This article is a *Plant Cell* Advance Online Publication. The date of its first appearance online is the official date of publication. The article has been edited and the authors have corrected proofs, but minor changes could be made before the final version is published. Posting this version online reduces the time to publication by several weeks.

PERSPECTIVE

Reassess the *t* Test: Interact with All Your Data via ANOVA

Siobhan M. Brady,[a] Meike Burow,[b] Wolfgang Busch,[c] Örjan Carlborg,[d] Katherine J. Denby,[e] Jane Glazebrook,[f] Eric S. Hamilton,[g] Stacey L. Harmer,[a] Elizabeth S. Haswell,[g] Julin N. Maloof,[a] Nathan M. Springer,[h] and Daniel J. Kliebenstein[b,i,1]

[a] Department of Plant Biology, University of California, Davis, California 95616
[b] DynaMo Center of Excellence, University of Copenhagen, DK-1871 Frederiksberg C, Denmark
[c] Gregor Mendel Institute, Austrian Academy of Sciences, Vienna Biocenter, 1030 Vienna, Austria
[d] Department of Clinical Sciences, Division of Computational Genetics, Swedish University of Agricultural Sciences, SE-75007 Uppsala, Sweden
[e] School of Life Sciences and Warwick Systems Biology Centre, University of Warwick, Coventry CV4 7AL, United Kingdom
[f] Department of Plant Biology, University of Minnesota, St. Paul, Minnesota 55108
[g] Department of Biology, Washington University in St. Louis, St. Louis, Missouri 63130
[h] Microbial and Plant Genomics Institute and Department of Plant Biology, University of Minnesota, St. Paul, Minnesota 55108
[i] Department of Plant Sciences, University of California, Davis, California 95616

Plant biology is rapidly entering an era where we have the ability to conduct intricate studies that investigate how a plant interacts with the entirety of its environment. This requires complex, large studies to measure how plant genotypes simultaneously interact with a diverse array of environmental stimuli. Successful interpretation of the results from these studies requires us to transition away from the traditional standard of conducting an array of pairwise *t* tests toward more general linear modeling structures, such as those provided by the extendable ANOVA framework. In this Perspective, we present arguments for making this transition and illustrate how it will help to avoid incorrect conclusions in factorial interaction studies (genotype × genotype, genotype × treatment, and treatment × treatment, or higher levels of interaction) that are becoming more prevalent in this new era of plant biology.

IDENTIFYING BIOLOGICAL INTERACTIONS BETWEEN AND AMONG TREATMENTS, GENOTYPES, AND ENVIRONMENTS IS CRITICAL IN PLANT SCIENCE

Testing interactions between and among treatments, genotypes, and environments (Table 1) is central to nearly every field of plant biology, from genetics tests for epistasis, to physiology tests for interactions of multiple treatments. Understanding how results translate from one condition to another requires us to determine how these variables interact with each other in the context of an experiment. These interactions between variables form the basis of integrative studies that aim to assess how genetic variation influences the response to a specific treatment or environment. As a result, numerous plant biology studies require robust statistical methods to test hypotheses about how two variables interact.

OVERUSE OF STUDENT'S *t* TESTS

Given the ubiquity of testing interactions, plant biologists are naturally well versed in the importance of assessing their data for statistical significance. Unfortunately, the analysis methods used are not always appropriate. A survey of three recent issues of *The Plant Journal* (Vol. 81, issues 1 to 3), *The Plant Cell* (Vol. 26, issues 10 to 12), and *Plant Physiology* (Vol. 166, issue 4, and Vol. 167, issues 1 and 2) showed that 83 of 185 articles (45%) relied solely upon pairwise *t* tests using a single trial of an experiment to analyze quantitative data, with the vast majority of studies involving multiple variables, experiments, or interactions. Of the remaining articles, 42 (23%) presented no quantitative data relevant to the statistics discussed in this article (i.e., modeling results or developmental pictures) and 41 (22%) reported quantitative data for which statistical analysis was either not conducted or not described. Finally, only 19 (10%) combined the data from multiple trials within an ANOVA to directly test for an interaction between two variables. Although we do not suggest that ANOVA would have been the best option in all of the above instances, we feel that these numbers indicate the extent to which ANOVA is underutilized in our community. The *t* test (described below) is familiar to most molecular biologists, is easy to perform and interpret, and is properly applied when there is no interaction among variables or variation across trials. However, given the increasing prevalence of experiments investigating the interaction among variables within a single study, the *t* test does not fully use the power of the experimental data or even directly test the hypothesis that two variables/components interact. Under these conditions, the trial-

[1] Address correspondence to kliebenstein@ucdavis.edu.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Daniel J. Kliebenstein (kliebenstein@ucdavis.edu).
www.plantcell.org/cgi/doi/10.1105/tpc.15.00238

