



Toward an Integrated Genome-Based Surveillance of *Salmonella enterica* in Germany

Laura Uelze¹, Natalie Becker², Maria Borowiak¹, Ulrich Busch³, Alexandra Dangel³, Carlus Deneke¹, Jennie Fischer¹, Antje Flieger⁴, Sabrina Hepner³, Ingrid Huber³, Ulrich Methner⁵, Jörg Linde⁵, Michael Pietsch⁴, Sandra Simon⁴, Andreas Sing³, Simon H. Tausch¹, Istvan Szabo¹ and Burkhard Malorny^{1*}

¹ Department of Biological Safety, German Federal Institute for Risk Assessment, Berlin, Germany, ² Department of Food, Feed and Commodities, Federal Office of Consumer Protection and Food Safety, Berlin, Germany, ³ Bavarian Health and Food Safety Authority, Oberschleißheim, Germany, ⁴ Unit of Enteropathogenic Bacteria and Legionella (FG11) – National Reference Centre for Salmonella and Other Bacterial Enteric Pathogens, Robert Koch Institute, Wernigerode, Germany, ⁵ Institute of Bacterial Infections and Zoonoses, Friedrich-Loeffler-Institut, Jena, Germany

OPEN ACCESS

Edited by:

Dario De Medici,
National Institute of Health (ISS), Italy

Reviewed by:

Jie Zheng,
United States Food and Drug
Administration, United States
Wesley Mattheus,
Sciensano, Belgium

*Correspondence:

Burkhard Malorny
burkhard.malorny@bfr.bund.de

Specialty section:

This article was submitted to
Food Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 07 November 2020

Accepted: 21 January 2021

Published: 10 February 2021

Citation:

Uelze L, Becker N, Borowiak M, Busch U, Dangel A, Deneke C, Fischer J, Flieger A, Hepner S, Huber I, Methner U, Linde J, Pietsch M, Simon S, Sing A, Tausch SH, Szabo I and Malorny B (2021) Toward an Integrated Genome-Based Surveillance of *Salmonella enterica* in Germany. *Front. Microbiol.* 12:626941. doi: 10.3389/fmicb.2021.626941

Despite extensive monitoring programs and preventative measures, *Salmonella* spp. continue to cause tens of thousands human infections per year, as well as many regional and international food-borne outbreaks, that are of great importance for public health and cause significant socio-economic costs. In Germany, salmonellosis is the second most common cause of bacterial diarrhea in humans and is associated with high hospitalization rates. Whole-genome sequencing (WGS) combined with data analysis is a high throughput technology with an unprecedented discriminatory power, which is particularly well suited for targeted pathogen monitoring, rapid cluster detection and assignment of possible infection sources. However, an effective implementation of WGS methods for large-scale microbial pathogen detection and surveillance has been hampered by the lack of standardized methods, uniform quality criteria and strategies for data sharing, all of which are essential for a successful interpretation of sequencing data from different sources. To overcome these challenges, the national GenoSalmSurv project aims to establish a working model for an integrated genome-based surveillance system of *Salmonella* spp. in Germany, based on a decentralized data analysis. Backbone of the model is the harmonization of laboratory procedures and sequencing protocols, the implementation of open-source bioinformatics tools for data analysis at each institution and the establishment of routine practices for cross-sectoral data sharing for a uniform result interpretation. With this model, we present a working solution for cross-sector interpretation of sequencing data from different sources (such as human, veterinarian, food, feed and environmental) and outline how a decentralized data analysis can contribute to a uniform cluster detection and facilitate outbreak investigations.

Keywords: *Salmonella*, surveillance, food-borne disease outbreak, whole genome sequencing, cgMLST

INTRODUCTION

The surveillance of zoonotic pathogens is an important task usually conducted by official authorities. The early detection of food-borne pathogens has a crucial role in reducing the major clinical and economic burden caused by food-borne disease outbreaks (Jain et al., 2019). Whole-genome sequencing (WGS), combined with data analysis, is a high throughput technology with an unprecedented discriminatory power, which is increasingly used for cluster detection, source tracking, outbreak investigation and surveillance. WGS data are highly informative, portable and standardizable (Deng et al., 2016). Although WGS is recognized as the most up-to-date methodology for the detection of infection clusters and its use is highly encouraged by international authorities (ECDC, 2016), efficient real-time surveillance using WGS requires the development and implementation of a functional cross-sectoral concept (covering public health, veterinarian, food, feed and environmental sectors). A successful concept should be designed to prospectively detect clusters of very similar isolates based on the WGS data and thus requires the use of powerful bioinformatics tools. The combination of sequence data analysis with relevant metadata and commodity chain information allows to trace transmission paths and to identify possible sources of outbreaks, thereby improving consumer protection and microbial food safety (Aarestrup et al., 2012; Moura et al., 2017; EFSA Panel on Biological Hazards, 2019).

Up to date, genome-based surveillance in most European countries, if implemented, is mainly realized through stand-alone solutions. As a consequence, different quality parameters and analysis processes are employed in different laboratories. This represents a major challenge for a cross-sectoral consolidation of data, which is needed for a consistent interpretation of results. Without a uniform quality of data sets and subsequent interpretation of data, it is not possible to reliably identify coherencies, assess risks and take measures to contain outbreaks and emerging epidemic clones.

A number of countries have established, or are currently establishing national WGS-based surveillance systems and platforms, such as the United Kingdom (Public Health of England) (Ashton et al., 2016) and Switzerland (Egli et al., 2018). At the same time, large international platforms, such

as Pathogenwatch¹, INNUENDO (Llarena et al., 2018), NCBI pathogen detection² and GenomeTrakr (Allard et al., 2016; Timme et al., 2018) have been developed to enable the analysis of multinational genome sequence data for the surveillance and investigation of cross-border transmissions and outbreaks.

All aforementioned platforms have in common that they require raw sequencing data (and often accompanying metadata) to be deposited in a central database in order to facilitate a centralized analysis of the genome data. Although an enhanced surveillance system would ideally encompass genome data on a global scale, the requirement for centralized data storage, combined with the existence of specific legislative responsibilities at local, regional or federal level, in practice often prevents or delays the participation of countries or institutions in these networks. Overall, there is no one-fits-all solution which can be easily adapted and which does not require central data storage and sharing.

