Supplement to Final Report

Characterisation of Potato Cyst Nematode Populations in Great Britain for sustainable crop management

Ref: 115R471

Report Author: Katarzyna Dybał

Date supplement submitted: 23/10/2020

Additional copies of this report and a list of other publications can be obtained from:

AHDB Potatoes
Stoneleigh Park
Kenilworth
Warwickshire
CV8 2TL

Tel: 02476 692051
Fax: 02476 478902
E-mail: Potatoes.Publications@ahdb.org.uk

Our reports, and lists of publications, are also available at potatoes.ahdb.org.uk

Next-generation sequencing

Mitotype diversity was determined in 101 of the survey samples collected between 2013 and 2016 from fields in England, previously used for ware potato production, and confirmed to be *G. pallida*. Samples were analysed to investigate the presence and composition of these mitotypes using SNPs as a molecular marker in sequencing methodology developed by Eves-van den Akker *et al*. (2015). They modified primers which amplify 310bp region of cyt-*b* by adding a distinctive 4bp barcode which was designed to allow the identification of individual survey samples during the data analysis. Additionally two 5' adenosines were added before barcode and actual forward or reverse primer sequences to act as a protecting buffer for the barcode in the samples metagenetic sequencing process. Sequencing of 310bp amplicon was carried out at Edinburgh Genomics (University of Edinburgh, Edinburgh, UK). After sequencing, bioinformatic analyses and data analysis were performed.

The paired-ends reads (over 14 million) were subjected to bioinformatic processes, which included trimming of the leading and trailing low quantity bases (Q-score 28) and selection for the minimum lengths of 150bp. The remaining 8.2 million reads (58.5%) were assembled using pair-end read merger (Pear-0.9.6) to select for overlapping pairs of reads that were then converted to Phred64 coding and next into FASTA format. The remaining 7.9 million (96.5%) reads were uniformed by changing the orientation to 5'-3' and reads from file with forward barcode and file with reverse barcode were concatenated. This reduced the number of reads to 7.8 million. Next, reads of a minimum length of 265bp containing the forward and reverse primer binding sites and barcodes no shorter than 4bp (but up to 6bp to account for the additional 5'

adenosines) were selected, which further reduced surviving reads to 7.6 million. All steps described above were performed to present the reads in a uniform format and could be compared with data from Eves-van den Akker *et al*. (2015).

It was noted that one of the previously described SNPs that was used to distinguish between types, lies within a stretch of homo-polymers, and coincided with a sequencing error (deletion) in this dataset at the position 118 of the 265bp cyt-*b* fragment. This region was deemed to be uninformative, and was excluded from further analyses by replacing with letter N. This correction to the data did not affect further analyses as the SNP at the position 118 is not the only descriptive SNP of mitotype 1, others can be used to distinguish it from types 2 and 3.

The data taken for further analyses, now reduced to 7.4 million reads, was found to contain, after the barcode sequences were excluded, 143,028 unique sequences (100% of all analysed reads) which were sorted from most to least replicated. From this selection, the top 11 most common unique sequences, which contained 72.7% of all analysed reads (Figure 1.) but represent only 0.008% of the unique sequences, were selected and extracted from the file containing all sequences. The remaining unique sequences are expected to mostly contain very rare PCR and/or sequencing errors.

