

# What Is the Minimal Optimal Sample Size for Plant Ecophysiological Studies?

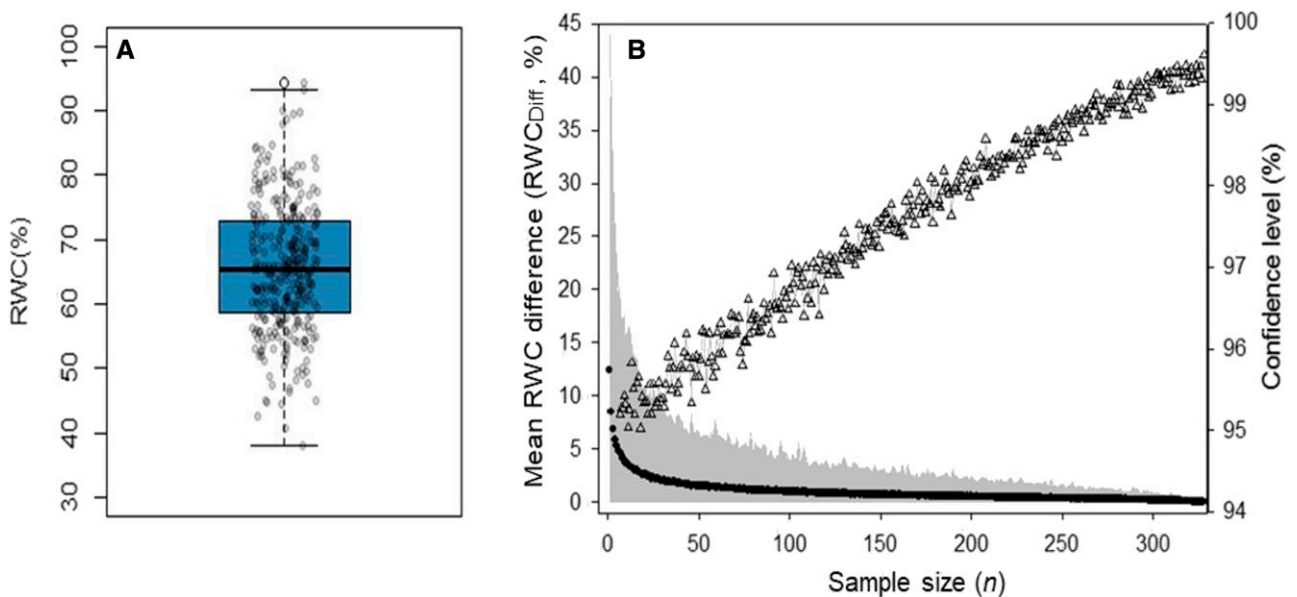
Dear Editor,

Although significant advances have been made toward elucidating the most appropriate sample size for studies in many fields (Meere and Mulchrone, 2003; Fiske et al., 2008; McDonald, 2008; Pérez-Harguindeguy et al., 2013; Hajian-Tilaki, 2014), choosing the minimal optimal sample size in plant physiology remains a challenge. This is particularly true for ecophysiology (i.e. research carried out in the field using physiological techniques), where both top-down and bottom-up approaches are required to understand not only the responses of individuals but also those of populations and ecosystems to a variety of environmental factors such as high temperature, drought, or salinity (Leroux and Loreau, 2015). Achieving high sample sizes is limited by both internal (e.g. intraspecific variability) and external factors (e.g. time, human capital, and funding). Here, we discuss the importance of key internal factors constraining sample size, individual heterogeneity, sample size representativity, and context-dependent variability, to provide suggestions to ascertain the minimal optimal sample size that is compatible with hypothesis testing in plant ecophysiological studies.

