Phomadidacte: a computer-aided training program for the assessment of phoma stem canker severity of oilseed rape

JN Aubertot\textsuperscript{a}, Y Sohbi\textsuperscript{b}, H Brun\textsuperscript{c}, A Penaud\textsuperscript{d}

\textsuperscript{a} UMR d’Agronomie INRA/INA P-G, BP01, 78850 Thiverval-Grignon, France, aubertot@grignon.inra.fr
\textsuperscript{b} UMR d’Agronomie INRA/INA P-G, BP01, 78850 Thiverval-Grignon, France, sohbi@grignon.inra.fr
\textsuperscript{c} UMR INRA/Agrocampus Rennes : Biologie des Organismes et des Populations appliquée à la Protection des Plantes
INRA Domaine de la Motte, BP 35327, 35653 Le Rheu Cedex, France, brun@rennes.inra.fr
\textsuperscript{d} CETIOM, Centre de Grignon, BP04, 78850 Thiverval Grignon, France, penaud@cetiom.fr

Abstract

Phoma stem canker (\textit{Leptosphaeria maculans}) is one of the main diseases that affect oilseed rape worldwide. It is important to use reliable methods to describe the disease which is often characterised by the visual assessment of the severity of cankers created by the pathogen at the bottom of the stems. In order to maintain standardisation of assessment keys, a rating scale has recently been developed in France. However, like many other rating systems for disease severity, the inter-rater reliability has to be improved. A computer-aided training program, named \textit{Phomadidacte}, has been developed to tackle this question. The basic principle of this program consists in having a user rate a number of pictures of cankered cross-sections that had previously been rated by a set of experts. Those two sets of disease assessments, are then compared for discrepancies. At the end of a training session, the program either suggests that the user keeps on training or declares a good match between the experts’ and the user’s ratings, based on two thresholds: the percentage of pictures correctly rated and the percentage of pictures for which the rate given by the user and the expert differ by more than one severity grade. At the end, a graph summarising the differences between the user’s and the experts’ grades is displayed and an ASCII file containing the data of the session is created. The user can then browse all the pictures to compare the grade that he or she gave with the experts’ grade. An online version of \textit{Phomadidacte} is also available. A “before and after training test” for one rater is presented in this paper to exemplify the use of \textit{Phomadidacte}. The method described in this paper to develop such a program is not specific to phoma stem canker and could therefore be successfully applied to other pathosystems.

Key words: \textit{Leptosphaeria maculans}, \textit{Brassica napus}, blackleg, canola, phytopathometry

1 Introduction

Phoma stem canker (blackleg), caused by \textit{Leptosphaeria maculans} (Desmaz.) Ces. & De Not. [anamorph \textit{Phoma lingam} (Tode: Fr.) Desm.], is a major oilseed rape (\textit{Brassica napus} L.) disease world-wide (West et al., 2001). Epidemics are initiated during autumn by air-borne ascospores released from infected stubbles of previous crops. Once in contact with plants, these ascospores germinate and produce leaf lesions. The fungus then grows systemically from the leaf lesions to the stem where it produces cankers responsible for yield losses that can be very important. Research and extension programs related to the pathosystem require reliable estimation of the disease’s symptoms, especially the severity of stem cankers. Studies related to genetics (e.g. Chèvre et al., 1997 or Pilet et al., 1998), chemical control (e.g. Khandura and Barbetti, 2002), cultural control (Aubertot et al, 2004a), epidemiology (e.g. Wherrett et al., 2004), yield losses (e.g. Zhou et al., 1999), diagnosis in commercial fields (such as defined by Doré et al., 1997; e.g. MacGee and Emmet, 1977) require the thorough characterisation of stem canker severity. However, this assessment is not easy to perform due to the heterogeneity of symptoms that can be
observed in fields. Indeed, cankers can widely vary in the extension, shape, and discolouration at the crown or basal stem levels. A rating scale widely used in France (Pierre and Regnault, 1982) has been recently modified in order to maintain standardisation of assessment keys (Aubertot et al., 2004b). It consists in defining 6 severity classes as a function of the percentage of the discoloured cross-section: 1, healthy plant, no visible lesion; 2, [0-25%] of discoloured cross-section; 3, [25-50%] of discoloured cross-section; 4, [50-75%] of discoloured cross-section; 5, [75-100%] of discoloured section; 6, section without any living tissue, plant lodged or broken at the crown level during sampling. Because inter-rater variability should always be minimised, a computer-aided training program has been developed to help teach how to perform severity assessment of phoma stem canker of oilseed rape using the scale proposed by Aubertot et al. (2004b). This program is named Phomadidacte. This paper first presents its structure, and then a training session is exemplified with a “before and after training test”.

