iVES OENO One vine & wine

DOI: https://doi.org/10.20870/oeno-one.2026.60.3.9871

OPEN ACCESS



SPECIAL ISSUE

How long is long enough? A case study on the stability of commercial and pre-commercial grapevine rootstock rankings from multi-year data

Luis Gonzaga Santesteban[1,2,*], Javier Eraso[3], Carlos Miranda[1,2], Maite Loidi[1], Santiago Yániz[1], Diana Marín[1], Haizea Romeo[1], Mónica Galar-Martínez[1,2], Maider Velaz[1,2], Nazareth Torres[1,2], Ana Villa-Llop[1,2,3]

[1] Department of Agronomy, Biotechnology and Food Science, Public University of Navarre, Pamplona, Navarra, Spain
[2] Institute for Multidisciplinary Research in Applied Biology (IMAB-UPNA), Public University of Navarre, Campus Arrosadia 31006 Pamplona, Spain
[3] Vitis Navarra S.A.T. Nursery, Larraga, Navarra, Spain

This article is a short communication article published in cooperation with the 16th International Terroir Congress and the 3rd ClimWine Symposium (July 5–9, 2026), hosted by the École Supérieure des Agricultures in Angers, France.

Guest editors: Cécile Coulon-Leroy and Etienne Neethling.

Article number: 9871

*correspondence:
gonzaga.santesteban@unavarra.es

Associate editor:
Cécile Coulon-Leroy

Received:
29 January 2026

Accepted:
08 April 2026

Published:
01 July 2026

This article is published under the **Creative Commons licence** (CC BY 4.0).

Use of all or part of the content of this article must mention the authors, the year of publication, the title, the name of the journal, the volume, the pages and the DOI in compliance with the information given above.

ABSTRACT

In viticulture, the evaluation of scion varieties and rootstocks relies on multi-site field trials that must be maintained and monitored for several years once the vineyards become productive. Such trials are costly and resource-intensive, so it would be valuable to know how long they must be maintained to yield meaningful results. However, we have little evidence of how long a trial must last to reach robust and reliable conclusions. To shed light on this question, we analysed cluster number, yield, pruning weight, and Ravaz index data from a 10-year Tempranillo trial that included both commercial and newly developed rootstocks to quantify how the number of evaluated years affects the robustness of rootstock performance rankings. The dataset was analysed using several complementary approaches. For each possible trial duration from two to ten years, all combinations of years were treated as partial series, and their rankings were compared with those obtained from the complete 10-year dataset. Rank-based statistics, including Kendall's τ and pairwise inversion rate, were used to quantify stability. To assess the robustness of conclusions across different trial durations, we also applied leave-one-year-out and leave-k-years-out simulations. The analyses of commercial rootstocks' performance revealed that approximately five years of evaluation may be sufficient for commercial rootstocks, whereas pre-commercial genotypes may require longer evaluation periods.

KEYWORDS: Terclim 2026, long-term evaluation, trial duration, stability, rootstock breeding programs, genotype-environment interaction

INTRODUCTION

Grapevine scion varieties and rootstocks are typically evaluated through multi-site field trials conducted over several years, which require substantial resources and long-term data collection (Santesteban *et al*., 2023; Pitt *et al*., 2018). However, despite their importance, local information on the performance of rootstocks and varieties beyond the first decade of production remains scarce. This lack of long-term evidence is relevant because apparent rankings among genotypes can shift as vines age, and early-year data may not accurately predict later performance (Pitt *et al*., 2018).

Although it is widely accepted that multi-year testing is essential to obtain reliable information on plant material performance, we still have little evidence of how long such trials should last to reach robust and reproducible conclusions. Previous studies in grapevine have shown that year-to-year variability and genotype × environment interactions can mask genetic differences when evaluations are short (Migicovsky *et al*., 2021), and that the influence of rootstocks on scion performance may only become fully apparent after many years in the field (Ollat, 2015; Ferlito *et al*., 2020). Yet, due to practical limitations, most comparative trials of grapevine plant material are typically assessed for a limited number of seasons. This limitation is particularly relevant when evaluation focuses on pre-commercial rootstocks, *i.e.*, new materials that have been bred to improve tolerance to biotic and abiotic stresses over the established ones (Marín *et al*., 2021; Chen *et al*., 2024), but whose long-term consistency under field conditions needs to be verified carefully before going into the market.

