ATTACHMENT 1
Analysis of seedling phenotyping of crown rot symptoms in wheat and barley

conducted by Alison Kelly
Department of Employment, Economic Development and Innovation,
Toowoomba QLD, 4350
61 7 4639 8831

Contents

Chapter 1

Seedling Ring Test

1.1 Background

The seedling ring test was designed to compare the different screening methods for crown rot developed at sites across Australia. The study seeks to quantify the variability in results between methods using a common set of thirty entries, comprised of wheat and barley lines with varying levels of response to crown rot infection. The ultimate aim of the project is to produce a robust and reliable method for high throughput crown rot screening.

The four methods in the study each implement a differing trial structure, layout and design, together with method of infection and subsequent scoring technique. These differences are documented in the report, and results for the level of crown rot infection of thirty genotypes under each methodology are compared.

1.2 Background to statistical analysis

Classical block analyses with the underlying assumption of normality are often used to analyse disease assessment trials. However, the type of measurement commonly made to quantify disease presence and severity do not naturally follow a normal distribution. Incidence data is binomial, formed from a count of the number of plants or leaves infected, out of a total possible number of plants/leaves assessed. Severity data is generally a rating, measured according to an underlying

scale for the degree of infection. Each measurement of incidence or severity is best analysed with an appropriate underlying distribution for residual variation. Coercion to meet the assumptions of normality through appropriate transformation of data is achievable and often used effectively.

Trialling methods for disease assessment may be structured with an underlying hierarchy for assessment as follows. Genotypes are grown in pots or field plots, and as such pots/plots form the experimental unit for the trial. These units are replicated and randomised to give a valid estimate of experimental variation. A number of plants may be assessed for each experimental unit, and sometimes mutiple leaves per plant are also assessed. This forms a nested hierarchy of plot/plant/leaf strata which should be accounted for in the analysis. Averages across leaves and plants are often calculated prior to analysis, and this is representative when formed across an equal number of measurements. Often the amount of information measured for each treatment is not equal, as a differing number of plants/leaves are assessed. In this case, an efficient analysis weights for the number of individuals assessed in each strata.

Due to the difficulties associated with disease screening, the level of measurement error can be greater than desireable. Some influencing factors may be due to environmental effects and variation in application of the methodology, together with the difficulty of objectively assessing the level of infection. Efficient statistical design and analysis can achieve gains in reducing experimental error. No statistical analysis can account for lack of experimental design, or resurrect results from a poorly conducted/non-reproducible screening method.

The next section documents the methodology, design and appropriate statistical analysis for each assessment method.

1.3 Experimental method and Statistical analysis

1.3.1 CSIRO droplet test

The CSIRO trial followed no experimental design, but contained three replicates of 30 entries, with each replicate processed in standard order (unrandomised). A replicate was comprised of trays containing 30 punnets, and five to twelve punnets were grown for each entry. As there is no record of the position of lines within a tray or the position of trays inside a glasshouse, trend effects could not

be explored for these data.

Each replicate of the trial was assessed at a different time;

Replicate	Planted	Inoculated	Assessed
1	11/06/2009	22/06/2009	28/07/2009
2	8/07/2009	21/07/2009	24/08/2009
3	15/07/2009	28/07/2009	31/08/2009

The measurement scale was an index formed as,
CR index = [browning length/plant height]*number of leaf sheaths infected

A summary of the data is given in Figure 1.1, and shows the range of the data and degree of skewness in each of the three measurements taken, and in the derived index. A linear model with the following terms was fitted to the severity index.
% infected = genotype + replicate + replicate:punnet + replicate:punnet:plant

Preliminary analysis indicated variance heterogeneity, so all subsequent analyses were conducted on a *log* scale. Caution should be taken in interpretation of these results as the absence of experimental design produces potentially biassed genotype effects.

1.3.2 DEEDI pot assessment

The DEEDI pot trial was designed as three replicates of 36 entries, containing the 30 ring test entries and an additional 6 standards, randomly laid out as a 3 column by 36 row square array. Up to ten plants were assessed for each pot, with three leaves assessed per plant. The trial was potted on 28/08/2009; inoculated on 4/09/2009; and scored on 18/09/2009.

The measurement scale was the percentage of leaf sheath lesioned, rated in 5% intervals, totalled across the first three leaves on each plant.

A summary of the raw data is given in Figure 1.2(a).

A linear model with the following terms was fitted to the severity measure.

