BLUP, a new paradigm in host-range determination

D.K. Berner *

Foreign Disease-Weed Science Research Unit, Agricultural Research Service, U.S. Department of Agriculture, Ft. Detrick, Frederick, MD 21702, USA

A R T I C L E   I N F O
Article history:
Received 7 October 2009
Accepted 21 January 2010
Available online 28 January 2010

Keywords:
BLUP
Biological control agents
Classical biological control
Disease prediction
Henderson's mixed model equations
Host-range testing
Insects, invasive weeds
MME
Non-target damage prediction
Plant pathogens

A B S T R A C T
The centrifugal phylogenetic method has been the basis of selection of non-target plants to test for specificity of biological control agents for the last 35 years. In the last decade there has been increased attention paid to modernizing the approach and basing selection of test plants on molecular phylogeny rather than taxonomic classification. Recently mixed model equations (MME) and best linear unbiased predictors (BLUPs) were used to determine the probable host-range of two plant pathogens proposed for classical biological control of Russian thistle. BLUPs were derived with the MME by incorporating disease ratings, variances, and relationship matrices computed from genetic (DNA) distances among plant species related to Russian thistle. Although this work focused on evaluating disease severity on related plant species, the MME can be used with any biological weed control agent or target as long as the evaluation criterion is quantitative (or appropriately transformed) and variances and molecular genetic relationships among test species can be obtained. Moreover, since BLUPs can be generated for species with no observed data, the MME are ideally suited to both evaluation of test plant species and construction of test plant lists based on molecular phylogenetic relationships and reactions of the species to the biological control agent. Many biological control practitioners may be unfamiliar with this methodology. The objectives of this manuscript are to familiarize biological control researchers and regulators with some of the requirements and advantages of the MME and the use of the MME to construct test plant lists.

1. Introduction

Determining host specificity of candidate classical biological control agents is necessary before introduction of the agent into a new country or ecosystem. For weed biological control agents scientists are challenged with how to best evaluate an agent's effect on non-target plant species to avoid discarding potentially beneficial organisms, for which considerable research has been expended, while ensuring that the agent has a sufficiently narrow host-range to be safe to release. For the last 35 years the selection of non-target plants to test for specificity of biological control agents has been based on Wapshere's centrifugal phylogenetic method (Wapshere, 1974). This method focuses on testing plant species taxonomically related to the target as well as unrelated taxa with morphological or biochemical similarities to the target weed. Also recommended are tests of a few species from other families within the same order as the target and representatives of species in related orders. This has been the basis for constructing test plant lists in the USA and is the standard for proposals to regulatory bodies in the USA. To date, this has been a cumbersome process, often involving tests of a large number of non-target species under conditions that make it difficult to predict results in the field. In recent years, modernization of Wapshere's approach has been proposed by several biological control researchers (Briese, 2003, 2005; Briese and Walker, 2008; Sheppard et al., 2003, 2005). The recommendation from a biological control workshop organized by the Co-operative Research Centre for Australian Weed Management was that molecular phylogeny, and not taxonomic classification, should now be used to compile host test lists (Sheppard et al., 2003, Briese (2005) and Briese and Walker (2008) re-state this recommendation and add that inclusion of distantly related “safe-guard” species in test plant lists should cease. Briese (2005) argues that the focus of test plant lists should be on describing the actual host-range of the biological control agent rather than determining what individual plants are safe.

An approach that can incorporate molecular phylogeny along with species' reactions to biological control agents, to both evaluate the host range of the agent and construct test plant lists, are the mixed model equations (MME) that generate Best Linear Unbiased Predictors (BLUPs) (Henderson, 1975, 1977) of plant species' reactions. The MME incorporate the genetic relationships among species, data on an agent's effect(s), and variances in effect(s) among species, to get broad-based predictions of these effects for each species. The value of this methodology in host-range determination of candidate biological control organisms is the ability to predict the effects of the organism on plant species relative to that of the target species and enable predictions of the effects of the
organism on species that cannot be tested, because they are either very rare or extremely difficult to grow. Thus the complete host-range, among both tested and not tested species, of a biological control agent can be predicted. This information can then be used to determine which species should be placed on test plant lists.

Recently, the MME were used effectively with DNA sequences of the internal transcribed spacer 1, 5.8S ribosomal RNA, and internal transcribed spacer 2 regions (hereafter referred to as ITS sequences), as measures of genetic relationships among species, to clarify the host ranges of the obligate parasitic rust fungus *Uromyces* *salsola* and the facultative parasitic fungus *Colletotrichum gloeosporioides* f. sp. *salsola* (CGS) for biological control of Russian thistle (*Salsola tragus* L.) (Berner et al. 2009a,b). The MME could be used in the same manner to determine host ranges of other biological control agents of weeds, and other targets, and construct test plant lists, but application of the MME to host-range testing is new and many biological control practitioners may be unfamiliar with the methodology and the output of the analysis. The objectives of this manuscript are to familiarize biological control researchers and regulators with some of the data requirements of the MME, some of the advantages of the MME in terms of output of the analyses, flexibility of the MME in using different data types, and the use of the MME to construct test plant lists.

2. Requirements of the MME

2.1. Data choices in host-range determination

Because Wapshere’s centrifugal phylogenetic method (Wapshere, 1974) was developed to construct and not evaluate test plant lists, Wapshere does not discuss methods to determine probability of susceptibility or specificity of a biological control agent. However, the MME are used to evaluate the reaction of non-target species to the control agent and then use this information to construct and re-evaluate test plant lists. Therefore, choice of type of evaluation data to collect is important. In general, quantitative data are most amenable to MME analysis, although binary and binomial data, discussed later, can be used. Frequently, however, test plant evaluations of biological control agents involve evaluation of susceptibility based on ordinal or nominal data that segregate into discrete categories, e.g., 0, 1, 2, 3, 4, etc. or “very susceptible”, “susceptible”, etc. The problem arises over subsequent analysis of these types of data, since ordinary least squares analysis (ANOVA) of ordinal or nominal data is inappropriate for several reasons (Schabenberger and Pierce, 2002a).