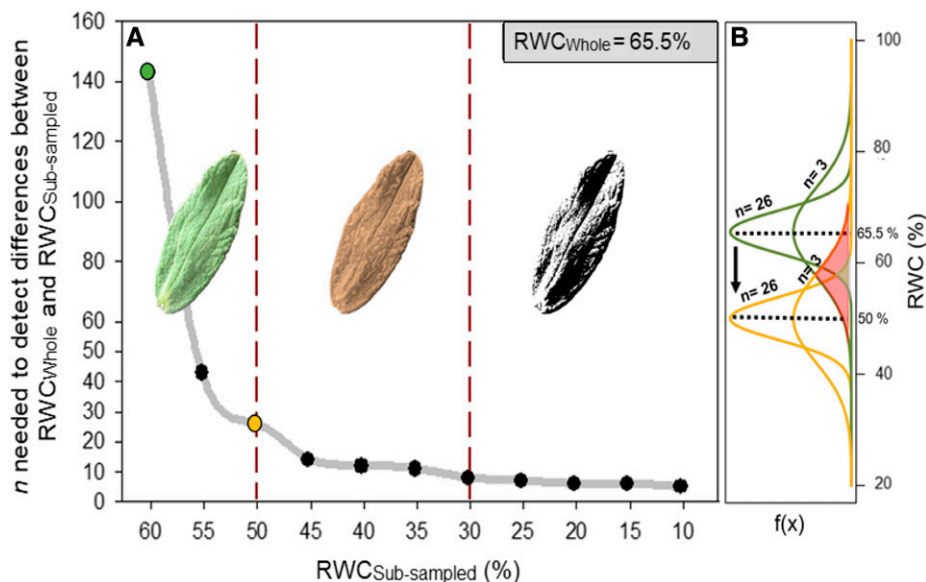
To illustrate our narrative, we employ the widely used functional trait relative water content (RWC).

The precision of any estimate is affected by its sampling variability, as well as by process variability, including environmental, inter- and intraindividual, and methodological variability (White, 2000). These sources of variability tend to be closely linked (Messier et al., 2010) and can influence the trueness of the measurement, thus constraining the statistical power necessary to detect potential intergroup significant differences (Bean et al., 2012). Although maximizing intraspecific sample size generally improves statistical power, the improvement is typically species- (Fiske et al., 2008) and trait-specific (Harmon and Losos, 2005), and may be limited by external factors such as time and funding.

When considering process variability (i.e. the combined effect of demographic, spatial, temporal, and individual variability), not only intraspecific but also inter- and intraindividual variability must be considered. Indeed, each species is characterized by a distinct array of functional trait values (Violle et al., 2007), which may vary at spatial and temporal scales (Messier et al., 2010). Individual differences may be



**Figure 1.** Sample size representativeness of RWC in a population of 328 individuals of *C. albidus*. A, Boxplot and scatter plot of the RWC data. The boxplot shows the median of the data between the first and the third quartiles; whiskers indicate the 10th and 90th percentiles. B, Percent mean difference between our population ( $n = 328$ ) and a population of variable sample size ( $n = 1 \dots 328$ ) after 10,000 iterations at each sample size (black dots). A sample size of  $n = 14$  corresponds to the inflection point (estimated via the segmented function from the SEGMENTED R package). The gray area represents the minimum and maximum mean differences between the populations, and white triangles represent the confidence level, that is, the percentage of iterations in which the compared samples were found nonsignificantly different among 10,000 iterations.



**Figure 2.** Sample size required to detect differences in RWC between subsampled populations of *C. albidus* (see Fig. 1) compared to the whole population containing 328 individuals. A, Dots represent the sample size necessary to detect significant differences at a  $P < 0.05$  in RWC between a whole population ( $n = 328$ ,  $RWC_{\text{Whole}}: 65.55\% \pm \text{SE} = 0.54$ ) and subsampled populations with different mean values, as represented by the x axis, while keeping the SD constant. Dashed vertical lines at RWC 50% and 30% represent severe water stress and the point of irreversible water loss, respectively. Sampled subpopulations were performed using 10,000 iterations.  $RWC_{\text{Whole}}$  is the mean RWC of the whole population, and  $RWC_{\text{Subsampled}}$  is the mean RWC of the subsampled populations. B, Comparison of the whole population to a subsampled, simulated population ( $RWC_{\text{Subsampled}}: 50\%$ ) with two different sample sizes representing these populations ( $n = 26$  and 3). The red-shaded area represents  $P > 0.05$ , whereas the green-shaded area represents  $P < 0.05$ .

detected by sampling among different individual plants, whereas intraindividual differences may emerge when sampling within the same individual at different spatio-temporal scales. For instance, Valladares et al. (2000) found significant differences for structural and physiological leaf traits (including leaf mass per unit area, photosynthetic capacity, and root:shoot ratio, among others) across a light gradient in growth chambers. Most recently, Aguilar-García et al. (2018) found a great deal of variation in floral traits as a function of their position around individuals of *Myrtillocactus geometrizans*. These findings emphasize the fact that sample size may affect the accuracy of estimates of ecophysiological traits (White, 2000; Bean et al., 2012). Furthermore, increased variability in the trait of interest (e.g. due to high phenotypic plasticity; West-Eberhard, 2003) may compromise the sample representativeness in a complex way.