As a consequence, in countries where sequence data is not shared in real-time between sectors, public health and food control laboratories, the potential of sequencing data is not fully exploited for cluster detection.

Here, we present the ‘GenoSalmSurv’ project (‘Integrated genome-based surveillance of *Salmonella*’), which targets these challenges and develops a practical approach for the genome-based cross-sectoral surveillance of the food-borne pathogen *Salmonella*. The project is funded by the German Federal Ministry of Health and part of a funding framework on integrated genome-based surveillance of zoonotic pathogens. The project is carried out by the Robert Koch Institute (RKI) hosting the National Reference Centre for *Salmonella* and other Bacterial Enteric Pathogens (NRC-*Salmonella*) (human sector) and the German Federal Institute for Risk Assessment (BfR) hosting the NRL-*Salmonella* (food, feed, animal, environmental sector), as well as the Bavarian Health and Food Safety Authority (LGL) as representative state authority (regional animal, food and human sector). Two further partners are associated, the ‘NRL for Salmonellosis in Cattle’ at the Friedrich-Loeffler-Institut (FLI) (epizootic disease sector) and the federal contact point for food-borne outbreaks, which is located at the Federal Office of Consumer Protection and Food Safety (BVL). The main goals of the project are (i) to establish standard procedures for the joint use of the high-throughput WGS data of *Salmonella* spp. from different sectors, collected from monitoring programs and surveillance/sentinel programs, and (ii) to supply open-source tools for the subsequent bioinformatics analysis, which enables strain comparison on a high resolution level. Two major challenges are thus addressed: On the one hand the establishment of laboratory parameters for WGS and on the other hand the development and harmonization of bioinformatics procedures. These involve all steps from genome assembly (including read and assembly QC), to *in silico* serotyping and automated cluster detection. Selected tools for the bioinformatics procedures are intended to be internationally recognized and scientifically validated and will be additionally evaluated for applicability.

Abbreviations: BfR, German Federal Institute for Risk Assessment; cgMLST, core genome multilocus sequence typing; EFSA, European Food Safety Authority; EC, European Commission; ECDC, European Centre for Disease Prevention and Control; EU, European Union; FAO, The Food and Agriculture Organization; FLI, Friedrich-Loeffler-Institut; GenoSalmSurv, German Ministry of Health funded project “Integrated genome-based surveillance of *Salmonella*”; IGS, integrated genome-based surveillance; LGL, Bavarian Health and Food Safety Authority; MLST, multilocus sequence typing; MLVA, multiple locus variable-number tandem repeat analysis; NGS, next generation sequencing; NRC-*Salmonella*, National Reference Centre for *Salmonella* and other Bacterial Enteric Pathogens (RKI); NRL-*Salmonella*, National reference laboratory for *Salmonella* (BfR); NRL for Salmonellosis in Cattle, National reference laboratory for Salmonellosis in Cattle (FLI); PCR, polymerase chain reaction; PFGE, pulsed-field gel electrophoresis; QC, quality control; RKI, Robert Koch Institute; S., *Salmonella enterica* subsp. *enterica*; SNP, single nucleotide polymorphism; UPGMA, unweighted pair group method with arithmetic mean; wgMLST, whole genome multilocus sequence typing; WGS, whole genome sequencing.

¹<https://pathogen.watch/>

²<https://www.ncbi.nlm.nih.gov/pathogens/>

A particular focus is on the open-source availability of the selected tools to provide wide accessibility and sustainability.

To ensure the practicability of the developed processes and protocols, the project conducts a proof-of-principle study (real-time application phase). During this 6-months period, about 2,300 *Salmonella* spp. isolates of human, animal and food origin, assigned to conventional typing, are additionally analyzed according to the workflow established in the GenoSalmSurv project. The participating partners (RKI, BfR, LGL) apply the developed procedures with the aim to detect emerging clusters and to identify relevant surveillance markers. To ensure a manageable scope, the focus is set on the serovars *Salmonella enterica* subsp. *enterica* (herein after abbreviated as *S.*) Enteritidis, *S. Typhimurium* and *S. Infantis*, which are not only responsible for most infections (70–80%) and outbreaks in humans (RKI, 2015, 2016, 2017), but also represent the most prevalent serovars in certain production animal species. Further, additional serovars of interest are included [e.g., *S. Choleraesuis*, known as swine-adapted serovar, but highly systemic in humans (Chiu et al., 2004)]. The insights gained from this proof-of-principle study will be evaluated in terms of quality control, user friendliness and data sharing strategies. However, this study might reveal further needs for improvements, covering prospective outbreak detection (e.g., effects on defining priorities and responsibilities, communication and potential increased risk management needs).

Here, we elucidate the past and current epidemiology in animal, food and humans, as well as important outbreaks of *Salmonella* in Germany. We present a working model ('GenoSalmSurv') for the establishment of a *Salmonella* surveillance, as well as source tracking system based on WGS data. Our approach for an integrated genome-based surveillance aims to strengthen the collaboration of public health and food safety authorities and allows authorities to rapidly detect *Salmonella* food-associated disease clusters, which supports a more focused epidemiological outbreak investigation.

MATERIALS AND METHODS

Reporting *Salmonella* Epidemiology and Outbreaks in Germany

The German Protection against Infection Act (IfSG)³ lays down the notifiable diseases and mandatory deadlines for reporting. The obligation to notify includes the suspicion of and disease from microbial food poisoning or acute infectious gastroenteritis, if two or more similar diseases occur for which an epidemic link is probable or suspected (§6 IfSG). Further, any direct or indirect evidence of certain pathogens (including *Salmonella* spp.) in patients shall be notified if indicating an acute infection (§7 IfSG). According to these legal requirements, public health laboratories or practitioners report to the local health authorities who evaluate the notifications and transmit the anonymized data to the state health departments. In 2001, the electronic reporting system