Appropriate non-parametric approaches can be used with ordinal or nominal data to generate estimates of susceptibility, but these approaches are cumbersome and the results often difficult to interpret, particularly with many species and, for example, a five-class (0–4) rating scale. A logistic regression approach would require deciding, a priori, which of the five classes was the cut-off class for susceptibility, e.g., >1 or >2, etc., and the resulting logit values or probabilities for this cut-off for each species would be difficult to interpret in terms of susceptibility, e.g., a probability of 0.6 for a rating greater than 2 would be difficult to interpret. In addition, non-parametric approaches lack the ability to incorporate genetic relatedness into estimates, which is the strength of the MME and BLUPs. An alternative is to rank ordinal data and then use mixed models ANOVA on the ranks (Shah and Madden, 2004) to generate least squares means of the ranks or incorporate genetic distance data with the rank data to generate BLUPs with the MME. Tables 1 and 2 in the discussion of the advantages of the MME are based on ranks of disease severity ratings.

2.2. Variances among species

To obtain variances among species for the trait of interest, e.g., disease severity rankings or other effects of a biological control agent, data for each plant within each species and replication are collected and the mean of each repetition for each species used in a mixed model ANOVA with the mixed procedure of Statistical Analysis System (SAS) software (SAS Institute Inc. 2004) and restricted maximum likelihood (REML) estimation with only the intercept as a fixed effect and species as a random effect. Species is a random effect because the individual plants tested were selected randomly from the entire population (Schabenberger and Pierce, 2002b). The results of this analysis provides the variance estimate for species which is subsequently used in the relationship (G) matrix.

2.3. Relationship matrix generation

A method for generating ITS sequences is described in Berner et al., 2009a. An alternative is to obtain ITS, or other DNA sequences, from a public database like GenBank at the National Center for Biotechnology Information (NCBI) in the U.S. The sequences, in either case, should be trimmed to include, if ITS sequences, only the ITS1, 5.8S ribosomal DNA, and ITS2 regions prior to alignment. The sequences, in PHYLIP format (Felsenstein, 1989), should then be put into the same order, usually alphabetical by species, as the data output from the variance estimation step. This will ensure that the observed data and sequence data match after construction of the relationship matrix. The ITS sequences are then aligned with the ClustalW2 tool (Larkin et al., 2007) available at the EMBL-European Bioinformatics Institute website (http://www.ebi.ac.uk/Tools/clustalw2/index.html). The output alignment file, in alphabetical (input) order and PHYLIP format, is then analyzed by TREE-PUZZLE (Schmidt et al., 2002) available for download at http://www.tree-puzzle.de/ to generate a matrix of pairwise maximum likelihood distances among species. An outlier species, distantly related to the bulk of the species tested, should be included in the TREE-PUZZLE program. The alignment and distance matrix generation steps are relatively trivial and require only 5–30 min to complete, depending on number of species and computer hardware. A further option from the TREE-PUZZLE program is to generate a phylogenetic tree among species; this is a computationally intense step and, depending on the number of species, may require more than 24 h CPU time.

The relationship (G) matrix among species is derived from the distance matrix. The matrix is transformed into a proximity (relationship) matrix by subtracting the values of each element from ‘1’ to put the relationship between a species and itself as ‘1’ and relationships with other species as decimal fractions of ‘1’. Each element in this proximity matrix is then multiplied by the variance for species obtained in the first REML step.

2.4. The MME

After sorting the observed data to arrive at an order to coincide with the order of the relationship matrix, the relationship matrix of variances and covariances is read into the Mixed procedure of SAS using the GDATA = option. The SAS model statements used are described in Berner et al., 2009a. In this model, only the intercept is considered a fixed effect while species are considered random effects with the specified relationship (G) matrix. The error (R) matrix and associated error variance is estimated by iteration (REML) in the final model.
Advantages of the MME over least squares methods

The mixed model methods of generating BLUPs and least squares means (lsmeans) for disease reaction to *U. salsolae* and CGS are detailed in Berner et al., 2009a, b. An abbreviated list of lsmeans from these publications is presented in Table 1 and the corresponding BLUPs in Table 2. Not all of the species evaluated are presented in Tables 1 and 2, but all of the species with BLUP values significantly different than zero for disease severity to CGS were included in both tables along with, for reference, an additional four species with the next highest, but non-significant, BLUPs for CGS. Ranks of disease ratings were analyzed by a weighted mixed model to generate lsmeans (Table 1), and then a genetic relationship matrix, based on ITS sequences, among species was incorporated and the data re-analyzed with the MME to generate BLUPs (Table 2). From Table 1, it is evident that quite a few species were not tested (NT), and this was because there was inadequate plant material for replicated tests; in some cases no plant material was available. In total, lsmeans were generated for 59 species inoculated with CGS and 46 species inoculated with *U. salsolae*. In Table 2 there are no not tested (NT) species, although four species were not analyzed because no ITS sequence could be obtained at the time of analysis. BLUPs were generated by the MME for 91 accessions (89 distinct species) evaluated for CGS and 64 species evaluated for *U. salsolae*.