PERSPECTIVE

Table 1. Definitions for This Article

Term	Definition
Additive	A condition in which the quantifiable phenotypic effect of allelic variation (mutations/polymorphisms) in nonallelic genes can be simply added to predict the phenotype of the polygenic mutant. This also applies to individual genes when distinguishing between heterozygotes and homozygotes.
Biological replicate	One of several samples where an organism with the same genotype is grown or treated with the same conditions and independently evaluated. Biological replicates may be sampled within a single experiment or in replicate experiments (trials); the power of an experiment is vastly improved by conducting replicate experiments.
Environment	Any and all growth conditions under which the plant is maintained. Some aspects of the environment may be varied in a controlled fashion depending on the experiment.
Epistasis	A condition in which the phenotypic effect of allelic variation (mutations/polymorphisms) across single genes does not predict the phenotype of the polygenic mutant.
Experiment	A test established to assess if a chosen factor of interest (genotype, treatment, environment, etc.) affects a particular phenotype.
Factor	An experimental element that is to be tested for its phenotypic contribution. For example, treatment (presence versus absence of a hormone), trial number (1, 2, etc.), or genotype (wild type versus mutant) are common factors. This is often equivalent to the term variable.
Genotype	The allelic state of all genes within each individual line being studied in an experiment
Interaction	A condition in which the combined effect of two simultaneous treatments, or the effect of a genotype and a treatment together, leads to a change in the phenotype that cannot be predicted from the isolated treatment or genotypic effects. This is similar to the definition of epistasis for two or more genes
P value	The probability that the data are consistent with the null hypothesis under a theoretical random set of experiments and trials. If the calculated P value is less than a given threshold, frequently 0.05, then the null hypothesis is rejected in favor of the alternative hypothesis. Typically in plant biology the null hypothesis is that the treatment or genotype has no effect on the observed phenotype and the alternative hypothesis is that there is an effect.
Phenotype	Any measurable property of a plant that is of interest to the researcher.
Redundancy	A specific form of epistasis in which single gene mutations have little to no known phenotypic effect, while the polygenic mutant displays an altered phenotype. This is a form of epistasis in which the single mutants do not fully predict the polygenic mutant phenotype.
Statistical power	The ability of a statistical test to reveal a statistical significance that actually exists within a set of data.
Technical replicate	Data points derived from multiple assays on the same sample. This is employed to assess the level of variance that is contributed by variance in technical aspects of measurement, i.e., from the equipment or experimental protocol. These should be averaged prior to statistical analysis or accounted for differently than biological replicates.
Treatment	Any manipulation of the external or internal environment that is applied by the experimenter to test the phenotypic response of the plant.
Trial	A completely independent growth and testing of plants conducted to assess the experimental question at hand and how reproducible the answer is across trials. Trials are typically separated in calendar times. Typically, experiments should contain multiple independent trials for which data sets can be combined.
Variation	Any measurable difference in phenotype between individuals; can be caused by changes in genotype, treatment, or stochastic processes.

specific *t* test can deliver false conclusions that could be avoided by combining all the data in an ANOVA.

LINEAR STATISTICAL MODELS SUCH AS ANOVA PROVIDE MORE POWER TO TEST THE HYPOTHESIS OF INTERACTION

Employing a linear statistical model like ANOVA can allow a researcher to include data from all experiments under analysis and make better use of existing experimental designs that seek to test interactions. Here, we use two simulated data sets to compare statistical analyses of multiple variables across multiple trials using *t* tests and ANOVA. We first show how ANOVA can prevent a false conclusion that would have been drawn if relying on the *t* test and then show how ANOVA can reveal a true conclusion that would have been dismissed by the *t* test. Finally, we illustrate the critical nature of the replication underlying the experiment and the importance of post-hoc graphical analysis of the data. Because all research programs rely on previous results to shape future experiments, preventing errors that can affect further studies will

PERSPECTIVE

require increased efforts to more accurately model the data. Transitioning to ANOVA and crucially using all available data can help to increase the productivity of a research group by diminishing inferential errors. We aim to demonstrate how changing our standard approach to statistical analysis of interactions could significantly benefit plant biologists beyond these specific instances.

t TEST CALCULATIONS

A *t* test assesses if the means of two groups differ; we can choose one of several different variant *t* tests depending upon assumptions about the variances and directionality. In the articles we surveyed, most used a *t* test that assumed equal variances, using the equation $t = \frac{\overline{X_1} - \overline{X_2}}{s_{X_1X_2} \cdot \sqrt{\frac{2}{n}}}$. In this equation, $\overline{X_1} - \overline{X_2}$ is the difference in the means between the two groups being compared (genotypes A and B or treatments 1 and 2), while $s_{X_1X_2}$ is the pooled within-group SD and n is the number of samples for each group (if the groups have equal sample sizes). The P value is then estimated by comparing the *t* value to the expected distribution of *t* across all experiments of a similar sample size (degrees of freedom), the *t*-distribution. As can be seen from the equation for a *t* test, it cannot be extended to more groups or factors, is limited to a single pairwise comparison at a time, and thus is not designed to test interactions.

Resources

t tests can be run in nearly any worksheet package (such as Excel), statistical package (such as R, SAS, or SPSS), and even a number of calculator applications.

ANOVA CALCULATIONS

ANOVA is a class of linear statistical models written in a linear algebraic form that is extensible and allows us to specifically test multiple variables and their interactions. For example, to compare genotypes across two treatments with a term for trial, the model to explain the phenotype (y_{gte}) for each specific genotype (g), treatment (t), and trial or environment (e), would be written as $y_{gte} = \mu + G_g + T_t + E_e + G_g \times T_t + \varepsilon_{gte}$, where μ represents a constant, G represents the contribution of the genotype (g), T represents the contribution of the treatment (t), and E represents the contribution of the environment (e), incorporating the variation in all of the trials. Finally, ε represents residual error. Thus, y_{gte} can represent all the phenotypes for every genotype in every treatment in every trial or, more simply, all the data. Another way to think of this is that the model hypothesizes that the phenotype (y_{gte}) may be altered by variation in the genotypes (G_g) and/or variation in the treatments (T_t) and/or variation in the trials (E_e) and an interaction of the genotypes with the treatment ($G_g \times T_t$) plus some error that can't be controlled (ε_{gte}). A researcher can add additional levels for each factor (e.g., multiple levels of a treatment or different mutant alleles of a gene) as well as other factors, and it is thus extendable to new experimental designs.

Any variation in phenotype is then partitioned into the contributions made by the various factors and significance is determined by an F-test where $F = \frac{\text{variance between treatments}}{\text{variance within treatments}}$. This essentially asks how the variation in the average phenotype between the genotypes in the experiment compares to the random variation in the experiment. We can then determine the P value for this factor's *F* value by comparison to the *F*-distribution, the predicted distribution of all possible *F* values across all similarly sized experiments. The significant benefit of ANOVA, or other linear models, is that including all the experimental factors that may affect the phenotype allows for a better estimation of the random variation and thus gives more precision to the *F* value than to the corresponding *t* value.

