OUTBREAKS
Challenges of investigating a large food-borne norovirus outbreak across all branches of a restaurant group in the United Kingdom, October 2016
Mari Morgan, Vicky Watts, David Allen, Daniele Curtis, Amir Kirolos, Neil Macdonald, Deb Morgan, Ayoub Saei, James Sedgwick, Janet Stevenson, Deborah Turbitt, Roberto Vivancos, Catriona Waugh, Chris Williams and Valerie Decraene

REVIEW
A new Borrelia on the block: Borrelia miyamotoi – a human health risk?
Sally Cutler, Muriel Vayssier-Taussat, Agustín Estrada-Peña, Aleksandar Potkonjak, Andrei Daniel Mihalca and Hervé Zeller

RESEARCH
Disease burden of varicella versus other vaccine-preventable diseases before introduction of vaccination into the national immunisation programme in the Netherlands
Alies van Lier, Brechje de Gier, Scott A McDonald, Marie-Josée J. Mangen, Maarten van Wijhe, Elisabeth A.M. Sanders, Mirjam E. Kretzschmar, Hans van Vliet and Hester E. de Melker

Molecular identification of the source of an uncommon tularemia outbreak, Germany, autumn 2016
Daniela Jacob, Kristin Köppen, Aleksandar Radonić, Berit Haldemann, Philipp Zanger, Klaus Heuner and Roland Grunow

Genetic diversity and delineation of Salmonella Agona outbreak strains by next generation sequencing, Bavaria, Germany, 1993 to 2018
Alexandra Dangel, Anja Berger, Ute Messelhäußer, Regina Konrad, Stefan Hörmansdorfer, Nikolaus Ackermann and Andreas Sing
Challenges of investigating a large food-borne norovirus outbreak across all branches of a restaurant group in the United Kingdom, October 2016

Mari Morgan1,2,3, Vicky Watts3,4, David Allen6,7,8, Daniele Curtis9, Amir Kirolos10, Neil Macdonald9, Ellie Maslen9, Deb Morgan12, Ayoub Saei13, James Sedgwick9, Janet Stevenson10, Deborah Turbitt5, Roberto Vivancos5, Catriona Waugh9, Chris Williams1, Valerie Decraene5

1. Health Protection, Public Health Wales NHS Trust, Cardiff, United Kingdom
2. European Programme for Intervention Epidemiology Training, Stockholm, Sweden
3. These authors share first authorship
4. United Kingdom Field Epidemiology Training Programme, Public Health England, London, United Kingdom
5. Field Service - Epidemiology, National Infection Service, Public Health England, Liverpool, United Kingdom
7. National Institute for Health Research Health Protection Research Unit in Gastrointestinal Infections, United Kingdom
8. Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, United Kingdom
10. Department of Public Health and Health Policy, NHS Lothian, Edinburgh, Scotland
12. Incidents & Resilience Team, Food Standards Agency, London, United Kingdom

Correspondence: Mari Morgan (mari.morgan@wales.nhs.uk)

During October and November 2016, over 1,000 customers and staff reported gastroenteritis after eating at all 23 branches of a restaurant group in the United Kingdom. The outbreak coincided with a new menu launch and norovirus was identified as the causative agent. We conducted four retrospective cohort studies; one among all restaurant staff and three in customers at four branches. We investigated the dishes consumed, reviewed recipes, interviewed chefs and inspected restaurants to identify common ingredients and preparation methods for implicated dishes. Investigations were complicated by three public health agencies concurrently conducting multiple analytical studies, the complex menu with many shared constituent ingredients and the high media attention. The likely source was a contaminated batch of a nationally distributed ingredient, but analytical studies were unable to implicate a single ingredient. The most likely vehicle was a new chipotle chilli product imported from outside the European Union, that was used uncooked in the implicated dishes. This outbreak exemplifies the possibility of rapid spread of infectious agents within a restaurant supply chain, following introduction of a contaminated ingredient. It underlines the importance of appropriate risk assessments and control measures being in place, particularly for new ingredients and ready-to-eat foods.

Background
Norovirus is the predominant cause of acute gastroenteritis worldwide [1], responsible for approximately one fifth of all cases [2]. In the United Kingdom (UK) there are estimated to be 3 million sporadic episodes annually [3] and although typically mild and self-limiting [4], financial costs to patients, health services and businesses are significant [5-7].

Transmission is via the faecal-oral route [1,8], through contaminated food or water consumption or direct contact with infected persons or contaminated environments [8]. Outbreaks have been linked to contaminated shellfish [9,10], fresh produce [11-14] and ready-to-eat foods (often via infected food handlers) [15-18] frequently in restaurant settings [9,18,19].

Outbreak detection
On 27 October 2016, Public Health England (PHE) received reports of diarrhoea and vomiting affecting 10 staff at one London branch of a restaurant group comprising 23 branches across England, Wales and Scotland (none in Northern Ireland). In the following
days, customers from all branches and staff members from 22 branches reported gastrointestinal symptoms to the company head office.

Initial investigations revealed that on 26 October 2016, a new menu was introduced at all branches with over 70 dishes, 12 of which had not been served before. There were no large-scale changes in personnel or management which coincided with the outbreak.

On 1 November 2016, PHE convened an incident management team (IMT) with representation from the other UK public health agencies (Public Health Wales and Health Protection Scotland) and a joint epidemiological investigations team was established.

Here, we describe the epidemiological investigations undertaken to identify the vehicle and source of the outbreak, implement appropriate control measures and the challenges we faced in the joint investigations.

**Methods**

**Epidemiological studies**

**Descriptive**

The company head office compiled data on customers and staff reporting gastrointestinal illness, which they shared with environmental health departments and the IMT. Staff illness was reported by symptom onset date and customer illness by date of restaurant visit. All data, without the application of any case definitions, were used to provide a crude description of the outbreak within and across all 23 branches.
Given all 23 branches were affected by gastrointestinal illness – often spread through infected food handlers and cross-contamination, it was hypothesised that a centrally distributed ingredient had become contaminated or there was a change in how the ingredient was prepared leading to contamination. To determine whether the implicated food items were the same between staff and customers of different branches, four retrospective cohort studies were conducted, one among staff of the whole restaurant group and three in customers of four branches.

The staff cohort study population included all staff employed at any UK branch between 22 and 31 October 2016. Staff tasting sessions of the 12 new menu dishes were held 24–26 October, before the public menu launch on 26 October. Staff could also eat from the main menu during their shifts. To examine the effect of exposure to food items at tasting sessions and minimise potential bias introduced through secondary transmission, in the analytical study, we included ‘early onset cases’ defined as study population members who developed vomiting or diarrhoea between 24 and 28 October and excluded ‘late onset cases’, who developed symptoms from 29 October. Individuals with a history of gastrointestinal illness since 17 October, or whose household contacts had gastrointestinal illness, were excluded.

We conducted customer cohort studies in: (i) Cardiff (branch 12), (ii) Edinburgh (branch 20) and (iii) London (branches 13 and 22), a convenience sample based on availability of epidemiological staff and customer contact details. Studies were conducted by public health agencies of the respective administrations (Public Health Wales, National Health Service (NHS) Lothian, PHE Field Service) and analysed separately. Data were combined and analysed as a single cross-administration customer cohort study, using standardised case definitions, to determine whether exposures associated with illness were common across different branches. Customer contact details were obtained from branch booking lists for the periods of interest; we did not attempt to identify customers who did not book.

Customer cases were defined as persons who ate at one of the four branches who developed severe diarrhoea (three or more episodes in 24 hours) or vomiting or two other symptoms (mild diarrhoea (less than three episodes in 24 hours), bloody stools, nausea, fever, stomach cramps and headache) within 72 hours of eating at a branch. Time frames differed by branch.
Any staff members who had dined as customers were excluded. Customers with household members reporting gastrointestinal illness within 7 days before symptom onset were also excluded.

**Questionnaire data**

We developed separate but similar online questionnaires for staff and each customer study. All included questions on demographics, dining dates, symptoms, symptom onset date and the available menu items.

The staff questionnaire included questions on dishes from both the tasting and full menu. Menu items consumed in the 72 hours before symptom onset were requested for staff cases and we requested those consumed between 24 and 28 October for non-cases. The questionnaire was distributed to all staff members via email from the company management on PHE’s behalf. Staff could complete the questionnaire between 4 and 10 November 2016 and a reminder was sent after 3 days. Staff symptoms collected during the staff cohort study were plotted by date of onset.

Customers who went to the Cardiff, Edinburgh or London (branch 22) between 26 and 28 October and to London (branch 13) between 27 and 29 October, were asked to complete the questionnaire. Timings were based on illness reports to the company and voluntary branch closure dates. Customers were contacted by telephone and email and, on consenting to participate, they were sent a link to complete the online questionnaire. They were also asked to forward it to their co-diners. Customer questionnaires were collected over ca 2 weeks and no reminder was sent.

**Ethical statement**

Ethical approval was not required as in the UK, public health agencies are able to access and use personal identifiable information for communicable disease outbreak investigations in the public interest. How the data would be utilised was outlined in the questionnaires and completion of the questionnaire was considered as implied consent.

**Data analysis**

We estimated risk ratios (RR) and odds ratios (OR) as measures of association between food items consumed and being a case. We used generalised linear models to identify factors independently associated with being a case.

In the staff cohort study, exposures for tasting and full menu items were analysed separately. Food items associated with illness (RR > 1 and p < 0.1) were included in Poisson regression models with robust standard errors, constructed separately for the tasting and full menus, using a backwards stepwise approach.

For the combined customer cohort study, menu items associated with illness (RR > 1.5, eaten by at least eight cases and with 95% confidence intervals (CI) that did not include 1 were included in a logistic regression model using a backwards stepwise approach. To investigate the influence of heterogeneity between customer studies conducted at different branches, we developed an additional mixed effects logistic regression model, to estimate the association between menu items consumed and illness.

Guided by results of the multivariable analyses, we created combined variables of dishes according to common ingredients or kitchen preparation area (staff only), based upon information gathered from restaurant visits, chef interviews and national recipes provided by the company management. We repeated univariable and multivariable analysis with the combined variables.

Analysis was conducted in Stata v14.2 (StataCorp, College Station, Texas, United States) and R v3.2.3 (R Foundation for statistical computing, Vienna, Austria).

**Other investigations**

Environmental health investigations were coordinated locally for each restaurant branch by environmental health officers (EHOs), who visited branches, reviewed food hygiene procedures, took food samples and coordinated collection of faecal samples. The Food Standards Agency (FSA) conducted food chain investigations into products highlighted by the IMT including fish, shellfish, coriander, radish, chipotle and chicken products. Initial microbiological investigations of human samples were conducted in local microbiology laboratories and norovirus-positive specimens referred to the PHE Virus Reference Department (London, England) for characterisation. Food samples were tested against national standards [20] by the PHE Food, Water and Environmental Microbiology Laboratories in England.

**Results**

**Epidemiological studies**

**Descriptive**

In total, 287 staff members and 825 customers reported gastrointestinal illness to the company. All branches had customer cases and all but one had staff cases reported. The first report of illness on 19 October was in a staff member, a growing number of staff reported illness until 25 October, when 16 staff from six branches reported illness. Staff reports peaked on 28 October, with 45 staff affected across 16 branches. The last reported illness onset date was 11 November 2016. Customers first reported illness following consumption of food at one branch on 15 October. Customer reports peaked on 29 October, with 210 customers affected across 20 branches. The last customer to report illness dined on 10 November 2016. Although there was variation between branches (Figure 1), the overall interquartile date range for staff and customer illness was 27 to 30 October.
Fifty-eight percent (589/1,029) of staff completed the questionnaire, of whom a fifth (21%, 125/587) were categorised as cases. The age range of cases was 16-55 years; 57% were female. Cases were reported from 21 of 23 branches; there was large variation in attack rates between branches (0–40%). There was a sharp reduction in case numbers after 1 November 2016.

The analytical study was restricted to 73 early onset cases (Figure 2), giving an attack rate of 12% (73/587). Three dishes were associated with illness in univariable analysis of each menu (Table 1) and salmon tostadas was associated with illness on the tasting menu. In multivariable analyses, illness was independently associated with consumption of salmon tostadas (RR: 2.17; 95% CI: 1.43–3.28) on the tasting menu and with chicken wings (RR: 1.75; 95% CI: 1.05–2.91) on the full menu (Table 1). Ingredient analysis showed that consuming dishes containing chipotle product A or B was independently associated with illness on analysis of each menu (RR: 2.17; 95% CI: 1.43–3.28) on the tasting menu and with chicken wings (RR: 1.75; 95% CI: 1.05–2.91) on the full menu (Table 1). Being female was also independently associated with illness in the multivariable models for each menu and in the ingredient-based analysis (Table 1).

Customer cohort studies

Of 159 customer responses in the combined customer cohort study (28 from Cardiff, 94 from Edinburgh and 37 from London), 37 (23%) were male and the age ranged from seven to 65 years. There were 58 (36%) confirmed cases; the attack rate ranged from 21% to 46%. Although seven menu items were associated with gastrointestinal illness in the univariable analysis (Table 2), chicken tostadas that were consumed by 57% (33/58) of cases, was the only menu item independently associated with illness in the multivariable analysis (OR: 20.65; 95% CI: 7.24–58.89). This was consistent with results of the three individual customer studies (Supplementary Table S1). Analysis of dishes containing the component ingredients of chicken tostadas identified that the ready-to-eat poached chicken (OR: 4.11; 95% CI: 1.95–8.66) and chipotle mayo (OR: 2.27; 95% CI: 1.06–4.88) were independently associated with illness. Both items were eaten by 76% (44/58) of cases. Univariable analysis of chilli components

### Table 1

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Number of cases exposed</th>
<th>Univariable analysis</th>
<th>Multivariable analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RR 95% CI p value</td>
<td>RR 95% CI p value</td>
</tr>
<tr>
<td>Winter tasting menu Model 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Menu item a</td>
<td>Salmon sashimi tostada</td>
<td>36</td>
<td>2.22 1.46–3.38 0.000</td>
</tr>
<tr>
<td></td>
<td>Hibiscus glazed wings b</td>
<td>43</td>
<td>2.01 1.30–3.09 0.001</td>
</tr>
<tr>
<td></td>
<td>Chicken taquito b</td>
<td>33</td>
<td>1.73 1.13–2.64 0.011</td>
</tr>
<tr>
<td></td>
<td>Huitlacoche empanada b</td>
<td>37</td>
<td>1.58 1.03–2.41 0.034</td>
</tr>
<tr>
<td>Demographics</td>
<td>Female</td>
<td>NA</td>
<td>NA NA NA</td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ingredient</td>
<td>Chipotle product A or B</td>
<td>50</td>
<td>2.27 1.42–3.60 0.000</td>
</tr>
<tr>
<td>Demographics</td>
<td>Female</td>
<td>NA</td>
<td>NA NA NA</td>
</tr>
<tr>
<td>Full menu Model 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Menu item a</td>
<td>Hibiscus glazed wings</td>
<td>37</td>
<td>2.11 1.39–3.22 0.000</td>
</tr>
<tr>
<td></td>
<td>Chicken taquito</td>
<td>28</td>
<td>1.94 1.26–2.98 0.003</td>
</tr>
<tr>
<td></td>
<td>Huitlacoche empanada</td>
<td>30</td>
<td>1.61 1.05–2.47 0.030</td>
</tr>
<tr>
<td></td>
<td>Salmon sashimi tostada</td>
<td>22</td>
<td>1.41 0.89–2.42 0.141</td>
</tr>
<tr>
<td>Demographics</td>
<td>Female</td>
<td>NA</td>
<td>NA NA NA</td>
</tr>
<tr>
<td>Model 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ingredient</td>
<td>Chipotle product A or B</td>
<td>52</td>
<td>2.03 1.26–3.27 0.003</td>
</tr>
<tr>
<td>Demographics</td>
<td>Female</td>
<td>NA</td>
<td>NA NA NA</td>
</tr>
</tbody>
</table>

CI: confidence interval; RR: risk ratio; NA: not applicable.

a Only menu items with a RR > 1.0 and p < 0.1 are displayed.

b Excluded from final multivariable model.

c RR > 1.0 and p > 0.1. result displayed because menu item was associated with illness on analysis of the other menu.
of the chipotle mayo showed both chipotle product A (RR: 2.12; 95% CI: 1.27–3.53) and chipotle product C (RR: 2.06; 95% CI: 1.24–3.44) were associated with illness; further differentiation in a multivariable model was not possible because of their frequent combined use.

The mixed effects logistic regression model showed that study site heterogeneity did not significantly influence the menu items associated with illness (estimated coefficient for chicken tostadas: 4.72; 95% CI: 1.81–7.63; p = 0.0015).

Other investigations and control measures
Norovirus genogroup II.6 (GII.6) was identified from 30 of 48 samples from staff. Standard indicator organisms were not detected in any of the food samples collected. All branches were compliant with standard hygiene regulations and EHOs reported standards to be satisfactory.

Recipes were the same for all restaurant branches. Chipotle chilli was the only ingredient in common between the two dishes implicated in the multivariable analysis of the staff cohort study and the one dish implicated by the combined customer study. Chipotle chilli was obtained from different chipotle chilli products in the three dishes. Recipes for both salmon and chicken tostadas included chipotle mayo, which contained uncooked tinned chipotle chilli product A and dried chipotle chilli product C. The recipe for the chicken wing glaze included paste chipotle chilli product B. EHOs identified that in some branches product B had been labelled as product A (Figure 3).

Trace-back investigations identified that chipotle product B was newly imported from outside the European Union. Cook-chill and food safety management for the two chicken suppliers were satisfactory and the product was also distributed to other restaurants in the UK. A central kitchen in London supplied some components of dishes for London restaurants only. Restaurant service kitchens in all branches were divided into three sections; all dishes implicated by multivariable analyses (staff and customer studies) were prepared in the salad section.

The company voluntarily closed 16 branches in the UK (the first on 26 October) and discarded fresh and partially used produce. In addition, the branches carried out deep cleaning of their restaurants to remove any norovirus contamination, including environmental fogging. Staff were advised to stay off work for 72 hours following their last gastrointestinal symptom and were offered paid sickness absence to encourage policy adherence. All potential vehicles of norovirus transmission identified by the epidemiological studies were removed from the menu.

Discussion
We describe the largest norovirus restaurant outbreak recorded in the UK to date. Coincident gastrointestinal

---

### Table 2

Univariable and multivariable analysis of menu items and ingredients eaten by customers, in combined customer cohort studies in norovirus outbreak in a restaurant group, United Kingdom, 2016 (n=58)

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Univariable analysis</th>
<th>Multivariable analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RR</td>
<td>95% CI</td>
</tr>
<tr>
<td><strong>Menu items</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chicken tostada</td>
<td>4.54</td>
<td>2.92–7.06</td>
</tr>
<tr>
<td>Pork burrito</td>
<td>2.08</td>
<td>1.35–3.18</td>
</tr>
<tr>
<td>Chicken taquito</td>
<td>1.82</td>
<td>1.16–2.85</td>
</tr>
<tr>
<td>Pork taco</td>
<td>1.73</td>
<td>1.15–2.60</td>
</tr>
<tr>
<td>Chicken taco</td>
<td>1.68</td>
<td>1.10–2.55</td>
</tr>
<tr>
<td>Chorizo quesadilla</td>
<td>1.58</td>
<td>1.01–2.47</td>
</tr>
<tr>
<td>Chicken quesadilla</td>
<td>1.57</td>
<td>1.04–2.38</td>
</tr>
<tr>
<td><strong>Ingredients</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ready-to-eat chicken</td>
<td>2.88</td>
<td>1.72–4.81</td>
</tr>
<tr>
<td>Chipotle mayo</td>
<td>2.17</td>
<td>1.30–3.62</td>
</tr>
<tr>
<td>Chipotle product A</td>
<td>2.12</td>
<td>1.27–3.53</td>
</tr>
<tr>
<td>Chipotle product C</td>
<td>2.06</td>
<td>1.24–3.44</td>
</tr>
<tr>
<td>Chipotle product B</td>
<td>1.10</td>
<td>0.50–2.43</td>
</tr>
</tbody>
</table>

CI: confidence interval; NA: not applicable; OR: odds ratio; RR: risk ratio.

a Total number of cases = 58.
b Only menu items with a RR > 1.5, CI that did not cross 1 and eaten by at least eight cases are displayed.
c Excluded from final multivariable model.
d Excluded because of collinearity in multivariable model.
e Not significant in univariable analysis, so excluded from multivariable model.

---
illness in multiple branches suggested a point source outbreak, although later in the outbreak staff cases were likely due to secondary transmission. Concurrence with a national menu change and the involvement of all 23 branches in the restaurant group, suggested the vehicle of transmission was a nationally distributed item that had either been newly introduced to the menu or had been modified for use in the new menu. The sharp reduction in cases suggested the vehicle had either been withdrawn, destroyed or used. There were no similar reports of illness associated with other UK food outlets, suggesting the contaminated product batch was unique to this restaurant group.