BLUPs for species that were not tested are a linear function of BLUPs of species that were tested (Henderson, 1975, 1977). That is: the BLUP of the untested species is the sum of (each BLUP of tested species, the BLUP of each untested species). As there were 59 species with observed data for CGS, and thus 59 BLUPs of tested species, the BLUP of each untested species represent the sum of (each of the 59 BLUPs times the relationship, between 0 and 1, of each tested species with the untested species). Thus, each untested species represent the sum of (each of the 59 BLUPs times the relationship, between 0 and 1, of each tested species with the untested species).

### Table 1

<table>
<thead>
<tr>
<th>Genus species</th>
<th>Colletotrichum gloeosporioides f. sp. salsolae</th>
<th>Uromyces salsolae</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Least squares mean estimates</td>
<td>Standard error of estimate</td>
</tr>
<tr>
<td><em>Salsola kali</em>-U.K.</td>
<td>285.40</td>
<td>41.47</td>
</tr>
<tr>
<td><em>Salsola tragus</em></td>
<td>277.82</td>
<td>37.46</td>
</tr>
<tr>
<td><em>Salsola collina</em></td>
<td>285.98</td>
<td>35.00</td>
</tr>
<tr>
<td><em>Salsola paulsenii</em></td>
<td>296.40</td>
<td>34.19</td>
</tr>
<tr>
<td><em>Salsola kali</em>-Akhani</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td><em>Salicornia bigelovii</em></td>
<td>247.06</td>
<td>17.30</td>
</tr>
<tr>
<td><em>Salicornia australis</em></td>
<td>111.24</td>
<td>12.72</td>
</tr>
<tr>
<td><em>Salicornia kali-Maui</em></td>
<td>111.08</td>
<td>18.00</td>
</tr>
<tr>
<td><em>Salicornia europaea</em></td>
<td>144.18</td>
<td>16.09</td>
</tr>
<tr>
<td><em>Sarcocornia italensis</em></td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td><em>Sarcocornia fruticosa</em></td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td><em>Bassia hyssopifolia</em></td>
<td>239.08</td>
<td>29.50</td>
</tr>
<tr>
<td><em>Bassia scoparia</em></td>
<td>110.60</td>
<td>10.35</td>
</tr>
<tr>
<td><em>Nitraria occidentalis</em></td>
<td>254.37</td>
<td>81.79</td>
</tr>
<tr>
<td><em>Halocnemum subaphyllus</em></td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td><em>Arthrocnemum glaucum</em></td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td><em>Bassia americana</em></td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td><em>Bassia prostrata</em></td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td><em>Suaeda calcifloris</em></td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td><em>Kalidium foliatum</em></td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td><em>Spinacia oleracea</em></td>
<td>180.01</td>
<td>8.93</td>
</tr>
<tr>
<td><em>Suaeda glauca</em></td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td><em>Polykennium majus</em></td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td><em>Suaeda occidentalis</em></td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td><em>Suaeda moquinii</em></td>
<td>110.93</td>
<td>17.97</td>
</tr>
<tr>
<td><em>Suaeda taxifolia</em></td>
<td>119.07</td>
<td>17.61</td>
</tr>
<tr>
<td><em>Suaeda maritima</em></td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td><em>Halocnemum strobilaceum</em></td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td><em>Suaeda vera</em></td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td><em>Halogeton glomeratus</em></td>
<td>110.39</td>
<td>17.93</td>
</tr>
<tr>
<td><em>Sesuvium maritimum</em></td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td><em>Salsola soda</em></td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td><em>Salsola orientalis</em></td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td><em>Allionvolea occidentalis</em></td>
<td>112.94</td>
<td>3.80</td>
</tr>
</tbody>
</table>

### Notes

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Species are arranged in order of descending BLUP values (Table 2) for <em>C. g. salsolae</em> disease severity.</td>
</tr>
<tr>
<td>b</td>
<td>Standard error of estimate based on fixed species effect plus intercept.</td>
</tr>
<tr>
<td>c</td>
<td>Pr &gt;</td>
</tr>
<tr>
<td>d</td>
<td>Estimated β = probability of committing a Type II error and not rejecting a false null hypothesis, H0: μ = 0 and H1: μ &gt; 0, i.e., declaring no significant difference from zero when a difference exists. Power is the probability of correctly rejecting a false null hypothesis.</td>
</tr>
<tr>
<td>e</td>
<td>Not significantly different than zero at P &lt; 0.05.</td>
</tr>
<tr>
<td>f</td>
<td>Not tested.</td>
</tr>
</tbody>
</table>

### Footnotes

a Species are arranged in order of descending BLUP values (Table 2) for *C. g. salsolae* disease severity.

b Standard error of estimate based on fixed species effect plus intercept.

c Pr > |t| based on solution vector of fixed species effect without intercept.

d Estimated β = probability of committing a Type II error and not rejecting a false null hypothesis, H0: μ = 0 and H1: μ > 0, i.e., declaring no significant difference from zero when a difference exists. Power is the probability of correctly rejecting a false null hypothesis.

e Not significantly different than zero at P < 0.05.

f Not tested.
for CGS in Table 1. For *U. salsolae*, 30 species had BLUPs compared to 17 with *lsmean* estimates.

