Detection of genetic variability in cattle infectivity for bovine tuberculosis (bTB)

E. Sánchez-Molano[1*], D. Madenci[1], S. Tsairidou[1], M. Winters[2], A.P. Mitchell[3], G. Banos[1,4] and A. Doeschl-Wilson[1]

[1]The Roslin Institute and R(D)SVS, Easter Bush Campus, Midlothian EH25 9RG, Edinburgh, UK; [2]Agriculture and Horticulture Development Board, Kenilworth, Warwickshire CV8 2TL, UK; [3]Animal and Plant Health Agency, Surrey KT15 3NB, UK; [4]Scotland's Rural College, Easter Bush Campus, Midlothian EH25 9RG, Edinburgh, UK; [*]enrique.sanchez-molano@roslin.ed.ac.uk

Abstract

Bovine tuberculosis (bTB) is a zoonotic disease primarily affecting cattle and leading to important economic losses. Eradication programmes in various countries focus on surveillance, movement restrictions and culling, with some considering also selective breeding. These breeding strategies focus on reducing animal susceptibility to the disease and thus, the potential of reducing disease transmission by selecting also animals with low infectivity remains unknown.

The first step in assessing this potential is to detect underlying genetic variation in infectivity. In the present study, the number of secondary cases attributed to a given index case in a breakdown was used as an infectivity phenotype, and a range of linear and generalised linear mixed sire models were used to estimate its genetic variance and heritability.

Substantial genetic variance was detected for this trait, suggesting that infectivity can be reduced via selective breeding, hence contributing to more efficient eradication strategies.

Introduction

Bovine tuberculosis (bTB) is a chronic disease caused by *Mycobacterium bovis*, a Gram-positive bacterium of the *Mycobacteriaceae* family. Although mainly found in cattle, it also affects many other mammals, including humans (Michel *et al*. 2010; Payeur 2014). In the UK, bTB is one of the major causes of economic losses in cattle industry, due to the costs of the testing programme and the compensation payments to farmers when animals are culled (Abernethy *et al.* 2013).

Given the costs associated with this disease, several programmes across different countries were set up to eradicate rather than to control bTB. These are mainly based on intensive test and culling regimes, and stringent biosecurity measures to limit the spread of the disease. Although in many cases these strategies reported successful results, in others, such as the UK, the disease has proven difficult to eradicate with these measures alone (Allen *et al.* 2018). Given this, the TB Advantage genetic index was introduced as a complementary strategy in the UK in 2016 (Banos *et al*. 2017). This index indicates the degree of resistance to bTB that a bull is expected to pass on to his offspring and thus, allowing to reduce bTB susceptibility via genetic selection. There is evidence for the presence of bTB super-spreaders, i.e. highly infectious animals responsible for transmitting the disease to many herd-members (O'Hare *et al*. 2014). Hence, a promising additional strategy to eradicate the disease may also be to consider infectivity as part of the selection index, allowing for the reduction of the propensity of infected animals to transmit the infection. Previous studies have demonstrated that including infectivity in the selection index could drastically accelerate reduction in risk and prevalence of disease (Tsairidou *et al*. 2018), however the genetic basis of bTB infectivity is still unknown.

Therefore, the purpose of this preliminary study was to assess the existence of underlying genetic variation for infectivity using real data from UK bTB breakdowns.

Materials & Methods

Breakdown data. Repeated bTB skin test records collected in 60-day intervals, from all animals in a herd, for 32,745 breakdowns, were available for the period between 1995 to 2017. From these data, only closed breakdowns were considered, and the first animal with a positive test result (reactor) or slaughtered after an inconclusive test (inconclusive reactor) was identified as a potential index case. An index case is defined as the first infected animal in a herd that transmits infection to its herd members. In this study, breakdowns with more than one index cases (reactors detected on the same first test date) were discarded. After quality control, the total number of breakdowns retained for analysis was 4,598, with each breakdown corresponding to one given index case.

Index cases with high infectivity would be expected to generate, on average, more secondary cases than index cases with low infectivity. Hence, for each breakdown, the response variable (numR) was calculated as the total number of secondary cases (non-index case reactors or slaughtered due to inconclusive test results) identified in a given period since the start of the breakdown (either 120 or 365 days). Based on previous estimates of the disease stage durations (O'Hare *et al*. 2021), the chosen periods were assumed to be sufficiently long for secondary cases to be detectable and sufficiently short to minimise bTB cases resulting from transmission between secondary cases.