Three separate chipotle products were ingredients of dishes associated with gastrointestinal illness. Chipotle product A was a tinned constituent ingredient of the chipotle mayo, independently associated with illness in the combined customer study and served uncooked with both salmon and chicken tostadas. However, it was not a new ingredient and it was biologically implausible as the vehicle since norovirus would be inactivated by the canning process [21]. Environmental health investigations identified a similar product known as chipotle product B, which had been labelled as chipotle product A in some branches; this was a constituent ingredient of the chicken wings implicated in the staff cohort study. It had been newly imported from outside the European Union for the new menu, was not tinned and was not cooked during initial processing. It is plausible, therefore, that chipotle product B may have been mistakenly used in the chipotle mayo in place of chipotle product A in some branches and this may explain the variation in attack rate. Although implicated in customer cohort studies, chipotle product C was not considered a likely vehicle as it was boiled for 15 minutes before use, which would have destroyed norovirus.

Customer cohort studies observed a stronger association between gastrointestinal illness and consumption of dishes containing poached, ready-to-eat, vacuum-packed chicken from a new supplier, which had been introduced for the menu change. This product was only used once the chicken from the previous supplier had run out – possibly explaining the staggered symptom onsets in branches. This chicken product was also used in other dishes, but was only served without further reheating in the chicken tostadas. The food chain investigations found no evidence to implicate this chicken product as a vehicle for norovirus transmission. In addition, it was supplied to other UK restaurants, but there were no reports of similar outbreaks or reports of gastrointestinal illness from staff responsible for processing the chicken (data not shown).

There are plausible routes of contamination for chipotle product B and the ready-to-eat chicken, before they were received and distributed around the restaurant group. Fresh produce has been implicated in multiple norovirus outbreaks [11-14]. Food items can become contaminated during cultivation, harvesting or processing, usually as a result of contact with contaminated sewage or infected food handlers. Ready-to-eat meat products have also been implicated as vehicles for norovirus outbreaks [22] contaminated directly by food handlers during processing [23]; the capacity for contamination via slicing equipment has also been demonstrated [22]. In this investigation, we were unable to test any of the potential vehicles identified for norovirus, as accredited tests are only available for limited food items, not implicated in this outbreak. Development of sensitive laboratory methods for testing food specimens for viruses during outbreaks would be valuable in future investigations.

Introduction of contaminated fresh produce [24-27] or ready-to-eat foods [28] from a single supplier has been implicated in several food-borne outbreaks. All dishes associated with gastrointestinal illness in the multivariable analysis were prepared in the salad section, meaning that cross-contamination there could have played a role in transmission. This finding may explain the variation found between the customer and staff studies, as well as between the customer cohort studies conducted in different branches. This finding may also partially explain why staff cases were almost twice as likely as non-cases to be female; studies have
shown that women are more likely to choose salad items than men [29].

This outbreak and others affecting restaurant chains [24,30,31] highlight the speed with which pathogens can spread over wide geographical areas, when one or more contaminated ingredients enter a national restaurant supply chain and appropriate risk assessment and controls are not in place. The public health implications could have been much more serious had this outbreak been caused by a more virulent pathogen than norovirus, which is generally a mild and self-limiting illness. Both food products implicated by the epidemiological investigations were highlighted to the restaurant group management and they have since reported working with suppliers to minimise the risk of further outbreaks.

**Challenges**

There were many challenging aspects to the epidemiological investigations. The descriptive data for the entire outbreak was collected by the company and had no case definitions applied. This means that secondary cases and gastrointestinal illness unrelated to the outbreak may have been included as cases. Reporting of gastrointestinal illness may also have been influenced by high media coverage, potentially explaining reports of illness before the introduction of the new menu. The menu included a large number of dishes and was complex; dishes had many ingredients and garnishes and ingredients were often common to multiple menu items. In addition, the menu was designed for customers to share dishes, many with similar names, which could have made it more difficult for customers to distinguish and accurately recall what they had eaten. The high media interest and publicity regarding the outbreak may have encouraged customers to exaggerate symptoms, potentially misclassifying non-cases as cases. Introducing paid sickness absence may similarly have inflated staff case numbers.

The incident management team was supplied with standardised recipe cards that detailed the ingredients used, preparation instructions and photos of the dishes, which supported a recipe-based cohort study design (used in the fenugreek sprout *Escherichia coli* O104:H4 outbreak in Germany [32]). However, within the 23 branches, there may have been local undocumented variation in how a dish was prepared within the branches, there may have been local differences in practice.

**Conclusions**

This outbreak demonstrates that an entire restaurant group can be affected within a short time frame by the introduction of a contaminated ingredient. The investigation highlights the challenges in identifying the vehicle of transmission from a large, complex menu with multiple ingredients used in numerous dishes with possible undocumented variation between branches. In hindsight, a more coordinated approach to conducting the epidemiological studies may have achieved a more coherent outcome. For example, a single study incorporating both staff and customers from across the restaurant group would have been more logistically challenging to set up, but results may have been easier to interpret. This is a learning point for cross-UK outbreak management for the future.

We recommend that multi-branch restaurants with central suppliers and kitchens are vigilant to the possibility of contaminated ingredients entering their supply chain and the potential for rapid spread of pathogens. Food business operators should ensure that appropriate hazard analysis and critical control point processes are in place, particularly for new ingredients and ready-to-eat foods and consider the potential for cross-contamination within preparation areas in risk assessments.

**Acknowledgements**

We would like to thank: the company for cooperating with the investigation, collecting data on the overall outbreak and sharing this with the IMT; Environmental health departments in Edinburgh, Cardiff and throughout England, who visited individual restaurant branches and advised on control measures; PHE Virus Reference Department, which conducted testing for norovirus and genotyping of the stool samples; Edinburgh Scientific Services, City of Edinburgh Council, who tested food samples in Scotland; PHE Food, Water and Environmental Microbiology Laboratory, which analysed the food samples; Sangeeta Rana, National Infection Service Field Service, Public Health England, who contributed to the customer analytical studies in London; Staff and customers of the restaurant group for completing the questionnaires; FETP/EPIET coordinators for reviewing the manuscript; Paul Cleary and Thomas Inns, Field Service-Epidemiology, National Infection Service, Public Health England, who advised on the multivariable analysis of the staff study; the Gastrointestinal, Emerging and Zoonoses Infections section, National Infection Service, Public Health England, for reviewing the manuscript; Members of the incident management team and epidemiological investigations cell for their input into the investigation.

**Conflict of interest**

None declared.

**Authors’ contributions**

Mari Morgan and Vicky Watts conducted epidemiological investigations and wrote the manuscript. All authors reviewed and commented on the manuscript. Chris Williams, Valérie Decraene and Roberto Vivancos provided epidemiological support for the investigations in terms of study design,


License, supplementary material and copyright

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence and indicate if changes were made.

Any supplementary material referenced in the article can be found in the online version.

This article is copyright of the authors or their affiliated institutions, 2019.
A new Borrelia on the block: *Borrelia miyamotoi* – a human health risk?

Sally Cutler¹, Muriel Vayssier-Taussat², Agustín Estrada-Peña³, Aleksandar Potkonjak⁴, Andrei Daniel Mihalca⁵, Hervé Zeller⁶

1. School of Health, Sport & Bioscience, University of East London, London, United Kingdom
2. INRA, UMR BIPAR INRA, ENVA, Anses, Maisons-Alfort, France
3. Department of Animal Health, Faculty of Veterinary Medicine, University of Zaragoza, Spain
4. Department of Veterinary Medicine, Faculty of Agriculture, University of Novi Sad, Serbia
5. Department of Parasitology and Parasitic Diseases, University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca, Romania
6. European Centre for Disease Prevention and Control, Solna, Sweden

Correspondence: Sally J. Cutler (s.cutler@uel.ac.uk)

Citation style for this article:

Article submitted on 04 Apr 2018 / accepted on 11 Mar 2019 / published on 02 May 2019

Background: *Borrelia miyamotoi* clusters phylogenetically among relapsing fever borreliae, but is transmitted by hard ticks. Recent recognition as a human pathogen has intensified research into its ecology and pathogenic potential. Aims: We aimed to provide a timely critical integrative evaluation of our knowledge on *B. miyamotoi*, to assess its public health relevance and guide future research.

Methods: This narrative review used peer-reviewed literature in English from January 1994 to December 2018. Results: *Borrelia miyamotoi* occurs in the world’s northern hemisphere where it co-circulates with *B. burgdorferi* sensu lato, which causes Lyme disease. The two borreliae have overlapping vertebrate and tick hosts. While ticks serve as vectors for both species, they are also reservoirs for *B. miyamotoi*. Three *B. miyamotoi* genotypes are described, but further diversity is being recognised. The lack of sufficient cultivable isolates and vertebrate models compromise investigation of human infection and its consequences. Our understanding mainly originates from limited case series. In these, human infections mostly present as influenza-like illness, with relapsing fever in sporadic cases and neurological disease reported in immunocompromised patients. Unspecific clinical presentation, also occasionally resulting from Lyme- or other co-infections, complicates diagnosis, likely contributing to under-reporting. Diagnostics mainly employ PCR and serology. *Borrelia miyamotoi* infections are treated with antimicrobials according to regimes used for Lyme disease.

Conclusions: With co-infection of tick-borne pathogens being commonplace, diagnostic improvements remain important. Developing in vivo models might allow more insight into human pathogenesis. Continued ecological and human case studies are key to better epidemiological understanding, guiding intervention strategies.

Introduction

*Borrelia miyamotoi*, which belongs to the relapsing fever clade within the *Borrelia* genus, was first described in 1994 with its detection in *Ixodes persulcatus* ticks in Japan [1]. It was named after Professor Kenji Miyamoto who initially reported this spirochaete from Hokkaido, Japan. Its potential to cause human disease was not realised until 2011 when Platonov and colleagues described a series of cases of *B. miyamotoi* infection in Russia [2].

It is now established that this spirochaete has a global distribution and co-circulates with the related agent of Lyme borreliosis (*B. burgdorferi* sensu lato (s.l.)), which uses the same tick species as vectors, albeit at a lower frequency [3-9]. Similar to the Lyme borreliae, where different tick species endemic to specific regions of the globe serve as vectors, *B. miyamotoi* is found in multiple tick species that reside constrained by compatible geo-ecological habitats [2-4,10-13].

The objectives of this review were threefold: (i) to collate the rapid expansion of research findings on *B. miyamotoi* and its ecological interactions; (ii) to review the public health significance of *B. miyamotoi* and to (iii) highlight knowledge gaps in our understanding of this microbe and its importance as a human pathogen, thus focussing direction for future research.

Methods

We performed a non-systematic narrative literature review. Literature searches were thus not fully exhaustive. Reports relating to the search term ‘*Borrelia miyamotoi*’ published in English and indexed in biomedical databases including EBSCO (Academic Search Complete), Scopus and Science Direct were sought.
(Figure 1). These dated from the first description of *B. miyamotoi* in January 1994 to December 2018. Duplicate records were removed and articles were further screened, first by reading the titles and abstracts, then the full reports. Prioritisation to those papers that contributed original knowledge to our understanding was given. Supplementary literature was used to further support discussions beyond primary searches where justifiable (Figure 1).

**Results**

**Epidemiology**

*Borrelia miyamotoi* is being increasingly documented from the world’s northern hemisphere. This spirochaete has been recorded in Canada as well on the east and west coasts of the United States (US). It has also been observed in numerous European countries (including the Czech Republic, Denmark, Estonia, France, Germany, Netherlands, Norway, Poland, Romania, Sweden and Switzerland) as well as in Russia through to Japan [3,10,14,15].

Surveys to detect *B. miyamotoi* have tended to reflect local research interest rather than being systematic epidemiological studies. These have however established that endemic areas for *B. miyamotoi* overlap with those for Lyme borreliae (*B. burgdorferi* s.l.), and during tick surveys, specimens co-infected with both spirochaetes have been identified [4,11,16]. In several countries, prevalence studies based upon individual ticks infected with borreliae have noted lower rates of ticks infected with *B. miyamotoi* than with Lyme-associated borreliae [5,17].

Among 20 studies found in this review, from countries reporting clinical cases and *B. miyamotoi* prevalence in ticks, infection rates in ticks ranged from 0.02 to 6.4%, although most studies reported a range between 1 and 2% (Table 1) [3,5,15,18,19]. Nevertheless, pockets of higher infection rates have been described. In Napa County, California, US, for example, 15.4% of adult ticks (10/65 *I. pacificus*) studied were infected with *B. miyamotoi* [3] compared with a background infection level of 1.4% (44/3,255) of nymphs of this same species [19]. Moreover, in a study in Kurgan, Russia, 16% of ticks (26/162 *I. persulcatus*) were found infected [2]. These rates might reflect either hyperendemic areas or local efficient transmission events.

In some countries, such as Mongolia, where relatively high levels of tick infections with *B. miyamotoi* (4.5%; 48/1,069 *I. persulcatus*) have been detected, no human cases were reported [10]. This was similar in the town of Hannover, Germany, where a tick-infection prevalence of 8.9% (45/505 *I. ricinus*) was reported but no infections in people [20]. On the other hand, on Hokkaido Island, Japan, where, depending on the tick species, 2% (71/3,532 *I. persulcatus*) and 4.3% (5/117 *I. pavlovskyi*) of ticks were infected, human cases did occur [4]. Human cases were also observed in the Irkutsk region, Russia, where an overall 2.9% prevalence of tick infection has been estimated [21] (Table 1). In Yekaterinburg and Izhevsk, where clinical cases were first described, tick (*I. persulcatus*) infection prevalence rates of 0.9% (4/442) and 6.3% (25/394) were respectively found. Estimated incidence of human infection with *B. miyamotoi* in Yekaterinburg is likely to be 1 per 100,000 per year, accounting for a quarter of tick-borne borreliosis cases within this province [2,22].

**Phylogeny and diversity of Borrelia miyamotoi**

The taxonomic definition of the genus *Borrelia* is currently debated with the suggestion that the Lyme disease associated members be removed from this genus, *Borreliaell*a [23]. This suggestion has not been met with universal approval, with some experts suggesting this division is premature given the current discovery of new spirochaetes [24,25]. *Borrelia miyamotoi* clusters among the relapsing fever spirochaetes, but unlike these, resides alongside *B. theileri* and *B. lones tari* that are also transmitted by hard ticks (Figure 2).

Until recently, it was believed that *B. miyamotoi* fell into three geographically separated clades (genotypes), namely the Asian (or Siberian), European and American, (Figure 2, Table 2), transmitted by *I. persulcatus*, *I. pavlovskyi* (*I. ovatus*), *I. ricinus* and *I. scapularis* or *I. pacificus*, respectively [26,27]. However, as knowledge of this particular spirochaete increases, more diversity within the species is being recognised. A slight sequence variation was observed among the Asian clade based upon 16S sequence data, typified by strain Y14T1 from *I. persulcatus* ticks (Figure 2) that revealed sequence divergence from other members within this clade [10]. Similarly, in the US, divergence was noted among sequence types of American *B. miyamotoi* associated with *I. pacificus* (Figure 2) [12,28].

Given the highly conserved nature of 16S, it is not generally considered a reliable marker for the delineation of *Borrelia* [29], as such, the level of heterogeneity among these variants might be considerably more than that inferred from 16S data alone.

A potential fourth clade was recently described for strains found in *I. ovatus* ticks in Japan. These showed clear separation from the established sequence types for 16S rRNA (Figure 2) and the clade was further confirmed using more discriminatory multilocus sequence typing (MLST) phylogenetic approaches [30]. Representatives of this potentially novel clade of *B. miyamotoi* were carried by up to 0.6% of ticks collected from Honshu Island, but were absent from those ticks collected in Hokkaido, Japan [4,10].

Strains belonging to different genotypes of *B. miyamotoi* have been observed to geographically overlap. Asian and European genotypes have been detected in Russia and Estonia [2,11]. In south-eastern Estonia, the Asian strains were detected in both *I. persulcatus* and *I.*
**Ecology**

In studies assessing single strains of *B. miyamotoi* and their associated tick species, the spirochaete has been detected among unfed larval ticks, suggesting successful transovarial (i.e. vertical) transmission to successive generations of ticks. Thus, ticks should be considered as both vectors and reservoirs for *B. miyamotoi*. Vertical transmission is estimated to have a frequency of less than 0.8 per tick generation, thus not sufficient to sustain *B. miyamotoi* for more than a few successive generations, in isolation of other transmission mechanisms [6,33]. That such mechanisms are needed has gained further support from laboratory-based *in vivo* infection studies suggesting a decline of tick infection by *B. miyamotoi* during their development, through trans-stadial moults [34]. Prevalence studies conducted with different *B. miyamotoi* tick instars revealed lower infection rates among larvae than among nymphal and adult ticks (which had equivalent rates) [35]. The higher infection rates in nymphs and adults possibly result from infection of ticks by feeding (i.e. horizontal transmission) later in their life cycle. This contrasts with a staged infection rate by instar reported for *B. burgdorferi* s.l. [11,17], whereby infection may occur predominantly by horizontal transmission (i.e. feeding).

Similar to other borreliae, prevalence of *B. miyamotoi* appears to be amplified through infection of certain avian and rodent species. Varied small vertebrate species have been shown to be competent reservoirs of infection by *B. miyamotoi* including mice (*Apodemus* spp. and *Peromyscus* spp.), voles (*Microtus* spp., *Myodes glareolus*, and *Microtus arvalis*), chipmunks (*Tamias* spp.), squirrels (*Sciuridae* spp.), European hedgehogs (*Erinaceus europaeus*) and raccoons (*Procyon* spp.) [6-8,15,31,36-39]. Avian species have not been extensively surveyed, but *B. miyamotoi* has been detected in blackbirds (*Turdus merula*), great tits (*Parus major*), common chaffinch (*Phylloscopus collybita*), song thrush (*Turdus philomelos*), European robin (*Erithacus rubecula*) and European greenfinch (*Chloris chloris*) albeit at low levels 0.6–8% of avian-removed ticks [15,31,40]. A report from Tennessee, US describes a surprisingly high level of *B. miyamotoi* infection among wild turkeys (*Meleagris gallopavo*) with 58% (35/60) birds sampled testing positive for *B. miyamotoi* [41]. These birds were heavily infested with *A. americanum* ticks (70%; 42/60), but none of the ticks were positive for *B. miyamotoi* [41].

Furthermore, larger vertebrates have been found with evidence of infection such as wild boar [42,43] but their ecological significance remains unclear. Interestingly, deer have been suggested to amplify *B. miyamotoi* tick infection rates in the US [44] and representatives of the Asian genotype of *B. miyamotoi* have been recovered from infected deer in the far East [32]. These observations contrast with the Lyme borreliae, where deer are not competent species to enhance transmission to ticks during feeding. Nevertheless, it must be noted

---

**Figure 1**
Flow chart of the search strategy and inclusion of identified articles, narrative literature review of *Borrelia miyamotoi*

Additional supplementary information was included as indicated to support discussion points as required. This included information on the pathobiology of other relapsing fever species.

*riocinus* while the European strains could only be found in *I. ricinus* [11].

The situation in Estonia also illustrates that a single tick species (e.g. *I. ricinus*) can harbour several genotypes [11], a phenomenon that has also been observed for *I. ovatus*, which can be infected by both the Asian and the new variant genotype [10]. On the other hand, some genotypes have been found in multiple tick species such as the American strain found in *I. scapularis* and *I. dentatus* [31], with its close variant present in *I. pacificus* [12]. Furthermore, there are reports of Asian clade representatives of *B. miyamotoi* in other types of ticks such as *Haemaphysalis concinna* [13] and *H. longicornis* [32], but the ecological role of these species for maintaining this spirochaete has not been fully elucidated. Thus vector specificity may not be as strict as previous data suggested.
that a study from Japan failed to demonstrate *B. miyamotoi* in deer, but instead reported another hard-tick vectored borrelial species resembling *B. theileri* and *B. lonestari* [45] that had been previously reported from Japanese *Haemaphysalis* spp. ticks [46] and *H. longicornis* from China [32]. A role for deer was also not corroborated by a study from the Netherlands and the European genotype [15]. Because the various studies reported here were based upon different *B. miyamotoi* genotypes, strain differences might explain the differences observed for transmission efficiency following ticks feeding upon deer.

**Relations between Borrelia miyamotoi and Borrelia burgdorferi s.l.**

In vertebrates, co-infection with *B. miyamotoi* and *B. burgdorferi* has been recorded, but the probability of dual infection appears no greater than that which would be expected independently by chance [6].

In ticks however, it remains a question whether co-infection happens at random or not. Some studies suggest that co-infection is more frequent than either infection alone [15]. Others however suggest that this happens by chance. For example, subsequent to a study finding 264/5,431 (4.9%) *B. burgdorferi* sensu stricto in nymphal *I. pacificus* ticks, a subset of 3,255 ticks was retested for *B. miyamotoi*. A total of 44 (1.4%) ticks were infected *B. miyamotoi* alone, and only one single individual with *B. burgdorferi* sensu stricto was also found co-infected with *B. miyamotoi* [19,47]. On the other hand, among *I. scapularis* ticks from Midwestern US, Hamer et al. report a 2.1 times lower level of co-infection with *B. burgdorferi* and *B. miyamotoi* (0.05% co-infection among 1,565 questing adult ticks) than expected by chance [48].

Tick surveillance demonstrates that *B. burgdorferi* s.l. is more prevalent than *B. miyamotoi*, approximating to a ratio of 10:1 or greater in several studies [5,6,48,49]. This is intriguing given that an American study suggests that both species infect ticks to equivalent levels achieving infection rates of around 2,000 spirochaetes for nymphs and 5,000 for adult ticks (sometimes higher in the case of *B. miyamotoi*) [6]. A European study, which quantified by PCR *B. miyamotoi* in feeding ticks removed from humans, even reported considerably higher borrelial tick loads by *B. miyamotoi* (mean count of 2.1 x 10^3) than by *B. burgdorferi* s.l. (with a mean of 4.5 x 10^2 for *B. afzelii* and 2.7 x 10^2 for *B. garinii*) [49]. Thus, infection prevalence among ticks is higher for *B. burgdorferi* s.l. [5,6,48,49], however, *B. miyamotoi*-infected ticks appear to have equivalent [6], or higher quantities of borreliae within their tissues [49].