3.2. The MME generate BLUPs for reactions of species, genera, tribes, etc.

In host-range determinations there is an assumption that the plant material being tested is representative of the species as a whole. The validity of this assumption is questionable, since the tested material is only an infinitesimally small sample of the species; cannot be proved to be representative; and the assumption does not take into account innate variability within any species. Of course many different sub-samples of species can be tested, but this still does not guarantee complete representation and might additionally complicate evaluation, e.g., if one sub-sample shows slight effects while other sub-samples do not. The large number of inter-specific relationships used in the generation of BLUPs place the disease reaction of each species in context, genetically, with all species analyzed. In the case of CGS, replicated data were available for 59 species, but the BLUPs of these species reflect not just the observed data but also 59 × 59 (59 per species) fractional replications based on the genetic inter-relationships among these species. Thus, BLUPs reflect the disease reactions of each species based on the disease reactions of the species themselves plus the disease reactions of all of the other inter-related species. This makes the BLUPs broad-based and reflective of what the disease reaction of each species, not just the plant material tested, would be expected to be. BLUPs for the species can then be grouped into other taxonomic associations, e.g., genera, tribes, etc., with the expectation that these broader groupings would be even more representative of what would be expected of each taxon. Differences in reaction to the control agent among species as well as these broader groupings would be expected to be. BLUPs for the species can then be grouped into other taxonomic associations, e.g., genera, tribes, etc., with the expectation that these broader groupings would be even more representative of what would be expected of each taxon. Differences in reaction to the control agent among species as well as these broader groupings would be expected to be. BLUPs for the species can then be grouped into other taxonomic associations, e.g., genera, tribes, etc., with the expectation that these broader groupings would be even more representative of what would be expected of each taxon. Differences in reaction to the control agent among species as well as these broader groupings would be expected to be. BLUPs for the species can then be grouped into other taxonomic associations, e.g., genera, tribes, etc., with the expectation that these broader groupings would be even more representative of what would be expected of each taxon.

3.3. The MME generate BLUPs that are conservative and robust

Because the MME can generate BLUPs for species that are not tested, some researchers and regulators might expect that BLUPs are less stringent than other susceptibility estimates. The opposite, in fact, is the case. In Tables 1 and 2 the criterion for susceptibility, based on individual *lsmeans* and BLUPs, was *P > |t|* ≤ 0.05 under the null hypothesis that the *lsmean* or BLUP was equal to zero. When 7 species had significant (*P > |t|* ≤ 0.05) *lsmeans* and were estimated

---

Table 2

<table>
<thead>
<tr>
<th>Genus species</th>
<th>Colletotrichum gloeosporioides f. sp. salsolae</th>
<th>Uromyces salsolae</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BLUP Standard error</td>
<td>Pr &gt;</td>
</tr>
<tr>
<td>Salicornia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salsola</td>
<td></td>
<td></td>
</tr>
<tr>
<td>salsolae</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salsola</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kali-Maui</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salsola</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kali</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salsola</td>
<td></td>
<td></td>
</tr>
<tr>
<td>australis</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>Arranged in order of descending BLUP values for <em>C. g. salsolae</em> disease severity.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best Linear Unbiased Predictor includes fixed intercept estimate.</td>
<td></td>
</tr>
<tr>
<td>Standard error based on BLUP of random species effect plus intercept.</td>
<td></td>
</tr>
<tr>
<td>Estimated β = probability of committing a Type II error and not rejecting a false null hypothesis.</td>
<td></td>
</tr>
<tr>
<td>H₀: μ = 0 and Hₐ: lμ &gt; 0, i.e., declaring no significant difference from zero when a difference exists. Power is the probability of correctly rejecting a false null hypothesis.</td>
<td></td>
</tr>
<tr>
<td>Not directly tested by inoculation.</td>
<td></td>
</tr>
<tr>
<td>Not significantly different than zero at <em>P</em> ≤ 0.05.</td>
<td></td>
</tr>
</tbody>
</table>

---
susceptible to CGS, 3 species were estimated susceptible to *U. salsolae* based on lsmeans (Table 1). Of the same species, 30 were predicted susceptible to CGS and 7 susceptible to *U. salsolae* based on BLUPs (Table 2). Eight species with non-significant lsmeans estimates had significant BLUPs for CGS. These were: *Salsola australis*, *Salsola kali* from Maui, *Salicornia europea*, *Bassia scoparia*, *Nitraba occidentalis*, *Suaeda moquinii*, *Suaeda taxifolia*, and *Halogeton glomeratus*. Of these, *S. europea*, *N. occidentalis*, *S. moquinii* and *S. taxifolia* are native N. American species (Kartesz and Meacham, 1999). *S. europaea* is a synonym of *S. maritima* (USDA, NRCS, 2008). For species with significant BLUPs for *U. salsolae* had non-significant lsmeans estimates. These were: *Salsola collina*, *Salsola paulsenii*, *S. australis*, and *S. kali* from Maui. None of these are native N. American species.

BLUPs generated by the MME are also robust and depend not only on the relatedness of the plant material analyzed but also on the disease reactions of the different species. Comparison of BLUPs for CGS and *U. salsolae* (Table 2) showed that 29 species in the tribes Atripliceae, Camphorosmeae, Halopepideae, Polycne-maeae, Salicornieae, Salsolae, and Suaedae were susceptible to CGS while only 7 species of the genus *Salsola* were susceptible to *U. salsolae*. Because CGS is a necrotophic fungus, less specificity was expected, and found, than with the obligate rust fungus, but species in the genus *Salsola* were the most susceptible to CGS.

3.4. The MME generate BLUPs that have low variability and high power

For both lsmeans and BLUPs, for each pathogen, a pseudo coefficient of variation (CV) based on standard error rather than standard deviation) was calculated as (standard error ÷ least squares mean (or BLUP)) × 100. The range in CVs for least squares means of all species inoculated with *U. salsolae* was 2.62–142.37% while the range in CVs for BLUPs of the species analyzed with the MME was 4.99–36.74% (Berner et al., 2009a). The range in CVs for the least squares means of all accessions inoculated with CGS was 3.31–407.87% while the range in CVs for BLUPs of all the accessions analyzed with the MME was 3.53–43.32% (Berner et al., 2009b). For both pathogens, the large range in CVs for lsmeans indicated extremely high variability in disease reaction within and among species while the lower CVs for the BLUPs indicated much lower inter- and intra-specific variability.

