Project Title: Virus-Enabled Replication and Remobilization of Transposable Elements and Host eccDNAs in Plant Genomes

Project Overview: This fundamental research project aims to explore the intricate interplay between viruses, transposable elements (TEs), and host extrachromosomal circular DNAs (eccDNAs) within plant genomes. The research seeks to uncover novel mechanisms underlying TE replication, remobilization, and their potential impact on plant genomic evolution. This project is expected to contribute significantly to our understanding of viral-host interactions and genome plasticity. The project will span four years and the TCI laboratory is the main contributor.

Data Generation and Collection: Data for this project will be generated through a combination of experimental and computational approaches, encompassing genomic sequencing, viral infection assays, TE mobility tracking, and bioinformatic analyses. Data collection will adhere to rigorous standards for reproducibility and quality control. Various experimental conditions and time points will be considered to capture the dynamic nature of viral-TE-host interactions.

Data Types:

1. **Experimental Data:** This category includes raw data, such as DNA sequencing results and eccDNA characterization.
2. **Metadata:** Detailed information about experimental protocols, sample handling, and bioinformatics pipelines will be meticulously documented for each dataset.

Data Organization: To ensure data accessibility, integrity, and security, a dedicated research database will be established. This database will utilize standardized naming conventions and metadata tagging for efficient retrieval. Data will be stored in widely accepted, non-proprietary formats to facilitate long-term accessibility.

Data Sharing and Accessibility:

1. **Intra-team Sharing:** Collaborators within the research team will have controlled access to project data via a secure network (shared drive). Regular team meetings will foster data interpretation and issue resolution.
2. **External Sharing:** In line with the principles of transparency and open science, anonymized datasets and associated metadata will be deposited in appropriate public repositories (e.g., NCBI, ENA, or Zenodo). Data will be made publicly accessible upon publication or following a mutually agreed-upon embargo period.

Data Security and Ethics: Data security measures will be implemented to safeguard sensitive information and adhere to ethical guidelines. Any personally identifiable information will be anonymized or pseudonymized as needed. Robust access controls, data encryption, and routine backups will be employed to maintain data integrity.

Data Preservation: All data generated during the project will be preserved for a minimum of ten years post-completion. Preservation will be secured through

institutional servers and redundant backups. In anticipation of personnel changes, clear documentation and knowledge transfer protocols will be established to guarantee the long-term preservation of data.

Data Disposal: At the conclusion of the data retention period, data will be disposed of according to institutional and legal directives. This process will include the secure deletion of electronic files and the proper disposal of physical records and samples.

Data Management Responsibilities: The Principal Investigator (PI) will oversee the overall data management strategy. Each team member will be assigned specific responsibilities for data collection, documentation, and sharing. A dedicated Data Manager will be appointed to supervise data organization and security.

Training and Documentation: All team members will receive training on data management best practices and ethical guidelines. Comprehensive documentation of data collection and handling procedures will be upheld throughout the project, ensuring transparency and reproducibility.

Data Management Review: Data management practices will undergo periodic review and adaptation during the project's duration to accommodate evolving technologies and project needs.

This Data Management Plan is integral to our commitment to rigorous scientific research, data fidelity, and collaboration within the academic community. It will guarantee the efficient management, dissemination, and preservation of invaluable data for the advancement of future research and scientific knowledge.

1. General Project Information	
Name Grant Holder & ORCID	**Hervé Vanderschuren**
Contributor name(s) (+ ORCID) & roles	**Hervé Vanderschuren/ PI** **Min Xu/ postdoc researcher**
Project number[1] & title	C1 vertigo
Please provide a short project description	Geminiviruses are a large family of plant infecting pathogen whose life cycles are intimately linked to the host plant DNA replication and epigenetic machineries are particularly suitable to advance our knowledge of the plant epigenetic immune response. The potential of epigenetic variation and TE-driven variation is underexploited as it globally remains at the somatic cell (exposed to pathogens) level. eccDNA and TE have different aspects of correlation. In this project, a series of new technologies is used to purify and characterize eccDNA, and elucidate the correlation among eccDNA, active TE and virus infection regulated by methylation.