Pedigree data. A full pedigree consisting of 910,170 records was available. From this data, two shorter pedigrees corresponding to the 4,598 index cases described above were extracted, comprising the last seven generations (58,753 records). These index cases corresponded to 2,544 different sires, of which 80.7% (2,054) were sires of at least two index cases.

Data analysis. The above data was analysed with sire models to estimate the genetic effect of the index cases on the total number of secondary cases per breakdown (numR) identified in either the first 120 or 365 days after the breakdown start date. A sire would be associated to index cases in different breakdowns, thus expected to account for potential across-herd genetic variation in susceptibility of the infected animals. Different distributions of the response variable were examined, thus estimating genetic variances and heritabilities in the latent scale.

Initial exploratory analyses were performed only for the 365 days period using ASReml (Gilmour *et al*. 2015), without including any fixed effects. However, given the skewed distribution of the number of secondary cases and the high proportion of breakdowns with zero secondary cases, the response variable was transformed to log(numR+1), which was assumed to be normally distributed.

A further analysis was performed through a generalised linear mixed model accounting for an over-dispersed Poisson distribution (de Villemereuil *et al*. 2016) of the response variable using ASReml (Gilmour *et al*. 2015). In this case, the response variable was transformed to numR+1, and the following covariates and fixed effects were included in the analysis: index case age in months at test, maximum herd size during breakdown, risk code (type of risk area), season at onset of breakdown and breakdown duration (in days).

One potential issue of using Poisson distributions (even if accounting for over-dispersion) is the fact that they account poorly for zero inflation. Therefore, additional distributions for the response variable were considered in the generalised linear mixed models.

A negative binomial distribution was firstly assumed, with the number of successes (φ) being 1 and, therefore, effectively resembling a geometric distribution. This analysis was done in

ASReml, with the response variable being transformed to numR+1, given the logarithmic link function and using all the previously described fixed effects and covariates.
An additional two-step model was also applied to the data, following the rationale of a Hurdle model (Mullahy 1986). In this analysis, after accounting for the above fixed effects and covariates, the data was assumed to follow a mixture of a Binomial and Poisson distributions. The response variable (numR) was firstly analysed in ASReml as a binary variable (presence or absence of secondary cases) and the probability of non-zero values (breakdowns with numR ≥ 1) was analysed following an over-dispersed Poisson model in ASReml.
A final model was tested using Markov Chain Monte Carlo techniques via the MCMCglmm package in R (Hadfield 2010). In this model, the response variable (numR) was assumed to follow a zero-inflated Poisson distribution (ZIP), and all previous fixed effects and covariates were included. The model was run for 200,000 iterations, plus 10,000 burn-in, with a thinning interval of 100. The prior assumed independent residual and random effect variances for the count and binary terms.

Results

The distribution of the number of secondary cases per breakdown in the first 365 days (Figure 1) was skewed, presenting an excess of zeros, and with a mean of 2.23 and variance of 34.37. About 0.8 % of the breakdowns presented more than 30 secondary cases during this interval, with the maximum being one breakdown with 108 secondary cases.

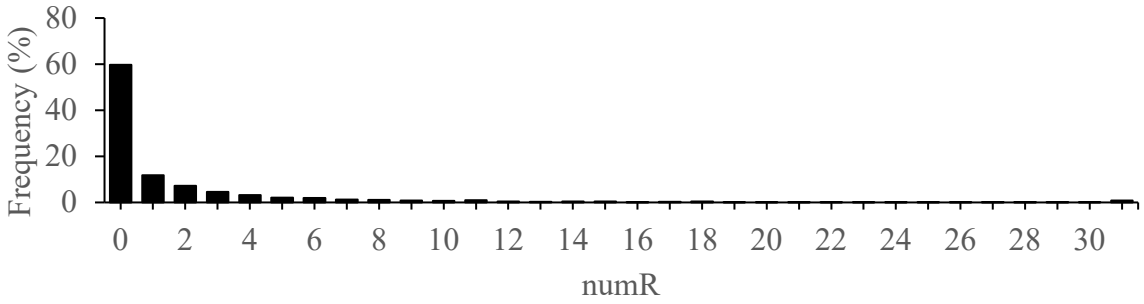

Figure 1. Distribution of the average count of secondary cases per sire.

Results for the estimates of genetic variance and heritabilities on the latent scale are presented in Table 1 for each of the two intervals considered. Heritability and genetic variances for the binomial part of the two-step models were not estimable.

Table 1. Heritabilities (h^2) and genetic variances (σ_a^2) in the latent scale for the different models and periods.