2 Materials and Methods

2.1 Development of a standard database for phoma stem canker severity

The first information needed when developing a computer-aided training program in the field of phytopathometry is a standard for the disease severity being studied. Usually, that type of program uses computer generated diagrams where the proportion of discoloured tissues is precisely defined. These approach has been successfully applied for numerous pathosystems (e.g. DISTRAIN; Tomerlin and Howell, 1988; Disease.Pro, Nutter and Schultz, 1995). However, internal symptoms such as phoma stem cankers are difficult to depict in computer generated diagrams because they vary widely in their discolouration. In addition, disease assessments can often be perturbed by contingencies such as multiple pest symptoms (other diseases or animal pest symptoms), or variability in phenotype of the sampled plants. This is why we chose to create a database of pictures rated by 8 experts from INRA (Institut National de la Recherche Agronomique), CETIOM (Centre Technique Interprofessionnel des Oléagineux Métropolitains), and GEVES (Groupe d’Etude et de contrôle des Variétés Et des Semences). The pictures were not chosen to represent typical symptoms, but to represent cases that are commonly encountered when assessing phoma stem canker severity in the field, such as small soil aggregates present on the cross-section being observed, or multiple symptoms present on the same cross-section (e.g. black scurf caused by Rhizoctonia solani, or galleries of rape stem weevil, Ceuthorhynchus napi Gyllenhall).

A large set of pictures was rated independently by each of the 8 experts. In order to synthesise this expertise, a picture was declared to be in a given severity class when the majority of the experts were in agreement (i.e. whenever at least 5 of the 8 experts agreed on the same severity class for a given picture). A set of 20 pictures for each of the 6 severity classes was thus constituted, leading to a standard database of 120 pictures.

2.2 Process flow chart of Phomadidacte

![Fig. 1 Process flow chart of Phomadidacte.]

The structure of Phomadidacte is presented Fig 1. After having specified his or her name, the user rates the 120 standardised pictures which appear in a random order (Fig 2). At the end of the rating stage, the distribution of the differences between the experts’ and the user’s grades is displayed and a short message interprets the results, either congratulating the user for being in good agreement with the experts or suggesting to carry on training before assessing phoma stem canker severity in experiments. The content
of the message depends on two criteria: the proportion of picture correctly rated and the proportion of picture with a difference greater than 1 class. In order to be congratulated, the user has to rate 75% of the pictures exactly like the experts and to have not more than 5% of picture rated differently from the experts by more than 1 severity class. These thresholds were defined using the intra-variability of the experts: each expert would satisfactorily fulfil these two conditions when compared with the other seven experts. The user then has the possibility to browse through the pictures to compare his grades to the ones of the experts. An output file is created with the results of the training session.

2.3 Illustration of a training session

In order to illustrate how Phomadidacte operate, a “before and after training test” was performed with a user that had never rated phoma stem canker severity before. The test consisted in comparing the accuracy of the rater before and after training with Phomadidacte. An initial training session consisted in viewing a set of 6 pictures (not used in the standard database) and sketches illustrating the 6 severity classes defined by Aubertot et al. (2004b). Phomadidacte was then used to test the accuracy of the rater for the “before” situation with a set of only 10 pictures per severity class (60 pictures in a random order in total). This rating session constituted a second training (with Phomadidacte), along with the possibility of browsing all the rated pictures and comparing the user’s and the experts’ grades. The accuracy of the user was then tested using a second set of 10 pictures per severity class (60 pictures in a random order in total, different from those used in the training session). The non-parametric Wilcoxon signed rank test of the Univariate procedure of SAS (Release 6.12 for Windows, SAS Institute Inc., 1989) was used to analyse whether the mean difference between the phoma stem canker severity classes given by the user and the experts was significantly different from zero (pairwise comparison).
3 Results