In this work, we analysed data from a 10-year Tempranillo trial that included both commercial and newly developed rootstocks to quantify how the number of evaluated years affects the robustness of rootstock performance rankings, as a case study of a practically important yet seldom explored aspect of long-term field experimentation.

MATERIALS AND METHODS

1. Plant material and experimental set-up

A long-term field trial with *Vitis vinifera* L. cv. Tempranillo was conducted at the Vitis Navarra nursery in Miranda de Arga (Navarre, Spain; 42° 27’ 50.6” N, 1° 48’ 10.6” W; 308 m.a.s.l.), within the Ebro Valley viticultural region. The site has a continental-Mediterranean climate, with mean annual rainfall around 350–400 mm. Soils are Quaternary alluvial, with sandy loam texture, moderate active lime (7.5–8.0 %), pH ≈ 8.6, and organic matter ≈ 2.0 %, showing adequate nutrient supply and no salinity risks.

The vineyard was planted in spring 2011 with bench-grafted Tempranillo vines (omega grafting), combining 12 commercial and 9 newly developed rootstocks (RG series). RG series were obtained from the first generation of a hybridisation programme via cross-pollination between 41 B MGt (*V. vinifera* × *V. berlandieri*, clone V14D) and 110 R (*V. rupestris* × *V. berlandieri*, clone 1D). Further description of new hybrids can be read in Marin *et al*. (2023). A randomised complete block design was used, with three replicates of ten vines per rootstock. Vines were trained to a unilateral Royat cordon, pruned to five two-node spurs, and not shoot-trimmed during the season. Plant spacing was 3.0 m × 1.0 m (3333 vines ha^{-1}). Drip irrigation was applied during the summer period (July – September) as a supplementary input (≈ 30–70 mm per season) to avoid severe water stress and was applied equally to all rootstocks. The inter-row was maintained with natural vegetation that was mown once or twice a year, and the under-vine strip kept weed-free with herbicides. Vineyard management has remained consistent since establishment; more details on the experimental layout can be found in Marín *et al*. (2019).

2. Agronomic characterisation

From 2014 (first harvest) to 2023, cluster number (CN), yield (Yield) and pruning weight (PW) were determined individually for each vine and averaged to calculate the value for each replicate. Each harvest, bunches of the ten vines from the three replicates were collected, counted and weighted to obtain yield per rootstock-replicate by using a handheld hanging scale (mod. CH 15K20, KERN & Sohn GmbH, Germany). Similarly, once leaves had completely fallen, vines were pruned, and the pruning weight was measured with the hanging scale. Yield and pruning weight were used to calculate the Ravaz index as a ratio between the two variables.

3. Statistical analysis

The dataset was analysed using a rank-based stability framework to quantify how the number of evaluated years affects the reliability of rootstock performance rankings for each variable. For each possible trial duration from two to ten years, all combinations of years were treated as partial series, and their rankings were compared with those derived from the complete 10-year dataset. To measure stability, we calculated two complementary statistics; (i) Kendall’s τ and (ii) pairwise inversion rate. Together, these metrics provide a direct and intuitive evaluation of ranking consistency across years and were selected because they focus on relative performance rather than absolute values and allow the assessment of the stability of rankings without relying on assumptions about the distribution of the data (*e.g.*, normality or homogeneity of variances).

Kendall’s rank correlation coefficient (τ) is a non-parametric measure of association specifically designed for ordinal data. Kendall’s τ evaluates the degree to which two rankings preserve the same pairwise ordering of items, and is well-suited to assess the stability of rootstock rankings derived from datasets of different temporal lengths. Formally, Kendall’s τ is defined as:

$$\tau = \frac{N_c - N_d}{\frac{1}{2}n(n-1)}$$

where is the number of concordant pairs, the number of discordant pairs, and n is the total number of ranked rootstocks. A concordant pair is defined as a pair of rootstocks whose relative order is identical in both rankings, whereas a discordant pair indicates a reversal in relative position. T values range from −1 (complete inversion of rankings) to +1 (perfect concordance), with values close to zero indicating weak or unstable agreement.