The answer to why we observe higher prevalence for *B. burgdorferi* s.l. compared with *B. miyamotoi* in ticks, might reside in the superior ability for *B. burgdorferi* to persist in the skin of its vertebrate host, potentially extending the window of transmission to other feeding ticks. The failure to detect *B. miyamotoi* in the skin of erythema migrans patients corroborates this theory [15]. Moreover, studies of *Peromyscus leucopus* mice during *I. scapularis* nymphal questing times have shown that Lyme-associated borreliae resided in the skin of their host, while *B. miyamotoi* gave a higher blood burden with density counts five times higher than *B. burgdorferi* s.l. [6]. While levels up to 251 spirochaetes of *B. miyamotoi* per mL blood have been observed, persistence in blood within an infected vertebrate appears comparable to *B. burgdorferi* s.l. [6,50].

The sympatric overlap of *B. burgdorferi* s.l. and *B. miyamotoi* begs the question as to how these closely related spirochaetes might interact. Remarkably, the prevalence of *B. miyamotoi* infection in ticks appears indifferent to variations in the habitat type, or to ecological influences that impact upon the prevalence of *B. burgdorferi* s.l. [19]. As nutritional requirements are likely to overlap, is there antagonism between these species, indifference or a more synergistic relationship? Competitive interactions have not been explored per se, however, emerging data suggest that these species appear indifferent to each other. A study of shared small rodent hosts suggests different seasonal peaks of infection, with *B. burgdorferi* s.l. rodent infection peaking in spring and being driven by nymphs, whereas *B. miyamotoi* rodent infection being predominant in summertime through larval ticks [51,52]. This separation might provide a means for both spirochaetes to co-exist [6]. Indeed, transovarial tick transmission of *B. miyamotoi* enables larval ticks to be infectious for their subsequent hosts. In contrast, tick larvae harbouring *B. burgdorferi* s.l. are rarely reported, suggesting infrequent vertical transmission in *B. burgdorferi* s.l. Among field-collected larvae, a study found *B. burgdorferi* s.l. in 0.62% as opposed to 2% for *B. miyamotoi* [53]. Some recent publications query whether previously reported low levels of *B. burgdorferi* s.l. vertical transmission constituted in fact detection of *B. miyamotoi* [54,55].

**Transmission to humans, magnitude of human infection and pathogenesis**

Emerging evidence supports presence of *B. miyamotoi* in tick salivary glands [34], with a study finding that over 88% of second generation progeny of infected *I. scapularis* nymphal ticks had *B. miyamotoi* in their salivary glands, as seen with other relapsing fever spirochaetes in their soft tick vectors [56]. Experimental mouse (CD-1 outbred Mus musculus mice) transmission studies using ticks derived from a field collected population infected with the American strain of *B. miyamotoi*, have demonstrated 10% infection after 14 hrs, rising to 31% at 48hrs and 63% by 72hrs [56]. Though not as rapid as the transmission of classical relapsing fever spirochaetes by their soft tick vectors [57], this demonstrates transmission of *B. miyamotoi* in the first day of tick attachment, contrasting with the days of attachment needed by Lyme borreliae to migrate from the tick midgut to salivary glands for transmission [57].
In this respect, in humans, transmission efficiency has been estimated at 8.3% [58], based upon a cohort of 24 persons bitten by PCR-positive ticks, of whom only two later developed compatible clinical disease [58]. By comparison, three of 68 (4.4%), humans bitten by *B. burgdorferi* s.l. infected ticks, developed disease (erythema migrans) [58,59].

In terms of the magnitude of human infections with *B. miyamotoi*, this depends to some extent on the frequency of tick bites among people and on the prevalence of tick infection. Ixodid ticks are holarctic in distribution and assessments of tick bites vary by country and location. In the Netherlands, it was estimated that tick bites amount to 71,980 per million inhabitants (2007) [60]. In the Irkutsk region in Russia 2014, tick bites were estimated at 12,500 per million people [21]. The prevalence of infection among ticks ranges from 0.5 to 6% in many regions of the northern hemisphere. Taken together with a transmission rate of 8.3% [58], it is likely that human infections are underestimated [61].

---

**Table 1**

<table>
<thead>
<tr>
<th>Location</th>
<th><em>Borrelia miyamotoi</em> cases with reference (year of publication)</th>
<th>Percentage of cases among persons studied</th>
<th>Reported infection prevalence in ticks</th>
<th>Clinical signs</th>
<th>Diagnostic method</th>
<th>References for prevalence in ticks and human cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russia</td>
<td>51/302 humans bitten by ticks (2011) [2]</td>
<td>16.9%</td>
<td><em>Ixodes ricinus</em> 0.8%</td>
<td>Fever, chills, sweating, headache, fatigue and vomiting (relapsing fever in 5)</td>
<td>PCR and serology</td>
<td>[2,21,104,105]</td>
</tr>
<tr>
<td>Izhevsk, Russia</td>
<td>2/24 B. miyamotoi infected-tick bites developed disease (2015) [58]</td>
<td>8.3%</td>
<td><em>I. persulcatus</em> 2.9–10.5%</td>
<td>Fever, chills, sweating, headache, fatigue, nausea, vomiting, dizziness.</td>
<td>PCR and serology</td>
<td>[58]</td>
</tr>
<tr>
<td>Yekaterinburg, Russia</td>
<td>71/459 tick-borne infection (including 1* also with Lyme borreliosis) (2018) [22]</td>
<td>15.5%</td>
<td><em>I. pavlovski</em> 6.4%</td>
<td>Clinical details not described.</td>
<td>PCR and serology</td>
<td>[2,22]</td>
</tr>
<tr>
<td>Hokkaido, Japan</td>
<td>2*408 Lyme borreliosis cases (2016) [75]</td>
<td>0.49%</td>
<td><em>I. persulcatus</em> 1.6–2%</td>
<td>Fever, myalgia, anorexia</td>
<td>PCR and serology</td>
<td>[4,10,75]</td>
</tr>
<tr>
<td>Japan</td>
<td>12/459 suspected Lyme borreliosis (2018) [64]</td>
<td>2.6%</td>
<td><em>I. papulosus</em> 4.3%</td>
<td>One case meningoencephalitis; clinical history not disclosed on remainder.</td>
<td>Serology</td>
<td>[4,10,64]</td>
</tr>
<tr>
<td>Hokkaido, Japan</td>
<td>1 case study (2017) [80]</td>
<td>NA</td>
<td><em>I. ovatus</em> 0.5%</td>
<td>Fever, macular erythematous rash, low blood pressure, thrombocytopenia.</td>
<td>Serology</td>
<td>[4,10,80]</td>
</tr>
<tr>
<td>Germany</td>
<td>1 case study suspected Lyme neuroborreliosis (2016) [72]</td>
<td>NA</td>
<td><em>I. ricinus</em> 1.2–2.4%</td>
<td>Lymphomatous meningitis (immunocompromised)</td>
<td>PCR, CXCL13 and microscopy</td>
<td>[3,5,72,82]</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1 case study (2013) [73]</td>
<td>NA</td>
<td><em>I. ricinus</em> 2.1–3.6%</td>
<td>Meningoencephalitis (immunocompromised)</td>
<td>Microscopy, PCR and equivocal serology</td>
<td>[15,73,106,107]</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1 case study (2018) [103]</td>
<td>NA</td>
<td><em>I. ricinus</em> 2.1–3.6%</td>
<td>Lymphadenopathy, leucopenia and thrombocytopenia (immunocompetent)</td>
<td>Serology</td>
<td>[15,103]</td>
</tr>
<tr>
<td>United States</td>
<td>97/11,515 acute febrile patients (2015) [78]</td>
<td>0.84%</td>
<td><em>I. scapularis</em> 0.02–3.1%</td>
<td>Fever, chills, myalgia, arthralgia, headaches, neutropenia, thrombocytopenia</td>
<td>PCR</td>
<td>[3,6,18,19,35,78,83,108]</td>
</tr>
</tbody>
</table>

NA: not applicable.

* Denotes dual-infected cases who also presented with erythema migrans lesions.

---
In Yekaterinburg Province, Russia in 2009, it was estimated B. miyamotoi infections occur in 1 per 100,000 inhabitants [2]. Studies from the Netherlands suggest that some 36,000 humans are bitten each year by B. miyamotoi (European strain) infected ticks [9]. Serological studies from the US appear to corroborate human exposure [62,63]. Despite this, reports of human infections are comparatively few; for example, retrospective serological evaluation of sera from cases suspected for Lyme disease only detected 19 indigenous cases in Japan between 2013 and 2017 [64].

Borreliae as a group, are renowned for their ability to persist within humans. Relapsing fever Borrelia undergo an elaborate series of antigenic variation coupled with other mechanisms such as binding factor H and evading the components of the complement cascade, enabling their persistence in blood or their human host over time. Likewise, B. miyamotoi possesses variable membrane proteins (vmp) that could facilitate relapse with different antigenic variants [65], and has the ability to bind factor H and related proteins, thus evading host complement-mediated degradation [66-68]. Clinical reports of relapse with B. miyamotoi infection are described in the literature, however surprisingly, this appears less frequently than seen with classical relapsing fever infections [2,69].

The borreliae have proven neurotropism potential, using the nervous system as an immunologically protected niche enabling persistence in their vertebrate host [70,71]. The precise mechanisms by which spirochaetes exploit this niche are poorly understood, but animal studies have shown that relapsing fever Borrelia can survive for 270 days within this site, thus enabling greater in vivo persistence [70]. Neurological sequelae have been a feature of infection among immunocompromised individuals (see clinical presentations below) [72-74]. An in vivo infection model is urgently needed to unravel the pathogenesis of human B. miyamotoi infection. A severe combined immune deficiency (SCID) mouse model has been established [34], but has limitations to extrapolate to infection in immunocompetent humans.

Clinical presentation

Clinical cases have been reported from Europe, Japan [4,64,75], China [13], Russia [2] and the US [76-78]. Infection with B. miyamotoi does not present with obvious hallmark signs. Instead, patients present with fever accompanied by non-specific influenza-like symptoms, such as chills, fatigue, headache, myalgia and arthralgia (Table 1) [9].

Although B. miyamotoi clusters within the relapsing fever group of spirochaetes, cases with the characteristic recurring febrile episodes interspersed with non-febrile intervals that typify classical relapsing fever have only been described sporadically [2,79]. In these, up to three febrile episodes have been recorded [2], however this might be an underestimation given that patients are typically managed with antimicrobial therapy upon diagnosis. A case who was retrospectively diagnosed following spontaneous recovery had a documented relapsing illness with two episodes, albeit with a lengthy 3-week afebrile period [69].

Furthermore, unlike relapsing fever spirochaetes, epistaxis, abortion, jaundice and major organ failure have not appeared as features of B. miyamotoi infection. Nevertheless, both B. miyamotoi and classical relapsing fever share fever, headaches, chills, myalgia, arthralgia, and nausea/vomiting.

Some differences in clinical presentation have been noted between US and Russian cases, particularly regarding the presence of thrombocytopenia documented in approximately half of American cases [78], but not reported from those in Russia. This differential clinical presentation may be an artefact given the recent description of thrombocytopenia in a Japanese clinical case infected with an Asian B. miyamotoi strain [80], akin to those reported from Russian cases.

In two studies, cases have been reported with erythema migrans [75,78], however it is likely that these had concomitant infection with B. burgdorferi s.l., thus representing co-infections. Indeed, one of these studies retrospectively sought presence of B. miyamotoi in sera from cases diagnosed with Lyme disease, whereas the other reported one case of erythema migrans among 51 B. miyamotoi patients who had an overall 14% co-infection rate with B. burgdorferi s.l. A larger study of 71 PCR-confirmed B. miyamotoi infected cases from Yekaterinburg, Russia, found only five of these cases with erythema migrans [22]. Analysis of blood for confirmation of pathogens disclosed that one case was co-infected with B. burgdorferi s.l., but it was concluded that the remaining cases with erythema migrans were realistically also co-infections given the poor sensitivity of PCR to detect Lyme disease using blood samples (as B. burgdorferi s.l. tends to be found in the skin) [22,49].

Borrelia miyamotoi, like other members of the borreliae have demonstrated their ability to result in neurological sequelae with descriptions of meningoencephalitis, albeit among immunocompromised individuals [64,72-74]. Unlike the acute febrile presentation described above, infections in these immunocompromised cases described to date, have shown a more insidious onset, often over several months. One case presented with memory deficits and disturbed gait, with lumbar puncture revealing pleocytosis and raised cerebrospinal fluid (CSF) protein [73].

Diagnostics

For immunocompromised patients, diagnosis with microscopy has been used to a certain extent [72-74], sometimes combined with immunofluorescence [74]. For example, in three case reports on immunocompromised patients infected with B. miyamotoi [72-74],
Spirochaetes in the CSF were retrospectively detected by microscopy in one case [73]. For the other two cases, *B. miyamotoi* was also identified by microscopy, albeit after concentration of the CSF sample and either Giemsa or acridine orange staining [72,74]. For all three cases, however, microscopy was complemented with PCR for diagnosis confirmation [72-74].

Generally, for overall patients, diagnostic approaches depend on the stage and duration of infection. During the acute phase of infection, the presence of the spirochaete in blood and CSF can be demonstrated using PCR and microscopy. The success of such methods rapidly reduces from the fourth day of disease correlated with depletion in the spirochaetal blood counts [22]. Clinical reports of relapse are described in the literature [2,69]. If this occurs, it is likely that direct detection using PCR or microscopy might again be valuable.

Use of concentration methods can improve the diagnostic sensitivity of these techniques [72,74,81]. Later in the clinical course, serology is the mainstay diagnostic option.

In terms of PCR/molecular diagnostics during acute stages of infection — and for assessment of non-human vertebrates or ticks —, real-time PCR assays for *B. miyamotoi* based upon either 16S rDNA [6] or the flagellin gene target have been described. The effectiveness of such assays for detection of the newly described *B. miyamotoi* variants nevertheless remains to be established [82]. Given the sympatric nature of *B. miyamotoi* and its closely related Lyme-associated borreliae, a logical strategy would be to utilise a multiplex assay able to screen for both pathogens simultaneously. A multiplex approach is also probably better suited to a diagnostic setting to avoid multistep methods and

---

**Figure 2**

Strain diversity of *Borrelia miyamotoi* based upon 16S rRNA sequences
reduce contamination risks that arise at each step. Some studies have suggested assays that differentiate conditions with overlapping clinical presenting features, such as *B. miyamotoi* infection and anaplasmosis [77] and have thus produced multiplex PCR assays to differentiate these infections [83].

Concerning serology, many studies have used the glycerophosphoryl diester phosphodiesterase (GlpQ) antigen expressed by members of the relapsing fever borreliae, but absent from *B. burgdorferi* s.l. [84]. Assays based upon GlpQ will not be *B. miyamotoi*-specific, but in areas where other members of the relapsing fever borreliae are not prevalent, such assays can be useful for population surveys and diagnosis on non-acute infections [85]. Besides the difficulties of GlpQ assays to distinguish *B. miyamotoi* from other relapsing fever borreliae, it is noteworthy that homologous proteins have been reported from both *Klebsiella pneumoniae* and *Salmonella enterica* [86]. In addition to lack of specificity, other reported limitations with the GlpQ antigen-based approach have been poor diagnostic sensitivity, with ability to only detect 28 of 36 convalescent samples from established cases [78].

As the use of vmps as antigens has also been explored for serodiagnosis [87], combinations of GlpQ together with mixtures of highly immunogenic vmps derived from *B. miyamotoi* have been evaluated as a way to improve the diagnostic efficacy [88]. This involved a comprehensive series of 182 PCR-confirmed Russian patients who were followed with sequential sera collected over several months post-infection. Notably, combinations of antigens provided superior sensitivity and/or specificity, with diagnostic titres for IgM, which were reached from 11 to 20 days post-disease onset and for IgG, from 21 to 50 days. This study used blood donor controls and additionally, controls with tick-borne encephalitis recruited from the same geographical region together with controls without tick exposure. Use of the combined antigens resulted in a sensitivity of 94.7% and specificity of 96.6% for IgM from 11 to 20 days post clinical presentation, thus providing important improvements over previous assays [88]. Assessment of duration of serological reactivity in seven of the 182 patients, showed that the IgM response waned within a year, while half (4/7) remained seropositive for IgG a year following disease. All individuals had been prescribed antimicrobial therapy.

On a cautionary note, the C6 ELISA used for diagnosing Lyme disease may additionally be positive in those infected by *B. miyamotoi* [69]. Furthermore, when using serological tests to detect *B. miyamotoi* in areas where multiple spirochaetes are endemic, the possibility of serological cross-reactivity between other relapsing fever borreliae such as *B. hermsii* should be kept in mind as this may obscure diagnosis [64,85].

Given the increasing recognition of co-infections among those with tick exposure, it might be prudent to take a more holistic approach and screen for a broader range of tick-borne pathogens than just borreliae [89]. Some diagnostic centres have taken this approach, but many do not yet have the resources for more comprehensive tick-borne pathogen screening.

Isolation of borreliae is always a challenge as these microbes are particularly fastidious to cultivate, requiring complex liquid medium. Isolation is typically confirmed by dark field microscopy. The Barbour–Stoermer–Kelly (BSK) commercially-available medium (BSK-H) used for *B. burgdorferi* s.l. is unreliable for growth of relapsing fever spirochaetes (data not shown) and when used for *B. miyamotoi*, is unable to sustain passage of this organism [90]. Growth of *B. miyamotoi* has been achieved using modifications of Kelly–Pettenkofer medium (MKP) [16,91] or in a variation of BSK described as BSK-M [4]. Other studies have used media modifications with inclusion of 50% serum in order to cultivate these spirochaetes [92].

Propagation of isolates has additionally been achieved using inoculation of SCID mice. Typically, these will show spirochaetes in blood films between 7 and 14 days post-infection [34].

**Treatment**

Therapy for *B. miyamotoi* infection has typically followed guidelines used for treatment of Lyme borreliosis. Only a few cultivable strains have been recovered to date, restricting evaluation of different clinical management regimes. Moreover limited *in vitro* susceptibility testing has been undertaken to verify the efficacy of different therapeutic protocols [93]. Koetsveld and co-authors, noted resistance to amoxicillin *in vitro* (16–128 mg/L) using two isolates of *B. miyamotoi* [93]. Interestingly, this feature was also shared by the relapsing fever spirochaete *B. hermsii* that was assessed in parallel. Despite these *in vitro* findings, a patient treated with amoxicillin (and sultamicillin) responded without complications [80]. No treatment failures have been reported to date, thus it is probable that the hypothesised susceptibility profile being analogous to the Lyme-associated species is supported. Standard methods are not applicable for evaluation of the susceptibility testing of borreliae given their need for liquid cultivation, microaerophilic conditions and coupled with their slow mean generation time [94-96]. Akin to other members of the genus and spirochaetal infections in general, a proportion of patients may develop a Jarisch–Herxheimer reaction (JHR) associated with a sudden exacerbation of clinical signs upon onset of treatment [97]. Though reported, JHR does not appear frequently for cases of acute *B. miyamotoi* infection [2,74].

**Future research directions**

Much of the data so far arise from studies designed and funded to look either at tick-borne diseases or more specifically Lyme-associated borreliae. Though valuable, these studies might be biased and not reflect
some of the different ecological driving factors underpinning the observed epidemiology of this spirochaete. As such, we still need to more specifically address *B. miyamotoi* epidemiology further. This is important both for risk assessment and for application of control/intervention strategies. Indeed, many questions remain regarding our understanding of the pathobiology of this spirochaete (Table 3).

Future epidemiological studies need to consider the different genotypes of *B. miyamotoi*, particularly in areas of geographical overlap such as in Estonia and Russia [2,11]. Ability to assess the epidemiology of *B. miyamotoi* has been enhanced by the development of several multiplex PCR diagnostic methods providing a more cost-effective means for high throughput screening of samples [6,82,83]. These approaches, which are tailored to detect genotypes already described, however run the risk of missing hitherto undescribed variants. Indeed, while three genotypes of *B. miyamotoi* (American, Asian and European) have been previously described, each associated with different tick vectors (Table 2) [3], recent studies have found more diversity within the *B. miyamotoi* species [10,12,28]. Future studies may shed light on how diverse *B. miyamotoi* genotype strains and variants behave. For example, it has not been comprehensively addressed if each of the *B. miyamotoi* genotypes or variants and their respective tick hosts are equally competent for transovarial transmission. Our knowledge to date is based upon studies that typically have only assessed one strain and tick species [33,54]. This information is currently extrapolated to others, but not underpinned by rigorous scientific enquiry.

Despite larger diversity recognised among *B. miyamotoi* strains, the species appears to be less heterogeneous than the *B. burgdorferi* s.l. complex, for which representatives sympatrically overlap with *B. miyamotoi*. Although it needs to be evaluated how strict new *B. miyamotoi* variants are to certain tick species, the variants identified to date have been found to be generally restricted to separate tick species, supporting the idea that, like for other *B. miyamotoi*spirochaetes, ticks still serve as both reservoirs and vectors for these [19]. Conversely, for the Lyme-associated species, ticks act mainly as vectors, while a plethora of vertebrates serve as reservoirs for horizontal infection, potentially driving diversity within this complex.

