

Combining Legacy Data from Heterogeneous Crop Trials to Identify Genotype By Environment

Interactions Using Model-Based Recursive Partitioning

David Brown ^{1,2}, Sebastien Carpentier ³, Sytze de Bruin ¹, Kauê de Sousa ⁴, Lewis Machida ⁵ and Jacob van Etten ⁴

¹Laboratory of Geo-Information Science and Remote Sensing, Wageningen University & Research, Wageningen, Netherlands; ²Bioversity International, Turrialba, Costa Rica; ³Bioversity International, Leuven, Belgium; ⁴Bioversity International, Rome, Italy; ⁵Bioversity International, Kampala, Uganda

Leave Empty
This space will be automatically filled with a QR code and number for easy sharing

Background

Crop variety trials are important to generate insights on variety environmental adaptation.

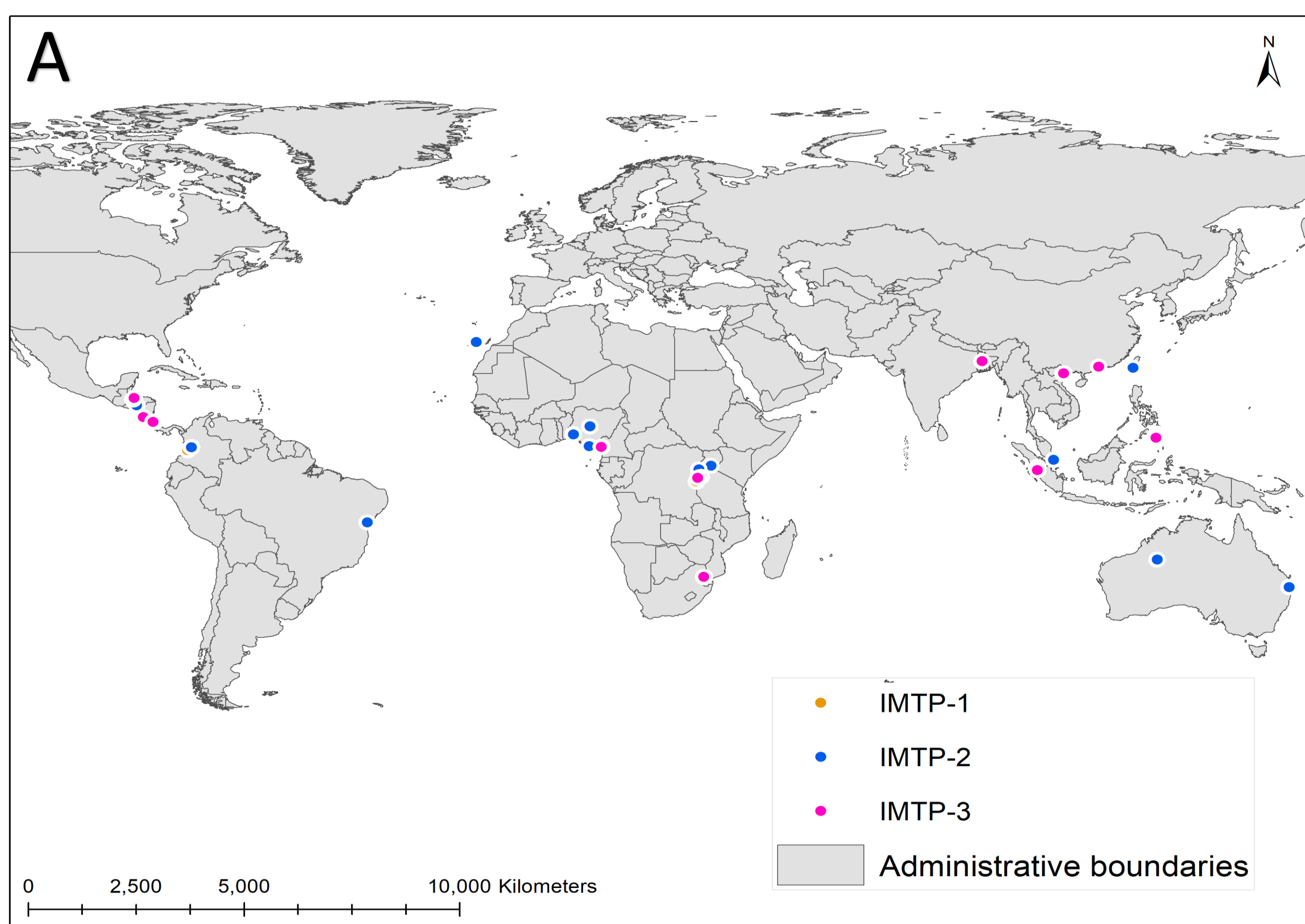
Aggregating data across various trials could generate more value from the same data, but this is often challenging due to data heterogeneity.