³Act on the Prevention and Control of Human Infectious Diseases in Man (IfSG) of 20 July 2002. BGBl. I p. 1045, as amended by article 5 of the Law of 19 June 2020. BGBl. I p. 1385. Available at: <https://www.gesetze-im-internet.de/ifsg/>

for surveillance of notifiable infectious diseases [SurvNet@RKI (Faensen et al., 2006)] was established. Notifications from the regional bodies are finally recorded and analyzed at the RKI, the German federal public health authority. Food-borne outbreaks can also be identified by consumer complaints to the food safety authorities (Figure 1). According to the Protection against Infection Act and the German Food, Feed and Consumer Goods Code⁴, there is a mutual information obligation between the public health and food safety sector on local level. Germany is a federal state where the competence for food-borne outbreak investigations lies with the local authorities. In the case of outbreaks spreading across jurisdictions, superior federal state authorities and federal authorities can be involved. The flow of information is depending on federal state structures and emergency contingency plans of the federal states. Cooperation between laboratories is necessary for subtyping of isolates, as well as comparison of typing data between the sectors. Depending on laboratory capacities in the 16 federal states, isolates for outbreak investigations are subtyped on regional or federal level. Public health laboratories voluntarily forward human isolates for typing to the NRC-*Salmonella* at the RKI. National Reference Centers are appointed for a 3-year period by the Federal Ministry of Health as competence centers in the field of laboratory science for a particular pathogen or group of pathogens. There is a regular evaluation by the scientific advisory board for public health microbiology and expert reviewers. Official food laboratories send isolates from food, production environments and primary production to the NRL-*Salmonella* at the BfR. In addition, the 'NRL-for *Salmonella* in Cattle' at the FLI investigates *Salmonella* isolates obtained from cattle in Germany. Both NRLs perform their work in accordance with the Zoonoses Directive 2003/99/EC⁵, the Zoonoses Regulation (EC) No 2160/2003⁶, and the §64 of the German Food, Feed and Consumer Goods Code. The typing results are communicated downstream within sectors (Figure 1). If there is an outbreak affecting more than one federal state, epidemiological information and concise typing data are shared between the sectors *via* the federal contact point for food-borne outbreaks, which is located at BVL.

WGS Tools Selected for the Working Model for the Surveillance of *Salmonella* spp.

For the decentralized initial bioinformatics analysis of raw sequencing data on site, a number of freely available, open-source

⁴Food, Feed and Consumer Goods Code in the version of the announcement of 3 June 2013. BGBl. I p. 1426, as amended by article 97 of the Regulation of 19 June 2020. BGBl. I p. 1328. Available at: <https://www.gesetze-im-internet.de/lfgb/LFGB.pdf>

⁵Directive 2003/99/EC of the European Parliament and of the Council of 17 November 2003 on the monitoring of zoonoses and zoonotic agents, amending Council Decision 90/424/EEC and repealing Council Directive 92/117/EEC. OJ L 325, 12 December 2003 pp. 31–40, as amended by Council Directive 2013/20/EU of 13 May 2013. OJ L 158 p. 234, 10 June 2013. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02003L0099-20130701&from=EN>

⁶Regulation (EC) No 2160/2003 of the European Parliament and of the Council of 17 November 2003 on the control of *Salmonella* and other specified food-borne zoonotic agents. OJ L 325, 12 December 20. Available at: <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:325:0001:0015:EN:PDF>

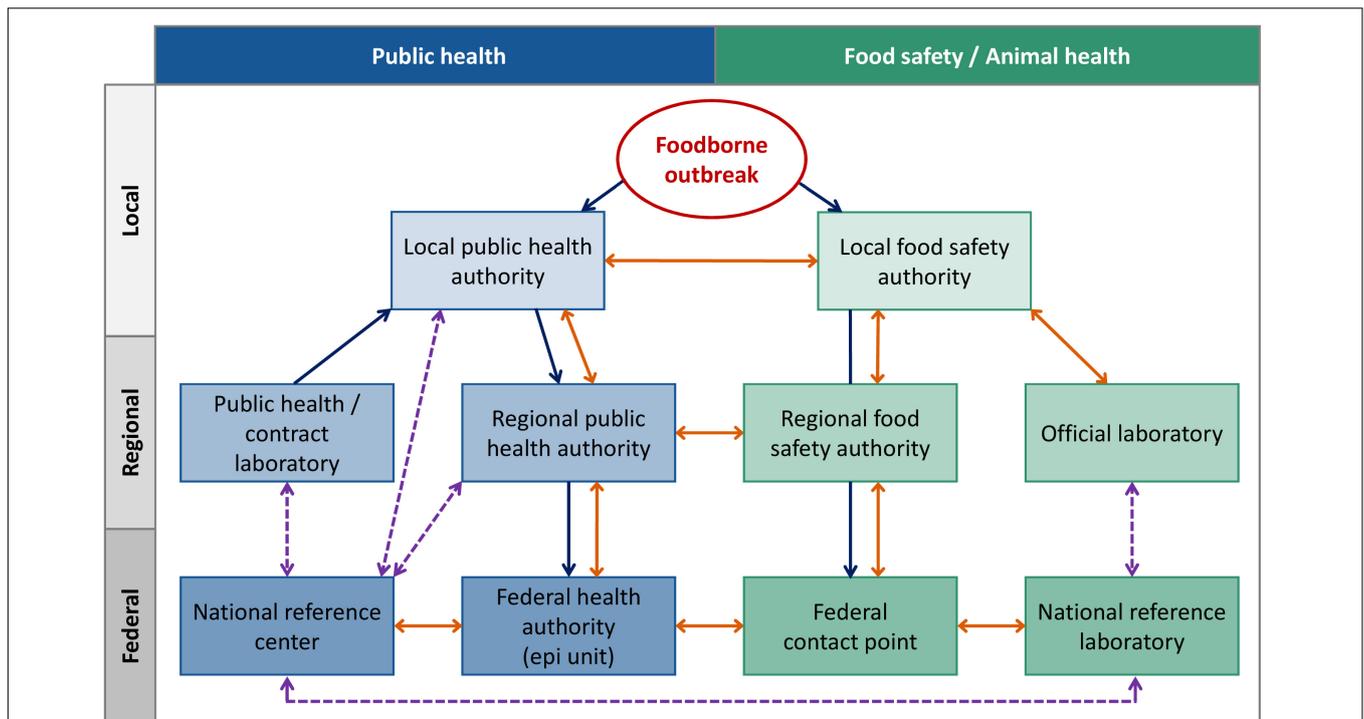


FIGURE 1 | Overview on reporting (blue), information exchange (orange), and subtyping activities (dashed line, purple) for food-borne outbreaks between the German public and animal health as well as food safety sector.