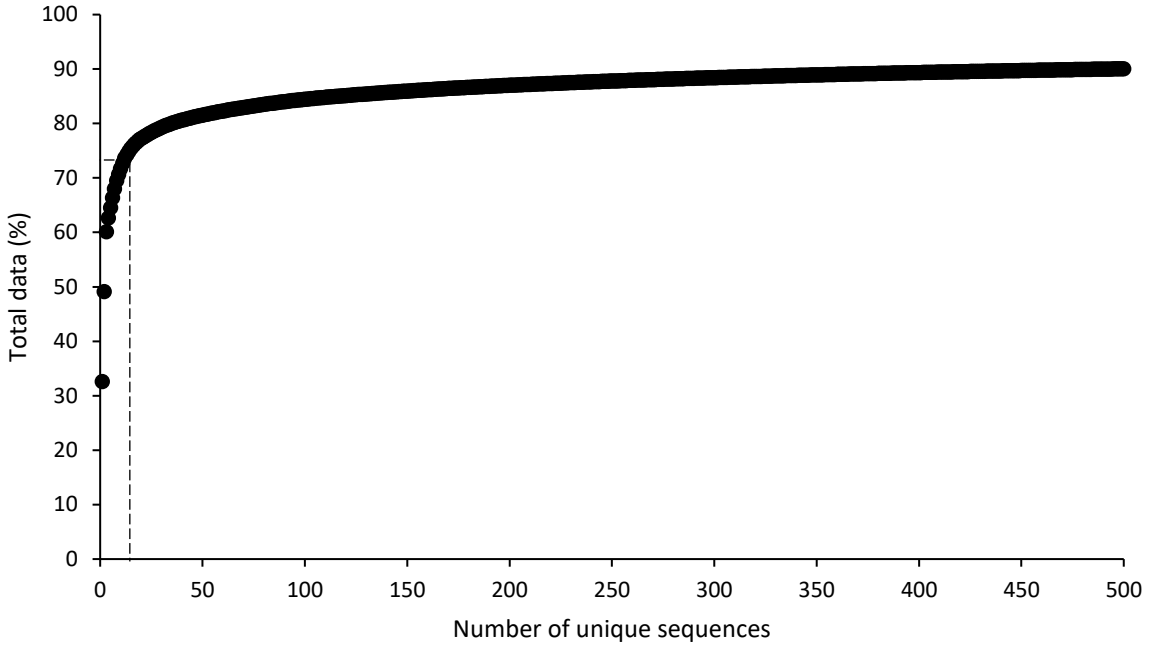

Figure 1. The top 500 most common unique sequences and their abundance as a percentage of the total data. The vertical dashed line highlights the proportion (%) of sequences accounted for by the top 11 most common unique sequences.

The top 11 most common unique sequences were numbered in order of frequency and compared to sequences describing the fragment of cyt-*b* gene from the three *G. pallida* mitotypes (plasmid Type 1, 2 and 3). The first most replicated unique sequence was identified as mitotype 3, which was also detected in the 4th, 6th, 7th and 8th most replicated unique sequences. The second most replicated unique sequence was identified as mitotype 2, which was also recognised in the 9th and 10th most replicated unique sequences. The third most replicated unique sequence was identified as mitotype 1, which was also detected in the 5th and 11th most replicated unique sequences (Figure 2.).