Power is the probability of correctly rejecting a false null hypothesis and is calculated as 1 – β where β (power) is the probability of committing a Type II error and not rejecting a false null hypothesis (Park, 2008), i.e., declaring no significant difference from zero when a difference exists or declaring a species not susceptible when it probably is. Although not commonly determined in host-range evaluation, power seems as important a criterion as Pr > |t| in host-range determination, since an erroneous determination of non-significant or not susceptible could lead to post-release non-target damage.

Estimated power values were calculated for each BLUP and lsmean for each pathogen. The GLMPOWER procedure of SAS (SAS, 2004) was run with predicted values of BLUPs and associated standard deviations and estimated lsmeans and associated standard deviations output from the respective mixed models procedures. The BLUPs and lsmeans of each species were weighted with the number of repetitions for each species and a list of power values corresponding to the range of standard deviations for each pathogen, in increments of one standard deviation, were output from the GLMPOWER procedure for each dataset. The total number of species analyzed by the MME or weighted mixed model analyses, in the case of lsmeans, for each pathogen was an input variable in this procedure since power depends on total sample size. Since power values could be generated by the SAS procedure only for those species with observed data, the standard deviations for BLUPs and lsmeans of each species for each pathogen were manually calculated based on the respective standard error and number of repetitions. For BLUPs generated for species with no observed data, 2.3 repetitions, corresponding to the average number of repetitions for species with observations, were used in the calculations. This biased lower power values for these species. The power values corresponding to these standard deviations for each species were then look up from the list output by GLMPOWER.

All of the species, except *N. occidentalis*, determined to have non-significant lsmeans, had power values above 0.75 for CGS (Table 1). The power value for the lsmean disease reaction of *N. occidentalis* to CGS was <0.10 indicating a high probability (β) of declaring no significant difference from zero when a difference existed. For the species determined to have non-significant lsmeans for *U. salsolae*, all power values were below 0.50 and some were <0.05 indicating high probability that significant differences from zero for these lsmean estimates existed but were not detected. Power values for 4 species determined to have non-significant BLUPs for CGS (Table 2) were above 0.75, indicating a high degree of confidence in the NS determination. The BLUP for *N. occidentalis* was highly significant (P < 0.01) as opposed to the non-significant *U. salsolae*. With the exception of one of the 23 species analyzed for *U. salsolae* and determined to have non-significant BLUPs (Table 2), all of the power values were above 0.75. These results indicate that BLUPs have considerably greater power than lsmeans and that BLUPs are the safer of the two estimates of non-susceptibility.

4. Use of the MME with binary and binomial data

For both CGS and *U. salsolae*, the disease severity data for each plant species were transformed into binary and, ultimately, binomial disease incidences typical of effect versus no effect data, i.e., “1” or “0”, and analyzed with the MME. To transform the data, any diseased plant within a repetition was assigned a value of 1, and non-diseased plants were assigned a value of 0. The sum of these values within each repetition was then divided by the total number of plants tested in each repetition to form a proportion of diseased plants in each repetition. Proportions of 0 and 1 were set to 0.01 and 0.99, respectively. These proportions were then converted to logit values: ln(proportion/(1 – proportion)), combined with the genetic relationships and variances among species, and analyzed with the MME to generate BLUPs of logit values. Odds ratios were generated by raising the natural logarithm to the power of the BLUP for each species, i.e., eBLUP. In this case the odds ratio is a probability of whether disease occurs or not. Odds ratios greater than one indicate that disease occurrence is the more likely outcome and the larger the odds ratio the greater the likelihood of disease (Deeks, 1996).

BLUPs of logit transformed disease incidence data, standard errors of the BLUPs, P > |t| values, and odds ratios for each plant species are presented in Table 3. The criterion for susceptibility was P > |t| < 0.05 for individual BLUPs. Based on BLUPs of logit values, 24 species were predicted susceptible to CGS and 8 to *U. salsolae*. There were six fewer species predicted susceptible to CGS and one more species predicted susceptible to *U. salsolae* based on BLUPs of disease incidence (logit values) than on BLUPs of disease severity (Table 2). This indicated that, for CGS, the disease severity on fewer susceptible plants was relatively high and resulted in more species being predicted susceptible based on disease severity. The opposite appeared the case with *U. salsolae* and *Halogeton glomeratus* which had significant (P > |t| = 0.0531) disease incidence but non-significant disease severity (Tables 3 and 2, respectively). Based on odds ratios, 15 species had odds ratios greater than one for CGS, indicating that disease occurrence was the more
likely outcome for these species. This was considerably fewer species predicted susceptible based on BLUPs of either disease incidence or disease severity (Tables 3 and 2, respectively). For *U. salolae*, 7 species had odds ratios greater than one. This was the same number of species predicted susceptible based on BLUPs of disease severity (Table 2) and one less than the number predicted by BLUPs of disease incidence (Table 3). Since different biological control agents can be expected to produce slightly different magnitudes of relative effects on different non-target species, as opposed to only the occurrence of an effect (incidence), it would be relatively simple to evaluate the agent on both incidence and severity of any effect and use the more conservative prediction, i.e., the one that produces more susceptible species.