Pairwise inversion rate was also calculated. This metric quantifies the proportion of rootstock pairs whose relative ordering changes when rankings are derived from partial datasets rather than from the reference series. Formally, the pairwise inversion rate is defined as:

$$IR = \frac{N_d}{\frac{1}{2}n(n-1)}$$

where is the number of discordant pairs between two rankings and is the total number of ranked rootstocks. A discordant pair corresponds to a pair of rootstocks whose relative positions are reversed when comparing the two rankings. The inversion rate ranges from 0 to 1, where 0 indicates perfect preservation of pairwise order (no inversions) and 1 indicates complete reversal of all pairwise relationships. In practical terms, the inversion rate represents the probability that the relative ranking of two rootstocks would be misclassified when a reduced number of seasons is used for evaluation.

To further assess the robustness of both Kendall's τ and the pairwise inversion rate, a resampling strategy was applied for calculations. For each trial duration, all possible combinations of years were considered, which implicitly encompasses leave-one-year-out and leave-k-years-out configurations. This approach links stability to the number of evaluated seasons and does not correspond to a single temporal sequence, which can be affected.

For all the calculations, analyses were performed separately for commercial and newly developed rootstocks to evaluate whether the time required to reach stable rankings differed between the two groups. All data analyses were conducted in R (R Core Team, 2025).

RESULTS

The rank-based analyses revealed that the stability of each variable-based rootstock rankings increased progressively with the number of evaluated years (Figure 1), as reflected by increasing Kendall's τ values. However, the rate at which high τ values were reached differed between variables. Taking $\tau \approx 0.85$ as an indicative threshold of high ranking stability, for cluster number and pruning weight (Figures 1b and 1c), this value of concordance was reached after 4 years, earlier than for yield (Figure 1a), where 5 years were needed and, specially, for Ravaz index, where the lowest concordance values were obtained after 7 years (Figure 1d). When the behaviour of commercial and newly developed (RG series) rootstocks was analysed, clear differences were found between these groups. For all variables except the Ravaz index, the ranking of commercial rootstocks showed higher Kendall's τ values, indicating greater stability of the results. Conversely, τ values for RG rootstocks required more years to stabilise, showing greater dispersion if the time series was short or intermediate. In fact, in this dataset, Kendall's τ values for RG rootstocks did not reach a clear plateau even after 10 years of evaluation.

The analysis of the pairwise inversion rate (Figure 2) showed a similar behaviour, as a progressive decrease was observed as the number of evaluated years increased. Similarly to the results obtained with Kendall's τ, the speed to reach low inversion rates was not the same for all the variables, cluster number and pruning having a more rapid reduction (Figures 2b and 2c), which indicates that misclassification of rootstock pairs for them was uncommon after having data for a few seasons. Yield showed an intermediate behaviour, while the Ravaz index had the highest inversion rates, reflecting greater instability (Figures 2a and 2d). When rootstocks were analysed by group, the commercial rootstocks tended to display lower inversion rates than the RG series, for all variables except for the Ravaz index, where inversion rates had similar values for both groups.

DISCUSSION

Long-term field trials are necessary to evaluate grapevine rootstocks and varieties, but they are time- and resource-intensive. These trials involve establishing vineyards with replicates and collecting data over multiple seasons once the vines reach full production. In practice, constraints related to cost, logistics, together with the need to generate publishable or transferable results, make researchers limit the duration of the comparative trials established. However, plant material (and especially rootstock effects) can change as vines age, so some years' results do not necessarily correspond to longer-term performance (Ollat *et al*., 2015; Ferlito *et al*., 2020). This concern has already been mentioned in mature rootstock trials, where yield rankings at young vineyards sometimes did not match those observed later (Pitt *et al*., 2018). In particular, this work showed in Shiraz and Cabernet-Sauvignon trials that the yields measured at three to six years of age did not correspond with those at 22–25 years, when some rootstocks exhibited age-related declines and others showed more stable long-term performance. In our work, although we have not extended rootstock evaluation to such an advanced age, we continuously evaluated rootstock performance for 10 years, starting from the first productive year, when vines were 3 years old.