Previous research has shown that ranking-based methods can deal with heterogeneous data from different trials to gain insights in average performance of genotypes¹, but this approach has not been used in combination with climatic covariates yet.

We propose the Bradley-Terry model² in combination with recursive partitioning (Bradley-Terry trees)³ for data synthesis⁴ of trial data to gain insights in the influence of environmental variation on cultivar performance.

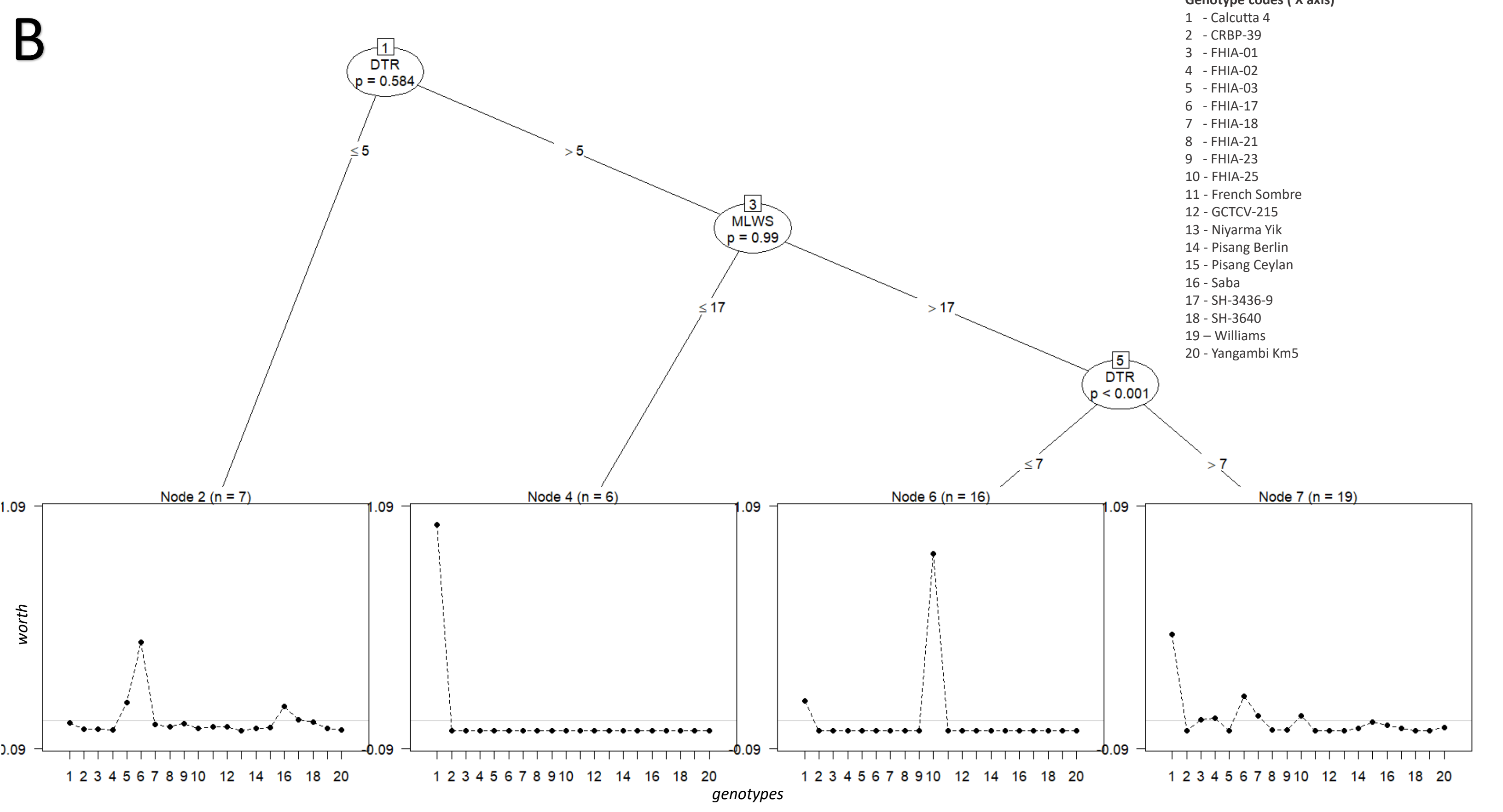
Materials and Methods

A data synthesis approach⁴ was used to aggregate and analyze legacy trial data from the three phases of International *Musa* Testing Program (IMTP)^{5,6,7} retrieved from the AgTrials database⁸. Each IMTP phase established trials at different locations (Figure A) and time frames, to test genotype reaction to black leaf streak disease. The original data were transformed into pairwise rankings to be analyzed with Bradley-Terry model. Temperature and precipitation data from the AgERA5⁹ database were used to generate climate covariates using the package climatrends¹⁰ in R¹¹. A Bradley-Terry tree model³ was fitted to the aggregated dataset, allowing to compare 20 genotypes.



Preliminary Results

The splitting variables selected by the recursive partitioning algorithm were diurnal temperature range (DTR) and maximum length of consecutive days with rain >= 1 mm (MLWS). The terminal nodes in the Bradley-Terry tree (Figure B) indicate differences in the genotypes' reaction to black leaf streak disease under different climatic conditions.



1. Simko, I., & Pechenick, D. A. (2010). Combining partially ranked data in plant breeding and biology: I. Rank aggregating methods. *International Journal of the Faculty of Agriculture and Biology, Warsaw University of Life Sciences, Poland*, 5(1), 41-55.
2. Bradley, R. A., & Terry, M. E. (1952). Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika*, 39(3/4), 324.