5. Use of the MME with multiple fixed effects and multiple variables

In the U.S. releases of plant pathogens for biological control of weeds is restricted, by regulation, to single isolates of the pathogen, and host-range determination of exotic weed pathogens is restricted to a quarantine bio-safety level 3 facility. Thus, host-range determination of exotic weed pathogens at the Foreign Disease-Weed Science Research Unit (FDWSRU) of USDA, ARS is based on one isolate of a pathogen in one environment. The MME used to generate BLUPs of plant species in Tables 1 and 3 reflect this and have only one fixed effect (Schabenberger and Pierce, 2002b), the intercept. However, in other biological control systems it might be possible and desirable to test multiple isolates, strains, or biotypes in multiple environments. The MME can readily accommodate these multiple fixed effects with either homogenous or heterogeneous variances (Henderson, 1975, 1977; Henderson and Quaas, 1976). To demonstrate this, above-ground oven-dried biomass data for plant species inoculated and not inoculated with the fungal pathogen *Phoma exigua* were analyzed with the MME. The results of this partial host-range test to evaluate the safety and effectiveness of the fungus to control Russian knapweed (*Acroptilon repens* (L.) DC. = *Rhaponticum repens* (L.) Hidalgo = *Centaurea repens*) are presented in Table 4. The relationship matrix was based on ITS sequences of the different plant species, as previously described, but the variances for the inoculated and not-inoculated treatment groups were unequal. This required that the G matrix of genetic relationships and variances among species be structured in compound symmetry to reflect the two different variances and the structure of the data. The resulting G matrix was:

\[
G = \begin{bmatrix}
G_1 & 0 \\
0 & G_2
\end{bmatrix}
\]
The difference between the control and inoculated treatment for the 18 species without fixed effects was highly significant. Results of the analysis are shown in Table 5. None of the BLUPs for the species were significantly greater than zero, and only the fixed effect intercept was significantly greater than zero. When combined with the selected fixed effects, both the inoculated and not-inoculated treatments on Acroptilon repens were highly significant. The difference between the control and inoculated treatment for Acroptilon repens was not significant. Although not a meaningful estimate in terms of inoculated and control treatments, the last estimate in Table 4 is meaningful if the two treatments were actually different isolates, strains, or environments and the objective was to predict the average effect. If the treatments were different environments, then Table 4 estimates could be read as “A. repens environment 1”, “A. repens environment 2”, “A. repens environment 1 minus 2”, “A. repens average environment effect”. There is no limit on the number of fixed effects, including covariates like temperature, dew period, etc., that can be included in the MME, but the G matrix structure could become complicated with many effects and their interactions.

It might also be desirable to test multiple variables measured on each experimental or sampling unit and generate a single BLUP from these variables for each entity of interest. Multiple variables with heterogeneous variances can be incorporated into the MME and weighted as scientifically or economically appropriate (Henderson and Quaas, 1976). To demonstrate this, the above-ground oven-dried biomass data on inoculated and not-inoculated plants, from the previous example, were reorganized as two fictitious variables measured on each plant in each repetition for each plant species. The means for each replication were used in the analysis. In order to use the Mixed procedure of SAS and the GDATA = option to do the analysis, the data needed to be appropriately coded to simultaneously analyze both variables. The SAS code for this is available in Wright (1998). The G matrix from the previous example, with separate variances for each variable, was used in the analysis. Linear estimable functions were written to generate estimates for selected plant species and variables, with variable 1 weighted twice that of variable 2. Results of the analysis are shown in Table 5. None of the BLUPs for the species were significantly greater than zero, but estimates for A. repens and C. tinctorius, that included the fixed effect intercept, were significantly greater than zero. There was no significant difference between these two species for the combined BLUPs of both variables. Because BLUPs are relative measures of realized values of random variables (Harville, 1976, 1977), BLUPs of disparate variables like feeding damage, eggs per leaf, etc. can be combined into a single weighted, or un-weighted, predictor for each entity of interest. For host-range determination tests of pathogens at FDWSRU, disease data are being combined with above-ground biomass data on each test plant to generate single BLUPs from both variables. Of course, multiple fixed effects, e.g., environments, strains, covariates, etc., can be combined with multiple variables in the MME to generate BLUPs of interest; the primary obstacle is the complexity of the G matrix with many fixed effects and multiple variables.

### Table 4

Solution vector of BLUPs, standard errors, Pr > |t| values in comparison with zero, for above-ground oven-dried biomass for plant species inoculated with *Phoma exigua* and not inoculated, with unique, unequal variances for each inoculation treatment.

| Genus speciesa | BLUPb | Standard errorc | Pr > |t|d |
|----------------|-------|----------------|------|
| Acroptilon repens | -0.1443 | 0.4833 | 0.7664 |
| Carduus tenuiflorus | -0.1036 | 0.4845 | 0.8316 |
| Carduus nutans | -0.0214 | 0.4852 | 0.9649 |
| Carthamus tinctorius | 0.0480 | 0.4843 | 0.9213 |
| Centaurea americana | -0.1505 | 0.4803 | 0.7553 |
| Centaurea calcitrana | 0.1473 | 0.4821 | 0.7612 |
| Centaurea cyanus | 0.1330 | 0.4818 | 0.7837 |
| Centaurea diffusa | -0.1464 | 0.4835 | 0.7633 |
| Centaurea jacea | -0.0458 | 0.4843 | 0.9249 |
| Centaurea melitensis | -0.1582 | 0.4822 | 0.8367 |
| Centaurea napifolia | -0.1621 | 0.4830 | 0.9547 |
| Centaurea rothrockii | -0.0998 | 0.4824 | 0.8288 |
| Centaurea solstitialis | -0.0330 | 0.4834 | 0.9060 |
| Centaurea sulfurea | -0.1049 | 0.4820 | 0.9337 |
| Cirsium fontinale | -0.0573 | 0.4819 | 0.9385 |
| Cirsium occidentale | 0.0403 | 0.4827 | 0.9031 |
| Cirsium vulgaris | 0.0373 | 0.4822 | 0.8367 |
| Cynara scolymus | 0.0590 | 0.4830 | 0.9457 |