Differences in ranking stability among variables showed that not all of them have the same response to the length of evaluation. In our conditions, cluster number and pruning weight reached stable rankings with fewer years than yield and Ravaz index. This behaviour can be related to the fact that under Mediterranean climates, yield is more sensitive than cluster number and pruning weight to vine water status (Uriarte *et al*., 2025), and interannual variations in rainfall can therefore have more impact on yield variations. In line with this, multi-experiment studies have shown that yield components respond differently to environmental drivers such

as precipitation, evapotranspiration and temperature (Ohana-Levi *et al.*, 2024), with yield being more strongly affected by these factors than more structurally determined traits. This differential sensitivity likely contributes to the lower temporal stability observed for yield-related variables. The slower convergence observed for the Ravaz index is consistent with its nature as a ratio between yield and pruning weight, which can amplify short-term variability. However, its temporal stability does not depend only on being a ratio, but also on the stability of its component traits; accordingly, when both yield and pruning weight show relatively consistent rankings, the Ravaz index also tends to stabilise more rapidly.

When the behaviour of commercial and newly developed (RG series) rootstocks is compared, we observed differences in the time required to rank rootstocks in terms of the variables

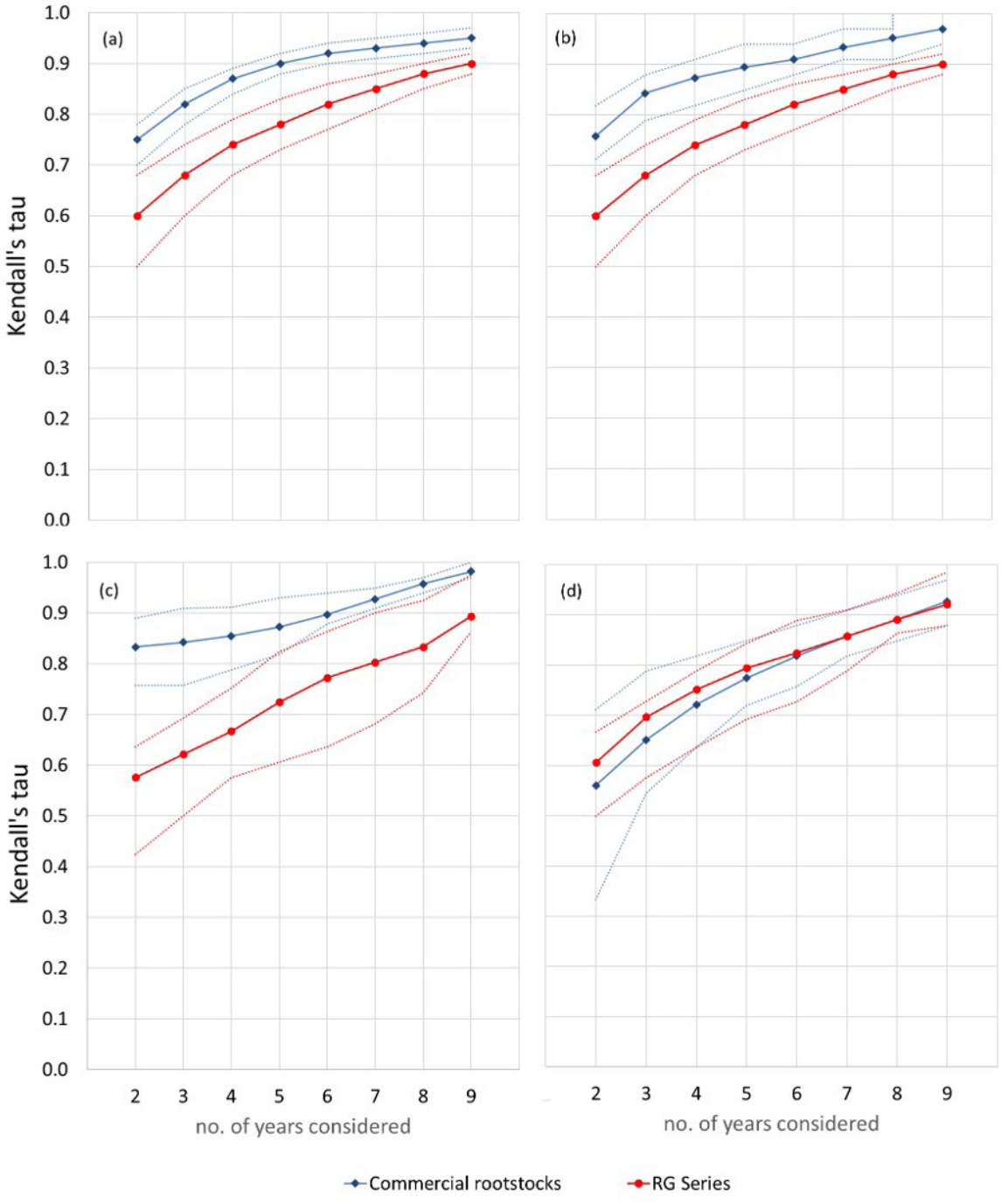


FIGURE 1. Stability of rootstock performance rankings according to Kendall's τ for (a) yield, (b) cluster number, (c) pruning weight, and (d) Ravaz index.