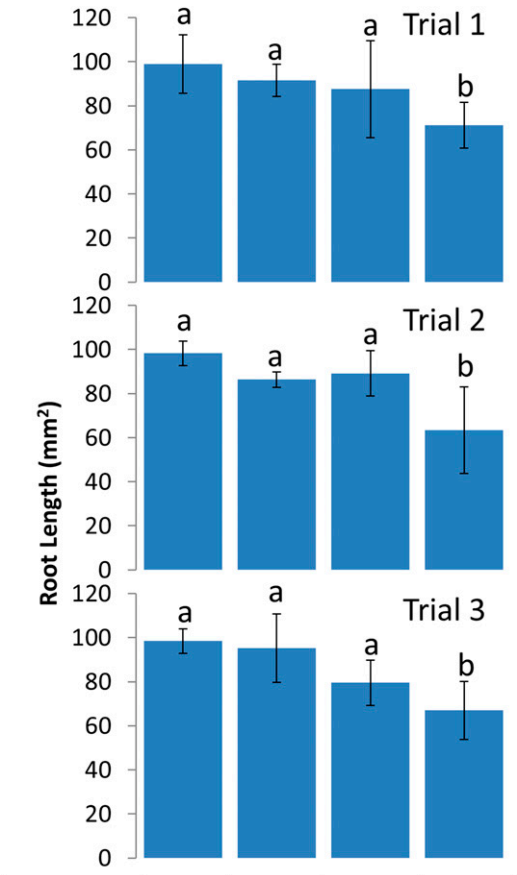
Extensions

The extensibility of the ANOVA linear model is one of its particular strengths. This includes the ability to add variables and generate different models to directly test a hypothesis. One-way ANOVA tests variation across a single variable, whereas two-way ANOVA assesses two variables and can include their potential interaction. This can be extended further. In addition to multiple variables, the variables can be classed as fixed or random effects. Fixed effects can be considered as variables that can be qualitatively grouped, such as genotypes. In contrast, random effects can be considered as variables that show quantitative variation or possibly a large range of levels that cannot be readily grouped into discrete classes. Both fixed and random effects can be included in the same model, leading to the generation of mixed models.

Resources

Simple one-way ANOVAs can be run in most worksheet packages such as Excel. However, the worksheet packages contain prepackaged structures that may not match the linear model to be tested. The application of two-way or more complex models requires the use of a publicly available statistical package such as R or a commercially available package such as SAS and SPSS. These statistical packages provide more flexibility and are thus more likely to be suitable for the types of analyses described here. Within R, the basic package allows for simple ANOVA analysis using the aov command (Crawley, 2014; R Development Core Team, 2014). A number of websites have information about basic statistics in R (http://www.statmethods.net/ and http://cran.r-project.org/). These include guides to beginning analysis (http://cran.r-project.org/web/packages/HSAUR/vignettes/Ch_analysis_of_variance.pdf). The basic R package also allows for the use of multivariate approaches like MANOVA. Advanced linear models will require the use of the CAR or LME4 packages, which allow the user to individualize the analysis to properly test the hypothesis using the experimental design command (Fox and Weisberg, 2011; Bates et al., 2014; Crawley, 2014).

PERSPECTIVE



Genotype 1	WT	KO	WT	KO
Genotype 2	WT	WT	KO	KO

ANOVA	Model Term			
Trial	G1	G2	G1 x G2	Trial
1	No	No	No	-
2	Yes	Yes	No	-
3	No	Yes	No	-
Combined	Yes	Yes	No	No

Figure 1. Case Study 1: Model of ANOVA Testing of How Two Genes May or May Not Interact.

Using the model described in the supplemental methods wherein two genes interact to modulate root length in a purely additive fashion, we randomly generated three independent trials involving three independent samples per genotype. These were then used to conduct *t* tests and ANOVA both within each trial and by combining all the data. The bar graphs show the standard representation of mean ± SD with letters showing if there is a significant difference from the control group using Student's *t* test comparisons of the mean. The summary table shows the ANOVA results (full results are shown in Supplemental Table 1). "No" means the term is not significant and "Yes" means it is significant. The ANOVA for each trial and combined across all trials is shown. Trial shows the trial term in the ANOVA combining all the trial data. KO, knockout genotype; WT, wild-type genotype. Data are shown in Supplemental Table 3.

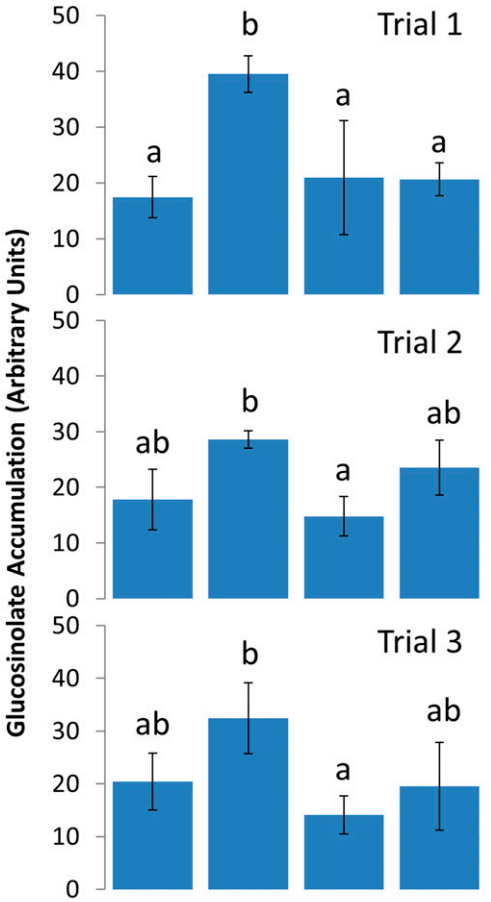
ANOVA PROPERLY TESTS GENETIC INTERACTIONS

Genetic analysis commonly tests if two genes show an epistatic interaction. This can examine, for example, whether two genes in a gene family have overlapping functions or if different genes interact epistatically to control a phenotype by functioning in the same pathway (for examples, see Lynch and Force, 2000; Carlborg et al., 2006; Freeling, 2009; Joseph et al., 2013; Mackay, 2014). For genes with redundant functions, both genes may need to be mutated to produce an altered phenotype. Unfortunately, testing genes for redundancy or epistasis by using pairwise *t* tests to compare mutant to wild type is prone to errors. Performing an ANOVA, which uses the data from all individuals and all trials of the experiment together, can describe the interaction more accurately. To illustrate this, we simulated two genes with a purely additive interaction with each single mutant leading to a reduction of 15 mm in root length in comparison to the wild type and the double mutant having a reduction of 30 mm root length (see Supplemental Methods for the fabricated data sets). Then we built a model that randomly drew samples from a normal distribution with homogenous variance for each class.

