excel file to be uploaded and click upload.

Select a excel file (xls,xlsx) to upload.
Choose File No file chosen

Upload

9. Upload sequences

To upload sequences and analysis results is necessary to have an FTP account, ask Segundo Fuentes for login instructions.

1. After login, you will see a data folder with raw and sample sub folders.
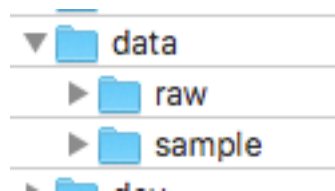
▼ 📁 data
  ▶ 📁 raw
  ▶ 📁 sample

2. Raw folder contains all raw data corresponding to each sample. Upload your raw data here. Name nomenclature is as follows:
SampleID.clean.fq.gz
Where SampleId must be the same as your excel sample ID.
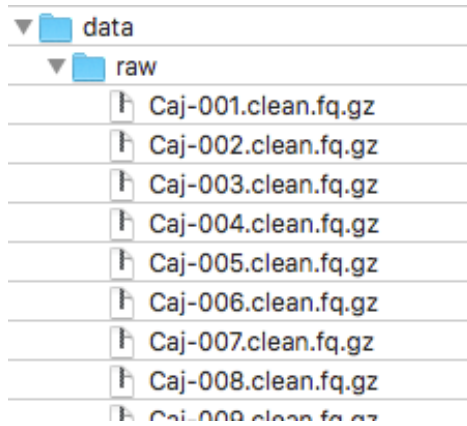
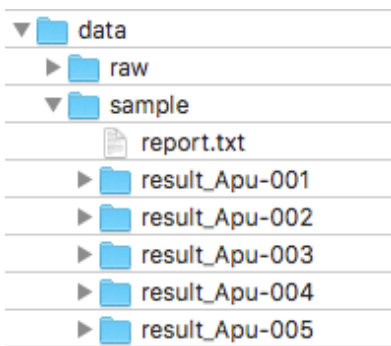3. Sample folder contains a report.txt file with sequence information as well as folders with individual sample results.

Report.txt is a tabular file with the following format. Where **#rsRNA** column corresponds to the sample ID

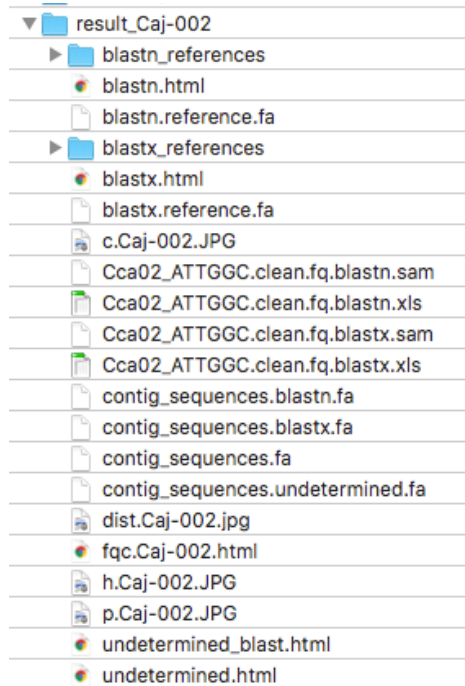| #sRNA	total	3Punmatch	3Pnull	3Pmatch	baseN	short	cleaned
Caj-014	4966975	116439	2448872	2401664	5320	225288	2171056
Caj-015	1233318	84476	555260	593582	572	89886	503124
Caj-017	5209043	100626	2729942	2378475	20366	141735	2216374
Caj-041	5179909	99232	2350233	2730444	25697	77353	2627394
Caj-042	5809490	172051	2868290	2769149	28780	94610	2645759
Caj-043	5938449	161252	2762535	3014662	27232	93183	2894247
Caj-048	10074845		217117	4888137	4969591	46861	170090	4752640

Folders correspond to each sample results. Naming convention is as follows:
result_sampleID
Where sampleID must be the same as in your excel sample ID.

result_sampleID folder must contains:

blastn.html	Correspond to Blast N result
blastx.html	Correspond to Blast X result
c.sampleID.JPG	Correspond to field picture
h.sampleID.JPG	Correspond to leaf picture
p.sampleID.JPG	Correspond to plant picture
fqc.sampleID.html	Correspond to fastQC report
dist.sampleId.jpg	Length distribution graph

```
▼ 📁 result_Caj-002
    ▶ 📁 blastn_references
      🔴 blastn.html
      📄 blastn.reference.fa
    ▶ 📁 blastx_references
      🔴 blastx.html
      📄 blastx.reference.fa
      🖼 c.Caj-002.JPG
      📄 Cca02_ATTGGC.clean.fq.blastn.sam
      📊 Cca02_ATTGGC.clean.fq.blastn.xls
      📄 Cca02_ATTGGC.clean.fq.blastx.sam
      📊 Cca02_ATTGGC.clean.fq.blastx.xls
      📄 contig_sequences.blastn.fa
      📄 contig_sequences.blastx.fa
      📄 contig_sequences.fa
      📄 contig_sequences.undetermined.fa
      🖼 dist.Caj-002.jpg
      🔴 fqc.Caj-002.html
      🖼 h.Caj-002.JPG
      🖼 p.Caj-002.JPG
      🔴 undetermined_blast.html
      🔴 undetermined.html
```

These files will be reflected in the following in the individual sample page:

Sample Cleaning ⌄

Download Clean Reads

sample ID	Total Reads	Reads to be cleaned	% to be cleaned	Final Clean Reads	% Final Clean Reads
Caj-002	3987250	2323719	58.28	1663531	41.72

Length Distribution ⌄

Sequence Length Distribution

Sequence Length (bp)

Virus Identified by nucleotide similarity (BLASTN) ›

Virus Identified by translated protein similarity (BLASTX) ›

FastQC ›