



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-02947072>

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.

Project	Plant code	Plant taxon	Geography	Origin date	Number of samplings	ELISA Detected pathogens	ELISA Doubtful pathogens	Extractions codes	Sequencing diagnostic	Created storages
EMERAMB	EM170231	courgette	FRANCE, Languedoc-Roussillon, Gard, Redessan	2017-07-03	1	WMV	CABYV, CMV, MNSV, MMWV, PRSV, SqMV, ZYMV	XWS170509	WMV_CI_groupe_EM05-270, 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	CABYV	CMV, MNSV, MMWV, PRSV, SqMV, WMV, ZYMV	XWS170510	CABYV_CP_groupe_EM160050, CABYV_RdRp_groupe_EM160050	EM170232-P01E01, EM170232-P01E01-J01, EM170232-P01E01-J02
EMERAMB	EM170233	courgette	FRANCE, Languedoc-Roussillon, Gard, Redessan	2017-07-03	1	CABYV	CMV, MNSV, MMWV, PRSV, SqMV, WMV, ZYMV	XWS170511	CABYV_CP_groupe_EM160050, CABYV_RdRp_groupe_EM160050	EM170233-P01E01, EM170233-P01E01-J01, EM170233-P01E01-J02
EMERAMB	EM170234	courgette	FRANCE, Languedoc-Roussillon, Gard, Redessan	2017-07-03	1	CABYV	CMV, MNSV, MMWV, PRSV, SqMV, WMV, ZYMV	XWS170512	CABYV_CP_groupe_EM160050, CABYV_RdRp_groupe_EM160050	EM170234-P01E01, EM170234-P01E01-J01, EM170234-P01E01-J02
EMERAMB	EM170235	courgette	FRANCE, Languedoc-Roussillon, Gard, Redessan	2017-07-03	1	CABYV	CMV, MNSV, MMWV, PRSV, SqMV, WMV, ZYMV	XWS170513		EM170235-P01E01, EM170235-P01E01-J01, EM170235-P01E01-J02
EMERAMB	EM170236	courgette	FRANCE, Languedoc-Roussillon, Gard, Redessan	2017-07-03	1	CABYV	CMV, MNSV, MMWV, PRSV, SqMV, WMV, ZYMV	XWS170514	CABYV_CP_groupe_EM160050, CABYV_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