Using an ANOVA approach to analyze randomly simulated data from all three independent trials, each with three biological replicates, identified the true underlying genetic model (Figure 1). It showed that each gene has an additive effect but finds no significant support (significance threshold is $P < 0.05$) for an interaction between the effects of gene 1 and gene 2, indicating no support for epistasis or redundancy of these two genes (Figure 1) (Li et al., 2008). To illustrate how combining all the data into a single analysis increased the power of the test, we also ran ANOVA within each separate trial (Figure 1; Supplemental Table 1). This trial-specific ANOVA made it difficult to draw any conclusion from the study, as the results were inconsistent between trials (Figure 1). The effects of the individual mutants were only significant in some of the tests using independent trials, which could have misled the researcher to conclude that each gene does not affect root length (Figure 1). Yet, the combined ANOVA explicitly showed that each genotype does affect root length.

Assessing reproducibility in the face of experimental variation requires the ability to assess the effect of variation among the trials. In the combined ANOVA, examining the "trial" term (E) revealed no statistical support for the three trials of the experiment being different (Figure 1). Thus, the researcher can conclude that the differences between the trials are not significant. If the trial variation is large, it is possible to use the modeling approach to directly address whether variation across trials interacts with the other terms to assess if the results are not reproducible across the trials. The ability of ANOVA to identify the correct underlying genetic model is further boosted by the inclusion of more data

PERSPECTIVE



Genotype	WT	WT	KO	KO
Treatment	Ctl	+	Ctl	+

ANOVA	Model Term			
Trial	Geno	Treat	G x T	Trial
1	No	Yes	Yes	-
2	No	Yes	No	-
3	Yes	Yes	No	-
Combined	Yes	Yes	Yes	No

Figure 2. Case Study 2: Model Showing When Combined Data Resolves Confusion.

Using the model described in the supplemental information wherein a mutation alters glucosinolate accumulation in response to a given treatment, we randomly generated three independent trials involving three independent samples per genotype × treatment class. These data were then used to conduct *t* tests and ANOVA both within each trial and by combining all the data. The bar graphs show the standard representation of mean ± SD with letters showing if there is a significant difference from the control group using Student's *t* test comparisons of the mean. The summary table shows the ANOVA results (full results are shown in Supplemental Table 2). "No" means the term is not significant and "Yes" means it is significant. The ANOVA for each trial and combined across all trials is shown. KO, knockout genotype; WT, wild-type genotype; Ctl, control treatment; Treat, alternative treatment. Data are shown in Supplemental Table 4.

points per genotype in the combined analysis while simultaneously showing that there is no difference between the trials.

The use of *t* tests to compare each mutant to the wild type within each trial led to a different conclusion. The trial-specific *t* tests suggested that only the double mutant differs significantly from the wild type (Figure 1). Thus, using *t* tests would lead the researcher to conclude, incorrectly, that gene 1 and gene 2 are redundant and that the results were similar in three independent trials. The reason the *t* test fails in this instance is that it treats the double mutant separately from the single mutants (i.e., each test in isolation) and thus does not test the effects of the two genotypes on each other, but instead tests the effect of the combined genotypes (the wild type versus the double knockout). As such, in this instance, the lack of a detected effect is not equivalent to the lack of an effect but is instead caused by not directly testing the interaction. We note in this instance that the observation that the three trials were not significantly different would have allowed the researcher to pool the data and conduct *t* tests of the individual genotypes. However, in this scenario, the researcher would not be able to test the interaction between genes. The ANOVA allows the effect of alleles to be tested across all the genotypes and trials and thereby enables the researcher to describe the underlying genetic model.

ANOVA PROPERLY TESTS A GENOTYPE × TREATMENT INTERACTION

In the previous case study, ANOVA identified the correct underlying mechanistic model while the isolated trial *t* test misled the researcher by statistically suggesting an interaction that was not present biologically. An isolated trial *t* test approach can also lead to incorrect rejection of an interaction hypothesis. If, for example, the treatment response was moderate and the mutant phenotype was not fully penetrant, typical results could resemble those shown in Figure 2 (Supplemental Table 2 and Supplemental Methods) (Kliebenstein et al., 1999, 2002; Li et al., 2014; Taylor-Teeples et al., 2015). To illustrate this example, we modeled how jasmonate application affects glucosinolate accumulation in plants mutated in a single transcription factor that quantitatively controls jasmonate perception (Sønderby et al., 2007; Sønderby et al., 2010; Li et al., 2014). In this case, using *t* tests may lead researchers to assume from the second and third trials that the mutant responded to jasmonate like the wild type, while ignoring the first trial as an outlier. Alternatively, researchers may run additional trials or, worse, conclude that no real effect was detected and drop the entire line of investigation. Combining the data in an

PERSPECTIVE

ANOVA provides statistical support for the interaction and argues that the transcription factor alters jasmonate responses (Figure 2; Supplemental Table 2). Taken together, Figures 1 and 2 illustrate that it is possible to be both positively and negatively misled by the isolated trial *t* test approach.

