



HAL
open science

ViroBase: An ad hoc relational database for the management and traceability of biological data

Catherine Wipf-Scheibel, Pauline Millot, Karine Berthier, Sylvain Piry

► To cite this version:

Catherine Wipf-Scheibel, Pauline Millot, Karine Berthier, Sylvain Piry. ViroBase: An ad hoc relational database for the management and traceability of biological data. 17. Rencontres de Virologie Végétale (RVV 2019), Jan 2019, Aussois, France. 127 p., 2019, 17èmes Rencontres de Virologie Végétale: RVV 2019. hal-02947072

HAL Id: hal-02947072

<https://hal.inrae.fr/hal-02947072v1>

Submitted on 23 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Catherine Wipf-Scheibel¹, Pauline Millot¹, Karine Berthier¹, Sylvain Piry^{1,2}

¹ Pathologie végétale, INRA, 84140 Montfavet, France ² UMR CBGP, INRA, 34988, Montferrier sur Lez, France

Databases are crucial tools for the traceability and accessibility of scientific data, from field sampling to data sharing. We developed the "ViroBase" database to manage and facilitate access to all information on samples collected and analyzed in our lab, such as: sample geographical origin, host plant species, observed symptoms, data and type of samplings, host and pathogen taxonomy, samples processing. "ViroBase" has been developed in 2015-2016 under the open source object-relational database system PostgreSQL. A user-friendly interface implemented using the framework web2py facilitates access to records through simple queries.

During summer 2017, we sampled 989 cultivated plants in south of France

That's a lot of plants, it would be great if we could have some kind of automatization to record data...

A georeferenced picture of each sample, identified with datamatrix, has been recorded with a GPS

GPS data downloaded into ViroBase:

plant code
plant taxon
plant coordinates
sampling date

8500 ELISA tests were performed on all plants to detect 16 virus species: 810 positive tests from 566 plants

That's a lot of ELISA tests, it would be great if we could have some kind of automatization to record data...

Plaque	Puits	Groupe	Type	échantillon	Longueur	Abs
EM17P007-CABVY-2h	B01	Analyse Blanc	Blanc_0001 1/6	485	0,11	485
EM17P007-CABVY-2h	C01	Analyse Blanc	Blanc_0001 2/6	485	0,1	485
EM17P007-CABVY-2h	D01	Analyse Blanc	Blanc_0001 3/6	485	0,1	485
EM17P007-CABVY-2h	E01	Analyse Blanc	Blanc_0001 4/6	485	0,1	485
EM17P007-CABVY-2h	F01	Analyse Blanc	Blanc_0001 5/6	485	0,1	485
EM17P007-CABVY-2h	G01	Analyse Blanc	Blanc_0001 6/6	485	0,1	485
EM17P007-CABVY-2h	D02	Analyse échantillon	EM170291-P01E01	405	0,12	1,41
EM17P007-CABVY-2h	E02	Analyse échantillon	EM170292-P01E01	405	0,11	1,91
EM17P007-CABVY-2h	F02	Analyse échantillon	EM170293-P01E01	405	0,12	2,65
EM17P007-CABVY-2h	G02	Analyse échantillon	EM170294-P01E01	405	1,07	3,14
EM17P007-CABVY-2h	H03	Analyse échantillon	EM170295-P01E01	405	1,16	0,18
EM17P007-CABVY-2h	C03	Analyse échantillon	EM170296-P01E01	405	2,19	0,68
EM17P007-CABVY-2h	D03	Analyse échantillon	EM170297-P01E01	405	0,56	1,4
EM17P007-CABVY-2h	E03	Analyse échantillon	EM170298-P01E01	405	1,89	1,92
EM17P007-CABVY-2h	F03	Analyse échantillon	EM170299-P01E01	405	0,38	2,63
EM17P007-CABVY-2h	G03	Analyse échantillon	EM170300-P01E01	405	1,1	3,16
EM17P007-CABVY-2h	H04	Analyse échantillon	EM170301-P01E01	405	0,12	0,16
EM17P007-CABVY-2h	I01	Analyse échantillon	EM170302-P01E01	405	0,12	0,7
EM17P007-CABVY-2h	J01	Analyse échantillon	EM170303-P01E01	405	0,12	1,38
EM17P007-CABVY-2h	K01	Analyse échantillon	EM170304-P01E01	405	0,12	1,94
EM17P007-CABVY-2h	L01	Analyse échantillon	EM170305-P01E01	405	0,12	2,62
EM17P007-CABVY-2h	M01	Analyse échantillon	EM170306-P01E01	405	0,12	3,18
EM17P007-CABVY-2h	N01	Analyse échantillon	EM170307-P01E01	405	0,15	0,14
EM17P007-CABVY-2h	O01	Analyse échantillon	EM170308-P01E01	405	4,19	0,72
EM17P007-CABVY-2h	P01	Analyse échantillon	EM170309-P01E01	405	0,12	1,36
EM17P007-CABVY-2h	Q01	Analyse échantillon	EM170310-P01E01	405	0,12	1,96
EM17P007-CABVY-2h	R01	Analyse échantillon	EM170311-P01E01	405	0,67	2,6
EM17P007-CABVY-2h	S01	Analyse échantillon	EM170312-P01E01	405	3,63	3,2

Automatic import of ELISA results into ViroBase

A user-friendly interface implemented in the framework web2py facilitates access to records through simple requests.