Solid lines represent mean Kendall's τ values calculated across all possible combinations of years for each trial duration and dashed lines indicate the percentiles 25 and 75.

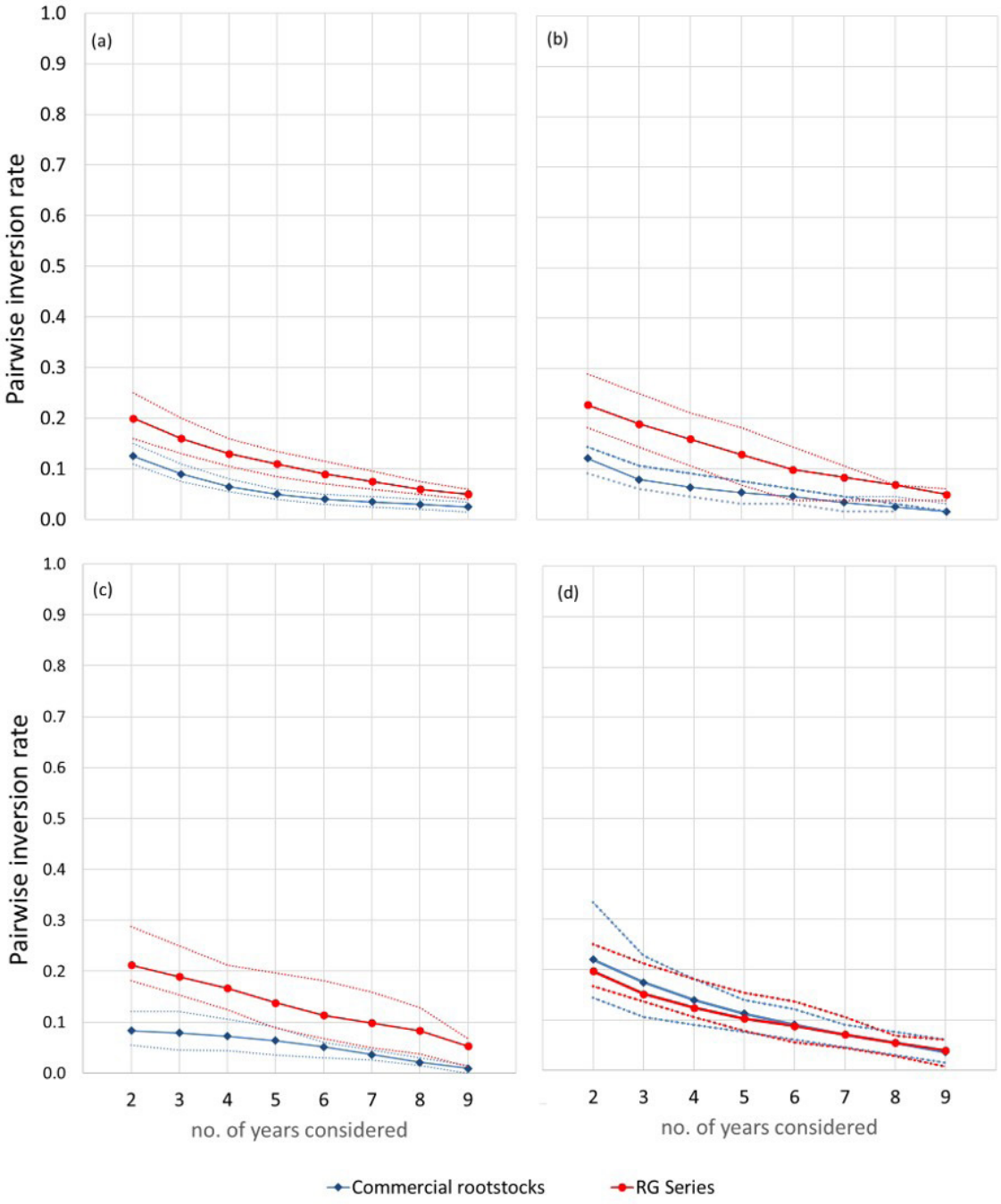


FIGURE 2. Stability of rootstock performance rankings according to the pairwise inversion rate for (a) yield, (b) cluster number, (c) pruning weight, and (d) Ravaz index.