CROSS-CHECK YOUR DATA ANALYSIS WITH VISUALIZATION

In both case studies discussed above, the ANOVA allowed the researcher to state whether or not they have evidence that their treatments, genotypes, and/or interactions influence their phenotype. However, the ANOVA does not provide visualization of the data that can enable a researcher to develop a model or hypothesis (for example, how a mutation affects root length). Thus, at this stage of analysis, it is very useful for a researcher to visualize the results by plotting the group means along with all of the individual data points to assess if the ANOVA and visual interpretation coalesce into a single conclusion. This visualization is easily conducted with the aid of bean or violin plots within R (Hintze and Nelson, 1998; Kampstra, 2008). Further analyses can include a comparison of the group means using post-hoc tests such as Tukey's or Welch's mean comparisons. Occasionally, the visual and statistical analyses do not agree. For instance, in trial 3 in Figure 2, the individual trial ANOVA provides no statistical support for an interaction but the means suggest an interaction. In such situations, exploring the data further can identify potential outliers due to technical errors such as genotype misclassification or misplanting. Additionally, further replicate trials should be conducted and/or the underlying hypothesis adjusted until the visualization and statistical analyses do agree. While it would ideally be possible to perform unlimited replicates to test all possible alternative hypotheses, this is not always the case given the numerous constraints on research. Thus, at a minimum, authors should fully describe any potential discrepancies or issues arising from comparisons of the visualization and statistical analyses. Finally, we note that all statistical approaches are susceptible to errors and as such the only way to cross-validate any result is to conduct an independent line of inquiry.

INDEPENDENT REPLICATION IS THE FOUNDATION OF ANY SUCCESSFUL HYPOTHESIS TEST

Appropriate and adequate independent replication within a biological experiment is essential to the successful use of any of the above approaches (Hurlbert, 1984; Vaux et al., 2012; Buttigieg and Ramette, 2014). A general reading of plant molecular literature shows that the most common strategy is to conduct three repeats in three different trials, although the nature of these biological repeats can vary widely. For example, a biological replicate may mean phenotyping three individuals per genotype per trial. The results from these three separate trials are typically analyzed independently, and one trial is shown in a figure with the researcher stating that similar results were found in other trials. This classical experimental design arose largely from molecular biological studies. A simple improvement is to double the replication from three to six or more individuals per trial if at all possible because increased levels of replication provide more power to any statistical approach.

In addition, the replicates performed need to be independent, meaning that there should be as little connection as possible between the samples. For example, leaves from different plants of a single genotype are more biologically independent from each other than leaves collected from the same plant (Schmid et al., 2003; Schmid et al., 2005). Measuring multiple leaves for a given plant increases accuracy in measuring the phenotype of that specific individual, but could be highly misleading if that individual plant is not representative of a genotype because of location, pathogen infection, or another factor. Thus, instead of repeatedly sampling the same individual, it is better to sample multiple individuals to minimize the potential for bias in the analysis. Even more independent are leaves from different plants where the seeds were obtained from different mothers. The researchers need to carefully determine what may influence a measurement and work to randomize sampling, or control the influences that they are not testing, to ensure that all replicates are as independent as possible (Elwell et al., 2011).

SUMMARY

The case studies presented above illustrate how *t* tests or trial-specific analyses can mislead a researcher. These examples only provide a small sample of all the possible instances when trial-specific pairwise *t* tests might lead to incorrect descriptions of the underlying mechanistic model. As a simple, but by no means all-inclusive, solution, we urge that researchers use ANOVA or other linear models (e.g., mixed models or random effect models) to directly test interactions and to include all the data in a single model. While there are many books or reviews that can be useful to learn and implement these statistical approaches, a readily available resource on this topic may be our colleagues in ecology, evolution, quantitative genetics, statistics, or mathematics, who can rapidly provide the necessary guidance to conduct the suggested analyses. The resources suggested here may also be a starting point for those new to this type of statistics. No matter how we get there, we anticipate that a shift in the field to employing linear models that include all the experimental data will help to improve the community's interpretation of experiments and generation of hypotheses.

Supplemental Data

Supplemental Table 1. Model of when *t* tests improperly inform about how two genes may or may not interact.

Supplemental Table 2. Model of when combined data resolves *t* test confusion.

Supplemental Table 3. Data from Figure 1.

Supplemental Table 4. Data from Figure 2.

Supplemental Methods. Journal analysis and computational models.

ACKNOWLEDGMENTS

We thank the following people for comments and suggestions that greatly improved this article: Fumiaki Katagiri, Stefan Schmollinger, Detlef Weigel, Cynthia Weinig, and additional anonymous reviewers. We apologize for the suggested topics that we were unable to include in this review because of editorial focus.

AUTHOR CONTRIBUTIONS

All authors contributed to the writing. D.J.K. conducted statistical modelling.

Received March 17, 2015; revised May 20, 2015; accepted July 17, 2015; published July 28, 2015.

REFERENCES

Bates, D., Maechler, M., Bolker, B., Walker, and S. (2014). lme4: Linear mixed-effects models using Eigen and S4. (http://CRAN.R-project.org/package=lme4).

Buttigieg, P.L., and Ramette, A. (2014). A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses. FEMS Microbiol. Ecol. **90:** 543–550.

Carlborg, O., Jacobsson, L., Ahgren, P., Siegel, P., and Andersson, L. (2006). Epistasis and the release of genetic variation during long-term selection. Nat. Genet. **38:** 418–420.

Crawley, M.J. (2014). Statistics: An Introduction Using R. (Wiley).

Elwell, A.L., Gronwall, D.S., Miller, N.D., Spalding, E.P., and Brooks, T.L.D. (2011). Separating parental environment from seed size effects on next generation growth and development in Arabidopsis. Plant Cell Environ. **34:** 291–301.

Fox, J. and Weisberg, S. (2011). An R Companion to Applied Regression. (Thousand Oaks, CA: SAGE).

Freeling, M. (2009). Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annu. Rev. Plant Biol. **60:** 433–453.

Hintze, J.L., and Nelson, R.D. (1998). Violin plots: a box plot-density trace synergism. Am. Stat. **52:** 181–184.

Hurlbert, S.H. (1984). Pseudoreplication and the design of ecological field experiments. Ecol. Monogr. **54:** 187–211.