This allows a quick extraction of information at each stage of the sample processing chain: results of serological tests or PCR amplifications, host infection status, pathogen DNA sequences, etc.

989 total RNA extractions and more than 3500 RT-PCR were performed to amplify 3 virus species on 1 to 3 markers

That's a lot of sequences, we need to find a way to analyze them

Variant calling analysis of NGS sequences was performed through a R pipeline based on the package DADA2 (Callahan et al. 2016)

High throughput sequencing was performed on amplicons using dual indexing (Galan et al. 2018) on MiSeq Illumina

Thanks to Virobase, all plant information can easily be extracted and used in other software. Here is an example of map, made with Google Earth. We can see all the different variants of WMV on the CI region!

Results by plant

Project A	Plant code	Plant taxon	Geography	Origin date	Number of samplings	ELISA Detected pathogens	ELISA Absent pathogens	ELISA Doubtful pathogens	Extractions codes	Sequencing diagnostic	Created storages
EMERAMB	EM170231	courgette	FRANCE, Languedoc-Roussillon, Gard, Redessan	2017-07-03	1	WMV	CABVY, CMV, MNSV, MWMV, PRSV, SqMV, ZYMV		XWS170509	WMV_CI_groupe_C05-270, WMV_CP_groupe_C05-270	EM170231-P01E01, EM170231-P01E01-J01, EM170231-P01E01-J02
EMERAMB	EM170232	courgette	FRANCE, Languedoc-Roussillon, Gard, Redessan	2017-07-03	1	CABVY	CMV, MNSV, MWMV, PRSV, SqMV, WMV, ZYMV		XWS170510	CABVY_CP_groupe_EM160050, CABVY_RdRp_groupe_EM160050	EM170232-P01E01, EM170232-P01E01-J01, EM170232-P01E01-J02
EMERAMB	EM170233	courgette	FRANCE, Languedoc-Roussillon, Gard, Redessan	2017-07-03	1	CABVY	CMV, MNSV, MWMV, PRSV, SqMV, WMV, ZYMV		XWS170511	CABVY_CP_groupe_EM160050, CABVY_RdRp_groupe_EM160050	EM170233-P01E01, EM170233-P01E01-J01, EM170233-P01E01-J02
EMERAMB	EM170234	courgette	FRANCE, Languedoc-Roussillon, Gard, Redessan	2017-07-03	1	CABVY	CMV, MNSV, MWMV, PRSV, SqMV, WMV, ZYMV		XWS170512	CABVY_CP_groupe_EM160050, CABVY_RdRp_groupe_EM160050	EM170234-P01E01, EM170234-P01E01-J01, EM170234-P01E01-J02
EMERAMB	EM170235	courgette	FRANCE, Languedoc-Roussillon, Gard, Redessan	2017-07-03	1	CABVY	CMV, MNSV, MWMV, PRSV, SqMV, WMV, ZYMV		XWS170513	CABVY_CP_groupe_EM160050, CABVY_RdRp_groupe_EM160050	EM170235-P01E01, EM170235-P01E01-J01, EM170235-P01E01-J02
EMERAMB	EM170236	courgette	FRANCE, Languedoc-Roussillon, Gard, Redessan	2017-07-03	1	CABVY	CMV, MNSV, MWMV, PRSV, SqMV, WMV, ZYMV		XWS170514	CABVY_CP_groupe_EM160050, CABVY_RdRp_groupe_EM160050	EM170236-P01E01, EM170236-P01E01-J01, EM170236-P01E01-J02

Pre-implemented queries within the interface provides various synthesis reports, such as a list of pathogen variant genotypes associated to geographical coordinates of their host plants.

Today, Virobase allows to manage various information including: 10456 samples records, originating from 4727 plants and corresponding to different storage process, 16402 ELISA test results, 7962 DNA sequences of virus (Sanger, NGS) characteristics for 2745 PCR primers...

Virobase and its dedicated interface facilitate the collaborative work, not only within the lab but also with external partners. Its conceptual framework will be soon extended to other biological models studied in our lab, such as bacteria and fungi.

REFERENCES

Callahan BJ, McMurdie PJ, Han AW, Johnson AJ, Holmes SP (2016). DADA2: High resolution sample inference from Illumina amplicon data. Nature Methods, 13(7), 581-3

Galan M, Pons J-B, Tournayre O, Pierre E, Leuchtmann M, Pontier D, Charbonnel N (2018). Metabarcoding for the parallel identification of several hundred predators and their prey: Application to bat species diet analysis. Mol Ecol Resour. 2018;18:474-489