pipelines have been customized for performing (i) quality control and assembly, (ii) bacterial characterization, (iii) cgMLST-based subtyping and (iv) phylogenetic analysis (Figure 2). The AQUAMIS pipeline (Assembly-based Quality Assessment for Microbial Isolate Sequencing)⁷ encompasses trimming and read-quality control (QC) using fastp (Chen et al., 2018), assembly using shovill⁸, automated reference search of complete NCBI Refseq genomes using mash (Ondov et al., 2016), and assembly-quality analysis using Quast⁹. Based on the draft assemblies produced by AQUAMIS, the Bakcharak pipeline¹⁰ conducts a characterization of the *Salmonella* genome. The Bakcharak pipeline runs NCBI amrfinder (Feldgarden et al., 2019) to detect antimicrobial resistance genes and ABRicate¹¹ to detect plasmid incompatibility groups from the plasmidfinder database (Carattoli et al., 2014), as well as virulence genes from the VFDB (Chen et al., 2004). Moreover, 7-gene MLST is performed using mlst¹² and the PUBMLST schemes (Jolley and Maiden, 2010). The *Salmonella* serovar is determined with the SISTR software (Yoshida et al., 2016). The chewieSnake pipeline¹³ performs cgMLST typing by implementing the assembly-based software chewBBACA (Silva et al., 2018). The alleles for a set of samples

are called using chewBBACA v2.0.12, allele profiles are then combined and an allele distance matrix and minimum spanning tree are calculated using GrapeTree v2.1 (Zhou et al., 2018). The cgMLST scheme for *Salmonella* was derived from Enterobase¹⁴. The AQUAMIS, Bakcharak and chewieSnake bioinformatic pipelines are version controlled and associated with release tags. All contained software versions are fully tracked with the conda software management tool.

RESULTS

Epidemiology of *Salmonella* in Germany

Salmonellosis is one of the most common bacterial food-borne illnesses worldwide, with approximately 153 million cases and 57,000 deaths per year (Healy and Bruce, 2017). Despite the remarkable and steady drop of reported human *Salmonella* infections from almost 195,000 in 1992 to about 14,000 in recent years, salmonellosis is, after campylobacteriosis, still the second most frequent bacterial food-borne disease in Germany. In the 5-year period from 2015 to 2019 (RKI, 2016, 2017, 2018, 2019) the number of reported cases ranged from 12,962 to 14,269, representing an incidence of 16 to 17 cases per 100,000 inhabitants. The highest incidence has always been found in children younger than 5 years. The majority of human salmonellosis cases in Germany develops gastrointestinal symptoms, such as (generally self-limiting) diarrhea and stomach

⁷https://gitlab.com/bfr_bioinformatics/AQUAMIS

⁸<https://github.com/tseemann/shovill>

⁹<https://github.com/ablab/quast>

¹⁰https://gitlab.com/bfr_bioinformatics/bakcharak

¹¹<https://github.com/tseemann/abricate>

¹²<https://github.com/tseemann/mlst>

¹³https://gitlab.com/bfr_bioinformatics/chewieSnake

¹⁴<http://enterobase.warwick.ac.uk/species/index/senterica>

were *S. Typhimurium* and its monophasic variant, whereas the most frequent serovars in sows were *S. Derby* followed by *S. Typhimurium*.

Within the National Zoonosis Monitoring Program, *Salmonella* was detected less frequently in fresh pork and beef than in fresh poultry meat in Germany. The dominating serovars in poultry products were *S. Infantis*, followed by *S. Paratyphi B d-tartrate fermenting (dT+)*, and *S. Typhimurium*, as well as its monophasic variant in pork (BVL, 2018a, 2019). Considering foods of plant origin, the NRL-*Salmonella* only sporadically receives isolates for further serotyping, indicating a rather low prevalence. This is confirmed by the EU-wide monitoring data which revealed a *Salmonella* detection rate of only 0.84% in ready to eat pre-cut fruits and vegetables (EFSA and ECDC, 2017), although outbreaks were detected (Table 1).

Table 2 shows the top five *Salmonella* serovars in food (all categories), laying hens, broilers, pigs and cattle serotyped in

2019 at the NRL-*Salmonella* and the 'NRL for Salmonellosis in Cattle' in Germany. Isolates from laying hens, pigs and cattle were collected at primary production stage, while strains from food originate from all production stages.

Salmonellosis Outbreaks in Germany

Salmonella enterica regularly causes large (supra-)regional food-borne outbreaks in Germany (Table 1) and thus is a pathogen of extraordinary public health and economic relevance (EFSA and ECDC, 2018). In the 5-year period from 2014 to 2018, approximately 260 salmonellosis outbreaks had been reported per year, comprising a median of 1,090 cases/year. The number of cases per outbreak ranged from 2 to 191. Occasionally, local outbreaks have been detected in day-care centers for children or retirement homes.

A high number of outbreaks are caused by the most prevalent serovars *S. Enteritidis* and *S. Typhimurium*. *S. Enteritidis* has

TABLE 1 | Selected *Salmonella* outbreaks in Germany since 2000.