This being said, vertebrates also play a role in the ecology of *B. miyamotoi*. In this respect, it is notable that high infection rates were reported in turkeys [41]. Studies of avian vertebrates and their role in the ecology of *B. miyamotoi* have largely excluded ground foraging gallinaceous species, which is surprising given their established role as a reservoir for *B. burgdorferi* s.l [98]. Looking into these avian species may provide missing pieces of the jigsaw of understanding the ecological niche of this spirochaete.

Coinfections by *B. burgdorferi* and *B. miyamotoi* have been observed in both vertebrates and ticks. Within ticks, a further level of complexity arises from the growing appreciation of how different microbes might facilitate survival or transmission. An ecological synergy is proposed whereby *B. burgdorferi* s.l. and *Babesia microti* benefit each other [99], but currently we have no knowledge of such interactions for *B. miyamotoi*. Importantly, we should consider how other pathogens present within ticks (including other *Borrelia*) will

<table>
<thead>
<tr>
<th>Tick species</th>
<th>Geographical range</th>
<th>Predominating <em>Borrelia miyamotoi</em> genotype (co-occurring variants)</th>
<th>Tick feeding preference</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Ixodes persulcatus</em></td>
<td>Baltic to Far East</td>
<td>Asian</td>
<td>Generalist</td>
<td>[2,11]</td>
</tr>
<tr>
<td><em>Ixodes pavlovskyi</em></td>
<td>Western Siberia and Far East</td>
<td>Asian</td>
<td>Ground foraging birds, small mammals</td>
<td>[4,104]</td>
</tr>
<tr>
<td><em>Ixodes ovatus</em></td>
<td>South East Asia</td>
<td>Asian (new Asian variant)</td>
<td>Generalist</td>
<td>[4,10]</td>
</tr>
<tr>
<td><em>Ixodes ricinus</em></td>
<td>Northern Sweden to north Africa, Ireland to Ural in Russia</td>
<td>European (Asian)</td>
<td>Generalist</td>
<td>[11,15]</td>
</tr>
<tr>
<td><em>Ixodes scapularis</em></td>
<td>North-eastern and upper Midwestern United States</td>
<td>American</td>
<td>Generalist</td>
<td>[33,44,54]</td>
</tr>
<tr>
<td><em>Ixodes pacificus</em></td>
<td>Pacific coast of United States</td>
<td>American (new American variant)</td>
<td>Generalist</td>
<td>[44]</td>
</tr>
<tr>
<td><em>Ixodes dentatus</em></td>
<td>Eastern United States</td>
<td>American</td>
<td>Rabbits, hares Birds (larvae and nymphs)</td>
<td>[31]</td>
</tr>
</tbody>
</table>
influence the survival, persistence and transmissibility of *B. miyamotoi*.

To anticipate human infection, consideration of the tick species likely to be encountered is important. An example might be the detected presence of *B. miyamotoi* within *I. dentatus* ticks that preferentially feed upon birds or lagomorphs and thus present a reduced risk of human infection [31]. Such information would enable instigation of risk prediction, modelling and targeted intervention approaches where justified.

Blood transfusion has been hypothesised as a risk factor for *B. miyamotoi* infection. Experimental studies have demonstrated the ability of *B. miyamotoi* to survive under conditions used for storage of blood transfusion products [100], raising concern that this could provide an additional source of infection (Table 3). Though theoretically possible, clinical cases presenting with acute signs, typically fever and associated non-specific influenza-like features, unlikely would present to donate blood [101]. Quantification of the numbers of spirochaetes present during human infection suggests that levels peak at ca $10^{3}$–$10^{4}$ *B. miyamotoi* copies per mL during the second to third day of illness, then rapidly wane until day eight [22]. These authors suggested that PCR diagnosis was unreliable after the fourth day of clinical signs which would suggest low risk of transfusion related infection.

An enigma of our current understanding is why we fail to see a correlation between clinical cases of *B. miyamotoi* infection and tick infection prevalence. Human cases have occurred in some areas with relatively low tick infection prevalence, like Yekaterinburg, Russia, while in other countries or areas with higher prevalence in ticks, such as Mongolia or Hannover in Germany, no cases were reported. The reasons are likely to be multifactorial. Tick factors such as feeding preferences or the role of vertebrate hosts in intensifying or negatively impacting transmission could perhaps account for the observations. Another explanation might be differential virulence in humans of the strains involved, as it remains to be solved if different strains/genotypes show differential virulence. Last, the absence of an evident relation between tick and human infection rates could also possibly result from lack of diagnostic ability, proximity to those with active research interest in tick-borne disease, or cases missed due to unspecific symptoms. The clinical features of *B. miyamotoi* indeed lack a diagnostic hallmark, and can potentially be misdiagnosed as granulocytic anaplasmosis [77,102]. Given these difficulties to recognise *B. miyamotoi* infections, it is likely that cases are under-reported. While spontaneous resolution of an infection with *B. miyamotoi* can occur without antimicrobial intervention [69,103], some clinical studies have noted severe infection, with considerable proportions of cases being hospitalised (24% [78]). Because when cases are diagnosed, they are likely to be promptly treated, the long-term consequences of human infection remain to be determined. While case studies of immunocompromised individuals report notable clinical features [64,72-74], the other risk factors that might influence the clinical progression towards disease are also not clear. Future detailed epidemiological studies of cases detected will enable a more complete clinical picture to unfold.

We also need to be aware of unusual presentations resulting from potential co-infection. The areas reporting cases are likely endemic for a variety of different tick-borne diseases. Clinically, presence of one tick-borne pathogen can obscure the concomitant presence of another. Consequently, it is a priority that those suspected of *B. miyamotoi* infection be also comprehensively screened for other tick-borne pathogens.

Diagnostic methods are nevertheless still widely considered ‘research tests’ and thus not generally available in more routine clinical settings. When these are possible, a delay in considering *B. miyamotoi* infection might further complicate their interpretation, due to the
poorly delineated clinical course of human infection. A further problem may be the variability of strains encountered which may affect diagnostic results.

Another problem relates to what is the most appropriate diagnostic sample. Ticks removed from patients have limited diagnostic value in that transmission may not have occurred and this might be only one of several ticks that might have bitten the individual in question. During acute infection, blood samples are a key sample to collect, however samples should be taken as early as possible and certainly before commencement of antimicrobial therapy. Collection of CSF is valuable in cases showing neurological features. This might require concentration of the specimen to improve diagnostic sensitivity [72,74]. Serum samples for serological investigation should be collected at all stages, with early sera used for assessment of seroconversion or increasing titre, and later samples for retrospective studies of prior exposure.

As for the therapeutic management of cases, a more comprehensive in vitro evaluation will be possible to guide this, once more isolates become available representing the diversity within this species.

Conclusion
Our understanding of B. miyamotoi and its ecology and infection potential have only recently started to unfold. The incidence of B. miyamotoi infection in humans is poorly explored, hampered by the lack of awareness and appropriate diagnostics. Our understanding of the clinical features of infection currently suggest a relatively mild infectious course without long-lasting sequelae for the majority of infected individuals, but with the caveat of being able to cause severe disease in the immunocompromised [72,74,86]. Current incidence is likely to be grossly under-reported, suggested by predicted tick bite exposure and tick prevalence data, as well as through the absence of a clinical presentation hallmark, making assessment of the impact of infection challenging.

Acknowledgements
We acknowledge ECDC for funding support through framework contract 01/24/04/2014-PROC/2014/023.

Conflict of interest
None declared.

Authors’ contributions
The initial review of the literature and compiling the review was undertaken by SJC and MVT. All authors (SC; MVT; AE-P; AP; AM & HZ) contributed towards the final preparation and review.

References


Disease burden of varicella versus other vaccine-preventable diseases before introduction of vaccination into the national immunisation programme in the Netherlands

Alies van Lier¹, Brechje de Gier¹, Scott A McDonald¹, Marie-Josée J. Mangen¹, Maarten van Wijhe±¹, Elisabeth A.M. Sanders±², Mirjam E. Kretzschmar±², Hans van Vliet¹, Hester E. de Melker², Elisabeth A.M. Sanders³, Hester E. de Melker²
1. Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, Netherlands
2. Department of Science and Environment, Roskilde University, Roskilde, Denmark
3. Department of Pediatric Immunology and Infectious Diseases, Wilhelmina’s Children Hospital, University Medical Center Utrecht (UMCU), Utrecht, Netherlands
4. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht (UMCU), Utrecht University, Utrecht, Netherlands
Correspondence: Alies van Lier (alies.van.lier@rivm.nl)

Citation style for this article:

Article submitted on 02 Jul 2018 / accepted on 21 Dec 2018 / published on 02 May 2019

Introduction: Estimating burden of disease (BoD) is an essential first step in the decision-making process on introducing new vaccines into national immunisation programmes (NIPs). For varicella, a common vaccine-preventable disease, BoD in the Netherlands was unknown. Aim: To assess national varicella BoD and compare it to BoD of other vaccine-preventable diseases before their introduction in the NIP. Methods: In this health estimates reporting study, BoD was expressed in disability-adjusted life years (DALYs) using methodology from the Burden of Communicable Diseases in Europe (BCoDE)-project. As no parameters/disease model for varicella (including herpes zoster) were available in the BCoDE toolkit, incidence, disease progression model and parameters were derived from seroprevalence, healthcare registries and published data. For most other diseases, BoD was estimated with existing BCoDE-parameters, adapted to the Netherlands if needed. Results: In 2017, the estimated BoD of varicella in the Netherlands was 1,800 (95% uncertainty interval (UI): 1,800–1,900) DALYs. Herpes zoster mainly contributed to this BoD (1,600 DALYs; 91%), which was generally lower than the BoD of most current NIP diseases in the year before their introduction into the NIP. However, BoD for varicella was higher than for rotavirus gastroenteritis (1,100; 95%UI: 440–2,200 DALYs) and meningococcal B disease (620; 95%UI: 490–770 DALYs), two other potential NIP candidates. Conclusions: When considering the introduction of a new vaccine in the NIP, BoD is usually estimated in isolation. The current approach assesses BoD in relation to other vaccine-preventable diseases’ BoD, which may help national advisory committees on immunisation and policymakers to set vaccination priorities.

Introduction
Routine childhood vaccination programmes started in the beginning of the 20th century and after new vaccines became available, have been extended ever since. The Dutch national immunisation programme (NIP) was officially launched in 1957 with universal childhood vaccination against poliomyelitis, diphtheria, tetanus and pertussis. However, mass vaccination programmes had already begun in 1953 with vaccination against diphtheria, and from 1954 onwards this was combined with tetanus and pertussis. The NIP gradually expanded with vaccination against rubella (1974: girls only; 1987: all children), measles (1976), mumps (1987), Haemophilus influenzae type b (Hib) disease (1993), meningococcal C disease (2002), pneumococcal disease (2006), human papillomavirus (HPV) infection (2010: girls only), and hepatitis B (2011; before 2011 risk group vaccination was in place). Vaccination against meningococcal W disease (using meningococcal ACWY vaccine) has been implemented as an outbreak measure in 2018/19 pending further advice from the Health Council of the Netherlands.

In several countries, additional childhood vaccines against highly common diseases like varicella (e.g. United States, Germany) and rotavirus gastroenteritis (e.g. Belgium, Germany, United Kingdom (UK)) have been included in the NIP [1,2]. In the UK, vaccination against meningococcal B disease was recently also implemented [2].
In the decision-making process on introduction of a new vaccine into the NIP, the first criterion used by the Health Council of the Netherlands is the burden of disease (BoD) at population and at individual level. The subsequent criteria taken into account cover the effectiveness and safety of vaccination, acceptability of vaccination, efficiency of vaccination (including cost-effectiveness), and priority of vaccination. Each criterion is formulated on the assumption that the previous one is met [3].

Population BoD can be high if a disease is severe for affected individuals and/or affects a large number of people [3]. Although varicella has a mild disease course for the vast majority of cases, severe complications and mortality may occur [4]. Furthermore, nearly everyone in the Netherlands encounters the varicella zoster virus (VZV) during early life [5] and consequently is also at risk of virus reactivation later in life, resulting in herpes zoster (HZ).

To assess the potential value of adding varicella vaccination to the NIP, it is insightful to compare the BoD of varicella to the BoD of other vaccine-preventable diseases, before vaccination against the latter was introduced into the NIP. Therefore, the objectives of this study were (i) to estimate the current BoD of varicella; and (ii) to compare this to BoD estimates of various vaccine-preventable diseases before their inclusion in the NIP; keeping in mind however, that the timing of introduction into the NIP differs by disease. This study can serve as an example in considerations to take for BoD when new vaccine candidates need to be assessed on eligibility for inclusion in a NIP.

Methods
In this health estimates reporting study, the BoD of the following vaccine-preventable diseases was estimated: diphtheria, pertussis, tetanus, poliomyelitis, measles, mumps, rubella, Hib disease, meningococcal C/W disease, pneumococcal disease, cervical cancer (HPV-infection), hepatitis B (current NIP diseases), meningococcal B disease, rotavirus gastroenteritis, and varicella (potential NIP candidates). The BoD of these diseases was estimated for the year prior to their introduction into the NIP, or for the year 2017 for potential NIP candidates, mainly using the Burden of Communicable Diseases in Europe (BCoDE) toolkit version 0.94 [6] and the parameters presented in Supplement 1. Additionally, when the incidence/BoD of a given disease was estimated to be higher in any of the five preceding years (when data were available), the BoD for the year with the highest incidence/BoD was also presented as an alternative, higher estimate of the BoD.

Disability-adjusted life years methodology
BoD was expressed in disability-adjusted life years (DALYs) [7] (or DALYs per 100,000 population). This composite measure combines morbidity (years lived with disability; YLD) and mortality (years of life lost; YLL) in a single measure of health loss, allowing comparison between diseases with varying severity and incidence (e.g. rare with high mortality vs common with short self-recovery). The underlying methodology, outcome trees, and clinical progression probabilities have been described elsewhere [8-11].

The BCoDE toolkit does not include rotavirus gastroenteritis, varicella and cervical cancer.

For rotavirus gastroenteritis, a model developed by Havelaar et al. specific for the Dutch situation was used [11].

For varicella, a new disease progression model was developed. Dutch data on VZV seroprevalence [12], general practitioner (GP) consultations [13], hospitalisations [14] and mortality [15] were used to determine the proportion of mild (no contact with healthcare)/moderate (GP consultation)/severe (hospitalisation) varicella, and the mortality risk. Congenital varicella syndrome was not included as Dutch VZV seroprevalence at childbearing age is close to 100% [5], and the syndrome only occurs in 0.4–2.0% of all children born to mothers with varicella during the first 20 weeks of gestation [4]. The BoD of varicella and HZ are often investigated separately; however, because these two diseases are so closely related, the BoD of HZ was also estimated. For HZ the disease model previously developed by Kristensen et al. [16] (using disability weights from Salomon et al. [17] and Kwong et al. [18]) for people aged<50 years was used. It was extended with parameters for people aged≥50 years and the Global Burden of Disease (GBD) 2010 life expectancy (LE) [19] was applied (Supplement 1, Tables A4.13/A1).

For cervical cancer the BoD estimates of McDonald et al. were used [20]. Note that the estimates derived using this method are not fully comparable to the other diseases: a different life table (Dutch LE 2014) and source for disability weights (Victorian BoD study) [21] were employed and BoD was computed from the number of registered cervical cancer cases and deaths and the HPV-attributable fraction, instead of deriving BoD from the number of incident HPV infections (which is unknown).

Except for cervical cancer and HZ, the European disability weights elicited by Haagsma et al. [22], and the GBD 2010 LE [19] were applied, in contrast to previous BoD estimates [9-11].

To test the validity of the standard case fatality parameters, historical mortality data were obtained. For diphtheria and tetanus, the original model estimated significantly fewer deaths than registered, whereas it estimated significantly more deaths than registered for poliomyelitis and measles. Therefore, for diphtheria, poliomyelitis and measles, YLL was estimated based on registered instead of estimated mortality, assuming one additional future measles death due to subacute
sclerosing panencephalitis (SSPE). For tetanus, we considered it likely that each reported case in 1953 (and 1950) – before availability of mechanical ventilation/intensive care – died of the disease [23] rather than using mortality data, which may be unreliable for tetanus in that period. Furthermore, the percentage of the paralytic form of poliomyelitis was based on notification data rather than literature-derived estimates (Supplement 1, Tables A4).

Incidence data

Incidence data were derived from various sources and adjusted for underestimation using multiplication factors (Supplement 1, Table A2/A3). The incidence of varicella was estimated based on transmission modelling of Dutch VZV seroprevalence data [12], the incidence of HZ was estimated based on incidence data of the Netherlands Institute for Health Services Research (NIVEL) [13]. For some diseases only cases with invasive disease caused by serotypes covered by the vaccine were included, i.e. Hib, meningococcus C/W/B and pneumococcus 1, 4, 5, 6B, 7F, 9V, 14, 18C, 19F, and 23F (PCV10 serotypes). For cervical cancer BoD was scaled by 71% because vaccine serotypes HPV-16/18 are estimated to be responsible for 71% of invasive cervical cancer [24]. For hepatitis B, only BoD attributable to new infections in 2010 was estimated; BoD due to infections acquired before the year 2010 was excluded.

**Figure 1**

Estimated disease burden of vaccine-preventable diseases in the year before introduction of vaccination into the national immunisation programme, or in 2017, with the years lived with disability and the years of life lost components shown separately, Netherlands, 1952–2017

---

CCa: cervical cancer (human papillomavirus (HPV)-16/18); DALY: disability-adjusted life year; Diph: diphtheria; HepB: hepatitis B; Hib: invasive Haemophilus influenzae type b disease; HZ: herpes zoster; Mea: measles; MenC/W/B: invasive meningococcal C/W/B disease; Mum: mumps; NIP: national immunisation programme; Pert: pertussis; Pneu: invasive pneumococcal disease (PCV10 types); Pol: poliomyelitis; Rota: rotavirus gastroenteritis; Rub: rubella; Tet: tetanus; Var: varicella; YLD: years lived with disability; YLL: years of life lost.

a Disease burden is expressed in DALYs.
b 1976 because there were no 1975 data.

Left of vertical dashed line: estimates for the year before inclusion in the NIP; right of vertical dashed line: estimates for 2017 for potential NIP candidates. Whiskers indicate 95% uncertainty intervals for DALYs. A higher estimate (orange dot) is given if the burden in DALYs in any of the 5 years preceding the year before inclusion in the NIP is higher than that year. In this case this higher estimate is the highest burden in DALYs in the five preceding years; see manuscript and Supplement 1 for all assumptions and limitations.
because the BoD associated with such infections can no longer be prevented by vaccination.

Cases with unknown age and/or sex were imputed using the univariate method. Prior to introduction of vaccination into the NIP, complete information was not always available on the distribution of incidence by sex and 5-year age group. For example, if the highest age group was ‘20 years or older’ these cases were assigned to the age group ‘20–24 years’ because for most diseases incidence was highest among young people in the pre-vaccination period. If sex was unknown, cases were divided equally between males and females (Supplement 1, Table A2).

Sensitivity analysis and uncertainty

In the baseline analysis the GBD 2010 LE was used. However, in the 1950s LE was ca 10 years lower than nowadays. To allow a comparison of health loss due to a disease across time, the BoD was also estimated with the year-specific Dutch LE [25] for the total population (no distinction men/women; Supplement 1, Table A1), except for cervical cancer for which the Dutch LE 2014 was used [20].

Statistical uncertainty (e.g. due to small sample size) was simulated using Monte Carlo techniques (5,000 iterations were run per disease model) and results are presented as the mean and 95% uncertainty intervals (UI) resulting from the stochastic simulations (see van Lier et al. [10]). DALY estimates were rounded to three
significant digits for numbers ≥ 10,000, to two significant digits for numbers between 10 and 10,000 and to one significant digit for numbers < 10.

Ethical statement
All data used in this study were aggregated and non-identifiable; therefore, ethical approval was not required.

Results
The estimated BoD of various vaccine-preventable diseases in the Netherlands in the year before introduction of vaccination into the NIP, or in 2017, Nederland, 1952–2017, are shown in Figure 1. BoD for varicella alone was estimated at 160 (95%UI: 160–160) DALYs but amounted to 1,800 (95%UI: 1,800–1,900) DALYs including HZ. This was higher than the BoD for rotavirus gastroenteritis (1,100; 95%UI: 440–2,200 DALYs) and meningococcal B disease (620; 95%UI: 490–770 DALYs), the two other potential NIP candidates.

In the year before introduction into the NIP, the estimated BoD was highest for poliomyelitis (27,800; 95%UI: 24,700–30,900 DALYs), rubella (16,900; 95%UI: 14,400–19,500 DALYs) and diphtheria (14,100; 95%UI: 11,500–17,000 DALYs).
Notwithstanding the possibility of severe complications, varicella is generally perceived as having a mild course (4.1 on a scale from 1 = not severe at all to 10 = very severe [29]). Various surveillance data – GP consultations, hospitalisations and mortality – reflect the severity of varicella [13,30,31], which contributes to the overall burden. However, as diseases and their consequences are heterogeneous, it is difficult to compare the overall burden of varicella with the overall burden of other vaccine-preventable diseases. We attempted to address this by using the composite health measure DALY. The current BoD of varicella in the Netherlands expressed in DALYs was relatively low compared with the BoD of most other vaccine-preventable diseases in the year before their introduction into the NIP. However, BoD of varicella and HZ combined was higher than for rotavirus gastroenteritis and meningococcal B disease, two other potential candidates for inclusion in the NIP.

