

## Applications of machine learning to identify drivers of *Phytophthora infestans* population diversity

ARACELI TORRO-GALIANA<sup>1</sup>, DAVID COOKE<sup>1</sup> & PETER SKELSEY<sup>1</sup>

<sup>1</sup> The James Hutton Institute, Errol Road DD2 5DA Dundee, Scotland, UK

### SUMMARY

*Phytophthora infestans*, the causal organism of late blight, is a devastating pathogen of potato and tomato crops that has been increasing in incidence and complexity worldwide. The objectives of this study were to analyse the genetic diversity of *P. infestans* populations in Great Britain (GB) in order to identify areas with a disproportionately high occurrence of non-clonal lineages. Analysis of >2000 late blight outbreaks during 2006-2017 revealed 163 non-clonal outbreaks in different postcode districts across GB. Spatial autocorrelation and space-time analyses revealed areas of positive and negative local spatial autocorrelation of non-clonal outbreaks (clusters and outliers). An Emerging Hot Spot Analysis identified postcode districts with increasing values of non-clonal outbreaks through time. These analyses provide the first quantitative evidence of space-time clustering of non-clonal outbreaks of late blight in GB. The results of this study will be used to guide future efforts to identify the drivers associated with increased diversity of the late blight pathogen in GB.

### KEYWORDS

Hotspot analysis, oospore, *Phytophthora infestans*, point pattern analysis, spatial statistics

### INTRODUCTION

Sexual reproduction of the potato late blight pathogen, *Phytophthora infestans*, directly affects control strategies of the disease. This is because oospores formed through sexual recombination can act as primary inoculum, surviving in the soil in the absence of the host (Smart and Fry, 2001; Montarry *et al.*, 2007). The sexual cycle also enhances fitness by spawning recombinant genotypes that may be more aggressive or resistant to fungicides (Cooke *et al.*, 2003; Andersson *et al.*, 2009; Leesutthiphonchai *et al.*, 2018). Additionally, the presence of oospores in soils can lead to an early epidemic during the growing season (Flier *et al.*, 2001; Hannukkala *et al.*, 2007; Andersson *et al.*, 2009; Cooke *et al.*, 2018).

In Great Britain (GB) the population of the pathogen remains dominated by clonal lineages reproducing asexually. However, differences in genotypic diversity have been found between populations in the northern and southern regions of potato production. Populations of

*P. infestans* in the north east of Scotland show a higher diversity of genotypes (Cooke *et al.*, 2016, 2018) and although there is not a direct evidence of soil-borne oospore as primary inoculum in British crops, evidence of sexual recombination in this area is accumulating.

A better understanding of the nature and causes of geographic variation in pathogen diversity would provide crucial information for improved management of the disease. In this study we applied the space-time pattern mining tools of ArcGIS to potato late blight outbreak data in GB during 2006-2017. The main objective was to explore the data for patterns and trends in the occurrence of non-clonal outbreaks in the northeast of Scotland and other areas of Great Britain.

## **MATERIALS AND METHODS**

### *Outbreak data acquisition*

Historical potato late blight outbreak data from the AHDB Potatoes 'Fight Against Blight' (FAB) campaign were obtained. These data are collected on a voluntary basis by Blight Scouts drawn from members of the industry who routinely walk potato fields during the season. Information on >2000 outbreaks from across GB during 2006 to 2017 were available: outbreak coordinates, date of observation, variety of potato cultivated, source and stage of the disease, the name of the scout who collected the sample, and the genotype of the pathogen. For confidentiality reasons the coordinates recorded for each outbreak are for a postcode district and not for a potato field.

### *Spatiotemporal analysis*

ArcGIS ® software was used to analyse patterns and trends in the distribution of non-clonal outbreaks and to generate maps for visualisation purposes.