Joseph, B., Corwin, J.A., Li, B., Atwell, S., and Kliebenstein, D.J. (2013). Cytoplasmic genetic variation and extensive cytonuclear interactions influence natural variation in the metabolome. eLife **2:** e00776.

Kampstra, P. (2008). Beanplot: A Boxplot Alternative for Visual Comparison of Distributions. J. Stat. Softw. **28:** 1–9.

Kliebenstein, D.J., Dietrich, R.A., Martin, A.C., Last, R.L., and Dangl, J.L. (1999). LSD1 regulates salicylic acid induction of copper zinc superoxide dismutase in *Arabidopsis thaliana.* Mol. Plant Microbe Interact. **12:** 1022–1026.

Kliebenstein, D.J., Lim, J.E., Landry, L.G., and Last, R.L. (2002). Arabidopsis UVR8 regulates ultraviolet-B signal transduction and tolerance and contains sequence similarity to human regulator of chromatin condensation 1. Plant Physiol. **130:** 234–243.

Li, B., Gaudinier, A., Tang, M., Taylor-Teeples, M., Nham, N.T., Ghaffari, C., Benson, D.S., Steinmann, M., Gray, J.A., Brady, S. M., and Kliebenstein, D.J. (2014). Promoter-based integration in plant defense regulation. Plant Physiol. **166:** 1803–1820.

Li, J., Hansen, B.G., Ober, J.A., Kliebenstein, D.J., and Halkier, B.A. (2008). Subclade of flavin-monooxygenases involved in aliphatic glucosinolate biosynthesis. Plant Physiol. **148:** 1721–1733.

Lynch, M., and Force, A.G. (2000). The origin of interspecific genomic incompatibility via gene duplication. Am. Nat. **156:** 590–605.

Mackay, T.F.C. (2014). Epistasis and quantitative traits: using model organisms to study gene-gene interactions. Nat. Rev. Genet. **15:** 22–33.

R Development Core Team (2014). R: A Language and Environment for Statistical Computing. (Vienna, Austria: R Foundation for Statistical Computing).

Schmid, K.J., Sørensen, T.R., Stracke, R., Torjek, O., Altmann, T., Mitchell-Olds, T., and Weisshaar, B. (2003). Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in *Arabidopsis thaliana*. Genome Res. **13:** 1250–1257.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D., and Lohmann, J.U. (2005). A gene expression map of *Arabidopsis thaliana* development. Nat. Genet. **37:** 501–506.

Sønderby, I.E., Burow, M., Rowe, H.C., Kliebenstein, D.J., and Halkier, B.A. (2010). A complex interplay of three R2R3 MYB transcription factors determines the profile of aliphatic glucosinolates in Arabidopsis. Plant Physiol. **153:** 348–363.

Sønderby, I.E., Hansen, B.G., Bjarnholt, N., Ticconi, C., Halkier, B. A., and Kliebenstein, D.J. (2007). A systems biology approach identifies a R2R3 MYB gene subfamily with distinct and overlapping functions in regulation of aliphatic glucosinolates. PLoS One **2:** e1322.

Taylor-Teeples, M., et al. (2015). An Arabidopsis gene regulatory network for secondary cell wall synthesis. Nature **517:** 571–575.

Vaux, D.L., Fidler, F., and Cumming, G. (2012). Replicates and repeats–what is the difference and is it significant? A brief discussion of statistics and experimental design. EMBO Rep. **13:** 291–296.

Trial 1	Df	SS	MS	F	P
Gene1	1	420	420	2.049	0.190
Gene2	1	749	749	3.656	0.092
G1 x G2	1	61	61	0.296	0.601
Residuals	8	1639	205		

Trial 2	Df	SS	MS	F	P
Gene1	1	1072	1072	8.022	0.022
Gene2	1	776	776	5.809	0.043
G1 x G2	1	144	144	1.075	0.330
Residuals	8	1069	134		

Trial 3	Df	SS	MS	F	P
Gene1	1	187	187	1.365	0.276
Gene2	1	1670	1670	12.2	0.008
G1 x G2	1	64	64	0.468	0.513
Residuals	8	1095	137		

All Trials	Df	SS	MS	F	P
Gene1	1	1492	1492	10.87	0.003
Gene2	1	3078	3078	22.42	< 0.001
Trial	2	61	31	0.223	0.802
G1 x G2	1	257	257	1.873	0.181
Residuals	30	4118	137		

Supplemental Table 1. Model of when t-tests improperly inform about how two genes may or may not interact. Using the model described in the supplemental information wherein two genes interact in a pure additive fashion, we randomly generated three independent trials involving three independent samples per genotype. These were then used to conduct *t*-test analysis by comparing the mutant to the WT. Additionally, ANOVA both by individual trial and by combining all the data were conducted and are presented. The ANOVA tables are shown for the same data. All variance in this model is set to random with no variance programmed between experiments.

Exp1	Df	SS	MS	F	P
Geno	1	176	176	5.126	0.053
Treat	1	353	353	10.261	0.013
G x T	1	375	375	10.894	0.011
Residuals	8	275	34		

Exp1	Df	SS	MS	F	P
Geno	1	49	49	2.885	0.128
Treat	1	286	286	16.731	0.003
G x T	1	3	3	0.187	0.676
Residuals	8	137	17		

Exp1	Df	SS	MS	F	P
Geno	1	277	277	7.100	0.029
Treat	1	226	226	5.800	0.043
G x T	1	32	32	0.800	0.390
Residuals	8	3118	39		

AllExp	Df	SS	MS	F	P
Geno	1	455	455	14.372	< 0.001
Treat	1	859	859	27.129	< 0.001
Exp	2	86	43	1.352	0.274
G x T	1	239	239	7.546	0.010
Residuals	30	949	32		

Supplemental Table 2. Model of when combined data resolves t-test confusion.
Using the model described in the supplemental information wherein a mutant partly affects the response to a given treatment, we randomly generated three independent trials involving three independent samples per genotype x treatment class. These were then used to conduct t-test analysis and ANOVA both by trial and by combining all the data. The ANOVA tables are shown for the same data. All variance in this model is set to random with no variance programmed between experiments.