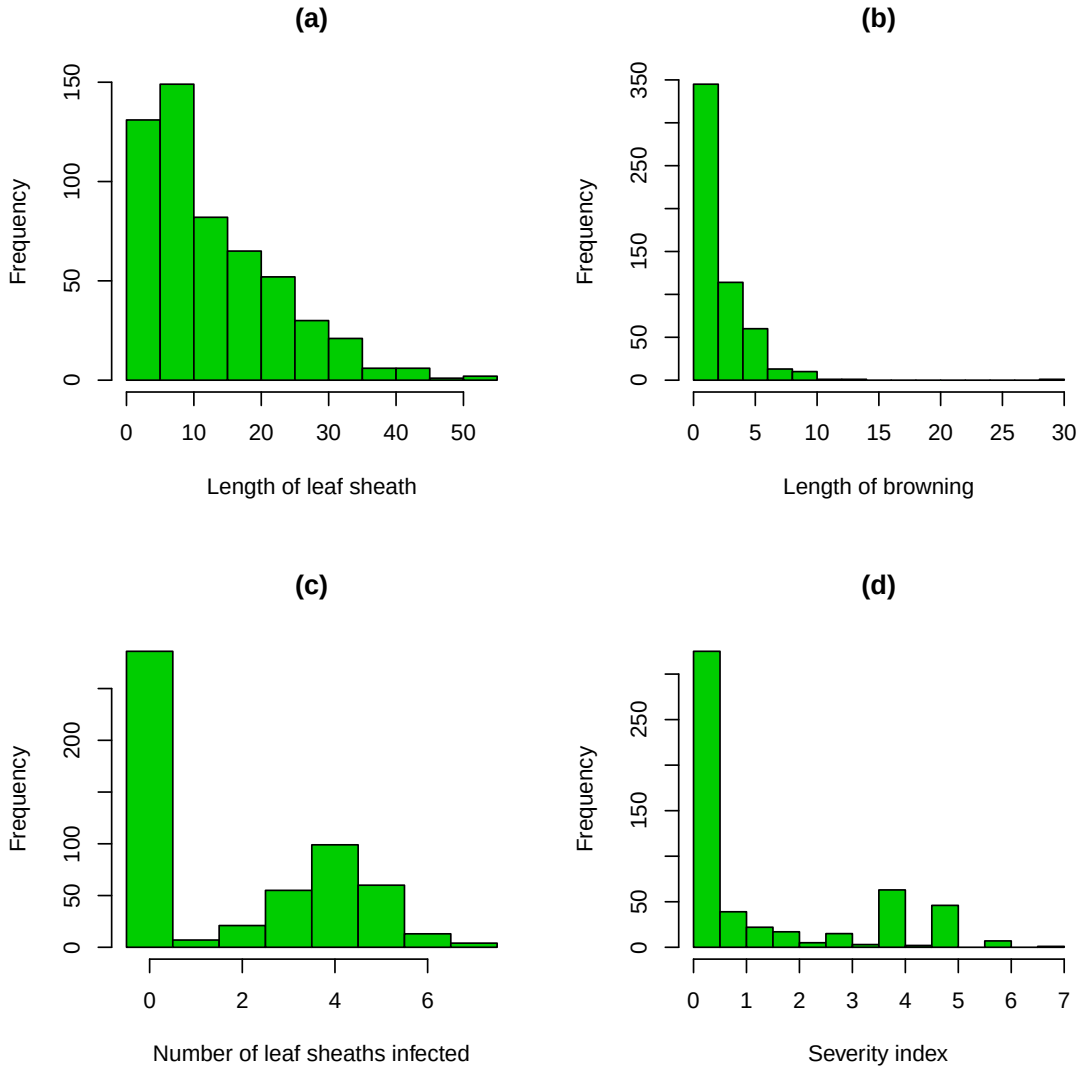

Figure 1.1: Histograms of raw data for CSIRO test; a) Length of leaf sheath, b) Length of browning, c) Number of leaf sheaths infected, d) Severity index

% browning = genotype + replicate + replicate:pot + replicate:pot:plant

1.3.3 SA Fusarium droplet test

The SA trial was designed as three replicates of 30 entries, randomly laid out as a 3 bay by 30 column square array. A replicate was comprised of two trays, each of which contained 15 pots. Up to six plants were assessed for each pot, with a varying number of 5-8 leaves assessed per plant. The trial was sown on 18/08/2009; inoculated on 01/09/2009; and scored on 13/09/2009.

The measurement scale was binomial, and recorded as the number of leaves infected out of the total number inspected for each plant. A second variable of sheath penetration was analysed. This measurement was formed as the number of infected leaves, averaged over plants.

The data provided for statistical analysis was % infected, averaged over leaves and plants. A linear model with the following terms was fitted to this average % infected data.
% infected = genotype + replicate + pot

A second analysis was undertaken for these data, as the measurement scale was actually binomial, where % of leaves infected were counted for a number of plants. A generalised linear mixed model with an underlying binomial distribution and logit link function was fitted to these data, including terms for *genotype, replicate* and *plant*. The second analysis is more appropriate for the data scale, and the results from this analysis are in the form of probabilities. The analysis methods are compared in the following section.

A summary of the data is given in Figure 1.2(c) for % leaves infected and in Figure 1.2(d) for sheath penetration..

1.3.4 UOS bench test

The UOS pot trial was designed as four replicates of 30 entries, randomly laid out as a 6 column by 20 row square array on one bench. Up to seven plants were grown in each pot.

Data was recorded for both incidence, number of plants infected out of total number of plants in pot; and severity, a rating scored for each pot on a 0-4 scale as follows:
0 = No lesion
1 = First internode partially lesioned
2 = First internode fully lesioned and second internode partially or fully lesioned
3 = More than two internodes lesioned
4 = Dead head (white head or no head) due to crown rot

The data for this assessment method was extremely skewed, and preliminary analysis indicated that average ratings over six plants displayed variance heterogeneity. More importantly, the underlying scale of measurement was ordinal, as infection was rated on a 0-4 scale. It is evident from Figure 1.2(c) that the majority of scores lay in category 1, with the first internode partially lesioned. In addition, 12 of the 30 lines had all plants rated as 1 adding no variability to the data, and hence were excluded from the statistical analysis.