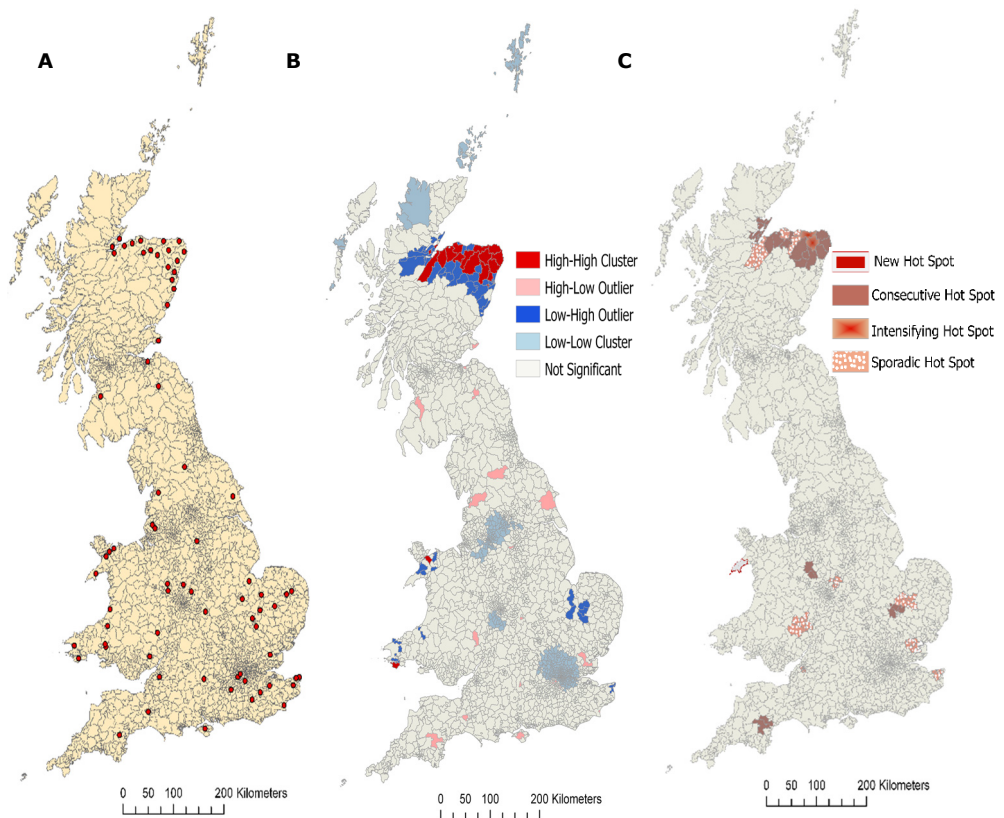
### *Spatial Cluster And Outlier Analysis*

In order to identify regional anomalies in the distribution of non-clonal outbreaks between postcode districts, a local autocorrelation index (Anselin Local Moran's I) was computed for the total number of non-clonal outbreaks in each postcode district. The Cluster And Outlier Analysis identifies statistically significant spatial clusters of high values (High-High), cluster of low values (Low-Low), outliers in which a high value is surrounded primarily by low values (High-Low), and outliers in which a low value is surrounded primarily by high values (Low-High). The analysis calculates a local Moran's I statistic (Anselin, 1995), a z-score (standard deviation) and a p-value (statistical probability). A positive value of the local Moran's I statistic indicates that a feature is a part of a cluster. A negative value of the statistic indicates that a feature is a part of an outlier. A 95% significance level (p-value<0.05) was used to indicate significant clusters or outliers of local autocorrelation.

### *Emerging Hot Spot Analysis*

Before running an Emerging Hot Spot Analysis, a netCDF (Network Common Data Form) space-time cube was generated by aggregating non-clonal outbreaks into bins defined by the postcode district and year. The netCDF file stores data in space (as latitude and longitude coordinates) and time. By assigning a count of non-clonal outbreaks to each bin in each postcode district a trend in count over time can be evaluated. The netCDF space-time cube was used as an input for an Emerging Hot Spot Analysis, to visualize and identify statistically significant increasing or decreasing trends in the occurrence of non-clonal outbreaks through the period 2006-2017.

The Emerging Hot Spot Analysis evaluates spatiotemporal patterns using two statistical measures: (1) the Getis-Ord  $G_i^*$  statistic (Ord and Getis, 1995) to identify the spatial and temporal clustering of non-clonal outbreaks, and (2) the Mann-Kendall trend test (Mann, 1945; Kendall and Gibbons, 1990) to evaluate temporal trends across the time period. The Getis-Ord  $G_i^*$  (space-time hot spot analysis) returns a z-score and p-value for each bin by comparing the number of non-clonal outbreaks in a given bin with the number of non-clonal outbreaks in the neighbouring bins. Then the hot and cold spots trends are evaluated using the Mann-Kendall trend test. The resultant trend z-score and p-value for each location with data and the space-time hot spot z-score and p-value for each bin are used to categorise significant ( $p$ -value $<0.05$ ) hot and cold spots. The analysis returns a map with different categories of increasing trend (hot spots) or decreasing trend (cold spots). Categories include new, consecutive, intensifying, persistent, diminishing, sporadic, oscillating and historical hot and cold spots. For the analysis, the 4 closest postcode districts were used to define neighbourhood size in space and temporal neighbours were defined using one prior time interval (1 year).



**Figure 1.** **A.** Postcode districts and distribution of non-clonal outbreaks in Great Britain (2006-2017). **B.** Cluster And Outlier Analysis (COA) results from non-clonal outbreak data (2006-2017). **C.** Emerging Hot Spot Analysis (EHSA) results from the space-time cube created using the non-clonal outbreak data

## RESULTS

### *Data summaries*

For the period 2006-2017, a total of 95 non-clonal outbreaks were registered in Scotland and 68 in England and Wales. In terms of the distribution of non-clonal outbreaks in Scotland, a large cluster of postcode districts with a high number of non-clonal outbreaks was identified in the north-eastern region (Figure 1A). In this region (Aberdeenshire, Moray and Inverness) a total of 90 outbreaks were recorded for the period. In contrast, in the south of Scotland (Fife, Angus, Edinburgh, Ayr and Scottish Borders) only 5 non-clonal outbreaks were registered.