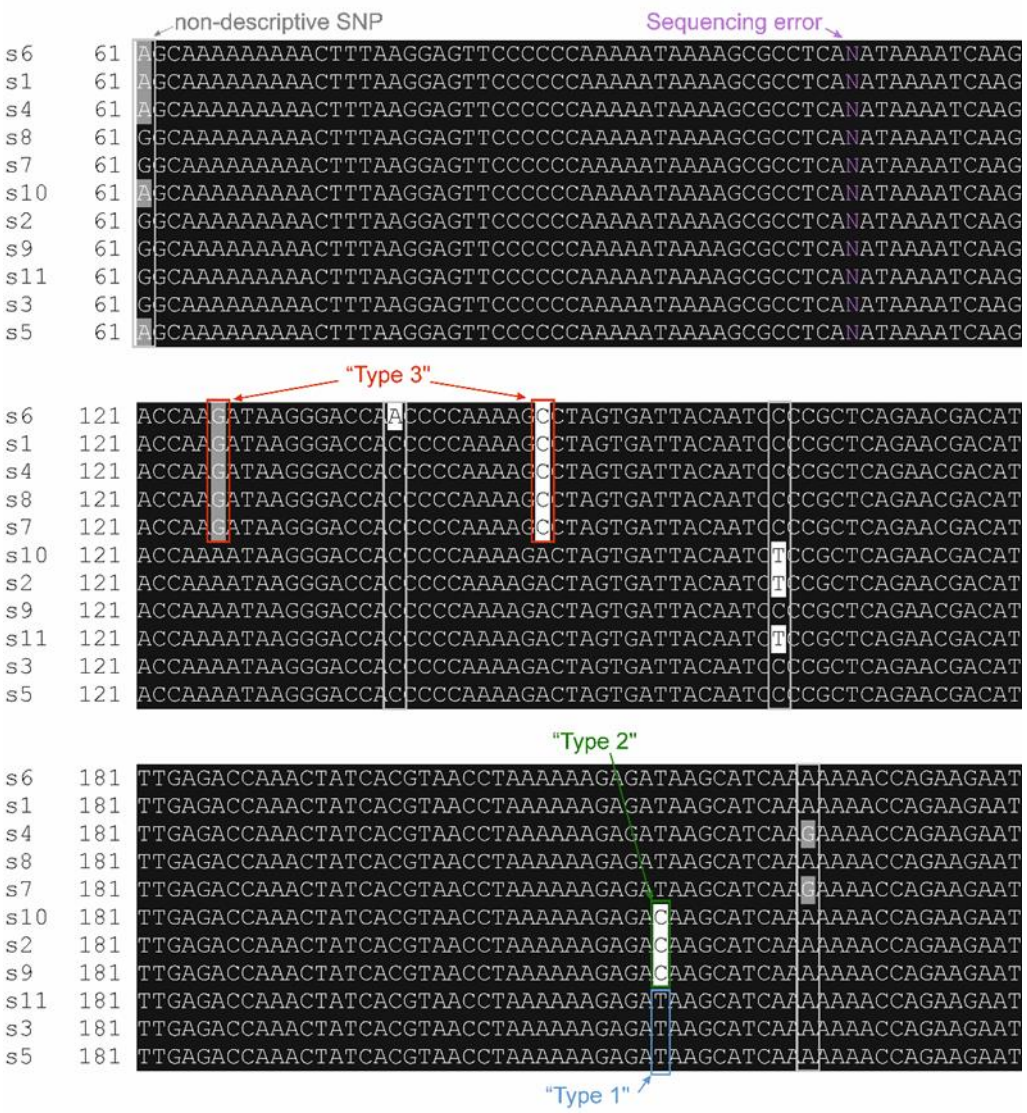

Figure 2. The fragment of the top 11 most common unique sequences showing the descriptive SNPs of each mitotype. A blue box indicates SNP descriptive of mitotype 1, a green box indicates SNP descriptive of mitotype 2, a red box indicates SNPs descriptive of mitotype 3 and a grey box indicate non-descriptive SNPs. Letter 'N' indicates corrected sequencing error.

Mitotype occurrence and distribution

Using a distinctive 4bp barcode pair (forward and reverse primer barcodes) assigned to each DNA sample, individual survey samples were identified and investigated for the presence of reads identified within the top 11 most common unique sequences descriptive of mitotypes 1, 2 and 3. From the survey samples, 61% contained a single mitotype and 39% contained a mixture of two mitotypes. Two survey samples were found to contain mitotype 1, mitotype 3 was detected in 57 survey samples and the remaining 38 survey samples were described as a mixture of mitotypes 1 and 3 (Figure 3.). Mitotype 2 was not found.