To illustrate how the sources of variability influence the sample size in an ecophysiological study, we quantified RWC across 328 individuals of *Cistus albidus* (Cistaceae) in four sites, located more than 300 m apart from each other, within a natural population in Spain (41.589N, 1.835E, 987 m a.s.l., Spain) from July 4 to 14, 2014. When defining our population as all 328 individuals,

the mean  $RWC_{\text{Whole}}$  was  $\mu = 65.55\% \pm 0.54 \text{ SE}$  (Fig. 1A). Indeed, although we found a great deal of individual variance, no significant differences in RWC were found as a function of sampled day and location (generalized linear model; site:  $F_{\text{df}=3} = 1.01$ ,  $P > 0.05$ ; day:  $F_{\text{df}=4} = 0.61$ ,  $P > 0.05$ ). To understand the relationship between intraspecific variability, sample size, and representativeness, we subsampled these data 10,000 times for increasing sample sizes ( $n = 1, 2, 3 \dots 328$ ), and then we contrasted the obtained mean value ( $RWC_{\text{Subsampled}}$ ) with the sampled whole-population mean ( $RWC_{\text{Whole}}$ ). The mean RWC difference ( $RWC_{\text{Diff}} = RWC_{\text{Whole}} - RWC_{\text{Subsampled}}$ ) between each of these simulated subsampled populations and the whole population for each sample size was calculated to determine the asymptote in this relationship, whereby no increase in the marginal benefit was achieved by increasing sample size (Fig. 1B). In spite of the fact that the inflection point in Figure 1B corresponded to  $n = 14$ , a sample size of four individuals was enough to guarantee sample representativeness at an accuracy of 95%. Strikingly, however, to achieve an accuracy of 99%, a sample size of 279 individuals was needed (Fig. 1B). Furthermore, in studies where multiple comparisons exist, larger sample sizes are required. As an example, we show that a sample size of 26 individuals would be required to detect significant differences between two populations with means of 65.55% of RWC and 50% at an accuracy of 95% (Fig. 2).

<sup>2</sup>Author for contact: smunne@ub.edu.

www.plantphysiol.org/cgi/doi/10.1104/pp.18.01001

**BOX 1. Recommendations to accurately determine the minimal optimal sample size of ecophysiological measurements in natural populations and to improve the quality of in situ estimates:**

- Run a pilot experiment on a reduced sample size and carry out power analysis (McDonald, 2008). The results of the power analysis will report how large the sample size should be.
- Implement jack-knife techniques on your data to estimate the inflection point of variance, which indicates the sample size beyond which increases in sampling effort do not significantly improve accuracy (See Figure 2).
- Whenever possible, quantify and report the variance components associated with sampling (i.e. spatial, temporal, and inter- and intraspecific variability), as governed by the experimental design.
- Endeavour to reduce methodological variability. This can be achieved by having only one highly trained person collecting the data, or training and centralising a team of fieldworkers, such that data are collected from individual plants within the same environmental and temporal conditions, and in the protocol and instruments.
- Take advantage of background noise; do not ignore it! Environmental conditions can vary from one day to the next, and harnessing and quantifying that variation can actually help contextualize the measured values (e.g. when all data are collected under a clear, sunny day after a month of drought, vs. a clear, sunny day but after a day of rains).