Supplemental Table 3. Data for Figure 1

	Trial	Genotype 1	Genotype 2	Phen
1	1	WT	WT	89.687
2	1	WT	WT	114
3	1	WT	WT	92.921
4	1	KO	WT	84.152
5	1	KO	WT	91.562
6	1	KO	WT	98.893
7	1	WT	KO	67.527
8	1	WT	KO	84.085
9	1	WT	KO	111.08
10	1	KO	KO	79.783
11	1	KO	KO	74.242
12	1	KO	KO	59.7
13	2	WT	WT	94.293
14	2	WT	WT	104.63
15	2	WT	WT	95.869
16	2	KO	WT	82.253
17	2	KO	WT	89.11
18	2	KO	WT	87.475
19	2	WT	KO	77.439
20	2	WT	KO	96.802
21	2	WT	KO	93.049
22	2	KO	KO	58.537
23	2	KO	KO	84.836
24	2	KO	KO	46.442
25	3	WT	WT	102.37
26	3	WT	WT	100.88
27	3	WT	WT	92.123
28	3	KO	WT	103.72
29	3	KO	WT	104.55
30	3	KO	WT	77.288
31	3	WT	KO	85.046
32	3	WT	KO	85.771
33	3	WT	KO	67.647
34	3	KO	KO	72.609
35	3	KO	KO	52.016
36	3	KO	KO	76.299

Supplemental Table 4. Data for Figure 2

	Trial	Genotype	Treatment	Phen
1	1	WT	Ctl	20.754
2	1	WT	Ctl	18.252
3	1	WT	Ctl	13.404
4	1	KO	Ctl	9.8154
5	1	KO	Ctl	23.316
6	1	KO	Ctl	29.811
7	1	WT	Treat	42.836
8	1	WT	Treat	39.42
9	1	WT	Treat	36.235
10	1	KO	Treat	21.164
11	1	KO	Treat	17.475
12	1	KO	Treat	23.314
13	2	WT	Ctl	20.027
14	2	WT	Ctl	11.646
15	2	WT	Ctl	21.753
16	2	KO	Ctl	13.068
17	2	KO	Ctl	18.875
18	2	KO	Ctl	12.411
19	2	WT	Treat	28.445
20	2	WT	Treat	30.225
21	2	WT	Treat	27.162
22	2	KO	Treat	27.213
23	2	KO	Treat	17.936
24	2	KO	Treat	25.415
25	3	WT	Ctl	23.614
26	3	WT	Ctl	14.245
27	3	WT	Ctl	23.521
28	3	KO	Ctl	10.02
29	3	KO	Ctl	16.621
30	3	KO	Ctl	15.66
31	3	WT	Treat	24.988
32	3	WT	Treat	38.09
33	3	WT	Treat	34.08
34	3	KO	Treat	27.163
35	3	KO	Treat	20.789
36	3	KO	Treat	10.672

Supplemental Data. Brady et al. (2015). Plant Cell 10.1105/tpc.15.00238

Supplemental Methods: Journal analysis and computational models

Journal Review: To assess the relative prevalence of different statistical approaches in the literature, all 185 research articles in three recent editions of Plant Journal (volume 81, issues 1-3), Plant Physiology (volume 166, issue 4, and volume 167 , issues 1-2) and The Plant Cell (volume 26, issues 10-12) were read and classified as to their optimal statistical use as follows:

1) 42 papers were classified as N/A indicating that no quantitative data were presented that was appropriate for the statistics discussed in this article (i.e. modelling results or developmental pictures).
2) 83 papers were classified as *t*-test indicating that the researchers solely relied upon pair-wise tests or only presented the results from one trial (experiment) without combining data across all trials.
3) 41 papers were classified as No Report indicating that the researchers either did not conduct statistics or did not state what was done in the paper.
4) 19 papers were classified as using ANOVA or another linear modelling approach with all data from all trials.