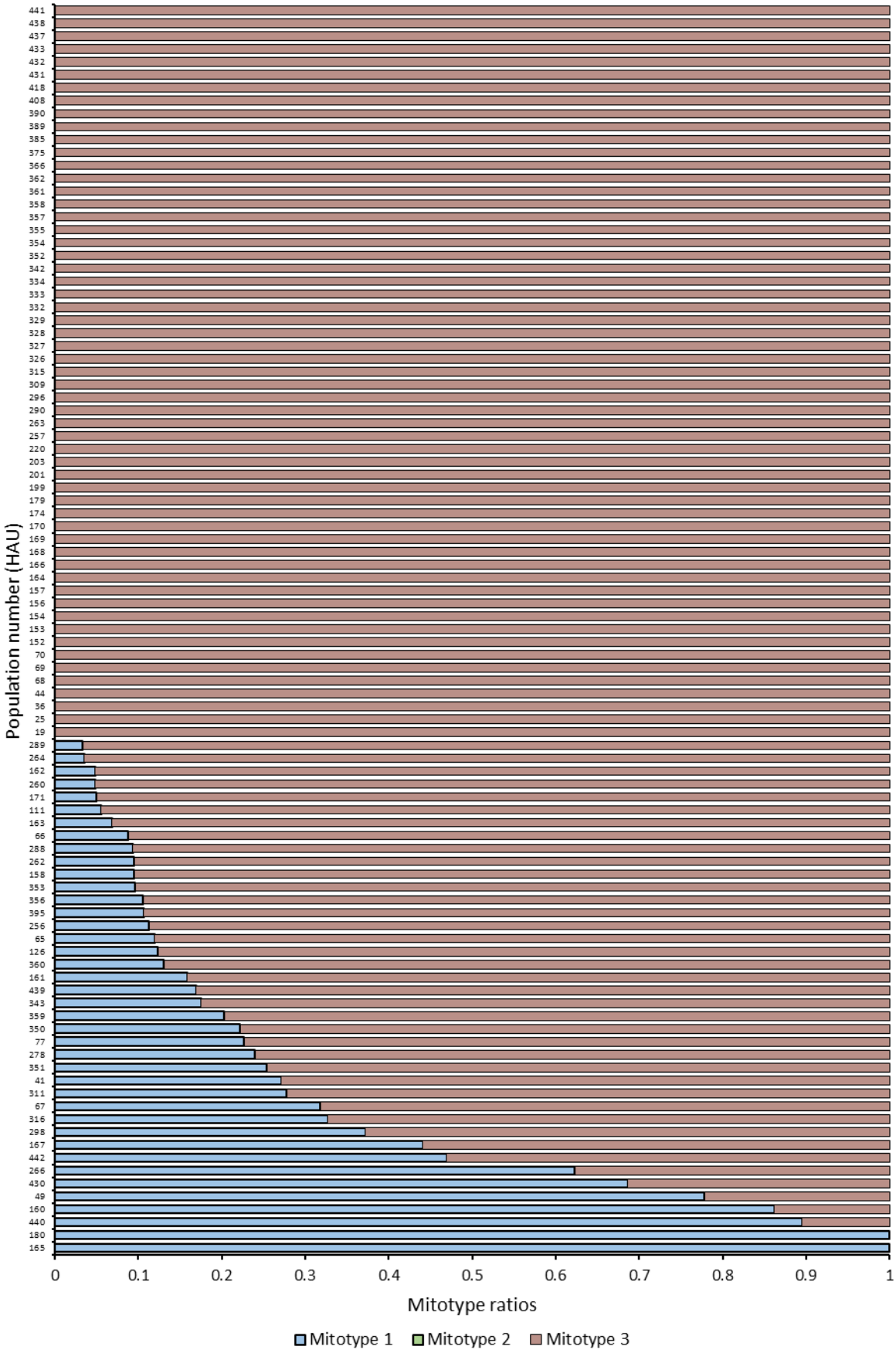

Figure 3. The presence and ratio of sequences descriptive of mitotypes 1, 2 and 3 in individual G. pallida field populations from England.

Information recorded in the survey's questionnaires showed that samples included in the next generation sequencing (NGS) analysis originated from East Midlands, East of England, North West, South East, South West, West Midlands and Yorkshire and the Humber. North East and Wales were not represented by survey samples included in NGS (Figure 4.).

