Editorial
Digital Twins of the Soil Microbiome for Climate Mitigation

Hussnain Mukhtar *, Rainer Ferdinand Wunderlich and Yu-Pin Lin

Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei 10617, Taiwan; rainer.wunderlich@gmail.com (R.F.W.); yplin@ntu.edu.tw (Y.-P.L.)
* Correspondence: mukhtar@ntu.edu.tw

Recent advances in computation power have enabled the creation of digital twins of the microbiome (DTM) to substantially curb soil greenhouse gases (GHG) emissions under global change conditions. For these twins to be more than large arrays of data, they must qualitatively replicate the microbial population dynamics as well as their complex underlying biogeochemical processes and terrestrial feedback to ensure the validity of the predictive metrics.

Since microbial metabolism is the largest source of the key biogenic greenhouse gases [1], carbon dioxide, methane, and nitrous oxide, there is vast potential for managing and manipulating soil microbial communities and processes to reduce the severity of the climate crisis. As microbial processes are embedded in terrestrial ecosystems and biogeochemical cycles, the domains of microbial and environmental genomics need to be complemented with soil and plant science as well as climate and ecosystem modeling. DTM would integrate the essential biological (e.g., genotype–phenotype map) and physicochemical processes (e.g., thawing permafrost) and ecological hypotheses (e.g., microbial cosmopolitanism and stress-gradient) into a computer simulation to predict the climate sensitivity of soil ecosystems and GHG fluxes. Such DTM, particularly when paired with machine learning constrained by the mechanistic modeling of environmental and ecological systems, could help to minimize gaseous nutrient loss, alleviate global warming, and improve soil fertility.

The digital twin concept originated from industrial engineering and originally aimed to optimize the design and structure of a complex machine [2]. Although the digital twins that are for natural processes are much more challenging to develop than those for engineered devices, this idea has recently been proposed in the medical field to diagnose and treat virus infections [3] as well as for the green transition of the Earth [4]. DTM would fully integrate the descriptive and predictive measurements of the environmental and microbial process to understand the archaean-, bacterial-, and fungal-induced cycling of macro- and micro-nutrients in the terrestrial ecosystem for the continual forecasting of GHG emissions (Figure 1) and for the evaluation of potential human interventions (e.g., chemical inhibitors or pH manipulation).

DTM combines a growing set of observations with mechanistic models to simulate the environmental, ecological, and microbial perspectives of a target system over space and time. On the one hand, DTM should be able to precisely simulate the steady and nonsteady states of earth systems and anthropogenic impacts, such as elevated CO2 and temperature or land-cover change across spatial and temporal scales. On the other hand, the DTM should be able to simulate the influence of novel or extreme environmental conditions on the microbial genome structure, including both short-term acclimatization and long-term adaptation, which result in different phenotypes and metabolic reaction rates, as well as their feedback on the earth system via soil nutrient dynamics and GHG fluxes.
Several necessary DTM pathways and processes (e.g., microbial respiration) can be simulated using existing process-based models and data assimilation tools—a combination of observed data sets (i.e., remote sensing data) with physical and numerical models [5,6]. Due to the high complexity and performance optimization of DTM, specific operations would be automatically executed when certain conditions are met (autonomous entity), while missing elements would be substituted via machine learning. Thus, DTM would start as a simple data assimilation framework that is gradually refined by a continuous self-learning process that results in maximized predictive performance. Ideally, a toolkit should be provided to facilitate the export of the summary statistics and data that are made accessible to non-expert stakeholders via a distributed cloud-based infrastructure.

Driven by data collected from sensors in real-time and historical aggregate information, DTM would be the most accurate reflection of physical and biological soil systems, including the production or consumption of GHG. It would reshape the traditional modeling framework via (1) the continual ingestion of additional information from all possible instruments, including satellites and smart sensor arrays, measuring variables such as soil moisture or chemical concentrations, (2) the integration of biogeophysical and socio-economic data and processes into a social-ecological system, (3) the simulation of multiple scenarios (e.g., climate trajectories) to facilitate our understanding of the future impacts of various risk mitigation and disaster management alternatives, (4) the inclusion of potential missing constraints identified by data-driven learning, and, finally, (5) the scalable complexity. These will help to improve the ability of models to make realistic and more reliable predictions and to enable targeted data collection across different soil ecosystems.

To implement fully functional DTM—which are currently in a hypothetical state—a number of issues need to be addressed. Firstly, the distribution of microbes within the different ecosystems are spatio-temporally complex and are still relatively poorly understood. For instance, with the discovery of ammonia-oxidizing archaea [7] as well as the anammox [8] and comammox [1] pathways, the complexity of nitrous oxide dynamics in the terrestrial ecosystem has been substantially increased. Furthermore, there is still no consensus on the geographical, ecological, and environmental drivers regulating the distribution and abundance of individual microbe taxa, while the relative importance of these parameters also varies across ecosystems and microbial groups [9,10]. Secondly, there would be substantial uncertainty due to errors propagating from observations and the parametrization of the models, specifically in terms of predictions regarding acclimatization or adaptive evolution.

The sampling rate and overall size of the data also matter. The former would be constrained by infrastructure limitations (e.g., bandwidth and interference), while the latter should be close to the optimum. Too few data may cause inaccurate predictions, whereas sample sizes beyond the optimum amount may cause excessive computational demands and diminish predictive performance. Moreover, a universal design and development platform is required to enhance the compatibility and entanglement of different models and data sets. Although individual groups may construct sub-models, the development
of comprehensive DTM will require the integration and independent validation of all of the sub-models.

DTM will revolutionize soil system simulations and their contributions to global warming via GHG emissions. These DTM could also be expanded to incorporate crop (and human) pathogens to minimize exposure or risk, rhizoremediation, or be coupled with other Earth system representations that have been already proposed [4], resulting in much more comprehensive models that account for the efficacy of different climate adaptation and mitigation strategies. Ultimately, DTM could serve as powerful decision support tools to harness the soil microbial potential to manipulate terrestrial ecosystems and mitigate the negative consequences of climate change.

**Author Contributions:** Conceptualization, H.M.; writing—original draft preparation, H.M., R.F.W. and Y.-P.L.; writing—review and editing, H.M., R.F.W. and Y.-P.L.; visualization, H.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**