Solution vector for fixed effects

| Intercept | 1.7241 | 0.4939 | 0.0026 |
| Control (not inoculated) | 0.1049 | 0.2426 | 0.6673 |

Selected estimates with intercept and inoculation treatment estimates

| Acroptilon repens control | 1.6847 | 0.2184 | <.0001 |
| Acroptilon repens inoculated | 1.5798 | 0.2184 | <.0001 |
| Acroptilon repens control minus inoculated | 0.1049 | 0.2426 | 0.6673 |
| Acroptilon repens treatment average | 1.6322 | 0.1816 | <.0001 |

---

a Species are arranged in alphabetical order.
b Best linear unbiased predictor without fixed effects estimates.
c Standard error based on BLUP of random species effect without fixed effects.
d Pr > |t| based on solution vector of random species effect without fixed effects.

where $G_1 = a 18 \times 18$ matrix of genetic relationships among the 18 plant species multiplied by the variance for the not-inoculated group; $0 = a 18 \times 18$ matrix of zeros; and $G_2 = a 18 \times 18$ matrix of genetic relationships among the 18 plant species multiplied by the variance for the inoculated group. The Mixed procedure of SAS and the GDATA = option was used to analyze the data and generate BLUPs for each plant species. Linear estimable functions were written to generate estimates for selected plant species in each inoculation group. BLUPs for the eighteen species, without fixed effects estimates, and the fixed effect solution vector are shown in Table 4. None of the BLUPs for the species were significantly greater than zero, and only the fixed effect intercept was significantly greater than zero. When combined with the selected fixed effects, both the inoculated and not-inoculated treatments on Acroptilon repens were highly significant. The difference between the control and inoculated treatment for Acroptilon repens was not significant. Although not a meaningful estimate in terms of inoculated and control treatments, the last estimate in Table 4 is meaningful if the two treatments were actually different isolates, strains, or environments and the objective was to predict the average effects. If the treatments were different environments, then Table 4 estimates could be read as “A. repens environment 1”, “A. repens environment 2”, “A. repens environment 1 minus 2”, “A. repens average environment effect”. There is no limit on the number of fixed effects, including covariates like temperature, dew period, etc., that can be included in the MME, but the $G$ matrix structure could become complicated with many effects and their interactions.

6. Use of the MME to construct test plant lists

The following steps are suggested to use the MME and BLUPs to effectively construct a test plant list: (1) A subset of 20–30 plant
species closely related to the target weed should first be evaluated, along with the target, in replicated tests, for any effect of an agent. (2) The results of this evaluation should be statistically analyzed to obtain a variance estimate among species (see Section 2.2). (3) DNA sequences of a highly conserved DNA region(s), like ITS sequences, should be obtained for these species and another 20–30 closely related species of interest. (4) A relationship matrix among genets, the MME could be used with any biological control agent with weeds as the primary target with examples based on pathogens. The agent has been delimited or until the host range has been determined too broad to warrant further evaluation and analysis. (5) Data from the additional 20–30 species that had no observations should be entered into the original dataset as missing values. (6) The relationship matrix should then be incorporated with the dataset and analyzed with the MME to generate BLUPs for all of the species (Berner et al., 2009a). (7) Those species with large positive BLUPs, but without direct evaluation data, should then be obtained and evaluated in replicated tests. (8) The data from these tests should be entered into the original dataset to replace the missing values. (9) Steps 2–8 should be repeated with additional species closely related to species with large positive BLUPs, and not just more species closely related to the target. These species could initially be represented only by DNA sequences, to generate BLUPs and indicate which related species would be expected to be susceptible. This process should then be repeated until the researcher(s) is satisfied that the expected host range, particularly among native and economically important plant species, of the agent has been delimited or until the host range has been determined too broad to warrant further evaluation and analysis.

7. Use of the MME with other control agents and other targets

Although the preceding sections have dealt, predominantly, with weeds as the primary target with examples based on pathogens, the MME could be used with any biological control agent on any target. The sole requirements are some measure of: the genetic relationships among target and non-target species, the reactions, in quantitative terms, of species to the biological control agent, and the variance(s) among species for the variable(s) of interest.
interest. With other agents and targets, it will be necessary to
determine what measure(s) of genetic relationships are most
appropriate. Berner et al. (2009a,b) constructed genetic relation-
ship matrices from ITS sequences of the target and non-target spe-
cies because these sequences were readily available in public
databases. As different types of DNA sequences become more readily
available in public databases, ITS sequences could readily be
combined with another type of sequence, e.g., chloroplast spacer
sequences, to improve the accuracy of the genetic relationship
matrices. Alternatively, to ensure that the sequences match with
the material tested, these sequences could be generated “in-house”
for each collection of the target and non-target that is tested.