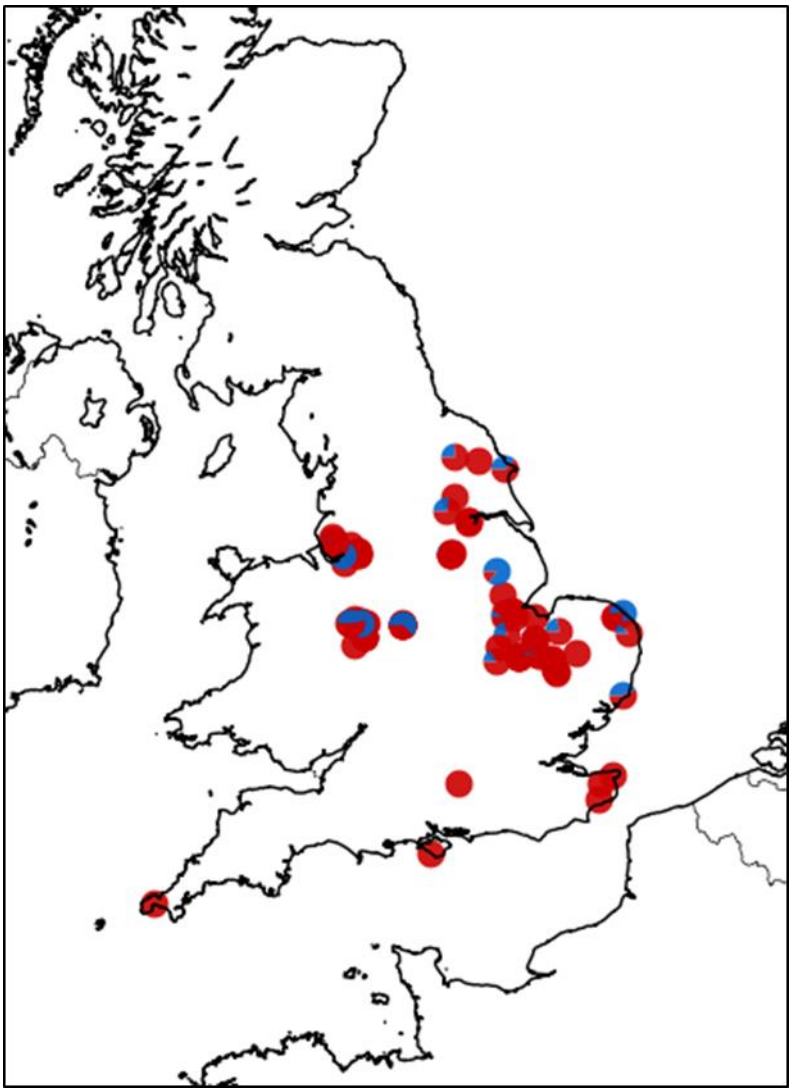

Figure 4. Distribution of *G. pallida* field populations from England characterised as mitotype 1 (blue shaded area of pie charts) or mitotype 3 (red shaded area of pie charts).

Potato cyst nematode populations identified as mitotype 3 were found in soil samples originating from all seven regions of England included in this study while

populations identified as mitotype 1 were detected solely in North West and West Midlands. The South East and South West were the only regions where populations described as a mixture of mitotypes 1 and 2 were not found (Table 1.).

Table 1. Field *G. pallida* populations characterised as mitotype 1, mitotype 3 or a mixture of mitotypes 1 and 3 from seven regions of England.

Country	Region	Total samples	Mitotype 1	Mitotype 3	Mix of mitotypes 1 and 3
England					
	East Midlands	19	0	11	8
	East England	25	0	15	10
	North West	13	1	8	4
	South East	5	0	5	0
	South West	1	0	1	0
	West Midlands	26	1	14	11
	Yorkshire and the Humber	8	0	3	5
Total		97	2	57	38

Overall, *G. pallida* field populations from England tested by NGS were found to be mitotype 1 (2% of samples), mitotype 3 (59% of samples) or a mixture of mitotypes 1 and 3 (39% of samples). Mitotype 2 was not detected in any of the populations tested, either solely or as a mix. In comparison, data obtained from a similar study conducted in Scotland (Eves-van den Akker *et al*., 2015), confirmed the presence of mitotype 2, albeit at a lower prevalence. Eves-van den Akker *et al*. (2015) also found that mitotype 3 was the most prevalent mitotype and mitotype 1 was the second most common, similar to the results presented here (Figure 5.). Furthermore, the Scottish study indicated that *c*. 79% of *G. pallida* populations contained a single mitotype, *c*. 18% contained a mixture of two mitotypes, and *c*. 2.3% contained a mixture of all three mitotypes. In this study, *G. pallida* populations belonged to either a single mitotype (61%) or a mixture of mitotypes 1 and 3 (39%). The combined detection of mitotypes 1 and 3 could be partly expected based on the finding by Eves-van den Akker *et al*. (2015)

who reported that a mixture of mitotypes 1 and 3 was approximately twice as likely to occur than mixtures of 1 and 2, or 2 and 3. The authors, following the hypothesis by Plantard *et al*. (2008) of geographically isolated original introductions from Peru, also concluded that the presence of mixed populations of *G. pallida* mitotypes within a field could give rise to novel hybrids.