Two analyses were completed for these data. The first assumed continuously measured data with an underlying normal distribution for the residuals, and involved a linear mixed model with terms for *genotype, replicate* and *pot*. An ordinal generalised linear mixed model with a probit link function was also fitted to these data, including terms for *genotype, replicate, pot* and *plant*. The second analysis is more appropriate for the data scale, and the results from this analysis are in the form of cumulative probabilities at threshold values across the ordinal scale. The analysis methods are compared in the following section.

The ultimate aim of this study was to compare the similarity in ranking of genotype performance between each of the methods. To achieve this, a multivariate analysis using each of the methods as traits was undertaken. Genotypes were fitted as random effects in the analysis, paralleling selection theory which states that Best Linear Unbiassed Predictions (BLUPs) give the minimum mean square error of prediction for genotype performance, and conversely maximum correlation to the 'true' underlying genetic effects. Hence, estimated correlations between the methods at the genetic level give the most reliable comparison of genotype ranks between methods. The approach is far superior to correlations between fixed effect means.

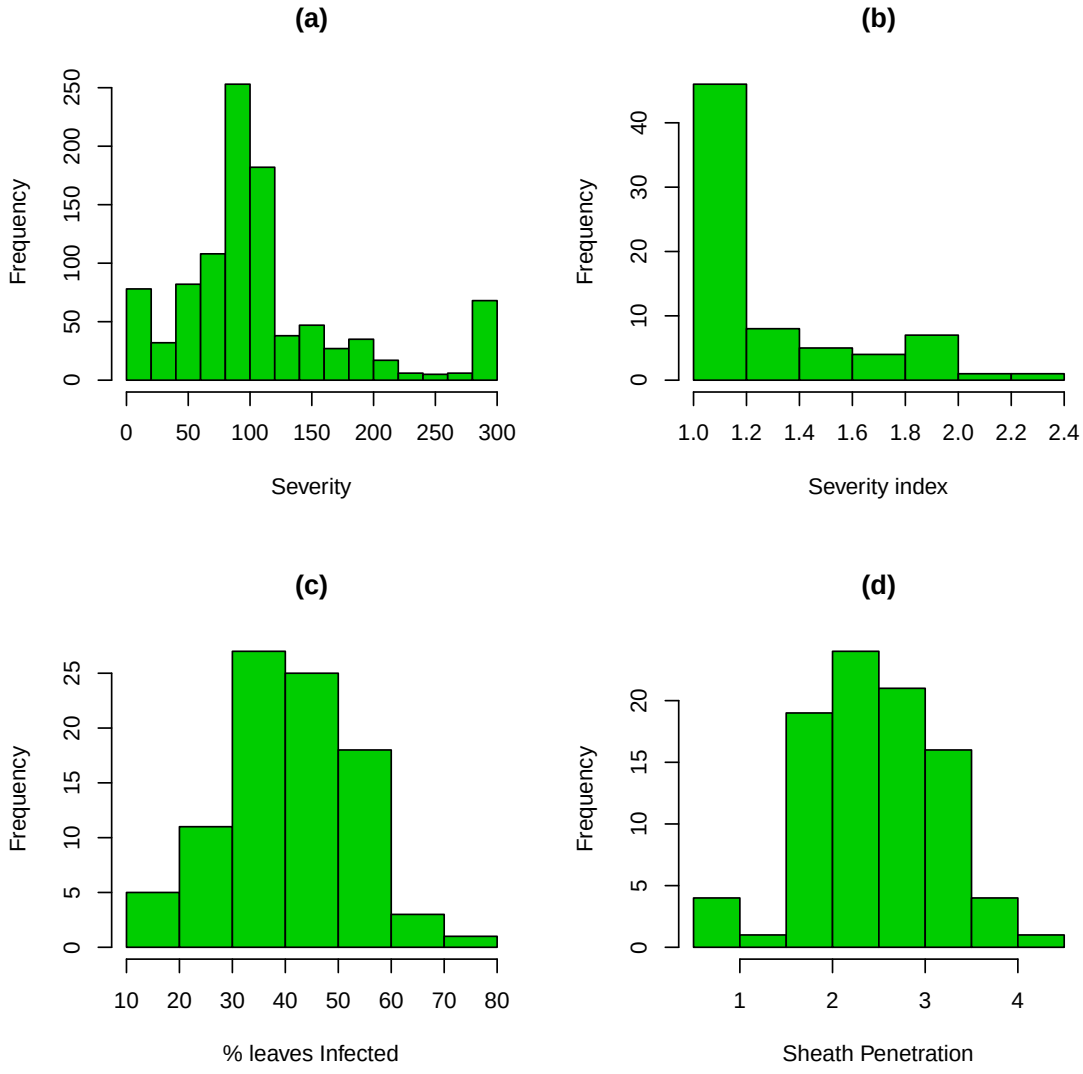

Figure 1.2: Histograms of raw data for three assessment methods; a) DEEDI pot test, b) UOS bench test, c) SA droplet test, d) SA droplet test

1.4 Results

The performance of genotypes in each method is presented in the following tables, and correlations between the methods are given at the end of this section.