Year	Serovar	Cases	Suspected vehicle of infection	Remarks	References
2018/2019	<i>S. Enteritidis</i>	>300	Eggs, egg products		Unpublished
2018	<i>S. Enteritidis</i>	30	'Spätzle' (fresh pasta with egg)		RKI (2019)
2018	<i>S. Panama</i>	28	Not identified	Retirement home	RKI (2019)
2017	<i>S. Enteritidis</i>	52	Not identified	Day-care center for children	RKI (2018)
2017	<i>S. Kottbus</i>	50	Ham	At the same time a smaller <i>S. Kottbus</i> outbreak related to quail eggs	Enkelmann et al. (2020)
2017	<i>S. Typhimurium</i>	31	Pork products		RKI (2018)
2017	<i>S. Enteritidis</i>	33	Bakery products		RKI (2018)
2016	<i>S. Enteritidis</i>	20	Not identified	Retirement home	RKI (2017)
2016/2017	<i>S. Vari</i>	47	Sesame products	Multinational, 13 German cases; novel serovar	Meinen et al. (2019)
2016	<i>S. Enteritidis</i>	14	Potato salad		RKI (2017)
2014	<i>S. Muenchen</i>	247	Pork products	Common source of 2013/2014 outbreaks in the primary production	Schielke et al. (2017)
2014	<i>S. Enteritidis</i>	>350	Eggs, egg products	Multinational, majority of cases in the United Kingdom	Dallman et al. (2016)
2014	<i>S. Typhimurium</i>	40	Ground pork		RKI (2015)
2014	<i>S. Bovismorbificans</i>	74	Sprouts	49 German cases, 25 in Switzerland	Knoblauch et al. (2015); RKI (2015)
2013/2014	<i>S. Derby</i>	145	Pork products ('Teewurst')	Hospitals and retirement homes	Simon et al. (2018)
2013	<i>S. Muenchen</i>	203	Pork products	Common source of 2013/2014 outbreaks in the primary production	Schielke et al. (2017)
2013	<i>S. Infantis</i>	267	Pork products		Schroeder et al. (2016)
2011	<i>S. Strathcona</i>	71	Tomatoes	Multinational, 21 German cases	Müller et al. (2016)
2011	<i>S. Newport</i>	106	Sprouts		Bayer et al. (2014)
2010	<i>S. Montevideo</i>	15	Food supplement	Hemp-based herbal food supplement	Stöcker et al. (2011)
2004	<i>S. Give</i>	115	Ground pork		Jansen et al. (2005)
2002/2003	<i>S. Agona</i>	42	Aniseed	Aniseed-fennel infusion for infants	Koch et al. (2005)
2001	<i>S. Oranienburg</i>	>400 (estimated)	Chocolate	Multinational	Werber et al. (2005)

TABLE 2 | Top five *Salmonella* serovars from German food, laying hens, broilers, pigs, and cattle isolates submitted in 2019 to NRLs for serotyping.

Food (all categories)		Laying hens		Broilers		Pigs		Cattle	
<i>n</i> = 750		<i>n</i> = 143		<i>n</i> = 52		<i>n</i> = 995		<i>n</i> = 307	
Serovar	%	Serovar	%	Serovar	%	Serovar	%	Serovar	%
<i>S. Typhimurium</i> , monophasic	14.5	<i>S. Enteritidis</i>	43.8	<i>S. Infantis</i>	23.1	<i>S. Typhimurium</i> , monophasic	40.7	<i>S. Typhimurium</i>	36.4
<i>S. Typhimurium</i>	13.6	<i>S. Typhimurium</i>	13.7	<i>S. Paratyphi B dT+</i>	17.3	<i>S. Typhimurium</i>	18.8	<i>S. Typhimurium</i> , monophasic	13.2
<i>S. Infantis</i>	12.8	<i>S. spp. I rough</i> ¹	5.5	<i>S. Enteritidis</i>	15.4	<i>S. Derby</i>	18.3	<i>S. Dublin</i>	27.4
<i>S. Derby</i>	10.0	<i>S. Mbandaka</i>	4.8	<i>S. Saintpaul</i>	5.8	<i>S. spp. I rough</i> ¹	4.6	<i>S. Enteritidis</i>	7.1
<i>S. spp. I rough</i> ¹	8.5	<i>S. Kiambu</i>	2.7	<i>S. Coeln</i>	5.8	<i>S. Infantis</i>	2.1	other serovars	15.9

¹*S. spp. I rough*: *Salmonella enterica* subspecies *enterica* harboring a rough lipopolysaccharide disabling serovar determination.

frequently been associated with eggs, egg products or egg dishes like Tiramisu and other (raw) egg-containing desserts, bakery products, 'Spätzle' (regional pasta-like specialty) or potato and pasta salads with mayonnaise. In contrast, *S. Typhimurium* outbreaks are often related to pork products, such as ground pork or different types of raw sausage (traditional meals in Germany). Both serovars might also be found in other food categories. These observations strongly reflect the situation in the EU (EFSA and ECDC, 2019).

In addition, food of plant origin like fruit and vegetables, herbs, spices and sprouts have been identified increasingly often as a source of *Salmonella* infections and outbreaks (Table 1). Foodstuffs contaminated with less frequent or even extremely rare serovars have also been associated with partly large (multinational) outbreaks.

In the summer of 2014, a multinational outbreak of *S. Enteritidis* was associated with an international egg distribution network (Dallman et al., 2016). The outbreak included more than 350 cases, reported in the United Kingdom, Germany, Austria, and Luxembourg. WGS investigations and phylogenetic analysis revealed a common ancestral relationship which indicated one particular German egg producer. The strength of the study was that WGS data in combination with information from the food supply network was applied. This approach enabled the trace-back to suspected sources across Europe categorizing the outbreak as multinational.

A few months earlier, one large salmonellosis outbreak, caused by *S. Derby* (a prevalent serovar in pigs), affected 145 primarily elderly people in hospitals and nursing homes (EFSA and ECDC, 2015; Simon et al., 2018) in the German states of Berlin and Brandenburg. Epidemiological investigations and additional microbiological evidence revealed raw fermented meat paste ("Teewurst") as the source. This point source outbreak was retrospectively investigated by WGS with special attention on the suitability of SNP and cgMLST based methods for cluster definition (Simon et al., 2018). In total, 55 isolates were selected (confirmed outbreak strains, probable outbreak strains and unrelated ones) and analyzed. Overall, the WGS approach confirmed the conventional typing results and additionally identified two of the seven probable outbreak strains as part of the outbreak.

Over a period of 2 years, during 2016 and 2017, a previously undescribed *Salmonella* serotype 11:z41:e,n,z15, now defined as *S. Vari*, caused a multi-national salmonellosis outbreak. The

first cases were reported in Greece (Mandilara et al., 2016), followed by cases in Germany, Czechia, Luxembourg, France, Serbia and the United Kingdom. Using WGS, a very close genetic relationship of isolates from different countries was observed and further epidemiological analysis revealed sesame products, processed in a Greek factory, as vehicles of infection. Samples from raw material and processed food acquired in the course of the event, tested positive for *S. Vari* and could be confidently attributed to human cases by WGS, taking into account the epidemiological data (Meinen et al., 2019).