Due to the limited number of samples, available from the English field sites, the presence of mitotype 2 cannot be completely excluded. The possibility of the mitotype not being detected becomes more likely when the low occurrence of mitotype 2 in Scotland, is taken into account. The absence of mitotype 2 in England may be due to (i) *G. pallida* mitotype 2 currently having a limited distribution following an introduction to Scotland or (ii) mitotypes being influenced by particular environmental conditions and/or management practices such as cultivar choice. Further tests are required on a larger scale to investigate this phenomenon and the biological significance of mitotype 2 and if confirmed measures need to be implemented to avoid the introduction of mitotype 2 into England.

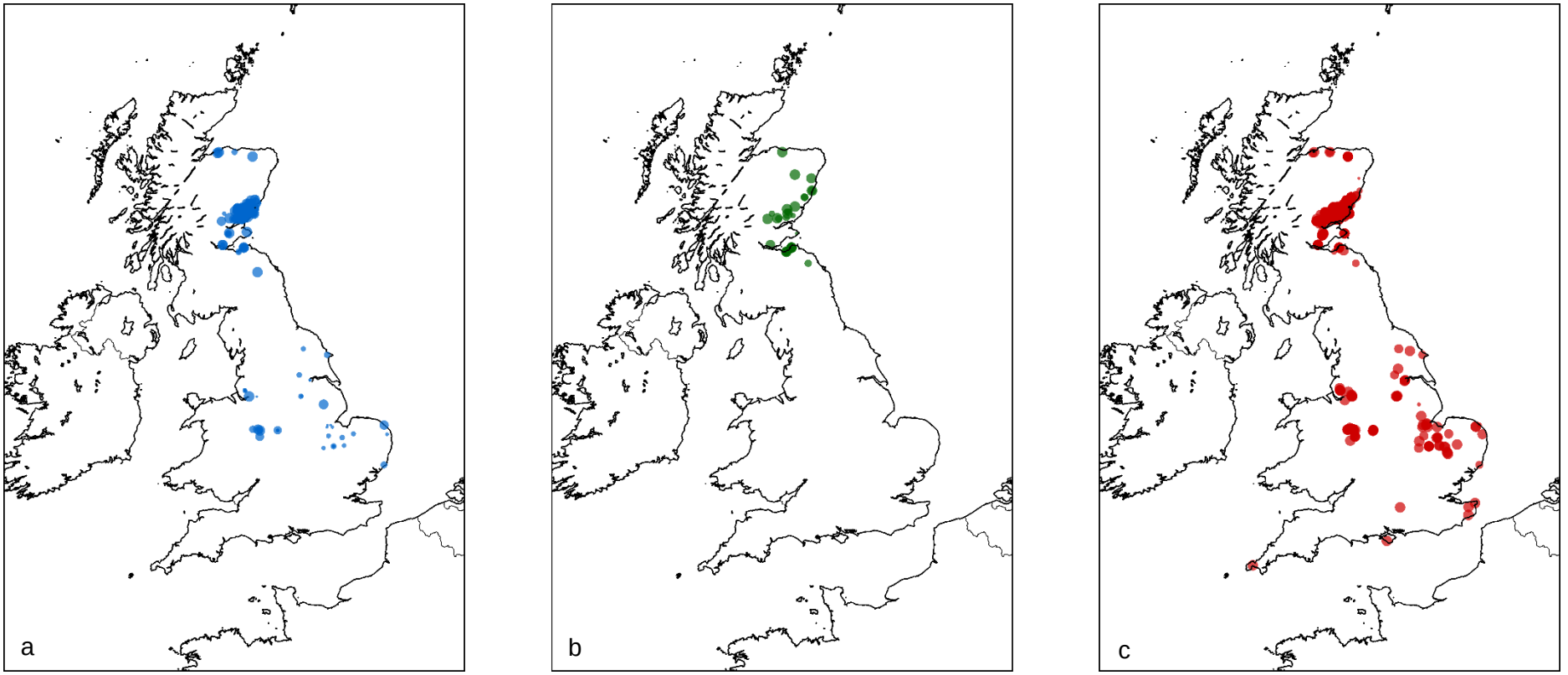

Figure 5. Distribution of *G. pallida* mitotypes from England and Scotland characterised as a) mitotype 1 (blue dots), b) mitotype 2 (green dots) and c) mitotype 3 (red dots).