Survey data showed that only a minority of both parents and professionals in the Netherlands favour universal varicella vaccination through the NIP [32-34]. In the most recent study, only 21% of professionals had a positive attitude towards universal varicella vaccination, and 28% of parents had a positive intention to vaccinate their own child against varicella if included within the NIP [34]. Justification for these opinions is that varicella was perceived as being neither important nor severe enough for vaccination to be needed [29,34]. The current BoD estimates may challenge these notions of professionals and parents on the usefulness of varicella vaccination.

Parents’ intentions were more positive towards universal vaccination against rotavirus gastroenteritis (38% positive) and especially meningococcal B disease (83% positive) [29]. According to Veldwijk et al. potential coverage for rotavirus vaccination would range between 23 and 86%, depending on vaccine scenario (vaccine effectiveness, protection duration) and implementation strategy [35].

While childhood vaccination against varicella may result in a reduction of varicella BoD, this might increase HZ BoD in the mid-term because it has been hypothesised that reduced VZV circulation reduces exogenous immune boosting, thereby increasing the probability of HZ [12]. On the other hand, vaccination against varicella will possibly diminish HZ among vaccinated individuals because the vaccine-strain is less likely to reactivate than the wild-type strain [36]. Therefore, cost-effectiveness of varicella vaccination in the Netherlands depends strongly on the impact on HZ and the economic time horizon. In the absence of exogenous immune boosting, vaccination with high coverage is expected to be cost-effective and may even be cost saving, while it is not expected to be cost-effective on reasonable time scales (<100 years) if immune boosting is present [12].

Alternatively, VZV BoD could be reduced by vaccination in middle-age against HZ [37] which could be marginally cost-effective in the Netherlands, depending on the vaccine price [38-40]. HZ vaccination would likely not affect the high VZV circulation among young children and thus maintain the benefits of early infection: less severe disease following primary infection and preventing susceptibility among women of reproductive age.
Diseases situated in the upper right quadrant of Figure 2 – most vaccine-preventable diseases included in the current NIP – have a relatively high BoD at both population and individual level, justifying inclusion in the NIP. Due to the limited number of cases, the Figure might generate more discussion regarding vaccination against hepatitis B and meningococcal W/B disease, but as the individual-level burden of these diseases is high, vaccination can still be relevant. With the availability of an affordable combination vaccine, hepatitis B vaccination was included in the NIP because there was more health gain with universal vaccination compared with the former risk group vaccination, at low additional costs [41]. Vaccination against meningococcal W disease was mainly introduced as an outbreak measure because of the sharp increasing incidence and high case fatality rate [42]. Diseases in the lower right quadrant (high BoD at population level but low BoD at the individual level) raise more discussion (most potential NIP diseases). Diseases in the lower left quadrant would only end up in the NIP when additional costs are very low or when vaccination is cost-saving. Although the BoD of mumps was modest at the time mumps vaccination was added to the NIP, it was included in the MMR vaccine through which vaccination against measles, rubella, and mumps could easily be combined [43]. Rabies (very severe and rare) is an example of a vaccine-preventable disease that would appear in the upper left quadrant.

BoD is the first criterion used by the Health Council of the Netherlands to determine a vaccine’s suitability for inclusion in the NIP. Other criteria cover effectiveness and safety of vaccination, acceptability of vaccination, efficiency of vaccination (including cost-effectiveness), and priority of vaccination [3]. In 2007, the council advised to further review the inclusion of vaccination against varicella, rotavirus gastroenteritis and meningococcal B disease once more information became available [44]. The current study provides valuable information on the BoD of these diseases. Recently, the council recommended vaccination against rotavirus gastroenteritis, while noting that universal vaccination is not cost-effective at current vaccine prices whereas risk-group vaccination is considered to be cost-saving [45,46].

The principal strength of this study is the utilisation of extensive historical data on the incidence and mortality of vaccine-preventable diseases. Furthermore, we applied a standardised BoD methodology using publicly available software and sets of outcome trees.

Interpretation of our findings should recognise several limitations.

First, the year of introduction of vaccination into the NIP differed across diseases; consequently it is not straightforward to compare the situation in the 1950s with the situation of today: healthcare and treatment options, immunisation status of the population and surveillance (notification criteria, laboratory testing) have changed significantly over this period, as has the population demographics. For example, the risk of dying from poliomyelitis or measles is very low nowadays and the proportion of elderly has increased. Regardless of the year of estimation, the same outcome tree was used, as relevant data were not available to adjust the clinical progression probabilities over time (with the exception of adjustments to case-fatality rates when the estimated number of deaths was considerably different from the registered number of deaths). As a consequence, BoD was probably underestimated for diseases for which vaccination was introduced many years ago. At the same time, BoD for these diseases was likely overestimated through use of the same LE (baseline) for all diseases, regardless of the year in which vaccination was introduced. Although LE at birth increased by approximately 10 years since the 1950s [25,47], we considered it unjust to value a life in the 1950s differently from a life today. Despite this concern, the sensitivity analysis showed that using year-specific Dutch LE did not have a large impact on the results.

A second limitation is that BoD was estimated for a single year, even though incidence can fluctuate over time (e.g. outbreaks) (Supplement 1, Figures A1.1–A1.17). An extreme example is diphtheria for which there were tens of thousands of cases annually during the Second World War period. There were fewer, reported cases however, in 1952 (n = 2,805), the year used in this study [48]. For poliomyelitis it is the other way around: BoD was estimated from 2,206 reported cases in 1956, whereas in 1957 only 203 cases were reported [48]. However, we showed that for most diseases, the BoD in the year before introduction of vaccination was highest or very similar to the highest BoD in any of the five preceding years, except from rubella and pertussis.

A third limitation is that only invasive Hib, meningococcal and pneumococcal disease were included in this estimation. Although vaccination was primarily introduced to prevent invasive disease, our estimates only cover a limited part of the total BoD (excluding for example pneumonia and otitis media) caused by these pathogens. A similar observation can be made for HPV: this vaccination is expected to also prevent cancers other than cervical cancer. In addition, cross-protection effects against serotypes not covered by the HPV vaccine [49] were not included.

Finally, the BoD might have been underestimated because of different reasons or assumptions. The BoD of diphtheria, tetanus and pertussis might be underestimated because vaccination already started before introduction of mass vaccination in 1953 [50]. Pertussis BoD was probably also underestimated due to underreporting of deaths [51] (used to estimate incidence), and the assumed age distribution of cases (outbreak 2012) which was probably not comparable to the year 1953 (i.e. higher age of infection in 2012 than in 1953.
as a result of the NIP). The BoD of invasive Hib disease is expected to be underestimated as BoI estimated more cases than were reported [52]. BoD of measles and rubella might be underestimated as well by using reported cases (corrected for underestimation based on outbreaks in 1999–2000 and 2013–14) while almost every child contracted these diseases in the pre-vaccination period [48]. Varicella BoD might be slightly overestimated because we based the incidence on seroprevalence data while some cases might be asymptomatic. However, underestimation of varicella BoD because of not including long-term sequelae due to congenital varicella syndrome is more likely.

Taken together, the BoD results presented in this manuscript must be seen as rough estimates: the exact value of these estimates is less relevant than the ratio of diseases to each other. The latter is less likely to change.

In conclusion, the present-day BoD of varicella – including HZ – in the Netherlands is somewhat lower than the BoD of most vaccine-preventable diseases before their inclusion in the NIP, but higher than the BoD of other potential NIP candidates. Based on established BoD estimation methods, the current approach provides a quantitative evidence base for decision-making regarding the inclusion of new vaccines – such as varicella vaccine – in NIPs. Whereas the introduction of a new vaccine into NIPs is usually assessed in isolation, the current analysis provides insight into the BoD of different diseases in relation to each other, which can be helpful for national advisory committees on immunisation (NITAGs) and policymakers to prioritise vaccination programmes.

Acknowledgements

We would like to thank Maartje Kristensen for the preparatory work on the development of the disease model for varicella, and Mirjam Knol (RIVM)/Arie van der Ende (NRLBM, Netherlands Reference laboratory for Bacterial Meningitis) (bacterial meningitis), Nicole van de Maas (RIVM) (diphtheria, pertussis, tetanus, and poliomyelitis), Irene Veldhuijzen (RIVM) (rubella, measles, mumps, and hepatitis B), Roan Pijnaar/Wilfrid van Pelt (RIVM) (rotavirus gastroenteritis), and Mark Nielen (NIVEL, Netherlands Institute for Health Services Research) (varicella, and herpes zoster) for providing incidence data and/or critical comments on different parameters.

We would like to acknowledge the BCoDE consortium members for developing the methodology and models for BoD estimation (in alphabetical order within institutions): National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands: Arie Havelaar; University Medical Centre Utrecht, Utrecht (UMCU), The Netherlands: John Brooke, Juanita Haagsma, Ardine de Witt; Bielefeld University, Bielefeld, Germany: Alexander Krämer, Paulo Pinheiro, Dietrich Plåß; University of Edinburgh, Edinburgh, United Kingdom: Eric Fève, Cheryl Gibbons; Catholic University of the Sacred Heart, Rome, Italy: Elisabetta Franco, Silvia Longhi, Walter Ricciardi, Chiara de Waure; UMIT–University for Health Sciences, Medical Informatics and Technology, Hall in Tirol, Austria: Beate Jahn, Nikolai Mühlberger, Uwe Siebert; National Institute for Health Development, Tallinn, Estonia: Taavi Lai, Ardo Matsi, Kristi Rüütel; European Centre for Disease Prevention and Control (ECDC), Stockholm, Sweden: Alessandro Cassini, Edoardo Colzani, Piotr Kramarz. Four of the authors were also members of the BCoDE consortium: National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands; Mirjam Kretzschmar, Alies van Lier, Scott McDonald; University Medical Centre Utrecht, Utrecht (UMCU), The Netherlands: Marie-Josée Mangen. We thank Daniel Lewandowski (NextPage Software) for the development of the BCoDE toolkit, the software used to calculate the burden of disease for this manuscript.

Funding: This study was supported by the National Institute for Public Health and the Environment (RIVM). No additional funding has been acquired.

Conflict of interest

None declared.

Authors’ contributions

AvL, BdG, SM, MJJM and MK were involved in the development of the different disease models and performed the BoD estimations; AvL estimated the burden of varicella and herpes zoster and the burden of all diseases before introduction of vaccination into the NIP (except from cervical cancer and rotavirus gastroenteritis), SM estimated the burden of cervical cancer, and MJJM estimated the burden of rotavirus gastroenteritis. MvW and HvV assisted in the search for historical incidence data and interpreted the data. AvL drafted the manuscript together with BdG, SM and MJJM. MvW, MK, ES, HvV and HdM interpreted the data and critically revised the manuscript. All authors read and approved the final manuscript. All authors have given final approval of the version to be published and are publicly responsible for its contents.

References


50. Hoogendoorn D. De inenting tegen diphtherie en de propaganda hiervoor. [Vaccination against diphtheria and propaganda to this end], Ned Tijdschr Geneeskd. 1954;98(26):1806-9. Dutch. PMID: 13185342


License, supplementary material and copyright

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence and indicate if changes were made.

Any supplementary material referenced in the article can be found in the online version.

This article is copyright of the authors or their affiliated institutions, 2019.
Molecular identification of the source of an uncommon tularaemia outbreak, Germany, autumn 2016

Daniela Jacob1,2, Kristin Köppen2,3, Aleksandar Radonić4, Berit Haldemann5, Philipp Zanger6,7,8, Klaus Heuner4,5, Roland Grunow1

1. Highly Pathogenic Microorganisms (ZBS 2), Centre for Biological Threats and Special Pathogens, Robert Koch Institute, Berlin, Germany
2. These authors contributed equally to this work
3. Cellular Interactions of Bacterial Pathogens, ZBS 2, Robert Koch Institute, Berlin, Germany
4. Genome Sequencing (MF 2), Methodology and Research Infrastructure, Robert Koch Institute, Berlin, Germany
5. Bioinformatics (MF 1), Methodology and Research Infrastructure, Robert Koch Institute, Berlin, Germany
6. Federal State Agency for Consumer & Health Protection Rhineland-Palatinate, Koblenz, Germany
7. Department of Infectious Diseases, Medical Microbiology and Hygiene, University Hospitals, Heidelberg, Germany
8. Heidelberg Institute of Global Health, Unit of Epidemiology and Biostatistics, University Hospitals, Heidelberg, Germany

Correspondence: Roland Grunow (GrunowR@rki.de)

Citation style for this article:

Background: In 2016, an uncommon outbreak of oropharyngeal tularaemia involving six human cases occurred in Germany, caused by drinking contaminated fresh must after a grape harvest. Aim: We describe the details of laboratory investigations leading to identification of the outbreak strain, its characterisation by next generation sequencing (NGS) and the finding of the possible source of contamination. Methods: We incubated wine samples in different media and on agar plates. NGS was performed on DNA isolated from young wine, sweet reserve and an outbreak case’s lymph node. A draft genome of the outbreak strain was generated. Vertebrate-specific PCRs using primers targeting the mitochondrial cytochrome b gene and product analyses by blast search were used to identify the putative source of must contamination. Results: No bacterial isolate could be obtained. Analysis of the draft genome sequence obtained from the sweet reserve attributed this sequence to Francisella tularensis subsp. holarctica, belonging to the B.12/B.34 phylogenetic clade (erythromycin-resistant biovar II). In addition, the DNA sequence obtained from the case’s isolate supported our hypothesis that infection was caused by drinking contaminated must. The vertebrate-specific cytochrome b sequence derived from the young wine and the sweet reserve could be assigned to Apodemus sylvaticus (wood mouse), suggesting that a wood mouse infected with F. tularensis may have contaminated the must. Conclusion: The discovered source of infection and the transmission scenario of F. tularensis in this outbreak have not been observed previously and suggest the need for additional hygienic precautionary measures when processing and consuming freshly pressed must.

Introduction
Francisella tularensis, a facultative intracellular Gram-negative bacterium, is the causative agent of tularaemia, a zoonotic disease. Outbreaks in humans are often associated with exposure to infected animals, contaminated water or aerosols, and different arthropod vectors [1-5]. The clinical manifestation mainly depends on the route of infection, and the two main subspecies—F. tularensis subsp. tularensis and subsp. holarctica—are clinically relevant [1,6].

In Germany, F. tularensis is endemic [7-14] and 20–40 cases of tularaemia are reported per year, with numbers increasing since 2005, indicating that tularaemia is a rare but re-emerging disease [7]. The only Francisella subspecies known to cause tularaemia in Germany is F. tularensis subsp. holarctica (Fth). In addition, a further Francisella species (Francisella sp. strain W12–1067, environmental isolate) was identified in Germany, but it is not yet known whether this species is pathogenic for humans [15].

On 2 October 2016, there was an unusual outbreak of oropharyngeal tularaemia involving six cases in a group of 29 persons attending a grape harvest in Rhineland-Palatine, Germany [16]. Grapes collected by a mechanical harvester were pressed at the winery and participants had the opportunity to consume the fresh must at the end of the harvest. Because tularaemia was not initially considered as a possible differential diagnosis, a delay of about 5 weeks occurred in confirming the diagnosis of tularaemia. One of the six serologically confirmed tularaemia cases had complicated tularaemia and was hospitalised with pharyngitis and cervical abscess-forming lymphadenopathy.
In order to identify and characterise the causative agent of the outbreak, we analysed the contaminated must-derived products, sweet reserve (SR) and young wine (YW), as the contaminated must was no longer available for investigation. The SR is fumigated must that already contains low amounts of ethanol and the YW is must with added yeasts to start the fermentation process. In addition, we used lymph node material (PL) from the patient with complicated tularemia for further investigation [16].

The aim of this report was to describe details of the laboratory investigations leading to the finding of the likely source of contamination of the must.

Moreover, by using next generation sequencing (NGS), we further characterised the outbreak strain and confirmed the presence of its DNA in the clinical material of one of the patients.

### Methods

#### DNA extraction

We obtained an aspirate lymph node sample from one of the six patients of the must-associated outbreak who had a protracted clinical course with abscess-forming lymphadenitis (Wetzstein N, Wolf T, personal communication, December 2016). The PL sample was collected directly after suspicion of the outbreak, i.e. 6 weeks after the event leading to the outbreak, which had not been recognised earlier. The contaminated must-derived products (sort 1A [16]), sweet reserve (RKI-sample number SR; A-856/3 (from sort 1A) and young wine (YW; A-856/2; from sort 1A), were collected by public authorities 3 weeks after the outbreak and were investigated. An additional YW (A-856/1; from sort 1B, pressed directly after sort 1A [16]) was also collected and studied. A total of 200 mL of each specimen were concentrated by centrifugation (45 min, $4,500 \times g$). Each pellet (SR, very thin pellet; YW, more pellet material, mainly yeast cells) was resuspended in 2 mL supernatant and further centrifuged at $20,000 \times g$ for 5 minutes. Each pellet was then resuspended in 100 µL DNA.
extractions were performed from the resuspended pellets (100 µL) and from 100 µL of aspirate (lymph node fluid) from the patient’s left neck lymph node (PL; A-877/1), according to the protocol for Gram-negative bacteria in the manufacturer’s instructions, using either the MoBio Power Soil DNA isolation kit (MoBio Laboratories, Carlsbad, California (CA), United States (US)) or the Qiagen DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany), respectively. During the extraction, at the point of the proteinase K digestion, a plasmid with an artificial sequence insert was added to each sample as an internal extraction and amplification control (pKoMa2 [17]). Elution of DNA was performed twice with 75 µL (SR, YW) and 50 µL (PL) of Aqua bidest, respectively.

Singleplex and multiplex real-time PCRs, RDI-PCR

Multiplex real-time PCR (5’ nuclease assay, TaqMan technology) targeting fopA and tul4 specific for F. tularensis, in combination with the extraction and amplification control targeting KoMa2, were performed with oligonucleotides and probes as described in Table 1. A singleplex real-time PCR assay was performed from the clinical human sample for the detection of c-myc as a process control (oligonucleotide and probe; Table 1). Both real-time PCR assays were run in a total volume of 25 µL, including 5 µL of DNA of the samples to be analysed. Samples were analysed in duplicate in each run. The reaction mix components were 6.25 µL TaqMan Environmental MasterMix 2.0 (ThermoFisher, Henningsdorf, Germany), 10 pmol/µL primers (0.75 µL each) and probes (0.25 µL each). Amplification was performed in an Applied Biosystems 7500 Real-Time PCR System (ThermoFisher Scientific, Langenselbold, Germany), each run with an initial denaturation step at 95 °C for 10 minutes, followed by 40 cycles containing a denaturation step at 95 °C for 15 seconds and a combined primer annealing and elongation step at 60 °C for 60 seconds.

The calculation of genome equivalents for tul4 and c-myc was done for the SR, YW and PL by using the plasmids of the TOPO TA vector cloning kit (Invitrogen, Karlsruhe, Germany), containing the respective target region for tul4 of F. tularensis or c-myc as quantitative standards. For this purpose, in each real-time PCR run, standard control plasmids at Table 2 canSNP analysis of the nucleotide sequences of Fth-Must and Fth-Patient, tularaemia outbreak, Germany, autumn 2016

<table>
<thead>
<tr>
<th>Clade or SNP</th>
<th>Ancestral</th>
<th>Derived</th>
<th>LVS</th>
<th>Must</th>
<th>Patient</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.4</td>
<td>AAATCC1GCAGCAAA</td>
<td>AAATCCaGCAGCAAA</td>
<td>A</td>
<td>A</td>
<td>NA</td>
</tr>
<tr>
<td>B.5</td>
<td>GCCAACAaCTTTAGCTGA</td>
<td>GCCACAAGtTTTAGCTGA</td>
<td>D</td>
<td>D</td>
<td>NA</td>
</tr>
<tr>
<td>B.6</td>
<td>CCGTGCtACAGAACTAT</td>
<td>CCGTGCTAtAGAAGCTAT</td>
<td>A</td>
<td>A</td>
<td>NA</td>
</tr>
<tr>
<td>B.12</td>
<td>GTCAATATACGAAATGTT</td>
<td>GTCAATATACGAAATGTT</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>B.72</td>
<td>CTTCAATAGAGCTTATCCT</td>
<td>CTTCAATAGACCCAGCTT</td>
<td>D</td>
<td>D</td>
<td>NA</td>
</tr>
<tr>
<td>B.71</td>
<td>GGTGCAATTTAGCTTA</td>
<td>GGTGCAATTTAGCTTA</td>
<td>D</td>
<td>D</td>
<td>NA</td>
</tr>
<tr>
<td>B.13</td>
<td>GCGGTACACTTAGATG</td>
<td>GCGGTACACTTAGATG</td>
<td>A</td>
<td>A</td>
<td>NA</td>
</tr>
<tr>
<td>B.39</td>
<td>CTTCAATCGaCGACCT</td>
<td>CTTCAATCGaCGACCT</td>
<td>A</td>
<td>A</td>
<td>NA</td>
</tr>
<tr>
<td>B.26</td>
<td>GGTGCCATTTGCTCT</td>
<td>GGTGCCATTTGCTCT</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>B.43</td>
<td>ATCTAGTGcCTTTCTCA</td>
<td>ATCTAGTGcCTTTCTCA</td>
<td>A</td>
<td>A</td>
<td>NA</td>
</tr>
<tr>
<td>B.23</td>
<td>CGCCCTCTAAGTATCTT</td>
<td>CGCCCTCTAAGTATCTT</td>
<td>A</td>
<td>A</td>
<td>NA</td>
</tr>
<tr>
<td>B.42</td>
<td>GGTGAAATGATGCAA</td>
<td>GGTGAAATGATGCAA</td>
<td>D</td>
<td>D</td>
<td>NA</td>
</tr>
<tr>
<td>B.21</td>
<td>TATATATAGGTAGTGC</td>
<td>TATATATAGGTAGTGC</td>
<td>A</td>
<td>A</td>
<td>NA</td>
</tr>
<tr>
<td>B.33</td>
<td>CGCCAAAGACCTACTT</td>
<td>CGCCAAAGACCTACTT</td>
<td>D</td>
<td>D</td>
<td>NA</td>
</tr>
<tr>
<td>B.34</td>
<td>GCTGATAcTAGAAGAA</td>
<td>GCTGATAcTAGAAGAA</td>
<td>A</td>
<td>D</td>
<td>NA</td>
</tr>
<tr>
<td>B.75</td>
<td>CTCTTAGGCCTAAAACC</td>
<td>CTCTTAGGCCTAAAACC</td>
<td>A</td>
<td>A</td>
<td>NA</td>
</tr>
</tbody>
</table>

A: ancetral; D: derived; Fth: F. tularensis subsp. holarctica; LVS: live vaccine strain; NA: no reads available; SNP: single nucleotide polymorphism.