Distrib.[1]	Response variable	365 days		120 days	
		σ_a^2	h^2	σ_a^2	h^2
Normal	log(numR+1)	0.004 ± 0.017	0.006 ± 0.021	-	-
OP	log(numR+1)	1.657 ± 0.107	0.599 ± 0.038	1.169 ± 0.085	0.674 ± 0.047
NB	log(numR+1)	0.094 ± 0.048	0.144 ± 0.072	0.151 ± 0.060	0.214 ± 0.083
TS-P	numR≥1	1.582 ± 0.154	0.408 ± 0.044	1.475 ± 0.161	0.591 ± 0.068
ZIP	numR	0.157 ± 0.004	0.066 ± 0.002	0.188 ± 0.004	0.086 ± 0.002

[1] Studied distributions were: Normal, Over-dispersed Poisson (OP), Negative Binomial (NB), Poisson part of the two-step model (TS-P) and Zero-inflated Poisson (ZIP).

Discussion

Under the assumption of a normal distribution for the transformed response variable, none of the analyses resulted in a significant genetic variance estimate. This was clearly due to the lack of normality in the response variable and the high frequency of zero secondary cases. Similarly, the use of generalised linear mixed models assuming over-dispersed Poisson distributions resulted in significant but overestimated heritabilities, likely due to the lack of concordance of these distributions with the observed in the data.

Both the negative binomial or the zero-inflated Poisson distribution models derived significant and relatively reliable heritability estimates; these estimates are in line with those observed for resistance to bTB in previous studies (Banos *et al*. 2017) as well as with those expected for other fitness-related traits (Gibson and Dechow 2018).

These results showed consistently higher heritabilities for secondary cases in the first 120 days than for the first 365 days. This was probably the consequence of the effects of the increased strength of the genetic component (the smaller the timeframe, the more likely that the infection of secondary case was transmitted by the index case) and the lower total phenotypic variance.

This study presents a simple approach to estimate the genetic component and heritability of bTB infectivity, resulting in low but substantial significant heritabilities for the studied trait. However, further research is required to validate these results. Given the length of the different disease stages reported in recent studies (O'Hare *et al*. 2021), a more accurate definition of secondary cases attributed to the index case is needed. Furthermore, additional models must be tested and used to take advantage of the data available from breakdowns with more than one index cases and to incorporate transmission from all infected cows.

References

Abernethy D.A., Upton P., Higgins I.M., McGrath G., Goodchild A.V., *et al*. (2013) Vet Record 172: 312-312. https://doi.org/10.1136/vr.100969.

Allen A.R., Skuce R.A., and Byrne A.W. (2018) Front Vet Sci 5:109. https://doi.org/10.3389/fvets.2018.00109.

Banos G., Winters M., Mrode R., Mitchell A.P., Bishop S.C., *et al*. (2017) J Dairy Sci 100:1272–1281. https://doi.org/10.3168/jds.2016-11897.

de Villemereuil P., Schielzeth H., Nakagawa S., and Morrissey M. (2016). Genetics, 204(3):1281–1294. https://doi.org/10.1534/genetics.115.186536.

Gibson K.D. and Dechow C.D. (2018) J Dairy Sci 101:1251–1257. https://doi.org/10.3168/jds.2017-13041.

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J., and Thompson R. (2015). ASReml User Guide Release 4.1 Functional Specification. VSN International Ltd, Hemel Hempstead, UK.

Hadfield J.D. (2010) J Stat Softw 33(2):1-22. https://doi.org/10.18637/jss.v033.i02

Michel A.L., Muller B., and van Helden P.D. (2010) Vet Microbiol 140:371–81. https://doi.org/10.1016/j.vetmic.2009.08.029.

Mullahy J. (1986) J Econom 33:341–365. https://doi.org/10.1016/0304-4076(86)90002-3.

O'Hare A., Orton R.J., Bessell P.R., and Kao R.R. (2014) Proc Biol Sci 281(1783): 20140248. https://doi.org/10.1098/rspb.2014.0248

O'Hare A., Balaz D., Wright D.M., McCormick C., McDowell S., et al. (2021) PLoS Comput Biol 17(6): e1009005. https://doi.org/10.1371/journal.pcbi.1009005.

Payeur J.B. (2014). Mycobacterium. In: Batt C.A. and Tortorello M.L. (eds) Encyclopedia of Food Microbiology. Academic Press, Cambridge, US, pp. 841-853.

Tsairidou S., Allen A., Banos G., Coffey M., Anacleto O., *et al*. (2018) Front Vet Sci 5:310. https://doi.org/10.3389/fvets.2018.00310.