In some cases, neither ITS sequences nor other sequences, which
are suitable for differentiating species relationships, are
suitable for constructing genetic relationship matrices. This is par-
icularly true when testing entities of a single species, e.g., crop cul-
tivars, that would be expected to have the same ITS sequence, or
other similar type of sequence. As in animal breeding (Henderson,
1975, 1977), coefficients of parentage derived from pedigreed infor-
cation can be used to construct genetic distance matrices, but
these may not exist or be particularly reliable for systems other
than animals (Van Becelaere et al., 2005). In these cases, other
DNA markers could be used to construct the necessary relationship
matrices (Bauer et al., 2006). These DNA markers include random
fragment length polymorphisms (RFLPs), amplified fragment
length polymorphisms (AFLPs), random amplified polymorphic
DNA (RAPDs), and simple sequence repeats (SSRs). Some compar-
sions of these markers for constructing genetic relationship matrices
are presented in Powell et al., (1996), Van Becelaere et al.
(2005), and Bauer et al., (2006). For fungi, Boss et al., (2007) used
inter simple sequence repeats (ISSRs) to fingerprint an isolate of
Stagonospora convolvuli (Convolvulus arvensis). Conceivably this approach
could be used in tandem with ITS sequences to construct genetic
relationship matrices among fungal species (ITS sequences) and
among fungal isolates within species (ISSRs) in biological control
of fungi/pathogens.

8. Conclusions

The MME can be used with ranked ordinal data, binary and
binomial data, and quantitative data to generate BLUPs based on
genetic relationships among species, the variance among species,
and the reactions of species to the biological control agent. BLUPs
generated by the MME are broadly applicable to species reaction,
are robust and more conservative than least squares means, have
lower variability and higher power, and are thus safer, than least
squares means, can be generated for more species than least
squares means, and can be generated for species without observed
data. The MME can also be used with multiple fixed effects, e.g.,
strains, environments, covariates, and multiple variables to gener-
ate single BLUPs for multiple variables under a variety of fixed ef-
facts. These qualities of the MME and BLUPs should be an
enticement to researchers and regulators to strongly consider use
of the MME in host-range evaluation. Moreover, since BLUPs can
be generated for species with no observed data, the MME are ide-
ally suited to the iterative evaluation of test plant species and con-
struction of test plant lists based on molecular phylogenetic
relationships and reactions of the species to the biological control
agent. Species determined susceptible with BLUPs, particularly na-
tive species, would need to be tested further to determine more
precisely the extent of any damage caused by the biological control
agent, but species with lower-value non-significant BLUPs and a
corresponding high power estimate could confidently be judged
non-susceptible. This applies to species tested by direct testing
and to species for which there were no observed data. Much of this
manuscript has dealt with weeds as the primary biological control
targets, but the MME can be used with any biological control agent
on any target. The primary consideration in application of the MME
to other targets is the type of DNA sequence to use to construct the
genetic relationship matrix.

Given the aforementioned qualities and strengths of the MME
in host-range determination, it would be instructive to use this ap-
proach to delimit host-range, ex post, of a released biological con-
rol agent that has some non-target effects in the area of release.
If the non-target species could be predicted, by the MME, to have
these effects, then this would be demonstrable evidence of the util-
ity and safety of the MME in host-range determination. To that end,
I invite any interested party, with access to historical host-range
determination data, to collaborate in predicting non-target effects
from a released biological control agent. Please contact me by
email if you are interested and have access to the data and permis-
sion to use it.

Acknowledgment

The author thanks Dr. William Bruckart, Research Plant Pathol-
ologist, USDA, ARS, Foreign Disease-Weed Science Research Unit, Ft.
Detrick, MD for reviewing the paper and making valuable sugges-
tions on organization and content.

References

Bauer, A.M., Reetz, T.C., Leon, J., 2006. Estimation of breeding values of inbred lines
using best linear unbiased prediction (BLUP) and genetic similarities. Crop Science 46,
2085–2091.
combining disease ratings and DNA sequences to determine host range of
Uromyces salsolae for biological control of Russian thistle. Biological Control 49,
68–76.
2009b. Best linear unbiased prediction of host range of the facultative parasite
Colletotrichum gloeosporioides f. sp. salsolae, a potential biological control
agent of Russian thistle. Biological Control 51, 158–168.
Boss, D., Maurhofer, M., Zala, D., Defago, G., Brunner, P.C., 2007. ISSR fingerprinting
for the assessment of the bioweed biocontrol agent Stagonospora convolvuli
LA39 after field release. Letters in Applied Microbiology 45, 244–251.
Briese, D.T., 2003. The centrifugal phylogenetic method used to select plants for
host-specificity testing of weed biological control agents: can and should it be
modernized. CRC Technical Series 7, 23–33.
Briese, D.T., 2005. Translating host-specificity test results into the real world: The
need to harmonize the yin and yang of current testing procedures. Biological
Control 35, 208–214.
Briese, D.T., Walker, A., 2008. Choosing the right plants to test: The host-specificity
of Langitarius sp. (Coleoptera: Chrysomelidae) a potential biological control
Cladistics 5, 164–166.
Harville, D.A., 1976. Extension of the Gauss-Markov theorem to include the
Harville, D.A., 1977. Maximum likelihood approaches to variance component
estimation and to related problems. Journal of the American Statistical
Association 72, 320–338.
Henderson, C.R., 1975. Use of all relatives in intraherd prediction of breeding values
Henderson, C.R., 1977. Best linear unbiased prediction of breeding values not in
North Carolina Botanical Garden, Chapel Hill, N.C.
Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam,
H., Valentin, F., Wallace, L.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J.,
2948.
Technical Working Paper. The University Information Technology Services
(UITS) Center for Statistical and Mathematical Computing, Indiana University.
Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingley, S., Rafaelski, A.,
1996. The comparison of RFLP, RAPD, AFLP, and SSR (microsatellite) markers for
germplasm analysis. Molecular Breeding 2, 225–238.

D.K. Berner / Biological Control 53 (2010) 143–152 151