A:3 Two SNPs (C → T; LVS position bp 78,650, FTL_0082 and A → G, LVS position bp 1,011,166, FTL_1056) were only found in eight sequenced Fth B.33 strains.

B.12 is associated with erythromycin resistance; in addition, rrl.1 and rrl.2 SNPs in the 23S RNA gene rrl (A453G and A2059C, respectively; Escherichia coli numbering) are directly associated with the erythromycin resistance of Fth strains of clade B.12 [30].
the concentration of $10^2$, $10^4$ and $10^6$ copies/25 µL for the different targets (tul4, fopA and c-myc) were added to generate target-specific standard curves that allow the calculation of the quantity of samples.

**PCR for Francisella tularensis subspecies differentiation**

The block PCR of the region of difference 1 (RD1-PCR) was used for the subspecies differentiation of *F. tularensis*. The PCR was carried out using the DreamTaq Polymerase (ThermoFisher, Hennigsdorf, Germany) with 15–100 ng of template DNA, according to the protocol described by Broekhuijsen et al. [18].

**Next generation sequencing**

For the sequencing of the YW and SR samples (both of sort 1A) and PL, Illumina sequencing in combination with Nextera XT library generation was used (Illumina, San Diego, CA, US). DNA was quantified by using the Qubit dsDNA HS Assay Kit (Life Technologies, Carlsbad, CA, US). Library size was determined by using the High Sensitivity DNA Analysis Kit (Illumina, San Diego, CA, US). Libraries were quantified by using the Qubit dsDNA HS Assay Kit (Life Technologies, Carlsbad, CA, US). Ten µL of the preprocessed DNA samples was sequenced 300 + 300 bases in paired-end mode.

**Mapping and generation of consensus sequence**

Sequence quality assessment and trimming was performed with our quality control pipeline QCumber, developed in house [19], which combines the external tools Trimmomatic [20] and FastQC [21]. Furthermore, the pipeline uses the tools Bowtie2 [22] and Kraken [23] to analyse the origin of sequenced reads. Results of the taxonomic classification performed by Kraken on a Kraken-customised database consisting of bacterial, archaeal, viral and fungal genomes were visualised with Krona [24].

Trimmed reads were then mapped to the reference genomes of the expected background organisms (for SR and YW: yeast genome (GCA_000146045.2) and grape genome (GCA_000003745.2); for the patient sample: human genome (GRCh38) using Bowtie2). Reads not mapping to the background organisms’ genomes were then mapped to the *F. tularensis* vaccine strain (LVS) (NC_007880.1) genome and a consensus sequence (draft genome sequence) was generated using Geneious (version R9.1.3 [25]; with a threshold of 75% and following the IUPAC code for ambiguities). The draft genome sequence generated from DNA isolated from SR (*Fth-Must*) has been submitted to GenBank (CP024807) and raw reads have been uploaded to the Sequence Read Archive (SRA; https://www.ncbi.nlm.nih.gov/sra; BioProject PRJNA417909).

**Phylogenetic analysis**

The draft genome sequence of *F. tularensis* was aligned with eight *F. tularensis* reference genomes (OSU18, FTNF002–00, FDC409, FSC162, FDC407, FSC200, LVS and FDC408) and four draft genomes of *F. tularensis* isolates (A-635, Fth-07, A-810/1 and A-663), representing main clades [26,27] using progressive Mauve alignment (MUSCLE 3.6). The phylogenetic tree was constructed by Geneious (version 10.0.5) using the neighbour-joining method (Tamura–Nei, outgroup: OSU18; 100 bootstrap replicates). The canSNP analysis was performed using published canSNP positions [26,27] to confirm the clade and subclade determination of the draft genome sequence of *F. tularensis*. Sequences used are given in Table 2.

**PCR detecting vertebrate cytochrome b and DNA cloning**

A vertebrate-specific PCR assay was used to identify mammalian DNA within all SR and YW samples. As described by Kent and Norris, the primers UNFOR03 and UNREV0125 (Table 1) specifically detect the mammalian mitochondrial cytochrome b gene [29]. Therefore, a PCR was performed using the TopTaq DNA polymerase (Qiagen, Hilden, Germany), according to manufacturer’s instructions (for each PCR reaction (50µL): 5µL of 10x TopTaq PCR Buffer, 1µL of dNTPs (Sigma Aldrich, St. Louis, Missouri, US), 10µL of 5x Q-Solution, 1µM UNFOR03, 1µM UNREV0125, 0.5µM of TopTaq DNA Polymerase and 22.5 µL of RNase-free water). Ten µL of the preprocessed DNA samples was used as a DNA template in each PCR reaction. As a positive control, 2 µL of sheep blood was directly pipetted into the PCR reaction. Chromosomal DNA of *Francisella* sp. strain W12–1067 and water served as negative controls. PCR amplification was done using a Thermocycler TRIO-Thermoblock (Biometra, Göttingen, Germany) involving initial denaturation (3 minutes,
94 °C), then 35 cycles including a denaturation step (30 seconds, 94 °C), an annealing step (30 seconds, 57 °C) and an extension step (1 minute, 72 °C), followed by a final extension (10 minutes, 72 °C). Subsequently, PCR products were separated according to their size in a gel electrophoresis. The expected PCR fragment (623 bp) was extracted using Wizard SV Gel and PCR Clean-Up System (Promega, Madison, Wisconsin (WI), US).

The isolated DNA was then cloned into a vector for subsequent sequencing of the insert DNA. For the cloning of the UNFOR403–UNREV1025 PCR product, the pGEM-T Easy Vector system was used according to the manufacturer’s instructions (Promega, Madison, Wisconsin (WI), US). Briefly, 5 µL of gel-purified PCR product was ligated overnight into pGEM T Easy vector using DNA T4 ligase. Next, 2 µL of the ligation reaction was transformed into chemical competent Escherichia coli cells (Top10 cells: ThermoFisher, Waltham, MA, US). Recombinant E. coli cells were selected by growing on Luria-Bertani (LB) agar plates containing 100 µg/mL ampicillin, 0.1 mM IPTG and 0.006% X-Gal. White clones were tested in PCR using the primer combination UNFOR403 and UNREV1025. Insert DNA of all clones was sequenced and a BLAST analysis was performed.

Results
Analyses of must products and patient samples
We incubated samples of SR and YW in medium T, on cystein-heart-blood-agar (CHAB) plates and on cystein-heart-blood-blood-agar (CBB) plates.
CHAB agar plates containing selective antibiotics (CHAB-PACCV) [27], but we were not able to obtain a *Francisella* strain. *Fth* DNA was detected in samples of SR and YW by real-time PCR targeting *tul4* and *fopA*. The subspecies was identified and confirmed by RD1-PCR analysis in both YW and SR (Figure 1A, lanes 1–3), as well as in the PL aspirate sample (Figure 1B, lane 6). The obtained DNA sample was used for NGS analysis to identify the outbreak strain. The PL sample and the SR and YW samples (sort 1A and 1B) were analysed by quantitative PCR analysis and revealed the following genome equivalents per mL: SR, $1.0 \times 10^2$ (sort 1A and 1B) were mapped to the yeast genome (GCA_000014604.5). Subsequently, unmapped reads were mapped to the grape genome (GCA_000003745.2). The unmapped reads obtained (YW: 3,674,761 and SR: 17,709,501) were mapped to the genome of *Fth* strain LVS (NC_007880.1). Of the reads of YW and SR, 1.9% (71,201) and 9.6% (1,696,328) were mapped to the *Francisella* genome. However, there was still a high number of unmapped and unclassified reads, mainly in the SR sample. The results demonstrate that both the YW and the SR were contaminated with *Francisella* DNA. Reads from the SR sample for the *Fth* LVS genome were used to generate a consensus sequence of *Fth*-SR (*Fth*-Must). The approximately $1.7 \times 10^6$ mapped reads are distributed over the whole genome of *Fth* LVS (1,696,328 reads 50–238 bp in length, mean: 187±53 bp; mean coverage: 167) providing a first draft *Fth* genome sequence (*Fth*-Must, 1,895,952 bp). This DNA sequence was aligned with whole genome sequences of different *Fth* strains and was used to generate a phylogenetic tree (Figure 2). The *Fth*-Must DNA was found to cluster with another human *Fth* isolate (A810–1) from Germany in the phylogenetic subclade B.34 (Figure 2). In addition, we performed an in silico analysis of the canSNP analysis scheme, confirming that isolate *Fth*-Must corresponded to the B.12 clade (erythromycin-resistant strains) and to subclade B.34 (Table 2). The affiliation to the B.12 clade was further confirmed by the identification of two SNPs within the 23S rRNA gene *rrl* that were recently found to be specific for strains belonging to this clade [30].

The alignment was performed using progressive Mauve alignment (MUSCLE 3.6) and the phylogenetic tree was built with Geneious (Geneious 10.0.5) (Tamura–Nei, Neighbour-Joining, 100 bootstrap replicates), setting *Fth* OSU18 strain as outgroup. The branch labels indicate the posterior probability and the branch length corresponds to the evolutionary distance (substitutions per site). For details of canSNP analysis B.X clusters and subclusters, see Table 2 and references [26–28]. German federal states: Brandenburg (BB); Mecklenburg-Western Pomerania (MV); Lower Saxony (NI); Rhineland-Palatinate (RP). Countries: France (FR), Germany (DE), Lithuania (LT), and Sweden (SE).

In addition, NGS sequencing of the PL-DNA generated 22,116,655 reads and, as expected, 98.6% of the reads mapped to the human genome sequence (GRCh38).
However, 691 and 690 reads (covering ca 107,073 bp of the whole \textit{Fth} LVS genome sequence) mapped to the genome of \textit{Fth} LVS and \textit{Fth}-Must, respectively. We analysed the obtained reads for the presence of canSNPs used for subtyping and found one read mapping to the B.12 canSNP and two reads mapping to two respective SNPs only found in strains of subclade B.33 (Table 2). In addition, three reads mapped to two of three copies (present at the genome of \textit{Francisella}) of SNP \textit{rrl}-2 in the \textit{rrl} gene (Table 2). The results confirmed that the isolated \textit{Francisella} DNA belonged to a \textit{Francisella} strain of clade B.12 associated with erythromycin resistance \cite{30} and at least to sub-clade B.33 (Figure 2), although the low coverage did not allow a clear identification. Further analyses could not be performed due to the restricted amount of DNA. Altogether, the results suggested that this patient was infected with the same \textit{Fth} strain that was identified in the SR and served as must during the grape harvest \cite{16}.

Identification of the putative contamination source

Despite the identification of the outbreak strain DNA, the question of how the must had been contaminated with \textit{Francisella} was still not answered. Since small rodents are occasionally found in mechanically harvested grapes \cite{16}, we used the unmapped reads of the last mapping step (using DNA from the SR, see above) and performed an additional mapping to a mouse genome (GCA_000001635.7). About 2,600 reads were found mapping to this mouse genome, but a definite identification of the species was not successful. However, based on the obtained results, we analysed the samples for the presence of vertebrate DNA to identify the species that may have contaminated the must with infectious \textit{Francisella}.

For this purpose, we used a vertebrate-specific primer pair to amplify specifically the mitochondrial cytochrome b gene if present in the must. The different sample DNAs (A-856/1–3) were used in the PCR reaction. A PCR band of the expected size (623 bp) could be detected in YW and SR samples of sort 1A.
Samples A-856/1–3 (lanes 1–3 and lanes 7–8) were analysed in PCR with primers detecting vertebrate cytochrome b (UNFOR403, UNREV6025, fragment size 623 bp). As a positive template control, sheep blood (PTC, lane 4) was used. Chromosomal DNA of Francisella sp. strain W12–1067 (NTC, lane 5) and water (NTC, lane 6) served as negative template controls. DNA ladder (M): GeneRuler 1 kb DNA Ladder. Sample A-856/1=YW sort 1B, A-856/2=YW sort 1A, A-856–3=SR sort 1A.

The alignment was performed using progressive Mauve alignment (MUSCLE 3.6) and the phylogenetic tree was built with Geneious (Geneious 10.0.5) (Tamura–Nei, neighbour-joining, 100 bootstrap replicates), setting Mus musculus as the outgroup. The branch length corresponds to the evolutionary distance (substitutions per site). cyt b: cytochrome b gene; Hap.: haplotype.

Discussion
Freshly pressed must was served to some of the 29 participants in the grape harvest [16]. Wine yeast was added to a large portion of the must for production of young wine and a smaller portion was fumigated and served as sweet reserve. In the SR and YW of sort 1A, Francisella DNA was detected at a high concentration and was confirmed to be Fth specific by PCR analysis (Figure 1). In sort 1B, pressed directly after sort 1A, only 440 genome equivalents per mL of F. tularensis was detected, suggesting a cross-contamination of this wine in the winepress [16].

NGS sequencing of DNA isolated from the YW (sort 1A) and SR revealed, as expected, more reads mapping to the yeast genome in the YW (8.6%) than in the SR (0.5%). In both unmapped reads, an equal but small proportion (ca 1%) of reads was mapped to the grape genome, suggesting that there was a low amount of grape DNA in the YW and SR. In addition, in the obtained unmapped reads, 1.9% and 9.6% of the YW and SR reads mapped to the Fth LVS genome, respectively, demonstrating that there was still a lot of Francisella DNA present in these samples. Half of the unmapped reads from the SR did not map to any known sequences in the Kraken customised database, and some reads mapped to different plant- and soil-associated bacteria, probably environmental and grape-associated bacteria.

After NGS sequencing of DNA from the SR, we were able to generate a consensus sequence (draft genome, Fth-Must) covering nearly the whole genome of Fth LVS. This draft genome may contain regions shared between different organisms in the sample. However, the distribution of mapped NGS reads, as well as the phylogenetic tree and canSNP analysis, demonstrated a good quality of the generated consensus sequence. The phylogenetic analysis of the Fth-Must DNA sequence revealed that the DNA belonged to a strain that clustered into subclade B.34 (Figure 2). Although the obtained reads of the NGS analysis of PL-DNA covered only a small part of the Fth-Must sequence, the results confirmed the hypothesis that the patient from the outbreak had been infected by drinking must. Further, in 2016, we identified a Francisella isolate (A-810/1, data not shown) from a patient in Lower Saxony who had contracted tularemia, confirming the presence of the subclade B.34 in Germany (Figure 2).

To identify the source that contaminated the must with Fth, we performed a vertebrate-specific cytochrome b gene PCR, a conserved mitochondrial gene used for phylogenetic investigations [31]. Sequencing of the obtained PCR products revealed a nucleotide sequence that was 99.8% identical to the cytochrome b gene of A. sylvaticus haplotype Germany-1/France, a wood mouse. A. sylvaticus is known to consume fruit, and it has been reported that this species can be infected with Francisella [32]. Different studies in Croatia, Germany, Hungary and Spain demonstrated that different small rodents, like Myodes (bank voles), Sorex (common shrew), Microtus (common vole), Muscardinus(common dormouse) and various species of Apodemus (A. flavicollis, A. agrarius and A. sylvaticus) can be infected by or can be carriers of Francisella tularensis [5,10,32-34]. The findings from these studies support our hypothesis of an infected wood mouse as the source that contaminated the must. Unfortunately, the late suspicion of the tularemia outbreak did not allow the isolation of the outbreak-causing strain from the must, the patient or the suspected mouse for further functional investigation. It should be emphasised that modern laboratory techniques made it possible that all genomic characterisation could be obtained from DNA only. More data on the occurrence of F. tularensis in the region, including in rodents...
and other wild animals, would be helpful for further risk assessment and greater awareness of tularemia during wine production.

Conclusion
Analysing this uncommon tularemia outbreak, we were able to determine a draft genome sequence of the responsible *Francisella* strain, although no isolate could be obtained. Using this draft genome, a phylogenetic analysis was successful. Some reads exhibiting specific canSNPs identified in the DNA extracted from a patient’s lymph node supported the finding of our previous cohort study that the patients were infected by consuming the fresh must. In addition, through the identification of the putative source of the contamination, we could propose a most likely route of transmission for this outbreak: The automatic harvester may have collected a wood mouse (or its carcass) infected with a high dose of *Fth*, then transferred it to the mash car, contaminating the mash, the press and finally 730 L of must—an infectious dose for humans. Subsequently, this must was served to a group of participants in the grape collection and six people contracted tularemia. Based on our results, it was suggested that additional hygienic precautions should be undertaken during wine harvesting and production. For example, rodent control should be put into practice throughout all steps of wine production and freshly pressed must for tasting should be produced from hand-picked instead of mechanically harvested wine grapes, since the latter is more difficult to control. As raw food products can be associated with a risk for infectious agents, pasteurisation before consumption is also recommended [35]. Further, our investigation shows that tularemia should be considered when individuals fall ill with relevant symptoms after a grape harvest event.

Acknowledgements
We would like to thank Donald Knautz, Christina Rothe and Gudrun Just-Nübling for sample collection and shipping and Kerstin Rydzewski, Iris Klein, Silke Becker, Petra Lochau and Julia Hinzmann for excellent technical support. Furthermore, we thank Piotr Wojciech Dabrowski from MF 1 for submission of the NGS and DNA sequence data to the European Nucleotide Archive.

Conflict of interest
None declared.

Authors’ contributions
KK, AR and BH: performed the experiments. DJ, KH and RG: analysed the data and drafted the manuscript. PZ: critically revised the manuscript.

References


License, supplementary material and copyright
This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence and indicate if changes were made.

Any supplementary material referenced in the article can be found in the online version.

This article is copyright of the authors or their affiliated institutions, 2019.
Genetic diversity and delineation of *Salmonella* Agona outbreak strains by next generation sequencing, Bavaria, Germany, 1993 to 2018

Alexandra Dangel¹-², Anja Berger¹-², Ute Messelhäußer¹, Regina Konrad¹, Stefan Hörmansdorfer¹, Nikolaus Ackermann¹, Andreas Sing¹

¹. Bavarian Health and Food Safety Authority (LGL), Oberschleissheim, Germany
². These authors contributed equally to this article

Correspondence: Alexandra Dangel (alexandra.dangel@lgl.bayern.de)

Citation style for this article:

**Background:** In 2017, a food-borne *Salmonella* Agona outbreak caused by infant milk products from a French supplier occurred in Europe. Simultaneously, *S*. Agona was detected in animal feed samples in Bavaria. **Aim:** Using next generation sequencing (NGS) and three data analysis methods, this study’s objectives were to verify clonality of the Bavarian feed strains, rule out their connection to the outbreak, explore the genetic diversity of Bavarian *S*. Agona isolates from 1993 to 2018 and compare the analysis approaches employed, for practicality and ability to delineate outbreaks caused by the genetically monomorphic Agona serovar. **Methods:** In this observational retrospective study, three 2017 Bavarian feed isolates were compared to a French outbreak isolate and 48 *S*. Agona isolates from our strain collections. The later included human, food, feed, veterinary and environmental isolates, of which 28 were epidemiologically outbreak related. All isolates were subjected to NGS and analysed by: (i) a publicly available species-specific core genome multilocus sequence typing (cgMLST) scheme, (ii) single nucleotide polymorphism phylogeny and (iii) an in-house serovar-specific cgMLST scheme. Using additional international *S*. Agona outbreak NGS data, the cluster resolution capacity of the two cgMLST schemes was assessed. **Results:** We could prove clonality of the feed isolates and exclude their relation to the French outbreak. All approaches confirmed former Bavarian epidemiological clusters. **Conclusion:** Even for *S*. Agona, species-level cgMLST can produce reasonable resolution, being standardisable by public health laboratories. For single samples or homogeneous sample sets, higher resolution by serovar-specific cgMLST or SNP genotyping can facilitate outbreak investigations.