In 2017, a multi-state outbreak of the rare serotype *S. Kottbus* in Germany involved 69 cases (Enkelmann et al., 2020). WGS enabled the differentiation of isolates belonging to the potential outbreak into three independent co-circulating clusters. For two of the identified clusters, the vehicle of transmission could be ascertained by epidemiological investigation: one was associated with raw smoked ham consumption, the other with quail egg consumption. For the third cluster no common food item was identified. Without WGS, attribution of all cases to just one event would have weakened the epidemiological evidence obtained from questionnaires and therefore impeded the identification of the vehicle of the first cluster outbreak and obscured the strong common exposure of quail egg consumption reported by the second cluster cases.

In 2017, a *S. Agona* outbreak among infants was observed, which involved 37 cases in France and two further international cases. Epidemiological and WGS investigation revealed infant milk products from a French supplier as the source (Jourdan-da Silva et al., 2018). Simultaneously, *S. Agona* was detected in three animal feed samples in Bavaria. In a retrospective WGS study, including the feed isolates and 48 additional *S. Agona* isolates (Bavarian isolates from 1993 to 2018), a connection between the feed isolates and the French outbreak was ruled out (Dangel et al., 2019). Furthermore, the WGS approach confirmed clusters that were previously identified by epidemiological investigations and detected additional ones.

The GenoSalmSurv Working Model for Cluster Detection and Bioinformatics Analysis Tools Used for Real-Time Surveillance

Beside the generation of high quality sequencing data, data sharing between laboratories is a necessary requirement for

efficient outbreak investigations (Aarestrup and Koopmans, 2016; FAO, 2016; Jagadeesan et al., 2019). However, any data sharing strategy needs to evolve around currently existing data protection regulations. Legal and ethical issues comprising confidentiality, data protection and intellectual property rights are of great concern (Pisani and AbouZahr, 2010; Roche et al., 2014). All parties involved in the handling of sequencing and metadata during outbreak and surveillance studies need to legally comply with such regulations. Previous studies have shown that many different interests hamper the sharing of data and these discrepancies still exist (Aarestrup and Koopmans, 2016; FAO, 2016). On European level, ECDC (European Centre for Disease Prevention and Control) and EFSA work closely together in outbreak detection and the EU legislation is considering the use of WGS data wherever an application is reasonable (EFSA Panel on Biological Hazards, 2019). The sharing of metadata along with genome sequence data, is a particular challenge as data protection is of high consideration in Germany. This restricts the exchange of sequence and metadata between federal states, as well as across sectors and thus complicates evaluation of outbreak analyses. To address these issues, joint working groups have begun to elaborate on WGS data sharing concepts. The aim is to reach a broad acceptance for data sharing embedded within a legal framework, in order to fully benefit from the application of WGS methodologies in food safety and public health.

Recognizing the existing barriers in data sharing, the GenoSalmSurv project has developed a practical technical solution for performing *Salmonella* real-time surveillance across all sectors. The model is based on a decentralized initial bioinformatics analysis of raw sequencing data on site, dispensing a direct exchange of raw sequence data between partners. Each partner utilizes their own local computing infrastructure and communicates results in a standardized format which facilitates data exchange (shrinking data exchange volume) and joint interpretation (harmonized results). A prerequisite for this approach is the use of harmonized analysis pipelines. Importantly, these pipelines must be easy to deploy, maintain and run on different (Linux) computers and servers. For the joint analysis, freely available, open-source pipelines have been customized for performing (i) quality control and assembly, (ii) bacterial characterization, (iii) cgMLST-based subtyping and (iv) phylogenetic analysis (Figure 2).

For the simplification of the local installation of all software components of the pipelines, the conda/bioconda software management tool is used, which ensures that every local system is equipped with the exact same software versions (Grüning et al., 2018).

For initial analyses of the bacterial isolate we deployed the AQUAMIS pipeline (Assembly-based Quality Assessment for Microbial Isolate Sequencing) (see text footnote 7). This snakemake pipeline (Koster and Rahmann, 2012) encompasses trimming and read-quality control (QC) using fastp (Chen et al., 2018), assembly using shovill (see text footnote 8), automated reference search of complete NCBI Refseq genomes using mash (Ondov et al., 2016), and assembly-quality analysis using Quast (see text footnote 9). The AQUAMIS pipeline produces a user-friendly html report, which summarizes essential quality

information. Important quality parameters, like the Q30 base fraction and the coverage depth are color-coded according to the traffic light system (green – good, yellow – sufficient, red – insufficient), which facilitates an easy visual first-pass inspection. A number of other quality parameters are calculated for a subsequent in-depth analysis of samples of interest.

Table 3 summaries important minimal quality control parameter for WGS of *Salmonella* isolates based on our experience with library preparation using the Nextera DNA Flex library preparation protocol with subsequent sequencing on an Illumina platform (MiSeq, NextSeq or iSeq). Other examples for minimal quality parameters determined for sequencing of bacterial isolates were previously published by Kozyreva et al. (2017) (Nextera XT library preparation and MiSeq sequencing).

The International Organization for Standardization (ISO) is currently in the process of developing a standard for WGS for typing and genomic characterization of food-borne bacteria¹⁷. Once publicly available, this ISO standard might contribute to a revision of quality parameters currently in place in different laboratories. Based on the draft assemblies produced by AQUAMIS, the Bakcharak pipeline (see text footnote 10) then conducts a thorough characterization of the *Salmonella* genome. The Bakcharak pipeline provides a detailed characterization for each sample including serotype, 7-gene MLST, resistome, plasmidome and virulome. Since all software tools and databases are standardized within the pipeline, all partners can straightforwardly arrive at comparable characterization results. The characterization results provide important additional information to the phylogeny, as they allow the detailed study of the transmission of particular resistance mechanisms and plasmids.

CgMLST analysis is performed on the draft assemblies with the chewieSnake pipeline (see text footnote 13) by implementing the assembly-based software chewBBACA (Silva et al., 2018). As a central feature, the nucleotide sequence of identified

¹⁷ISO/CD 23418. Available at: <https://www.iso.org/standard/75509.html>

TABLE 3 | Minimal quality parameters for Illumina sequencing of *Salmonella* isolates.