3. Strobl, C., Wickelmaier, F., & Zeileis, A. (2011). Accounting for Individual Differences in Bradley-Terry Models by Means of Recursive Partitioning. *Journal of Educational and Behavioral Statistics*, 36(2), 135-153.
4. Brown, D., Van den Bergh, I., de Bruin, S., Machida, L., & van Etten, J. (2020). Data synthesis for crop variety evaluation. A review. *Agronomy for Sustainable Development*, 40(4), 25. doi:10.1007/s13593-020-00630-7
5. Jones, D., & Tézenas du Montcel, H. (1994). *Final Report for UNDP/World Bank on the Results of the International Musa Testing Programme (Phase I)*. Retrieved from Montpellier, France:
6. Orjeda, G. (2000). *Evaluating Bananas: A Global Partnership: Results of IMTP Phase II*. Montpellier, France: Bioversity International.
7. Crichton, R., & Van den Bergh, I. (2014). *Publication of the results of the International Musa Testing Programme, Phase III (IMTP III)*. Paper presented at the XXIX International Horticultural Congress on Horticulture: Sustaining Lives, Livelihoods and Landscapes (IHC2014): IX 1114.
8. Hyman, G., Espinosa, H., Camargo, P., Abreu, D., Devare, M., Arnaud, E., . . . Traore, S. (2017). Improving agricultural knowledge management: The AgTrials experience [version 2; peer review: 2 approved]. *F1000Research*, 6(317). doi:10.12688/f1000research.11179.2
9. Copernicus Climate Change Service. (2019). Agrometeorological indicators from 1979 up to 2019 derived from reanalysis (Publication no. 10.24381/cds.6c68c9bb). AgERA5 Retrieved 2019, from ECMWF <https://cds.climate.copernicus.eu/cdsapp#!/dataset/sis-agrometeorological-indicators?tab=overview>
10. de Sousa, K., van Etten, J., & Solberg, S. Ø. (2020). climatrends: Climate Variability Indices for Ecological Modelling (Version 0.1.6). Retrieved from <https://CRAN.R-project.org/package=climatrends>
11. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