The results for severity index from the CSIRO droplet method are given in Table 1.1. The predicted results and standard error of difference are on the *log* scale, and transformed means are also given on the data scale.

The results for crown rot severity from the DEEDI pot test are given in Table 1.2. The predicted results and standard error of difference are on the data scale, which is percentage of leaf sheath lesioned, totalled over the first three leaves.

The results for crown rot severity from the SA droplet method are given in Table 1.3. The predicted results and standard error of difference are from a model assuming normality. When treating the data as purely binomial, predictions of probability of infection are listed as the second data column in the Table. Both analysis methods give nearly identical predictions of genotype performance. For the second variable of sheath penetration, the predicted results and standard error of difference are given in column 3. When treating the data as counts with an underlying poisson distribution, predictions of sheath count are listed as the fourth data column in the Table 1.3. Both analysis methods again give nearly identical predictions of genotype performance.

The results for crown rot severity rating from the UOS bench test are given in Table 1.4. The predicted means and standard error of difference are on the log scale, and the backtransformed means are on the data scale, which is a rating from 0-4. This method was the only one that did not detect significant differences between genotypes. One feature of the data was the lack of variation for a number of genotypes, ie all plants with rating 1. These lines were excluded from the analysis, and later inserted back with a constant mean, and no variance, (see 0^* in Table 1.4). Of further concern was the lack of discrimination between genotypes for this method. To explore whether this was caused by the unrealistic assumption of normality for these ordinal data, a second analysis was undertaken, modelling the data as counts in ordinal categories, and estimating the cumulative probability of being in each of the rating categories from 1-4. This analysis also detected little genetic discrimination, as all genotypes had a high probability of being rated in category 1, (see Probability column in Table 1.4).

The ultimate aim of this study was to compare the similarity in ranking genotype

Table 1.1: Results for the analysis of crown rot severity index from the CSIRO Droplet test on the Ring set entries

Entry	*Severity index* Log scale	*Severity index* Data scale
1	0*	0*
2	0*	0*
3	0.283	0.327
4	0.544	0.723
5	0.727	1.068
6	1.01	1.737
7	0.590	0.804
8	0.227	0.254
9	0.396	0.486
10	0.955	1.599
11	0.507	0.661
12	0.686	0.986
13	0.754	1.126
14	0.451	0.570
15	0.419	0.521
16	0.350	0.419
17	0.666	0.946
18	0.671	0.956
19	0.790	1.203
20	0.349	0.418
21	0.048	0.0496
22	0.167	0.182
23	0.671	0.956
24	0.450	0.568
25	0.430	0.538
26	0*	0*
27	0.383	0.466
28	0.024	0.0248
29	0.108	0.114
30	0.749	1.115
LSD(5%)	0.418	

Table 1.2: Results of analysis of the crown rot severity data from the DEEDI pot test on ring set entries

Entry	*Severity*
1	249.5
2	161.5
3	63.9
4	103.5
5	86.8
6	111.2
7	116.5
8	67.5
9	144.8
10	116.5
11	115.0
12	102.4
13	85.0
14	175.8
15	112.5
16	115.3
17	76.0
18	90.0
19	79.3
20	77.0
21	146.3
22	95.3
23	104.7
24	98.7
25	91.3
26	139.4
27	55.4
28	48.3
29	50.8
30	88.5
LSD(5%)	39.2

Table 1.3: Results for two analysis models for the crown rot data from the SA Droplet test on ring set entries

Entry	*% leaves Infected* Mean	*% leaves Infected* Proportion	*Sheath penetration* Mean	*Sheath penetration* Count
1	44.5	0.44	2.70	2.69
2	32.3	0.34	2.25	2.36
3	31.7	0.32	2.00	1.99
4	50.9	0.51	3.11	3.11
5	47.7	0.48	3.00	3.00
6	37.8	0.37	2.22	2.21
7	43.2	0.42	2.47	2.44
8	50.7	0.51	2.71	2.70
9	50.7	0.50	2.89	2.85
10	42.7	0.43	3.07	3.06
11	33.0	0.33	2.31	2.23
12	39.4	0.39	2.67	2.66
13	35.9	0.36	2.39	2.40
14	65.7	0.67	3.71	3.73
15	39.4	0.39	2.33	2.34
16	38.9	0.39	2.42	2.41
17	31.9	0.32	1.84	1.82
18	32.5	0.32	2.07	2.06
19	30.6	0.31	2.00	2.00
20	50.6	0.51	3.22	3.21
21	34.2	0.33	1.99	1.96
22	43.3	0.44	2.89	2.86
23	25.1	0.25	1.61	1.60
24	31.2	0.32	2.02	2.04
25	35.4	0.35	2.36	2.34
26	54.7	0.55	3.63	3.63
27	26.7	0.27	1.67	1.65
28	46.9	0.47	2.58	2.58
29	50.3	0.50	3.06	3.05
30	46.7	0.46	2.81	2.79
LSD(5%)	14.3		0.91	

Table 1.4: Results for two analysis models for the crown rot severity score from the UOS bench test on ring set entries