Before using *Phomadidacte*, the difference between the classes given by the user and the experts was statistically different from 0 (mean of the difference = -0.32; p < 1.0x10^{-4}; n = 60). The user agreed with the experts for 55% of the pictures, he overestimated the severity of the symptoms by one class in 38% of the cases and underestimated it by one class in 7% of the cases. No difference greater than one class was observed. After using *Phomadidacte*, the difference between the classes given by the user and the experts was also statistically different from 0 (mean of the difference = 0.25; p = 3.0x10^{-4}; n = 60). The user agreed with the experts for 71% of the pictures, he overestimated the severity of the symptoms by one class in 2% and underestimated it by one class in 27% of the cases. Like the “before training” situation, no difference greater than one class was observed.

Before training with *Phomadidacte*, the user generally overestimated by one class the first and fourth severity class and underestimated the fifth severity class (Fig 4). After training, the first, second and sixth severity class were correctly rated, but the third, fourth and fifth severity class were systematically underestimated (Fig 5).
Fig. 5 Distribution of the differences of phoma stem canker severity classes between Experts and user after training with Phomadidate by severity class of phoma stem canker. White bars: Experts’ rate-User’s rate= –1, black bars: = 0, gray bars: = +1.

4 Discussion

Using Phomadidacte improved the percentage of times that the user gave the same rating as the experts from 55% to 71% for the experiment conducted. However, even if the number of matches with the experts increased, the rater underestimated the severity of phoma stem canker for intermediate classes 3, 4 and 5. Reasonably, Phomadidacte advised the user to carry on training before assessing phoma stem canker severity in experiments. The purpose of this example was to illustrate how Phomadidacte operates and how its efficacy to train raters can be assessed. A real assessment of this efficacy would require a larger set of users since there is certainly a great variability of efficacy of Phomadidacte within a population of raters. Some raters would greatly improve their assessments after being trained, whereas others might be quite accurate from the beginning. In addition, the efficacy of Phomadidacte has been evaluated on half of the photo database available. Training with the whole standard database should certainly be more efficient.

The main difficulty when developing computer-aided training programs for disease severity assessment lies in the establishment of a standard to calibrate raters (Cooke, 1998). Image analysis could appear an objective way to quantify disease severity. However, since a threshold has to be used to distinguish healthy from diseased tissues, image analysis also suffer from a lack of objectiveness (Nutter et al., 1993). The originality of Phomadidacte consists not only in using pictures of symptoms to train raters, but also in using the expertise of a set of specialists to develop a standard. This approach should not be seen as an alternative to computer-aided training programs that use generated diagrams, but as a complementary method. Such programs aim at training to assess the percentage of a diseased area on plant organs (leaves, generally) for contrasted objects (e.g. orange pustules of brown rust on green leaves of wheat), whereas symptoms like cankers present a wide range of discoloration and shape that are very difficult to mimic with computer generated diagrams. This is why the presented approach is of interest for various other pathosystems. The drawback of this method is that the standard used is actually an agreement, and not really a true measurement of the disease severity. This is why it is important to constitute the standard with a panel of experts as wide as possible to ensure the representativeness of the proposed standard. At present, a French version of Phomadidacte is available online. An English version is being developed. Requests to use Phomadidacte online can be submitted to aubertot@grignon.inra.fr.
5 Acknowledgements

The authors thank D Le Floch and R Gosse for their technical assistance, the 8 experts from INRA, CETIOM, and GEVES for having accepted to rate hundreds of pictures of cankered cross-sections. They are also grateful to S Pineau, secretary of the UMR d’Agronomie INRA/INA P-G who accepted to perform the “before and after training test” to assess the efficacy of Phomadidacte. We thank Karyn Aubertot who thoughtfully checked the English text.

6 References