In conclusion, although the choice of sample size is usually constrained by external factors (e.g. funding, time, and portable, fast technology), researchers must carefully keep in mind the magnitude of the effects of internal factors (e.g. environmental variability, inter-individual variability, and method accuracy) on the power of analysis conferred by those sample sizes. This is particularly important when researchers account for the total variability that may constrain sample representativity and statistical power to optimize experimental designs and report results that are in fact most representative of natural settings. For instance, in terms of the representativity of an estimate of RWC in our natural population of *C. albidus* (a situation that also may be relevant to other traits and species), a minimal optimal sample size of four may suffice to obtain a good proxy for the actual value in natural settings, but differences between populations are not detected unless more than 20 individuals are sampled in a single time point:  $n = 26$  in our case study. Therefore, we recommend carefully considering the effects of

sampling size in ecophysiological studies to best depict phenomena in nature. We urge the scientific community to go well beyond the standard of  $n = 3$  individuals still found in several recent examples in the literature and to carefully consider the sampling size adequate for each functional trait and species. To that end, we provide a few recommendations in Box 1.

Marina Pérez-Llorca,<sup>a,1</sup>

Erola Fenollosa,<sup>a</sup> Roberto Salguero-Gómez,<sup>b,c,d</sup> and Sergi Munné-Bosch<sup>a,2</sup>

<sup>a</sup>Department of Evolutionary Biology, Ecology and Environmental Sciences, University of Barcelona, 08028 Barcelona, Spain

<sup>b</sup>Department of Zoology, University of Oxford, Oxford OX1 3SZ Oxford, United Kingdom

<sup>c</sup>Evolutionary Biodemography Laboratory, Max Planck Institute for Demographic Research, Rostock 18057, Germany

<sup>d</sup>Centre for Biodiversity and Conservation Science, University of Queensland, St Lucia, Queensland 4072, Australia

ORCID ID: 0000-0001-6523-6848 (S.M.-B.)

#### LITERATURE CITED

- Aguilar-García SA, Figueroa-Castro DM, Valverde PL, Vite F (2018) Effect of flower orientation on the male and female traits of *Myrtillocactus geometrizans* (Cactaceae). *Plant Biol (Stuttg)* 20: 531–536
- Bean WT, Stafford R, Brashares S (2012) The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. *Ecography* 35: 250–258
- Fiske IJ, Bruna EM, Bolker BM (2008) Effects of sample size on estimates of population growth rates calculated with matrix models. *PLoS One* 3: e3080
- Hajian-Tilaki K (2014) Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform* 48: 193–204
- Harmon LJ, Losos JB (2005) The effect of intraspecific sample size on Type I and Type II error rates in comparative studies. *Evolution* 59: 2705–2710
- Leroux S, Loreau M (2015) Theoretical perspectives on bottom-up and top-down interactions across ecosystems. In T Hanley, K La Pierre, eds, *Trophic Ecology: Bottom-Up and Top-Down Interactions across Aquatic and Terrestrial Systems*. Cambridge University Press, Cambridge, UK
- McDonald JH (2008) *Handbook of Biological Statistics*. Sparkly House Publishing, Baltimore, MD
- Meere PA, Mulchrone KF (2003) Effect of sample size on geological strain estimation from passively deformed clastic sedimentary rocks. *J Struct Geol* 25: 1587–1595
- Messier J, McGill BJ, Lechowicz MJ (2010) How do traits vary across ecological scales? A case for trait-based ecology. *Ecol Lett* 13: 838–848
- Pérez-Harguindeguy N, Díaz S, Garnier E, Poorter H, Jaureguiberry P, Bret-Harte MS, Cornwell WK, Craine JM, Gurvich DE, Urcelay C, (2013) New handbook for standardised measurement of plant functional traits worldwide. *Aust J Bot* 64: 715–716
- Valladares F, Wright SJ, Lasso E, Kitajima K, Pearcy RW (2000) Plastic phenotypic response to light of 16 congeneric shrubs from a Panamanian rainforest. *Ecology* 81: 1925–1936
- Violle C, Navas M-L, Vile D, Kazakou E, Fortunel C, Hummel I, Garnier E (2007) Let the concept of trait be functional! *Oikos* 116: 882–892
- West-Eberhard MJ (2003) *Developmental Plasticity and Evolution*. Oxford University Press, New York
- White GC (2000) Population viability analysis: data requirements and essential analyses. In L Boitani, TK Fuller, eds, *Research Techniques in Animal Ecology: Controversies and Consequences*. Columbia University Press, New York, pp 288–331