Entry	*Score* log scale	*Score* data scale	*Probability* of rating 1
1	0.298	1.35	0.8591
2	0*	1*	0.9573
3	0.213	1.24	0.8772
4	0*	1*	0.9678
5	0.306	1.36	0.8421
6	0.275	1.32	0.8658
7	0.275	1.32	0.8503
8	0*	1*	0.9673
9	0.0456	1.05	0.9511
10	0.347	1.41	0.8115
11	0.0656	1.07	0.9403
12	0*	1*	0.9679
13	0.0841	1.09	0.9375
14	0.292	1.34	0.846
15	0.147	1.16	0.9134
16	0.130	1.14	0.9213
17	0*	1*	0.9683
18	0*	1*	0.9667
19	0*	1*	0.9683
20	0.300	1.35	0.8282
21	0*	1*	0.9678
22	0*	1*	0.9678
23	0.0456	1.05	0.9518
24	0.0656	1.07	0.9403
25	0.101	1.11	0.9418
26	0*	1*	0.9679
27	0.0456	1.05	0.9517
28	0*	1*	0.9683
29	0*	1*	0.9678
30	0.412	1.51	0.7814
LSD(5%)	0.38		

Table 1.5: Genetic correlation between methods for the crown rot ring set entries - wheat lines only

DEEDI	0.57		
SA infected	-0.20	0.54	
SA sheath	-0.02	0.61	-
	CSIRO	DEEDI	SA infected

performance between each of the methods. Table 1.5 lists the genetic correlations between each of the methods. The UOS bench test had to be excluded from this comparison as it did not discriminate between the genotypes, (ie it had no genetic variance from which to estimate a covariance with other methods). As some of the methods appeared to have difficulty infecting barley, and because there were also issues with establishment for one of the barley lines, the correlations are only based on the wheat data.

Pairwise plots between BLUPs from this analysis are given in Figure 1.3, 1.4, 1.5, 1.6 and 1.7 .

1.5 Discussion

Some statistical points for consideration:

- *Incidence vs severity rating*: The level of correlation between methods is generally poor, and we can see that there are substantial differences in ranking of genotypes between each of the methods. The lack of correlation between the SA assessment with CSIRO may not be surprising even though both methods are droplet based. The SA method measures the percentage of leaves infected, whereas the CSIRO method attempts to quantify the proportion of sheath infected, scaled up for the total number of infected leaves. DEEDI also attempts to measure the percentage of sheath lesioned or browned, and is moderately correlated with both the CSIRO and SA method.

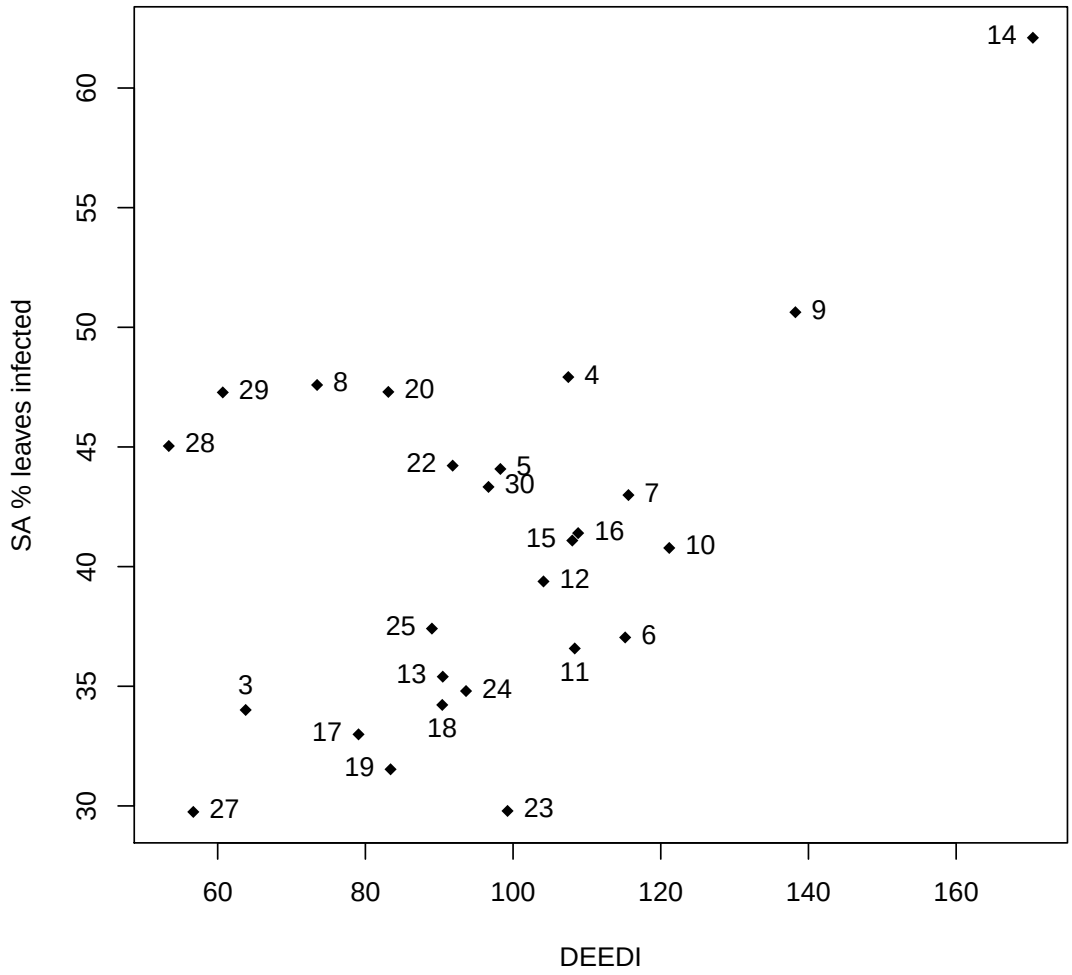

Figure 1.3: Plots of BLUPs of genetic effects between SA % leaves infected and DEEDI lesioned