Parameter	Passing threshold
% of bases with a quality score > Q30 (%Q30)	2 × 301 bp: >70% 2 × 251 bp: >75% 2 × 151 bp: >80% 2 × 76 bp: >85%
Cluster passing filter	>75%
PhiX aligned	0.5–10%
PhiX error rate	<6%
Number of bases after trimming (per sample)	>150,000,000
Fraction of closest NCBI reference covered	>0.9
Number of contigs in <i>de novo</i> assembly	<300
N50 of the <i>de novo</i> assembly	>20,000
Average coverage of the <i>de novo</i> assembly	>30
Assembly length	4.5–5.5 Mb

allele variants (including novel alleles) are converted to allele hashes. This ensures that identical allele variants are always assigned the same allele number, irrespective of analysis location and execution. Hence, when all partners analyze sequenced isolates in the same manner, i.e., using the same pipelines, reference databases (e.g., cgMLST scheme) and underlying software programs, the exchange of hashed allele profiles is feasible and sufficient for the joint interpretation of the analysis results. In particular, after exchange of the shared hashed allele profiles a joint allele distance matrix and a common minimum spanning tree can be calculated (Figure 3). Based on the joined allele distance matrix, samples are hierarchically clustered and assigned into cluster groups using a threshold of 10 AD. A unique and stable cluster name is assigned to each cluster group, thus providing an intercommunicable cluster nomenclature. Finally, all clustering information is summarized in a clustering report that is shared between the partners. The minimum spanning tree can be easily visualized together with the characterization results and important metadata information in Grapetree (Zhou et al., 2018).

The exchange of metadata and their unambiguous correlation to sequence data are crucial for a successful evaluation of transmission dynamics. Within the GenoSalmSurv project, the exchange of metadata is limited to isolate source attributes (categorized as shown Table 4), name of the state which forwarded the isolate, isolate submission date, isolate sampling date and information about phenotypical resistance (if available). Metadata information is securely stored throughout the project on our own server infrastructure (hosted at the BfR), physically apart from the cgMLST allele database. Cloud access to the metadata information is password protected and read/write permission is granted to all project partners. New data is added to the interactive table, whenever new allele profiles are exchanged. A consistent isolate naming scheme guarantees an automatic matching of metadata information to nodes of the phylogenetic tree.

Interpretation of Cluster Assignment

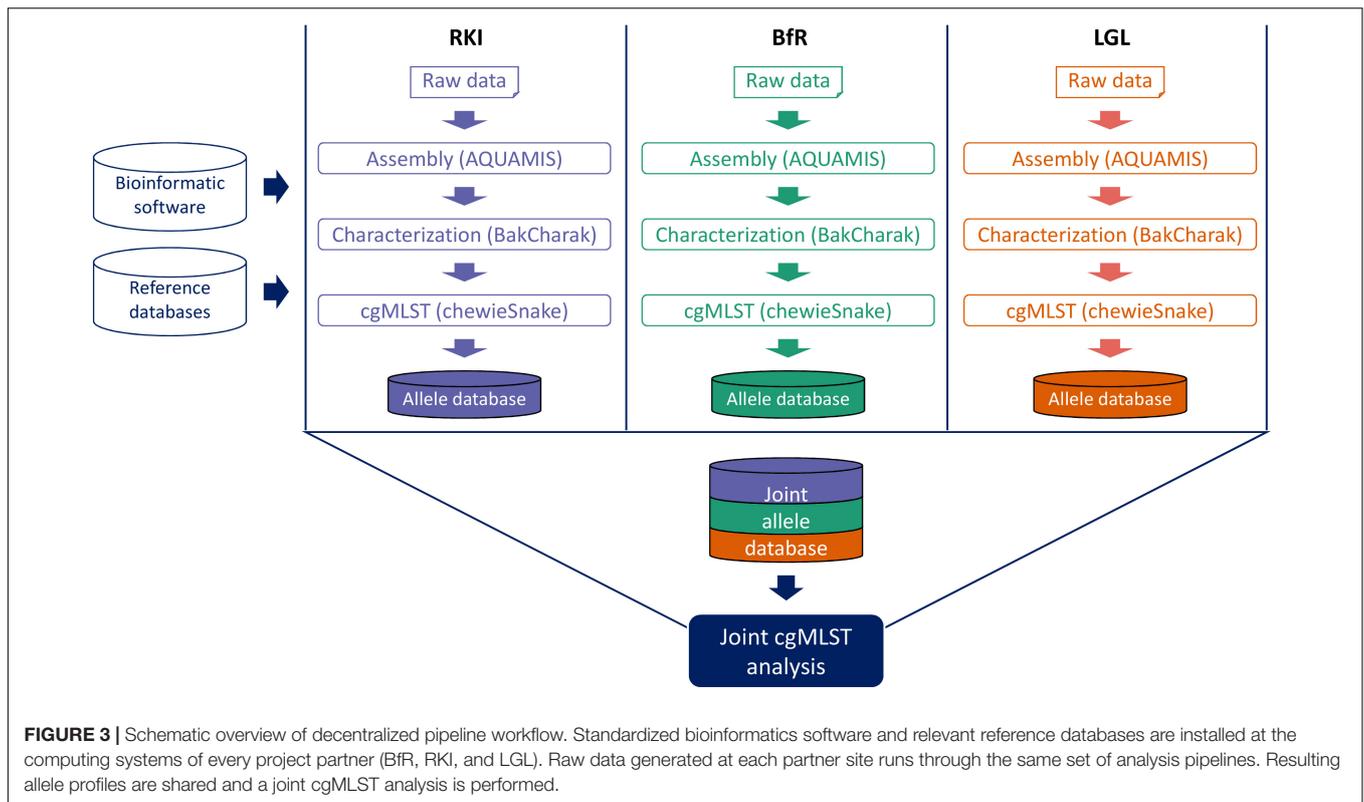
The prospective monitoring of *Salmonella* spp. using WGS, in which sequenced genomes are continuously added to previously sequenced genomes, can generate actionable cluster information for epidemiologists. However, the formation and the interpretation of WGS clusters must be consistent and robust over time. Differences between genomes are usually calculated by pairwise comparison of the isolates. The resulting matrix contains the number of allele differences between two isolates (Timme et al., 2018). For visualization, a tree is built using a distance-based hierarchical method [e.g., unweighted pair group method with arithmetic mean (UPGMA)] based on cgMLST/whole-genome MLST (wgMLST) analysis or a non-hierarchical method (e.g., maximum-likelihood) based on SNP analysis (Schürch et al., 2018; Uelze et al., 2020). For a first screening, many laboratories apply a specific allele or SNP threshold value for allocation of isolates to a WGS cluster. For *Salmonella*, practical thresholds for interpreting clusters in regulatory applications have been published on SNP counts (Pightling et al., 2018) and cgMLST alleles (Besser et al., 2019). A suggested working definition for *Salmonella* clusters includes ≥ 3 cases within a