**Introduction**
Salmonellosis is one of the most common food-borne human diseases. It is often transmitted via contaminated meat, eggs or seafood products. Moreover, due to the robustness of *Salmonella* spp., dried products like herbs or spices have also proven their potential as vehicle of infection. Of more than 2,600 different serovars of *S. enterica*, only a few non-typhoidal serovars are responsible for most human infections. In *S. enterica* subsp. *enterica*, these serovars include for example Enteridjis, Typhimurium and Agona. In Europe, *S*. Agona is far from leading the list of pathogenic serovars, as cases of *S*. Enteridjis and *S*. Typhimurium are much more numerous [1]. Globally, *S*. Agona is a common pathogen and food-borne outbreaks connected to it have been consistently reported in several countries. Examples are a 2002–03 outbreak, caused by aniseed-fennel-caraway tea products affecting 77 patients in Germany [2,3], a 2008 outbreak connected to meat products of one supplier causing 163 infections in 10 different countries with most cases in the United Kingdom [4], a 2011 multi-state outbreak in the United States (US) caused by fresh papaya resulting in more than 100 infected patients [5], or a point-sourced outbreak caused by tuna sushi in Sydney, Australia in 2015 [6].

In outbreak investigations, serotyping and phage typing have been used for decades in many laboratories including reference laboratories. Serotyping is still serving as a gold-standard technique for routine typing. In combination with other typing techniques like phage typing, it may be suited for the investigation of small, geographically limited outbreaks [7]. However, many serovars are polyphyletic and serotyping sometimes confounds genetically unrelated isolates and thus does not recognise evolutionary groupings in some cases. Therefore, attempts were made some
## Table 1a
Characteristics of sequenced Bavarian *Salmonella* Agona isolates and French representative outbreak isolate, Germany, 1993–2018 (n = 52)

<table>
<thead>
<tr>
<th>Sample name</th>
<th>Year</th>
<th>Material</th>
<th>Isolation source</th>
<th>Country/state</th>
<th>Country of origin/travel history</th>
<th>S. <em>enterica</em> cgMLST CT</th>
<th>cgMLST cluster</th>
<th>Epidemiological link</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERR2219379a</td>
<td>2017</td>
<td>NA</td>
<td>Human</td>
<td>France</td>
<td>NA</td>
<td>704</td>
<td>ND</td>
<td>NA</td>
</tr>
<tr>
<td>SA0001</td>
<td>2011</td>
<td>Stool</td>
<td>Human</td>
<td>Germany/Bavaria</td>
<td>Iraq</td>
<td>1195</td>
<td>2</td>
<td>Travel cluster</td>
</tr>
<tr>
<td>SA0002</td>
<td>2011</td>
<td>Stool</td>
<td>Human</td>
<td>Germany/Bavaria</td>
<td>Iraq</td>
<td>1215</td>
<td>ND</td>
<td>NA</td>
</tr>
<tr>
<td>SA0004</td>
<td>2012</td>
<td>Stool</td>
<td>Human</td>
<td>Germany/Bavaria</td>
<td>Afghanistan</td>
<td>1195</td>
<td>2</td>
<td>Travel cluster</td>
</tr>
<tr>
<td>SA0005</td>
<td>2012</td>
<td>Stool</td>
<td>Human</td>
<td>Germany/Bavaria</td>
<td>Madagascar</td>
<td>1209</td>
<td>None</td>
<td>NA</td>
</tr>
<tr>
<td>SA0006</td>
<td>2013</td>
<td>Stool</td>
<td>Human</td>
<td>Germany/Bavaria</td>
<td>Germany</td>
<td>1210</td>
<td>ND</td>
<td>NA</td>
</tr>
<tr>
<td>SA0007</td>
<td>2013</td>
<td>Stool</td>
<td>Human</td>
<td>Germany/Bavaria</td>
<td>Russia</td>
<td>1211</td>
<td>ND</td>
<td>NA</td>
</tr>
<tr>
<td>SA0008</td>
<td>2013</td>
<td>Bacterial strain</td>
<td>Human</td>
<td>Germany/Bavaria</td>
<td>Germany</td>
<td>1212</td>
<td>ND</td>
<td>NA</td>
</tr>
<tr>
<td>SA0009</td>
<td>2013</td>
<td>Bacterial strain</td>
<td>Human</td>
<td>Germany/Bavaria</td>
<td>Germany</td>
<td>1199</td>
<td>6</td>
<td>NA</td>
</tr>
<tr>
<td>SA0011</td>
<td>2014</td>
<td>Stool</td>
<td>Human</td>
<td>Germany/Bavaria</td>
<td>Unspecified foreign country</td>
<td>1214</td>
<td>8</td>
<td>NA</td>
</tr>
<tr>
<td>SA0012</td>
<td>2014</td>
<td>Stool</td>
<td>Human</td>
<td>Germany/Bavaria</td>
<td>Syria</td>
<td>1195</td>
<td>2</td>
<td>Travel cluster</td>
</tr>
<tr>
<td>SA0013</td>
<td>2014</td>
<td>Stool</td>
<td>Human</td>
<td>Germany/Bavaria</td>
<td>Syria</td>
<td>1195</td>
<td>2</td>
<td>Travel cluster</td>
</tr>
<tr>
<td>SA0015</td>
<td>2014</td>
<td>Stool</td>
<td>Human</td>
<td>Germany/Bavaria</td>
<td>Syria</td>
<td>1194</td>
<td>ND</td>
<td>NA</td>
</tr>
<tr>
<td>SA0016</td>
<td>2014</td>
<td>Stool</td>
<td>Human</td>
<td>Germany/Bavaria</td>
<td>Nigeria</td>
<td>1195</td>
<td>2</td>
<td>Travel cluster</td>
</tr>
<tr>
<td>SA0017</td>
<td>2015</td>
<td>Stool</td>
<td>Human</td>
<td>Germany/Bavaria</td>
<td>Kosovo under UN Security Council Resolution 1244</td>
<td>1196</td>
<td>ND</td>
<td>NA</td>
</tr>
<tr>
<td>SA0018b</td>
<td>2015</td>
<td>Stool</td>
<td>Human</td>
<td>Germany/Bavaria</td>
<td>Eritrea</td>
<td>1214</td>
<td>8b</td>
<td>NA</td>
</tr>
<tr>
<td>SA0019</td>
<td>2015</td>
<td>Bacterial strain</td>
<td>Human</td>
<td>Germany/Bavaria</td>
<td>Colombia</td>
<td>1197</td>
<td>ND</td>
<td>NA</td>
</tr>
<tr>
<td>SA0031c</td>
<td>2018</td>
<td>Nutritional supplement</td>
<td>Nutritional supplement</td>
<td>Germany/Bavaria</td>
<td>Colombia</td>
<td>1513</td>
<td>NDc</td>
<td>NA</td>
</tr>
<tr>
<td>SA0032</td>
<td>2017</td>
<td>Stool</td>
<td>Human</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1198</td>
<td>ND</td>
<td>NA</td>
</tr>
<tr>
<td>SA0033</td>
<td>2017</td>
<td>Stool</td>
<td>Human</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1124</td>
<td>ND</td>
<td>NA</td>
</tr>
<tr>
<td>SA0034</td>
<td>2017</td>
<td>Stool</td>
<td>Human</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1199</td>
<td>6</td>
<td>NA</td>
</tr>
<tr>
<td>SA0035</td>
<td>2017</td>
<td>Cattle</td>
<td>Cattle</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1200</td>
<td>ND</td>
<td>NA</td>
</tr>
<tr>
<td>SA0036</td>
<td>2015</td>
<td>Chicken faeces</td>
<td>Environment</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1174</td>
<td>4</td>
<td>Chicken cluster</td>
</tr>
<tr>
<td>SA0037d,c</td>
<td>2015</td>
<td>Cattle</td>
<td>Cattle</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>NDc</td>
<td>NDd,c</td>
<td>NA</td>
</tr>
<tr>
<td>SA0038</td>
<td>2015</td>
<td>Chicken faeces</td>
<td>Environment</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1134</td>
<td>4</td>
<td>Chicken cluster</td>
</tr>
<tr>
<td>SA0039</td>
<td>2015</td>
<td>Chicken faeces</td>
<td>Environment</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1134</td>
<td>4</td>
<td>Chicken cluster</td>
</tr>
<tr>
<td>SA0040</td>
<td>2015</td>
<td>Chicken</td>
<td>Environment</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1201</td>
<td>ND</td>
<td>NA</td>
</tr>
<tr>
<td>SA0041</td>
<td>2008</td>
<td>Black pepper</td>
<td>Spices</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1202</td>
<td>ND</td>
<td>NA</td>
</tr>
<tr>
<td>SA0042</td>
<td>2017</td>
<td>Animal feed</td>
<td>Animal feed</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1193</td>
<td>3</td>
<td>Feed cluster</td>
</tr>
<tr>
<td>SA0043</td>
<td>2017</td>
<td>Animal feed</td>
<td>Animal feed</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1193</td>
<td>3</td>
<td>Feed cluster</td>
</tr>
<tr>
<td>SA0044</td>
<td>2017</td>
<td>Animal feed</td>
<td>Animal feed</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1193</td>
<td>3</td>
<td>Feed cluster</td>
</tr>
<tr>
<td>SA0045</td>
<td>2003</td>
<td>Digestive tea</td>
<td>Tea/raw tea</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1203</td>
<td>1</td>
<td>Tea outbreak</td>
</tr>
<tr>
<td>SA0046</td>
<td>2003</td>
<td>Digestive tea</td>
<td>Tea/raw tea</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1203</td>
<td>1</td>
<td>Tea outbreak</td>
</tr>
<tr>
<td>SA0047</td>
<td>2003</td>
<td>Cough and bronchial tea</td>
<td>Tea/raw tea</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1203</td>
<td>1</td>
<td>Tea outbreak</td>
</tr>
<tr>
<td>SA0048</td>
<td>1994</td>
<td>Turkey leg</td>
<td>Food</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1204</td>
<td>ND</td>
<td>NA</td>
</tr>
<tr>
<td>SA0050</td>
<td>1994</td>
<td>Shredded coconut</td>
<td>Food</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1205</td>
<td>5</td>
<td>Coconut cluster</td>
</tr>
<tr>
<td>SA0051</td>
<td>1994</td>
<td>Shredded coconut</td>
<td>Food</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1207</td>
<td>7</td>
<td>NA</td>
</tr>
<tr>
<td>SA0052</td>
<td>1994</td>
<td>Shredded coconut</td>
<td>Food</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1205</td>
<td>5</td>
<td>Coconut cluster</td>
</tr>
</tbody>
</table>

cgMLST: core genome multilocus sequence typing; CT: cluster type; NA: no information available; ND: none detected; UN: United Nations.

a Representative dataset of the French outbreak.
b Excluded from cgMLST with *S. Agona* scheme.
c Excluded from cgMLST with *S. enterica* scheme.
years ago to replace this technique by molecular typing methods such as multilocus sequence typing (MLST), which is able to recognise evolutionary relationships with higher resolution [8]. Furthermore, since many years, pulsed-field gel electrophoresis (PFGE), classifying bacteria based on their universal band pattern after chromosomal restriction, is globally used as a standard molecular technique in outbreak investigations [9-11]. However, despite advantages of molecular techniques, traditional serotyping is still universally used and provides an important historical context.

Beyond PFGE and MLST, variable number of tandem repeats (VNTRs) proved to be suitable molecular targets for assessing genetic polymorphisms within bacterial species [12,13]. The multilocus variable number of tandem repeats (MLVA) technique, as a form of VNTR typing showed increased analysis depth in outbreak investigations and proved to be suitable for important Salmonella serovars such as Enteritidis, Typhimurium or Dublin [14-16]. However, the variability of protocols and targets hindered comprehensive standardisation, although efforts towards this are ongoing [17,18]. All these techniques provide reliable first level classification and are discriminative enough to investigate epidemiologically well-defined outbreaks.

Nonetheless, in the meantime, a number of studies have shown that whole genome sequencing (WGS) gives the highest resolution for outbreak investigation, especially if case distribution is diffuse with respect to geographical area or time frame of occurrence [7,19,20]. For implementation of WGS in S. Agona outbreak investigations and molecular surveillance, it has to be taken into account that this serovar is monophyletic, more homogeneous, as well as evolutionarily younger than most other well-investigated pathogenic serovars [21,22].

In 2017, three feed samples (rapeseed meal) of a factory in the district of Lower Bavaria were submitted to the Bavarian Health and Food Safety Authority. Culture-based species and serovar identification detected S. Agona in all three samples. In December 2017, an outbreak of the same serovar was reported in France, attributable to 37 French cases and two international cases, caused by infant milk products of a French supplier and traceable to one single French production facility [23,24].

The aim of the current WGS investigation, using next generation sequencing (NGS) was to verify potential clonality of the Bavarian isolates, to exclude any connection between the Bavarian feed samples and the simultaneous French outbreak and to gain a more precise insight into the genetic diversity of S. Agona collected in Bavaria over the past 25 years.

### Table 1b

<table>
<thead>
<tr>
<th>Sample name</th>
<th>Year</th>
<th>Material</th>
<th>Isolation source</th>
<th>Country</th>
<th>Country of origin/travel history</th>
<th>cgMLST CT</th>
<th>cgMLST cluster</th>
<th>Epidemiological link</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA0053</td>
<td>2003</td>
<td>Aniseed</td>
<td>Tea/raw tea</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1203</td>
<td>1</td>
<td>Tea outbreak</td>
</tr>
<tr>
<td>SA0054</td>
<td>2003</td>
<td>Children calming tea</td>
<td>Tea/raw tea</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1203</td>
<td>1</td>
<td>Tea outbreak</td>
</tr>
<tr>
<td>SA0055</td>
<td>2003</td>
<td>Children calming tea</td>
<td>Tea/raw tea</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1203</td>
<td>1</td>
<td>Tea outbreak</td>
</tr>
<tr>
<td>SA0056</td>
<td>2003</td>
<td>Aniseed</td>
<td>Tea/raw tea</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1203</td>
<td>1</td>
<td>Tea outbreak</td>
</tr>
<tr>
<td>SA0057</td>
<td>2003</td>
<td>Aniseed organic</td>
<td>Tea/raw tea</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1203</td>
<td>1*</td>
<td>Tea outbreak</td>
</tr>
<tr>
<td>SA0058</td>
<td>2003</td>
<td>Aniseed organic</td>
<td>Tea/raw tea</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1203</td>
<td>1</td>
<td>Tea outbreak</td>
</tr>
<tr>
<td>SA0059</td>
<td>2003</td>
<td>Pectoral tea</td>
<td>Tea/raw tea</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1203</td>
<td>1</td>
<td>Tea outbreak</td>
</tr>
<tr>
<td>SA0060</td>
<td>2003</td>
<td>Flatulence tea</td>
<td>Tea/raw tea</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1203</td>
<td>1</td>
<td>Tea outbreak</td>
</tr>
<tr>
<td>SA0061</td>
<td>2003</td>
<td>Pectoral and cough tea</td>
<td>Tea/raw tea</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1203</td>
<td>1</td>
<td>Tea outbreak</td>
</tr>
<tr>
<td>SA0062</td>
<td>2003</td>
<td>Pectoral and cough tea</td>
<td>Tea/raw tea</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1203</td>
<td>1</td>
<td>Tea outbreak</td>
</tr>
<tr>
<td>SA0063</td>
<td>2003</td>
<td>Aniseed</td>
<td>Tea/raw tea</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1203</td>
<td>1</td>
<td>Tea outbreak</td>
</tr>
<tr>
<td>SA0064</td>
<td>2003</td>
<td>Aniseed</td>
<td>Tea/raw tea</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1203</td>
<td>1</td>
<td>Tea outbreak</td>
</tr>
<tr>
<td>SA0065</td>
<td>1993</td>
<td>Turkey</td>
<td>Food</td>
<td>Germany/Bavaria</td>
<td>NA</td>
<td>1207</td>
<td>7</td>
<td>NA</td>
</tr>
<tr>
<td>SA0066</td>
<td>2018</td>
<td>Stool</td>
<td>Human</td>
<td>Germany/Bavaria</td>
<td>Thailand, Cambodia, Vietnam</td>
<td>1206</td>
<td>ND</td>
<td>NA</td>
</tr>
</tbody>
</table>

cgMLST: core genome multilocus sequence typing; CT: cluster type; NA: no information available; ND: none detected; UN: United Nations.

* Excluded from cgMLST with S. Agona scheme.
Figure 1
Minimum spanning tree of the core genome multilocus sequence type allelic profiles of Salmonella Agona strains, including 51 Bavarian isolates, a representative isolate of an infant-milk-caused outbreak in France and the reference strain SL-483, Germany, 1993–2018 (n = 53 isolates).

cgMLST: core genome multilocus sequence typing.

Bavarian S. Agona samples have the prefix SA. The representative isolate of the infant-milk outbreak in France [23] is ERR219379. The National Center for Biotechnology Information (NCBI) reference genome of strain SL-483 has GenBank accession number NC_011149.1.

The analysis and tree were obtained with the public Ridom-SeqSphere+ -integrated S. enterica cgMLST scheme of 3,002 target loci.

On the tree, allele distances between samples are indicated. Clusters of samples with maximum seven alleles distance are shaded in grey. Samples are colour coded by their isolation source, as given in the legend.
Methods

Isolates and sequence data used in the analyses
For this observational retrospective study, we used 48 S. Agona isolates dated from 1993 to 2018 from our strain collections to investigate the three isolates of the Bavarian rapeseed meal from 2017 in a wider context. The total 51 isolates were all the S. Agona isolates available for us. As many diagnostic laboratories exist in Bavaria, our isolates were not representative for the occurrence of S. Agona in this federal state, where, for example, 99 cases of human infections had been officially notified in the 2011 to 2018 period alone.

The 48 isolates that we employed were either from humans, food, feed, animals or the environment. A total of 28 thereof were outbreak-related. Outbreak isolates belonged to four epidemiologically-linked events represented by 15, three, three and two isolates as well as an additional five isolates with a suspected epidemiological connection.

The 48 isolates from 1993 to 2018 were studied by WGS together with the three 2017 isolates of the Bavarian rapeseed feed. Raw NGS data of the published representative isolate of the French outbreak [24], available under European Nucleotide Archive (ENA) accession ERR2219379, were also added to the bioinformatics analyses.

For the evaluation of necessary analysis depth, 70 NGS analyses were added to the data analysis as well. To distinguish the subset of isolates derived from our strain collections, or data thereof, from the data published by Zhou et al. 2013 [21], available under National Center for Biotechnology Information (NCBI) bioproject PRJEB1944, were added to the data analysis as well.

To distinguish the subset of isolates derived from our strain collections, or data thereof, from the data published by Zhou et al. [21], we further refer to our 51 S. Agona isolates/strains as ‘Bavarian’.

Species and serovar identification

All Bavarian S. Agona strains were cultured on Columbia sheep blood agar (Oxoid, Wesel, Germany) and identified by matrix-assisted laser desorption/ionisation time-of-flight mass spectroscopy (MALDI-TOF MS; Bruker, Bremen, Germany). Somatic (O) and flagellar (H) antigens were identified by using slide agglutination (antisera provided by Sfin, Berlin, Germany) according to the White–Kauffmann–Le Minor scheme [25].

Next generation sequencing

Salmonella Agona isolates were freshly grown on blood agar plates. One inoculation loop of bacterial material was suspended in 50 µL phosphate buffered saline (PBS) and cells were pre-treated with 1 µg lysozyme for 15 min at 37°C followed by a 2 hour incubation step at 65°C with 200 µL incorporation buffer, 30 µL 20 mg/mL Proteinase K and 10 µL 10 mg/mL ribonuclease (RNase) A (all reagents from Promega, Mannheim, Germany). Genomic DNA (gDNA) was then isolated with the Maxwell 16 LEV Blood DNA Kit on the Maxwell 16 instrument (Promega, Mannheim, Germany) according to manufacturer’s instructions with Tris buffer for gDNA elution.

Whole genome libraries for NGS were prepared using the Nextera XT kit (Illumina, San Diego, California, US). Next generation sequencing was performed on the Illumina MiSeq with 2x250 bp paired-end reads. Sequencing runs were evaluated for quality using the Illumina SAV Software.

Sequencing data were uploaded to the NCBI sequence read archive (SRA) [26], under BioProject PRJNA473689.

Multilocus sequence typing analyses

Core genome MLST (cgMLST) of reads was performed with Ridom SeqSphere+ Software (Ridom, Munster, Germany [27]) with default settings for trimming and velvet assembly. For the assignment of cgMLST alleles, two different schemes were used: (i) a publicly-available S. enterica (species level) cgMLST scheme designed by Enterobase and (ii) an in-house serovar-specific cgMLST scheme.

Enterobase-designed Salmonella enterica core genome multilocus typing scheme

The publicly available species-specific SeqSphere+ software-implemented S. enterica cgMLST scheme with the 3,002 target loci developed as Salmonella cgMLST v2 scheme by Enterobase was employed to analyse sequencing data [28,29].

In-house serovar-specific Salmonella Agona core genome multilocus typing scheme

An in-house developed serovar-specific S. Agona scheme, based on reference genome NC_011149.1 of strain SL-483 and query genomes NC_022991.1, NZ_CP015024.1, NZ_CP011259.1, was generated using the following default filter thresholds.