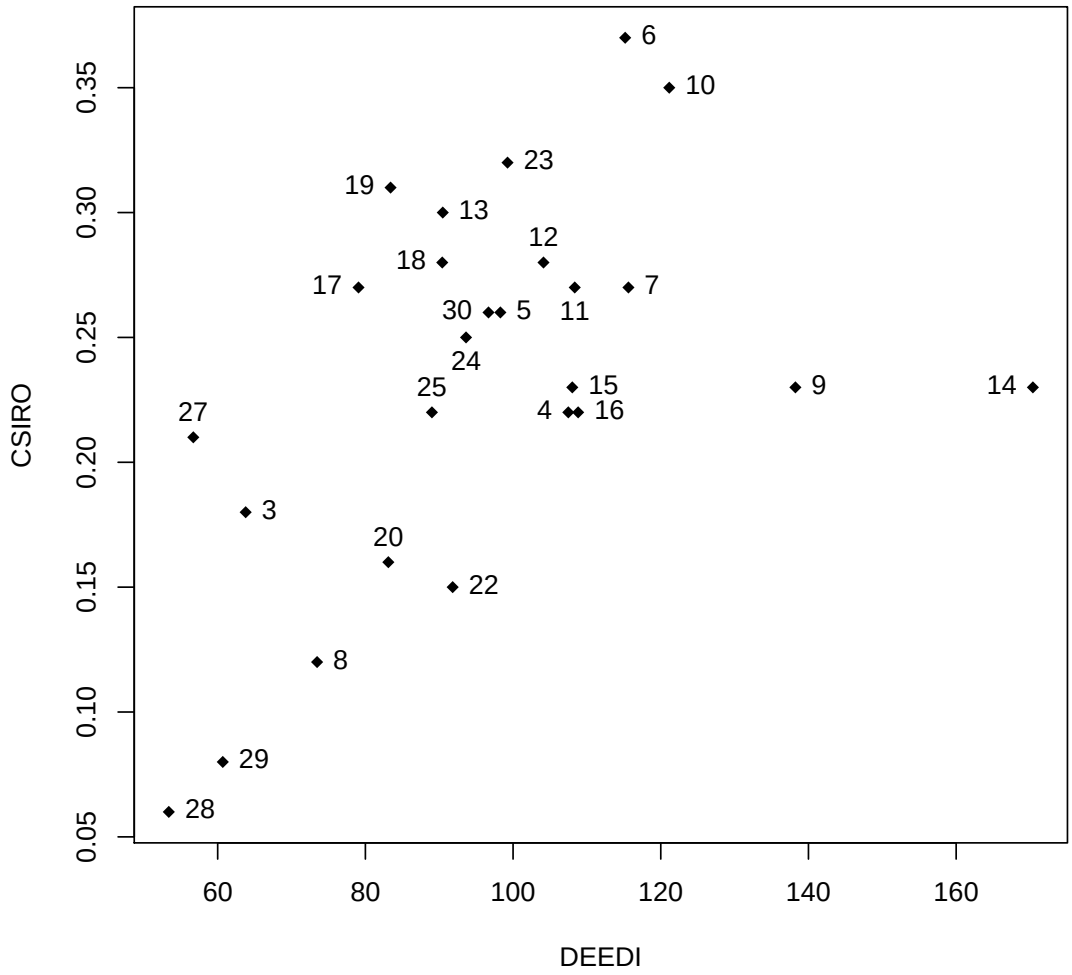

Figure 1.4: Plots of BLUPs of genetic effects between CSIRO severity index and DEEDI lesioned