Solid lines represent mean Kendall's τ values calculated across all possible combinations of years for each trial duration and dashed lines indicate the percentiles 25 and 75.

considered. In this regard, commercial materials had a more stable behaviour, and, therefore, a shorter time is required to characterise them. Although it is plausible that some of these newly developed materials are more sensitive to environmental changes than long-established commercial rootstocks, this factor alone is unlikely to explain the observed patterns. In our trial, some RG rootstocks exhibited a gradual decline in productivity as vines aged (data not shown), which has probably contributed to the slower stabilisation of their behaviour. A hypothesis that can explain this pattern can be related to the process by which rootstocks become commercially established. Over decades of use, rootstocks that show long-term declines in performance are less likely to be kept and widely adopted, *i.e.*, less likely to remain commercial. As a result, commercial rootstocks represent material that has already gone through a filtering process. In contrast, newly bred rootstocks have been evaluated for a more limited time and in a more limited range of environments. Thus, the evaluation of this RG series together with their parental lines (41 B and 110 R) showed genetic diversity in terms of growth, yield components, and phenolic and industrial maturity parameters, but no differential pattern between the commercial rootstocks and the new

hybrids over four years of evaluation (Marín *et al*., 2023). Nevertheless, previous studies have shown that rootstock effects on scion growth and productivity can evolve with vine age, reflecting adjustments in root system development, hydraulic functioning, or carbon allocation or even partial lack of compatibility (Ollat *et al*., 2015; Ferlito *et al*., 2020). The fact that there were no differences between commercial and RG series in terms of Ravaz index stability reinforces this hypothesis, since the changes associated with this potential decline apply similarly to yield and vegetative growth, making the Ravaz index less responsive to this trend.

Overall, our analysis of the number of years required to obtain reliable and reproducible rankings of rootstock performance based on the case study presented here provides a useful reference to help guide decisions on trial duration. For commercial rootstocks, evaluations conducted for around five years can be sufficiently informative, and although they may not reflect declines that can appear at a longer term, they are useful for vineyards in a similar age range. On the contrary, longer-term evaluations are needed for lesser-known materials, where changes in behaviour can appear relatively early in the vineyard's lifetime. In any case, it should be noted that our analysis is based on a single long-term dataset under specific environmental and management conditions. Although the patterns observed are consistent and agronomically meaningful, extending this type of analysis to multiple datasets across environments and genetic backgrounds would be valuable. This aggregated dataset could allow assessing the generality of our findings, and complement the rank-based approach we used here with model-based or probabilistic statistical methods to quantify the uncertainty and separate the effects of year-to-year variability from intrinsic differences among rootstocks.

ACKNOWLEDGEMENTS

This long-term evaluation was made possible thanks to the concurrent support of several funding sources, including the Government of Navarra's programmes Vit-Foot, Vit-Feet and Best-Feet; the Spanish National Program for Research (UPGRAPE, PID2021-123305OB-C32 funded by MICIN/AEI 10.13039/501100011033 and by the European Union Next GenerationEU/PTR), the CIEN-CDTI initiative LowpHwine, and the cross-border cooperation projects EFA 324/19 – VITES QUALITAS and EFA 033/01 – VITRES (both co-financed at 65 % by the European Regional Development Fund through the Interreg V-A Spain–France–Andorra programme). We are also deeply grateful to all the technical and field staff at Vitis Navarra for their dedication and support in maintaining the trial over the full evaluation period.