For the reference genome filter thresholds: (i) minimum length: 60 bases; (ii) start codon and single stop codon required at beginning and end of gene; (iii) homologous/paralogous gene filter, excluding multiple copies of a gene with basic local alignment search tool (BLAST) overlap ≥ 100 bp or identity ≥ 90%; (iv) overlap filter, excluding overlap with other genes > 4 bases.

For the query genome filters thresholds: (i) start and stop codon required at beginning and end of gene; (ii) BLAST hit locus overlap 100% and identity > 90% in every query genome; (iii) BLAST options: word size = 11, mismatch penalty = −1, match reward = 1, gap open costs = 5, gap extension costs = 2.

Thereby, the final scheme resulted in 4,111 target loci. New alleles and sequence types (ST) were submitted to the nomenclature server for the public scheme.
Figure 2
Maximum likelihood tree resulting from the whole genome single nucleotide polymorphism-based phylogenetic analysis of Salmonella Agona strains including 51 Bavarian isolates, a representative isolate of an infant milk-caused outbreak in France and the reference strain SL-483, Germany, 1993–2018 (n = 53 isolates)

cgMLST: core genome multilocus sequence typing.

Bavarian S. Agona samples have the prefix SA. The representative isolate of the infant milk outbreak in France [23] is ERR2219379. The National Center for Biotechnology Information (NCBI) reference genome of strain SL-483 has GenBank accession number NC_011149.1 and figures in the tree as ‘reference’.

In the tree, samples are colour coded according to isolation source, S. Agona cgMLST cluster and collection year as given in the legend.

The general scale bar indicates 0.001 substitutions per site (597 SNPs), based on an alignment of 596,849 positions.

SNP distance bars between specific samples are indicated on the right side. Blue vertical SNP scale bars indicate epidemiologically-linked cgMLST clusters, blue vertical dashed lines indicate cgMLST clusters without epidemiological link and black vertical lines indicate exemplary distances between non-clustered samples.
Assessment of relationships between isolates’ core genome multilocus sequence type allelic profiles by minimum spanning trees
cgMLST typing results were visualised in minimum spanning trees (MSTs), excluding all samples in the respective scheme, not fulfilling ‘good target’ quality control (QC) for >90% of the scheme’s target loci (>90% of targets present in the isolate, same length as reference +/-3 triplets, without ambiguities and without frame shifts in consensus, Table 1). A cluster was defined as a group of closely related cgMLST-analysed isolates in both schemes with a single-linkage threshold of ≤7 alleles. This was the default distance threshold for the software-implemented public S. enterica scheme and was adopted for direct comparison for the S. Agona scheme. During typing with the S. enterica cgMLST scheme, the SeqSphere+ software assigns an existing cluster type (CT) to each isolate with an allelic distance ≤7 to an already established CT founder profile on the central nomenclature server. Otherwise, a new CT is established, uploaded and the isolate becomes the founder of this CT [30,31].

Multilocus sequence typing
In silico MLST analysis of NGS data was performed with the standard seven-gene target scheme [32].

Single nucleotide polymorphism phylogeny
Whole genome (wg) single nucleotide polymorphism (SNP)-based phylogeny was calculated using the run_snp_pipeline script of the PHEnix pipeline by Public Health England [33]. It includes trimming with trimmomatic [34], mapping to the reference genome NC_011149.1 of strain SL-483 with bwa-mem mapping [35] with default settings, variant calling and filtering (frequency ≥ 0.9, mapping quality score ≥30, read depth ≥10) by Genome Analysis Toolkit (GATK)2 Unified Genotyper [36]. Variant calls for all SNP positions passing filters and all positions not passing filters were extracted and SNPs were concatenated to alignments with the vcf2fasta-script from the same pipeline, allowing ≤90% missing data per sample and ≤20% missing data per position throughout all samples. Maximum likelihood (ML) trees were generated from SNP alignments by RaxML [37], including 100 bootstrap replicates.

Results
Next generation sequencing of the 51 Bavarian S. Agona isolates (named with prefix SA), collected between 1993 and 2018 revealed high quality reads and reference genome coverage of 28–171 fold. In our bioinformatics data analyses, we included the published NGS raw data (ENA accession: ERR2219379) of an isolate from a case of the infant-milk-caused outbreak originating from a France-based manufacturer [24]. This isolate served as a representative of the French outbreak. All isolates were typed by in silico MLST, resulting in ST13, the typical ST for serovar Agona [8].

In a next step, the isolates were typed with Ridom SeqSphere+ software with the public S. enterica cgMLST scheme, consisting of Enterobase-developed 3,002 target loci (Figure 1). Additionally, wg SNP analysis was performed to investigate the phylogenetic relationship in highest possible resolution (Figure 2).

The MST from results of the S. enterica cgMLST scheme, including all samples exceeding the target-QC cut-off, revealed in total eight clusters with maximum six alleles difference (Figure 1). The four most relevant clusters (1–4) comprised three to 15 samples with a maximum within cluster difference of zero to five alleles (Table 2). Each of the remaining four clusters (5–8) included two samples and internal distances ranging from zero (cluster 7) to six alleles (cluster 8).

None of the 51 Bavarian isolates collected from 1993–2018 clustered with the representative sample from the recent outbreak in France. The French representative sample ERR2219379 differed from the Bavarian samples in at least 15 alleles/40 SNPs. This difference was observed to distinguish unrelated isolates and clusters from different years or with different epidemiological origins throughout the whole sample set (Figure 1, Figure 2, Table 1).

The three Bavarian feed isolates collected in 2017 from rapeseed cake (SA0042, SA0043, SA0044) built up a distinct cluster with two alleles maximum distance (Figure 1, cluster 3), corresponding to two SNPs in the wg SNP analysis (Figure 2). They differed from the nearest neighbours outside the clusters in at least 17 alleles/37 SNPs (Table 2).

Furthermore, some of the Bavarian strains, isolated in former years, aggregated in epidemiologically described clusters, although the corresponding isolates available in our strain collection for sequencing mainly covered food or veterinary samples and corresponding human isolates were not available for analysis. Most eye-catching, the isolates of the biggest cluster with 15 samples (Figure 1, cluster 1) were all isolated from tea or raw tea products, connected to a diffuse outbreak caused by aniseed-fennel-caraway tea products in 2002–03. The outbreak investigation at that time identified contaminated raw tea imported from Turkey as source [2,3]. All isolates of this cluster were closely connected with zero to one allele single-linkage distance (Figure 1) and three alleles or five SNPs (Figure 2) maximum distance within the cluster. They showed at least 10 alleles/26 SNPs to all other non-connected isolates (Table 2).

The five isolates of cluster 2 with single-linkage distances of zero to two alleles and maximum intra-cluster distance of five alleles or seven SNPs were obtained from asylum seekers between 2011 and 2015 (Table 1). Details of their travel history regarding countries or period are unknown. Hence the slightly higher
variation of their allelic/SNP distances than in other point-sourced epidemiologically linked clusters is not surprising. Cluster 4 consists of three samples with zero alleles/SNPs difference to each other but at least 20 alleles/45 SNPs difference to the other samples. Isolates in this cluster shared a clear epidemiological link, as samples originated from chicken faecal samples, collected in 2015 in different laying hen flocks of one Bavarian egg producer.

Clusters 5, 6, 7 and 8 only consist of two isolates each, respectively. Cluster 5 was built up from two shredded coconut samples from 1994 with four alleles/six SNPs difference which likely have an epidemiological link although information on their origin or supplier is not available. Isolates from two human patients without epidemiological link clustered together with two alleles (cluster 6) and seven SNPs distance. Cluster 7 came from two genetically identical strains from food samples from 1993–94 (shredded coconut and turkey) for which no epidemiological link is known. One additional pair of closely related human samples (SA0011 and SA0018) with reported travel history to Eritrea (SA0018) and an unspecified country (SA0011) built cluster 8 with six alleles and eight SNPs difference.

To evaluate the needed resolution of the cgMLST analysis regarding analysed genomic content and consequent cluster demarcation, the species level typing with the public S. enterica cgMLST scheme, consisting of Enterobase-developed 3,002 target loci was compared with an in-house generated ad hoc serovar Agona-specific cgMLST scheme with 4,111 target loci (Figure 3). Typing with both schemes was performed using the SeqSphere+ software algorithms. As no empiric threshold was defined prior the analysis for the serovar-specific scheme it was used with the same thresholds as the species-specific scheme. For the comparison of the data analysis approaches, 70 published NGS raw datasets of isolates described by Zhou

<table>
<thead>
<tr>
<th>Measure</th>
<th>Method</th>
<th>Cluster 1 (number of isolates)</th>
<th>Cluster 2 (number of isolates)</th>
<th>Cluster 3 (number of isolates)</th>
<th>Cluster 4 (number of isolates)</th>
<th>Cluster 5 (number of isolates)</th>
<th>Cluster 6 (number of isolates)</th>
<th>Cluster 7 (number of isolates)</th>
<th>Cluster 8 (number of isolates)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Cluster ID</td>
<td>Epidemiological link</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within cluster distance: median</td>
<td>Salmonella enterica</td>
<td>(15) Tea outbreak</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S. Agona cgMLST</td>
<td>(5) Travel cluster</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SNP phylogeny</td>
<td>(3) Feed cluster</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3) Chicken cluster</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2) None detected</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min distance to closest neighbour</td>
<td>S. enterica cgMLST</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S. Agona cgMLST</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SNP phylogeny</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nearest neighbour outside cluster in</td>
<td>S. enterica cgMLST</td>
<td>SA0040, SA0050 (in cluster 5)</td>
<td>SA0007</td>
<td>SA0053, SA0054, SA0055, SA0058</td>
<td>SA0055 (in cluster 1)</td>
<td>SA0053, SA0054, SA0055, SA0058</td>
<td>SA0053 (in clusters)</td>
<td>SA0055, SA0056 (all clusters)</td>
<td>SA0015</td>
</tr>
<tr>
<td>question</td>
<td>S. Agona cgMLST</td>
<td>SA0007, SA0050 (in clusters 5)</td>
<td>SA0047 and SA0058 (both in clusters), SA0007</td>
<td>SA0047 and SA0058 (both in clusters), SA0007</td>
<td>SA0007</td>
<td>SA0053, SA0055, SA0056 (all clusters)</td>
<td>SA0053 (in clusters)</td>
<td>SA0055, SA0056 (all clusters)</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>SNP phylogeny</td>
<td>SA0050 (in cluster 5)</td>
<td>SA0047 (in cluster 5)</td>
<td>SA0007</td>
<td>SA0057 (in cluster 1)</td>
<td>SA0007</td>
<td>SA0057 (in cluster 1)</td>
<td>SA0007</td>
<td>SA0018 (in cluster 8)</td>
</tr>
</tbody>
</table>

**cgMLST:** core genome multilocus sequence typing; max: maximum; min: minimum; SNP: single nucleotide polymorphism.

a Cluster 8 not detected in S. Agona cgMLST due to exclusion of sample SA0018.
et al. [21] were added in both approaches to widen the view on cluster definition and concordance by genomic and epidemiological data on a supra-regional level. This dataset covered S. Agona isolates from 1952 to 2010 covering several European outbreaks. The results of the public S. enterica cgMLST scheme (Figure 3A) are very similar to those of the in-house S. Agona cgMLST scheme (Figure 3B) and SNP-profiling (Figure 2). All Bavarian and published outbreak clusters were detected by both schemes (Figure 3A and B). Therefore, the usage of the public S. enterica scheme was generally evaluated as practical for S. Agona outbreak investigations.

As SA0018 was excluded from the serovar-specific cgMLST due to shortfall below the 90% good target threshold for the S. Agona scheme target loci, cluster 8 detected in species-specific cgMLST (Figure 3A) was not apparent in serovar-specific cgMLST (Figure 3B).

Due to a smaller number of target loci in the public cgMLST scheme, allele distances between samples were generally lower than with the serovar-specific scheme. Two epidemiologically unrelated clusters linked by an unrelated Irish environmental isolate (ERS180349) (Figure 3B, clusters 2 and 12) could not be delimited clearly in the S. entericascheme with a default cluster threshold of seven alleles (Figure 3A, cluster 2 +12). Those two clusters – one consisting of Bavarian patient isolates with travel history to African or Arabian countries described above (cluster 2; Table 1 and 2) and one described by Zhou et al. (Figure 3B, cluster 12), originating from a large multi-country 2008–09 outbreak originating from Ireland [21] – could more clearly be separated with the in-house cgMLST S. Agona scheme (Figure 3B, clusters 2 and 12), using the same cluster threshold. In the species-level cgMLST scheme, the software assigns CTs for all samples. These were different for most isolates of the two unrelated clusters, suggesting different two sources for these two clusters, additional to the epidemiological information.

Discussion

Our high resolution WGS-based analysis approach by either wg SNP-profiling or species- or serovar-specific cgMLST delivered good and reliable results for typing S. Agona isolates and evaluating their affiliation to, or delineation from, outbreak clusters. We could thereby exclude a connection between the Bavarian feed sample isolates from 2017 and a large outbreak due to infant milk products, which occurred at the same time in France and other European countries. Furthermore, we could confirm S. Agona epidemiological clusters from former years as well as identify previously unrecognised clusters. As recently shown for serovar S. Enteritidis [38], cgMLST was a standardisable and easily applicable to S. Agona and could reach an analysis depth comparable to wg SNP profiling. Recently, cgMLST was published as a tool with reasonable resolution for investigations of Salmonella population structure [29]. Thereby, the extensively curated S. enterica based scheme was made available on a publicly accessible website [28]. The same scheme was implemented in the Ridom SeqSphere+Software too, although with own allele calling and clustering rules and was used in this study to compare the resolution of species- and serovar-specific typing with our in-house developed serovar-specific scheme. It could be shown that even on species level all outbreaks were cleared identified by both schemes. Only two independent outbreak clusters could not be delimited clearly anymore by the publicly available scheme, due to the lower number of species-specific target loci and the used cluster threshold.

Of course, more focused typing schemes or approaches like serovar-level cgMLST or wg SNP genotyping deliver more detailed typing results catching at least the portion of the serovar-specific diversity manifested in the core- or reference-genome. This is particularly true for monomorphic organisms such as S. Agona, for which mobile elements that are mostly not covered by core genome or single reference-based approaches generate a lot of diversity [21], or for situations when single samples are to be inferred as part of specific clusters or not, especially when epidemiological information is not fully discriminative. Consequently, the serovar-specific in-house scheme developed for this investigation, worked very well for the analysed sample set. However, due to the lack of target curation and its creation from only one reference and three query genomes, it will need further optimisation and testing with diverse sample sets for suitability on a broader scale. This can be assumed from the fact that four isolates had to be excluded from typing due to shortfall below the 90% good target threshold using the in-house scheme, but only one isolate using the publicly available scheme (Table 1). The use of a publicly available scheme whose target loci underwent extensive curation, like the Enterobase S. enterica scheme [29], can render suitable and reliable results even in cases where serotyping has not yet been performed or in the event of the lack of a specific reference.

Generally, cluster definition is based on a single-linkage allelic difference threshold. The software implements a proposed default threshold of seven alleles for the S. enterica scheme. This was adopted for the S. Agona scheme for comparison and as no empirically tested threshold was established for this scheme before. However, as shown with the high-resolution in-house S. Agona cgMLST scheme, clusters with a clear epidemiological link did not show more than four alleles and six SNPs difference. Therefore, a cluster with a distance above four, but below the threshold of seven alleles (e.g. Figure 3B, clusters 6 and 11), may contain falsely grouped samples. Thus, the cluster thresholds in both schemes could be decreased down from seven based on the analysed sample set.
Minimum spanning trees obtained with different typing schemes (Panels A and B) of the core genome multilocus sequence type allelic profiles of 51 Bavarian *Salmonella* Agona isolates, a representative isolate of an infant milk outbreak in France, the reference strain SL-483 and 70 isolates from various European outbreaks (n = 123 isolates)\(^a\)

A. Result from the Ridom-SeqSphere + integrated *S. enterica* cgMLST scheme of 3,002 target loci

B. Result from the in-house *S. Agona* serovar-specific cgMLST scheme of 4,111 target loci

\(\text{cgMLST: core genome multilocus sequence typing.}\)

\(\text{\(a\) Due to shortfall below the 90% good target threshold for the } S. \text{ Agona scheme target loci, SA0018 does not figure in panel B, which shows a total of 122 isolates.}\)

Bavarian *S. Agona* samples have the prefix SA. The representative isolate of the infant milk outbreak in France [23] is ERR2219379. The National Center for Biotechnology Information (NCBI) reference genome of strain SL-483 has GenBank accession number NC_011149.1. The 70 isolates from various European outbreaks are published in a separate study [21].

Allele distances between samples are indicated and clusters of samples with minimum spanning distances of zero to seven alleles are shaded in grey. Samples are colour coded by their isolation source, as given in the legend.
However, theoretically a cluster threshold should be lower in a scheme with less loci like the *S. enterica* scheme (3,002 loci) than in a scheme with more loci like the *S. Agonas* scheme (4,111 loci). A decreased cluster threshold would be in accordance with the special genetic characteristics of *S. Agona*, which emerged more recently and is more monomorphic than most other serovars [21]. However, while facilitating cluster delimitation in *S. Agona*, a decreased threshold would, especially in the *S. enterica* scheme, very likely impair clustering in the case of more heterogeneous serovars or when linked isolates have evolved over a longer time period.

Cluster types assigned by the publicly available *S. enterica* scheme are intended to roughly classify genetically similar samples in an easy way. This is very helpful and an important step towards standardisation, also in terms of inter-laboratory communication and for alerts concerning the detection of a known CT. In the case of the two epidemiologically unrelated clusters 2 and 12 (Figure 3B), which were not clearly delimited by the public *S. enterica* scheme (Figure 3A, Cluster 2 + 12), the software-assigned CT values for each isolate were mainly distinct for the two clusters (cluster 2: CT-1195, cluster 12: CT-25). Hence, using the CT as a simplification measure to distinguish between clusters from different sources helped, although the clusters were difficult to resolve at least with a MST and the applied threshold for single-linkage clustering. However, due to the complexity of NGS data the logical principle on which such a simplification, like using the CT value, is based, always has to be considered for its correct use, as this can also result in false inclusion into or exclusion from a cluster. Indeed, the CTs are assigned depending on an isolate’s proximity in terms of allelic distance to a specific CT founder allele profile or by incremental expansion of the nomenclature when a new isolate exhibits a not yet known and sufficiently different allele combination [30,31]. The incremental nomenclature expansion together with the fact that only the distance to the nearest CT founder is reflected in the CT assignment of new isolates can lead to the assignment of the same CT to isolates with allelic distances above the threshold or of different CTs to isolates below a cluster threshold. This may be for example the case for isolate ERS180350, correctly grouped within the Irish outbreak cluster of 2008 with common CT-25 (Figure 3B, cluster 12), but being tagged with a deviant CT (CT-1224). Furthermore, microbial evolution can be different in certain epidemiological niches which cannot fully be covered by such simplification.

Due to these limitations, outbreak investigations should generally avoid relying solely on simplifications like CT values, but also include visual inspection of cluster formation in trees to avoid overlooking of connected samples. Classifications by clusters or simplified measures such as CTs should always be interpreted with caution, especially when allelic or SNP distances near an empirically tested cluster threshold occur. Moreover, particularly, but not only, for organisms with specific genetic characteristics such as *S. Agona*, empirical as well as epidemiological data should always be taken into account in addition to the molecular data [39].

Our analysis also shows that if a common reference or close genetic relationship for a specific set of isolates is known, high resolution approaches facilitate analysis and enable clear grouping of individual isolates in dispute, whereas the versatility of genetically broader approaches enables more standardised results for more heterogeneous sets of isolates. To reduce interpretation complexity and further extend fast and easy usability of NGS approaches for public health analyses, further standardisation and harmonisation between laboratories are nevertheless required. Many researchers and authorities have already realised this, but the implementation and ongoing optimisation will be a huge effort in the forthcoming years.

Concluding, consistent with results obtained for other serovars and species, wg SNP profiling as well as serovar- or species-specific cgMLST can be used for reasonable, reproducible and reliable high resolution classification of *S. Agona* WGS data to detect outbreak clusters. We showed this with a representative dataset from regional and international sources covering human, food, feed, veterinary and environmental isolates and thereby various types of focus areas of public health authorities. With this approach, relationships between past or international cases could also be inferred using representative public data. We also highlighted the importance and supportive power of epidemiological sample data and an integrated view on both molecular and epidemiological data. Importantly, NGS results still need careful evaluation, as their interpretation approaches often have to be a trade-off between highest resolution and versatility. Standardisation and harmonisation on an international level will further improve using the surplus of information coming from NGS in molecular surveillance.

**Acknowledgements**

The authors would like to thank all the colleagues in the public health microbiology and molecular biology laboratories as well as in the food bacteriology laboratories, contributing with excellent laboratory work going beyond their routine diagnostic tasks.

**Conflict of interest**

None declared.

**Authors’ contributions**

AD, AB, UM, RK, SH, NA, AS designed the study. UM and SH provided isolates from culture collections. AB and UM performed and supervised the bacteriology work. AD and RK performed and supervised the molecular analyses. AD analysed the NGS data. AD and AB drafted and revised the


License, supplementary material and copyright

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence and indicate if changes were made.

Any supplementary material referenced in the article can be found in the online version.

This article is copyright of the authors or their affiliated institutions, 2019.