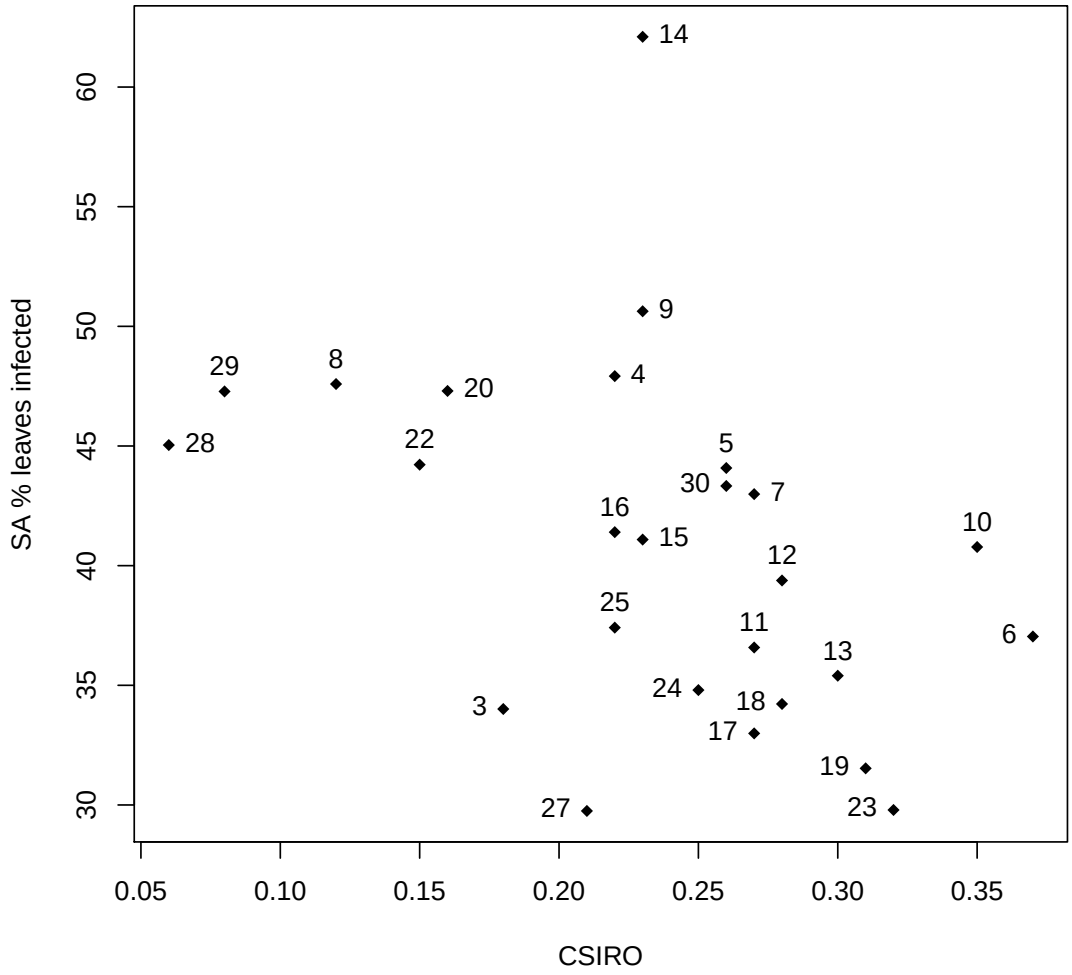

Figure 1.5: Plots of BLUPs of genetic effects between CSIRO severity index and SA % leaves infected

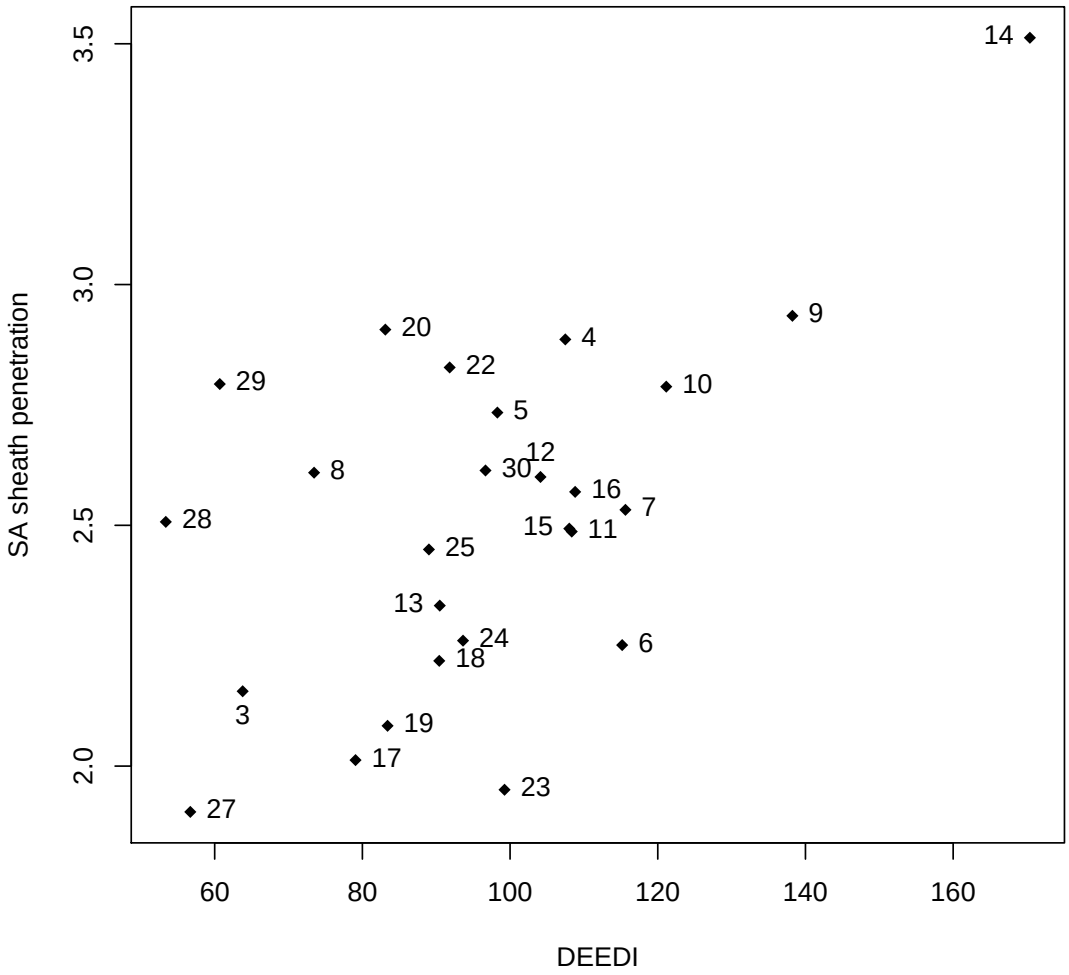

Figure 1.6: Plots of BLUPs of genetic effects between SA sheath penetration and DEEDI lesioned