REFERENCES

Chen, Y., Fei, Y., Howell, K., Chen, D., Clingeleffer, P., & Zhang, P. (2024). Rootstocks for grapevines now and into the future: selection of rootstocks based on drought tolerance, soil nutrient availability, and soil pH. *Australian Journal of Grape and Wine Research*, 2024(1), 6704238. https://doi.org/10.1155/2024/6704238 https://doi.org/10.1155/2024/6704238

Ferlito, F., Distefano, G., Gentile, A., Allegra, M., Lakso, A. N., & Nicolosi, E. (2020). Scion–rootstock interactions influence the growth and behaviour of the grapevine root system in a heavy clay soil. *Australian Journal of Grape and Wine Research*, 26(1), 68-78. https://doi.org/10.1111/ajgw.12415

Marín, D., Armengol, J., Carbonell-Bejerano, P., Escalona, J. M., Gramaje, D., Hernández-Montes, E., Intrigliolo, D.S., Martínez-Zapater, J.M., Medrano, H., Mirás-Avalos, J.M., Palomares-Rius, J.E., Romero-Azorín, P., Savé, R., Santesteban, L.G. & De Herralde, F. (2021). Challenges of viticulture adaptation to global change: tackling the issue from the roots. *Australian Journal of Grape and Wine Research*, 27(1), 8-25. https://doi.org/10.1111/ajgw.12463

Marín, D., García, R., Eraso, J., Urrestarazu, J., Miranda, C., Royo, J. B., Abad, F.J. & Santesteban, L. G. (2019). Evaluation of the agronomic performance of 'Syrah' and 'Tempranillo' when grafted on 12 rootstocks. *Vitis*, *58,* 111-118. https://doi.org/10.5073/vitis.2019.58.special-issue.111-118

Marín, D., Miranda, C., Abad, F.J., Urrestarazu, J., Mayor, B., Villa-Llop, A., & Santesteban, L.G. (2023). Agronomic evaluation of eight 41 B × 110 richter grapevine genotypes as rootstock candidates for Mediterranean viticulture, *Horticultural Plant Journal*, 9 (4), 720-728, https://doi.org/10.1016/j.hpj.2022.10.002.

Migicovsky, Z., Cousins, P., Jordan, L. M., Myles, S., Striegler, R. K., Verdegaal, P., & Chitwood, D. H. (2021). Grapevine rootstocks affect growth-related scion phenotypes. *Plant Direct*, 5(5), e00324. https://doi.org/10.1002/pld3.324

Ohana-Levi, N., Cohen, Y., Munitz, S., Michaelovsky, R., Ferman Mintz, D., Hagag, N., Getz, Y., & Netzer, Y. (2024). The response of yield, number of clusters, and cluster weight to meteorological factors and irrigation practices in grapevines: A multi-experiment study. *Scientia Horticulturae*, 326, 112761. https://doi.org/10.1016/j.scienta.2023.112761.

Ollat, N., Peccoux, A., Papura, D., Esmenjaud, D., Marguerit, E., Tandonnet, J. P., Bordenave, L., Cookson, S.J., Barrieu, F., Rossdeutsch, L., Lecourt, J., Lauvergeat, V., Vivin, P., Bert, P.-F. & Delrot, S. (2015). Rootstocks as a component of adaptation to environment. *In:* Grapevine in a changing environment: a molecular and ecophysiological perspective, 68-108. https://doi.org/10.1002/9781118735985.ch4

Pitt, T., Skewes, M., Stevens, R., Tan, J., Nicholas, P., & McCarthy, M. (2018). Longevity and sustained performance of rootstocks for Australian vineyards. *Wine & Viticulture Journal* 33, 47-50.

R Core Team. (2025). *R: A language and environment for statistical computing* (Version 4.4.2) [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Santesteban, L. G., Rekarte, I., Torres, N., Galar-Martínez, M., Villa-Llop, A., Visconti, F., Intrigliolo, D.S., Escalona, J.M., De Herralde, F. & Miranda, C. (2023). The role of rootstocks for grape growing adaptation to climate change. Meta-analysis of the research conducted in Spanish viticulture. *OENO One*, 57(2), 283–290. https://doi.org/10.20870/oeno-one.2023.57.2.7439

Uriarte, D., Santesteban, L. G., Mirás-Avalos, J. M., Buesa, I., Cancela, J. J., Chacón, J. L., Escalona, J.M., Intrigliolo, D.S., Lampreave, M., Montoro, A., Rivacoba, L., Visconti, F., Yuste, J. & Miranda, C. (2025). Quantifying the effects of water status on grapevine vegetative growth, yield, and grape composition through a collaborative analysis. *Australian Journal of Grape and Wine Research*, 2025(1), 1588228. https://doi.org/10.1155/ajgw/1588228