R code for the simulated models

```
###R model for Figure 1
###Redundancy or Epistasis Model

###Data matrix creation for trial 1

WT=rnorm(3, mean=100, sd=10) ##WT genotype
M1=rnorm(3, mean=85, sd=10) ##Single mutant in Gene 1
M2=rnorm(3, mean=85, sd=10) ##Single mutant in Gene 2
DM=rnorm(3, mean=70, sd=10) ##Double mutant in Gene 1 and Gene 2
M1G <- c("WT", "WT", "WT", "M1", "M1", "M1", "WT", "WT", "WT", "M1", "M1", "M1")
```

```
M2G <- c("WT", "WT", "WT", "WT", "WT", "WT", "M2", "M2", "M2", "M2", "M2", "M2")
Exp <- c("1", "1", "1", "1", "1", "1", "1", "1", "1", "1", "1", "1")
Phen <- c(WT, M1, M2, DM)
str(Phen)
Rep1 <- data.frame(Exp,M1G,M2G,Phen)
Rep1$M1G <- factor(Rep1$M1G)
Rep1$M2G <- factor(Rep1$M2G)
Rep1$Exp <- factor(Rep1$Exp)
str(Rep1)

###ANOVA analysis for trial 1
Int1 <- aov(Phen ~ M1G + M2G + M1G:M2G , data=Rep1)
out<-capture.output(summary(Int1))
cat(out,file="C:\\R Work\\Full.txt",sep="\n", append=TRUE)

###Data matrix creation for trial 2
WT=rnorm(3, mean=100, sd=10)
M1=rnorm(3, mean=85, sd=10)
M2=rnorm(3, mean=85, sd=10)
DM=rnorm(3, mean=70, sd=10)
Exp <- c("2", "2", "2", "2", "2", "2", "2", "2", "2", "2", "2", "2")
Phen <- c(WT, M1, M2, DM)
str(Phen)
Rep2 <- data.frame(Exp,M1G,M2G,Phen)
Rep2$M1G <- factor(Rep2$M1G)
Rep2$M2G <- factor(Rep2$M2G)
Rep2$Exp <- factor(Rep2$Exp)
str(Rep2)
```

```
###ANOVA analysis for trial 2
Int2 <- aov(Phen ~ M1G + M2G + M1G:M2G , data=Rep2)
out<-capture.output(summary(Int2))
cat(out,file="C:\\R Work\\Full.txt",sep="\n", append=TRUE)

###Data matrix creation for trial 3
WT=rnorm(3, mean=100, sd=10)
M1=rnorm(3, mean=85, sd=10)
M2=rnorm(3, mean=85, sd=10)
DM=rnorm(3, mean=70, sd=10)
Exp <- c("3", "3", "3", "3", "3", "3", "3", "3", "3", "3", "3", "3")
Phen <- c(WT, M1, M2, DM)
str(Phen)
Rep3 <- data.frame(Exp,M1G,M2G,Phen)
Rep3$M1G <- factor(Rep3$M1G)
Rep3$M2G <- factor(Rep3$M2G)
Rep3$Exp <- factor(Rep3$Exp)
str(Rep3)

###ANOVA analysis for trial 3
Int3 <- aov(Phen ~ M1G + M2G + M1G:M2G , data=Rep3)
out<-capture.output(summary(Int3))
cat(out,file="C:\\R Work\\Full.txt",sep="\n", append=TRUE)

###Combine data matrices for full data analysis
AllExp <- rbind(Rep1, Rep2, Rep3)
str(AllExp)

####ANOVA analysis of all data from all trials of the experiment
```

```
Full <- aov(Phen ~ M1G + M2G + M1G:M2G + Exp , data=AllExp)
write.table(AllExp,"C:\\R Work\\ AllExp.csv",sep=",")
out<-capture.output(summary(Full))
cat(out,file="C:\\R Work\\Full.txt",sep="\n", append=TRUE)
```

```
################R model for Figure 2#############################
###Datamatrix creation for trial 1
WT=rnorm(3, mean=20, sd=5) ###WT untreated
M1=rnorm(3, mean=16, sd=5) ###Mutant untreated
M2=rnorm(3, mean=32, sd=5) ###WT treated
DM=rnorm(3, mean=22, sd=5) ###Mutant treated

M1G <- c("WT", "WT", "WT", "M1", "M1", "M1", "WT", "WT", "WT", "M1", "M1", "M1")
M2G <- c("WT", "WT", "WT", "WT", "WT", "WT", "M2", "M2", "M2", "M2", "M2", "M2")
Exp <- c("1", "1", "1", "1", "1", "1", "1", "1", "1", "1", "1", "1")
Phen <- c(WT, M1, M2, DM)
str(Phen)
Rep1 <- data.frame(Exp,M1G,M2G,Phen)
Rep1$M1G <- factor(Rep1$M1G)
Rep1$M2G <- factor(Rep1$M2G)
Rep1$Exp <- factor(Rep1$Exp)
str(Rep1)

###ANOVA analysis of trial 1
Int1 <- aov(Phen ~ M1G + M2G + M1G:M2G , data=Rep1)
out<-capture.output(summary(Int1))
```

```
cat(out,file="C:\\R Work\\Full.txt",sep="\n", append=TRUE)

###Datamatrix creation for trial 2
WT=rnorm(3, mean=20, sd=5)
M1=rnorm(3, mean=16, sd=5)
M2=rnorm(3, mean=32, sd=5)
DM=rnorm(3, mean=22, sd=5)
Exp <- c("2", "2", "2", "2", "2", "2", "2", "2", "2", "2", "2", "2")
Phen <- c(WT, M1, M2, DM)
str(Phen)
Rep2 <- data.frame(Exp,M1G,M2G,Phen)
Rep2$M1G <- factor(Rep2$M1G)
Rep2$M2G <- factor(Rep2$M2G)
Rep2$Exp <- factor(Rep2$Exp)
str(Rep2)

###ANOVA analysis of trial 2
Int2 <- aov(Phen ~ M1G + M2G + M1G:M2G , data=Rep2)
out<-capture.output(summary(Int2))
cat(out,file="C:\\R Work\\Full.txt",sep="\n", append=TRUE)

###Datamatrix creation for trial 3
WT=rnorm(3, mean=20, sd=5)
M1=rnorm(3, mean=16, sd=5)
M2=rnorm(3, mean=32, sd=5)
DM=rnorm(3, mean=22, sd=5)
Exp <- c("3", "3", "3", "3", "3", "3", "3", "3", "3", "3", "3", "3")
Phen <- c(WT, M1, M2, DM)
str(Phen)
```

```
Rep3 <- data.frame(Exp,M1G,M2G,Phen)
Rep3$M1G <- factor(Rep3$M1G)
Rep3$M2G <- factor(Rep3$M2G)
Rep3$Exp <- factor(Rep3$Exp)
str(Rep3)

###ANOVA analysis of trial 3
Int3 <- aov(Phen ~ M1G + M2G + M1G:M2G , data=Rep3)
out<-capture.output(summary(Int3))
cat(out,file="C:\\R Work\\Full.txt",sep="\n", append=TRUE)

###Datamatrix creation for combining all trials
AllExp <- rbind(Rep1, Rep2, Rep3)
str(AllExp)

###ANOVA analysis of all trials combined
Full <- aov(Phen ~ M1G + M2G + M1G:M2G + Exp , data=AllExp)
write.table(AllExp,"C:\\R Work\\ AllExp.csv",sep=",")
out<-capture.output(summary(Full))
cat(out,file="C:\\R Work\\Full.txt",sep="\n", append=TRUE)
```

Reassess the *t* Test: Interact with All Your Data via ANOVA

Siobhan M. Brady, Meike Burow, Wolfgang Busch, Örjan Carlborg, Katherine J. Denby, Jane Glazebrook, Eric S. Hamilton, Stacey L. Harmer, Elizabeth S. Haswell, Julin N. Maloof, Nathan M. Springer and Daniel J. Kliebenstein

Plant Cell; originally published online July 28, 2015;
DOI 10.1105/tpc.15.00238