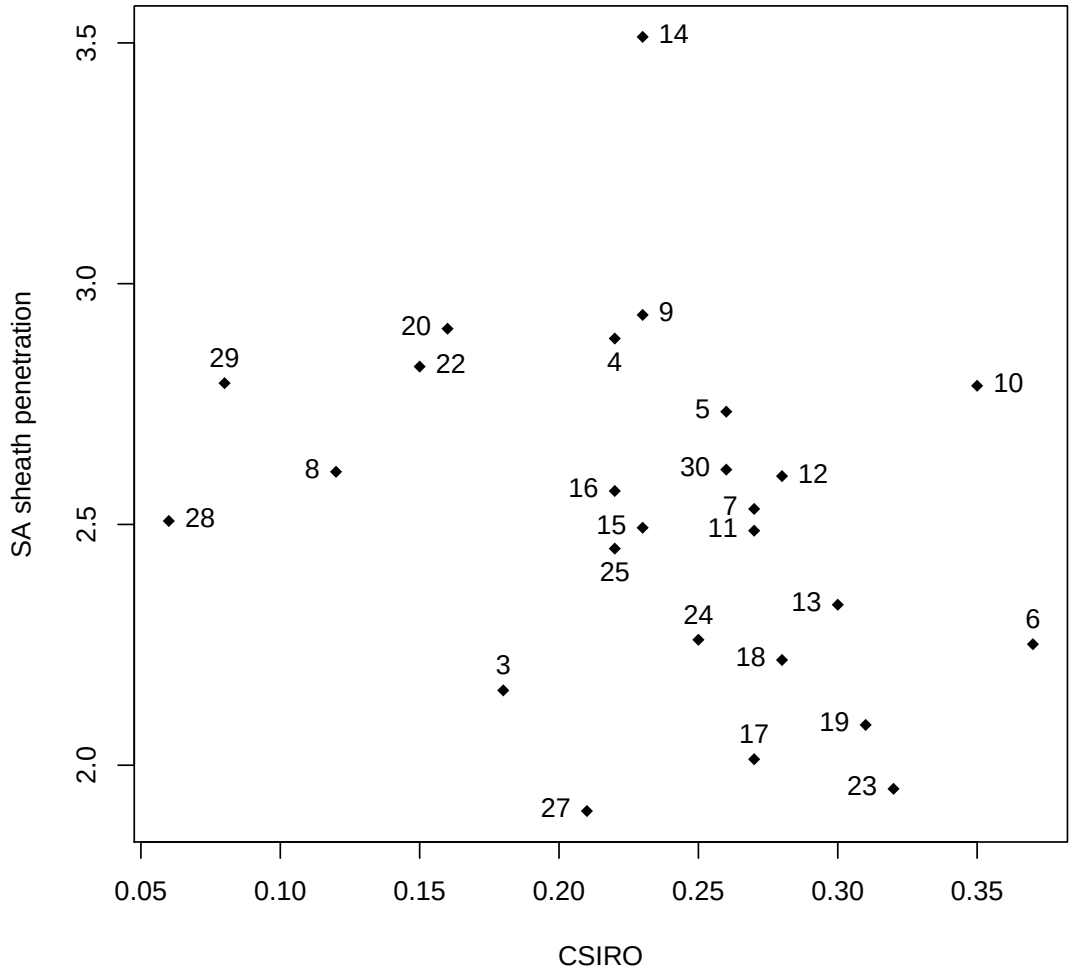

Figure 1.7: Plots of BLUPs of genetic effects between CSIRO severity index and SA sheath penetration

- *UOS rating scale of 0-4*: It appears that the rating scale for UOS screening does not adequately differentiate between levels of seedling infection. The scale may be appropriate for adult plant screening where a greater number of internodes are lesioned, but, based on these data, the rating scale does not adequately detect differences between seedlings of genotypes with a range of crown rot infection.

- *Measurement variation*: The LSD figure in the tables indicates the value above which we can consider the level of disease for genotypes to be different, and is one measure of precision for each method. This figure is given on the data scale for SA and DEEDI, and we can see that these methods can detect differences above approximately 14% plants infected for SA incidence, and 13% of one leaf sheath lesioned for the DEEDI severity measure. It is more difficult to assess precision of the CSIRO method as the analysis was conducted on the *log* scale and hence errors are also on this scale. An approximate precision figure for % browning for the CSIRO method is 0.52 on the scale of the index, which is the % browning scaled up by the number of leaves infected. The precision measure for the UOS results is also on the *log* scale and is greater than all differences between genotype scores.

- *Lack of experimental design for CSIRO method*: In one sense it is interesting to note how each method is routinely conducted. However, the absence of a randomised design for this method raises questions as to the robustness and validity of results. I strongly recommend that experimental design should be implemented in this methodology.