### *Spatial Cluster And Outlier Analysis*

From the Cluster And Outlier Analysis, statistically significant clusters of high or low values as well as outliers with values that are statistically different than their neighbouring postcode districts were identified. Interestingly, a large cluster of postcode districts with a high number of non-clonal outbreaks surrounded by postcode districts with a high incidence of non-clonal outbreaks (high-high cluster) was found in the north east of Scotland, in Aberdeenshire, Moray and Inverness (Figure 1B).

### *Emerging Hot Spot Analysis*

The Emerging Hot Spot Analysis revealed postcode districts with increasing values of non-clonal outbreaks through time. Particularly, a large cluster of consecutive, intensifying and sporadic hot spots was found in the northeast of Scotland. The results revealed two intensifying postcode districts in Aberdeenshire (AB45 and AB53 postcode districts). A new hot spot was identified in Wales, and a few isolated consecutive and sporadic hotspots were revealed in the south of England (Figure 1C).

## CONCLUSIONS

This study provides the first quantitative evidence of space-time clustering of non-clonal outbreaks of potato late blight in GB. The causes and consequences of the presence of statistically significant clusters of postcode districts with high and increasing incidence of non-clonal outbreaks in the north east of Scotland are not yet certain. According to Cooke (2018), the introduction of new genotypes to this northern Scottish region from areas with higher genotypic diversity is limited, because this area is a seed potato producing region and the import of new genotypes from other countries is restricted. Soil-borne oospores as primary inoculum is therefore the most likely source of genotypic variation in this area. Several isolated non-clonal hot spots were also identified in the south (England and Wales), and a more intensive investigation regarding the source of this diversity is required.

This study provided an overview of the space-time pattern of the genotypic diversity of the pathogen. A more exhaustive analysis of the causes and consequences of these hotspots or clusters of non-clonal outbreaks is the subject of a future study. The areas of intensifying diversity identified in this study have been intensively sampled, and a wide variety of machine learning techniques will be applied to these data, together with potato crop distribution and meteorological data, to identify the drivers of pathogen diversity in GB.

## ACKNOWLEDGEMENTS

We thank the Mylnefield Trust for funding this research, and also gratefully acknowledge AHDB Potatoes and the Fight Against Blight scouts for providing the outbreak data.

## REFERENCES

- Andersson, B., A.-K. Widmark, J.E. Yuen, B. Nielsen, S. Ravnskov, G.J.T. Kessel, A. Evenhuis, L.J. Turkensteen, J.G. Hansen, A. Lehtinen, A. Hermansen, M.-B. Brurberg and B. Nordskog (2009). The role of oospores in the epidemiology of potato late blight. *Acta Horticulturae* 834, 61–68.
- Anselin, L. (1995). Local Indicators of Spatial Association-LISA. *Geographical Analysis* 27, 93–115.
- Cooke, D.E.L. (2018). Final Report GB Late Blight Population Monitoring 2014 to 2017. Agricultural and Horticulture Development Board, 1-37.
- Cooke, D.E.L., E. Randall, L. Sullivan and A.K. Lees (2018). The origins and implications of a novel population of *Phytophthora infestans* on potato crops in Scotland. *Proceedings Crop Protection in Northern Britain 2018*, 203–208.
- Cooke, D.E.L., P. Skelsey, S.R. Dancey, E. Randall, L. Sullivan and A.K. Lees (2016). Potato late blight research update; populations, primary inoculum and predictions. *Proceedings Crop Protection in Northern Britain 2016*, 211–216.
- Cooke, D.E.L., V. Young, P.R.J. Birch, R. Toth, F. Gourlay, J.P. Day, S.F. Carnegie and J.M. Duncan (2003). Phenotypic and genotypic diversity of *Phytophthora infestans* populations in Scotland (1995–97). *Plant Pathology* 52, 181–192.
- Flier, W.G., G.J.T. Kessel, G.B.M. Van den Bosch and L.J. Turkensteen (2001). Impact of new populations of *Phytophthora infestans* on integrated late blight management. Sixth Workshop of an European Network for Development of an Integrated Control Strategy of Potato Late Blight, 193–201.
- Hannukkala, A.O., T. Kaukoranta, A. Lehtinen and A. Rahkonen (2007). Late-blight epidemics on potato in Finland, 1933–2002; increased and earlier occurrence of epidemics associated with climate change and lack of rotation. *Plant Pathology* 56, 167–176.
- Kendall, M.G. and J.D. Gibbons (1990). *Rank Correlation Methods*. Griffin, London.
- Leesutthiphonchai, W., A.L. Vu, A.M.V. Ah-Fong and H.S. Judelson (2018). How Does *Phytophthora infestans* Evade Control Efforts? Modern Insight Into the Late Blight Disease. *Phytopathology* 108, 916–924.
- Mann, H.B. (1945). Nonparametric Tests Against Trend. *Econometrica* 13, 245–259
- Montarry, J., R. Corbiere and D. Andrivon (2007). Is there a trade-off between aggressiveness and overwinter survival in *Phytophthora infestans*?. *Functional Ecology* 21, 603–610.
- Ord, J.K. and A. Getis (1995). Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis* 27, 286–306.
- Smart, C.D. and W.E. Fry (2001). Invasions by the late blight pathogen: Renewed sex and enhanced fitness. *Biological Invasions* 3, 235–243.