[1] "Project number" refers to the institutional project number. This question is optional since not every institution has an internal project number different from the GrantID. Applicants can only provide one project number.

2. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data[2].

				Only for digital data	*Only for digital data*	*Only for digital data*	*Only for physical data*
Dataset Name	Description	New or Reused	Digital or Physical	Digital Data Type	Digital Data Format	Digital Data Volume (MB, GB, TB)	Physical Volume
Constructs Genomic-seq eccDNA-seq pictures qPCR	Plasimids ONT-seq data Gel and fluorescence qPCR	☒ Generate new data ☒ Reuse existing data	☒ Digital ☒ Physical	☒ Observational ☒ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☐ .csv ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: ☐ NA	☐ < 100 MB ☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	In 2ml cryotubes and 1.5ml Eppendorf tubes

[2] Add rows for each dataset you want to describe.

GUIDANCE: *DATA CAN BE DIGITAL OR PHYSICAL (FOR EXAMPLE BIOBANK, BIOLOGICAL SAMPLES, …). DATA TYPE: DATA ARE OFTEN GROUPED BY TYPE (OBSERVATIONAL, EXPERIMENTAL ETC.), FORMAT AND/OR COLLECTION/GENERATION METHOD.* *EXAMPLES OF DATA TYPES: OBSERVATIONAL (E.G. SURVEY RESULTS, SENSOR READINGS, SENSORY OBSERVATIONS); EXPERIMENTAL (E.G. MICROSCOPY, SPECTROSCOPY, CHROMATOGRAMS, GENE SEQUENCES); COMPILED/AGGREGATED DATA[3] (E.G. TEXT & DATA MINING, DERIVED VARIABLES, 3D MODELLING); SIMULATION DATA (E.G. CLIMATE MODELS); SOFTWARE, ETC.* *EXAMPLES OF DATA FORMATS: TABULAR DATA (.POR,. SPSS, STRUCTURED TEXT OR MARK-UP FILE XML, .TAB, .CSV), TEXTUAL DATA (.RTF, .XML, .TXT), GEOSPATIAL DATA (.DWG,. GML, ..), IMAGE DATA, AUDIO DATA, VIDEO DATA, DOCUMENTATION & COMPUTATIONAL SCRIPT.* *DIGITAL DATA VOLUME: PLEASE ESTIMATE THE UPPER LIMIT OF THE VOLUME OF THE DATA PER DATASET OR DATA TYPE.* *PHYSICAL VOLUME: PLEASE ESTIMATE THE PHYSICAL VOLUME OF THE RESEARCH MATERIALS (FOR EXAMPLE THE NUMBER OF RELEVANT BIOLOGICAL SAMPLES THAT NEED TO BE STORED AND PRESERVED DURING THE PROJECT AND/OR AFTER).*	
If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type.	
Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, please describe these issues further and refer to specific datasets or data types when appropriate.	☐ Yes, human subject data ☐ Yes, animal data ☐ Yes, dual use ☒ No If yes, please describe:

[3] These data are generated by combining multiple existing datasets.

Will you process personal data[4]? If so, briefly describe the kind of personal data you will use. Please refer to specific datasets or data types when appropriate. If available, add the reference to your file in your host institution's privacy register.	☐ Yes ☒ No If yes: - Short description of the kind of personal data that will be used: - Privacy Registry Reference:
Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.	☐ Yes ☒ No If yes, please comment:
Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements, research collaboration agreements)? If so, please explain to what data they relate and what restrictions are in place.	☐ Yes ☒ No If yes, please explain:
Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain to what data they relate and which restrictions will be asserted.	☐ Yes ☒ No If yes, please explain:

[4] See Glossary Flemish Standard Data Management Plan

3. Documentation and Metadata	
Clearly describe what approach will be followed to capture the accompanying information necessary to keep **data understandable and usable**, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).	Certain DNA related software like snapgene, IGV will be used for opening DNA sequences and analysis Labnote will have description of file type and storage position.
Will a metadata standard be used to make it easier to **find and reuse the data**? If so, please specify which metadata standard will be used. If not, please specify which metadata will be created to make the data easier to find and reuse. *REPOSITORIES COULD ASK TO DELIVER METADATA IN A CERTAIN FORMAT, WITH SPECIFIED ONTOLOGIES AND VOCABULARIES, I.E. STANDARD LISTS WITH UNIQUE IDENTIFIERS.*	☒ Yes ☐ No If yes, please specify (where appropriate per dataset or data type) which metadata standard will be used: We will follow NCBI requirements/standards to make metadata If no, please specify (where appropriate per dataset or data type) which metadata will be created:

4. Data Storage & Back-up during the Research Project	
Where will the data be stored?	In the lab computer and J drive and NCBI

How will the data be backed up? *What storage and backup procedures will be in place to prevent data loss? Describe the locations, storage media and procedures that will be used for storing and backing up digital and non-digital data during research.*[5] *Refer to institution-specific policies regarding backup procedures when appropriate.*	Will be both store in lab computer and online platform
Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.	☒ Yes ☐ No If yes, please specify concisely: Lab computer and NCBI If no, please specify:
How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons? *Clearly describe the measures (in terms of physical security, network security, and security of computer systems and files) that will be taken to ensure that stored and transferred data are safe.* [5]	The access is limited only to authorized persons
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	500euro, by the project

[5] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

5. Data Preservation after the end of the Research Project	
Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).	Plasmids and glycerol stock will be kept in -80 freezer for more than 5 years. Sequencing data will be kept for more than 5 years. DNA and RNA samples will not be kept for so long time because it is consumable, can be regenerate and there is limited space. Temporary data for single experiment that doesn't work will not be kept for long time.
Where will these data be archived (stored and curated for the long-term)?	In -80 freezer and in J drive
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	

6. Data Sharing and Reuse	
Will the data (or part of the data) be made available for reuse after/during the project? Please explain per dataset or data type which data will be made available. *NOTE THAT 'AVAILABLE' DOES NOT NECESSARILY MEAN THAT THE DATA SET BECOMES OPENLY AVAILABLE, CONDITIONS FOR ACCESS AND USE MAY APPLY. AVAILABILITY IN THIS QUESTION THUS ENTAILS BOTH OPEN & RESTRICTED ACCESS. FOR MORE INFORMATION: HTTPS://WIKI.SURFNET.NL/DISPLAY/STANDARDS/INFO-EU-REPO/#INFOEUREPO-ACCESSRIGHTS*	☒ Yes, in an Open Access repository ☐ Yes, in a restricted access repository (after approval, institutional access only, …) ☐ No (closed access) ☐ Other, please specify:
If access is restricted, please specify who will be able to access the data and under what conditions.	
Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain per dataset or data type where appropriate.	☐ Yes, privacy aspects ☐ Yes, intellectual property rights ☐ Yes, ethical aspects ☐ Yes, aspects of dual use ☐ Yes, other ☒ No If yes, please specify:
Where will the data be made available? If already known, please provide a repository per dataset or data type.	Not known yet

When will the data be made available? *This could be a specific date (dd/mm/yyyy) or an indication such as 'upon publication of research results'.*	Upon publication of research results
Which data usage licenses are you going to provide? If none, please explain why. *A data usage license indicates whether the data can be reused or not and under what conditions. If no licence is granted, the data are in a grey zone and cannot be legally reused. Do note that you may only release data under a licence chosen by yourself if it does not already fall under another licence that might prohibit that.* *Example Answer: E.g. "Data from the project that can be shared will be made available under a Creative Commons Attribution license (CC-BY 4.0), so that users have to give credit to the original data creators."* [6]	Sequencing Data from this project will be shared on public platform when the research results are published and the physical materials will be kept in lab and only sharing when the researchers think necessary.
Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here. *Indicate whether you intend to add a persistent and unique identifier in order to identify and retrieve the data.*	☐ Yes ☒ No If yes:
What are the expected costs for data sharing? How will these costs be covered?	

[6] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Hervé Vanderschuren and Min Xu
Who will manage data storage and backup during the research project?	Hervé Vanderschuren and Min Xu
Who will manage data preservation and sharing?	Hervé Vanderschuren and Min Xu
Who will update and implement this DMP?	Hervé Vanderschuren and Min Xu