60-day period with 0–10 alleles difference (depending on the applied typing scheme), where at least two cases are related by at most five differences (Besser et al., 2019). However, this liberal allele difference might not prove useful for very clonal serovars, e.g., *S. Enteritidis*, or point source outbreaks. Furthermore, the sole setting of a distance threshold for inclusion or exclusion of isolates to an outbreak should be used with caution and not without considering the epidemiological context (EFSA Panel on Biological Hazards, 2019). Generally, cases with one or more epidemiological links and near-identical pathogen genome sequences, can be supposed to be infected by the same source or vehicle. Individual cases within a WGS cluster for whom no spatio-temporal link could be established are more difficult to allocate to a distinct outbreak event.

The evolution rates of certain *Salmonella* serovars over time have been estimated recently to be approximately 1 to 5 bases per year per genome (Hawkey et al., 2013; Deng et al., 2014; Phillips et al., 2016; Leekitcharoenphon et al., 2019). Such mutation rates are calculated by temporally structured sequence analyses within a Bayesian framework based on sets of isolates giving an indication for the rate of mutational changes within a population over the time (Leekitcharoenphon et al., 2019). Nevertheless, recognizing the dimension of an outbreak, possibly running over years and consisting of several WGS subclusters, requires extensive consideration of epidemiological data and an understanding of the biology of the pathogen itself, as well as the environmental conditions in which the pathogen persists. For example, stress responses to harsh conditions such as chemical substances (e.g., disinfection), temperature, acid, dryness or the host environment lead to prolonged generation time, increased mutability, or reduction of a particular *Salmonella* population (Jayeola et al., 2020). The persistence of *Salmonella* spp. for long periods of time in dry food production environments has been observed (Hoffmann et al., 2020). Further, it is known that *Salmonella* spp. are capable of biofilm formation. Therefore, the biological peculiarities of particular *Salmonella* strains may confound data interpretation during outbreak investigations.

DISCUSSION

Whole genome sequencing technologies have pushed the development for bacterial genome comparisons and advanced typing approaches. The successful identification of (cross-border) transmission and outbreaks analyses requires the comparative analyses of genome data across sectors and borders. A number of international WGS-based surveillance platforms (Uelze et al., 2020) have been developed up to date, such as Pathogenwatch (see text footnote 1), NCBI pathogen detection (see text footnote 2) and GenomeTrakr (Timme et al., 2018) which utilize either cgMLST or SNP typing approaches to identify food-borne pathogen outbreaks. All three platforms require that raw sequence data is available in a central data storage, followed by a centralized analysis and cluster detection of the genome data. Currently, these platforms represent the most promising solution for a global food-borne disease surveillance strategy. However, many countries and institutions only hesitantly participate



in these networks (especially where the sharing of metadata along with genome sequence data is concerned), often due to existing data protection guidelines, regulations, respective research interests or privacy concerns. This results in many missed opportunities for important cluster detection and wastes

the potential which is encompassed by the generation of valuable sequence data. Although we strongly encourage the participation in international surveillance networks, we recognize the specific barriers which frequently prevent large-scale, real-time data sharing.

TABLE 4 | Details of minimal isolate source attributes that are exchanged between project partner.

Food ¹	Farm animal ¹	Human ¹	Animal ¹	Environment ¹	Feed ¹
Pork	Laying hen	Blood	Wild animal	Slurry	Feed plant origin
Beef	Broiler	Stool	Domestic animal	Water	Feed animal origin
Poultry meat	Turkey	Urine	Zoo animal	Other	Other
Other/mixed meat	Other poultry	Other	Laboratory animal		
Eggs	Cattle	Unknown	Other		
Other foods animal origin	Pig		Unknown		
Other foods plant origin	Other				
Other	Unknown				
Unknown					

¹Isolates which cannot be assigned to one of the listed attributes can be categorized as 'other' or 'unknown.' A free text field for additional information is available.

To address this issue, we have developed the GenoSalmSurv project as a working model for an integrated genome-based surveillance of *Salmonella enterica*, which permits reliable cluster identification without sequence data exchange.

Our concept is based on a decentralized analysis of hashed cgMLST allele profiles with open-source bioinformatics tools. Crucially this approach eliminates the need for a large centrally organized and synchronized cgMLST scheme database and sequence database and simplifies the data sharing process, as the exchange of small text files is sufficient for the compilation of data for cluster identification and analysis. In comparison to the aforementioned central analysis platforms, our approach improves scalability and speed of analysis, as time- and resource-consuming data storage, management and processing needs are distributed, and upload and download times are minimized. These are major advantages compared to a central system, which is continuously required to expand, in order to meet the needs of the exponentially growing amount of sequencing data, which places increasingly higher demands for storage and processing of these data on the central computing/storage unit.

Furthermore, as the exchange of data is limited to the essential information necessary for cluster analysis (i.e., the cgMLST allele profiles) our approach provides a practical solution for laboratories which are restricted in their ability to share/exchange

sequence data, for example during outbreak situations. An exchange of the raw sequencing data – for instance by uploading to open-access databases – can then be accomplished at a later time point and is highly recommended for in-depth analysis, independent confirmation and integration with international surveillance networks. In addition, the integration of comprehensive and standardized epidemiological data is of utmost importance for reliable data interpretation.

Overall, our objective is to provide accessible, “easy to handle” and standardized tools for the high-resolution method of genome analysis for cluster detection, pathogen monitoring and infection source analysis to all authorities involved in the surveillance of *Salmonella*. Our working model is built around practical solutions for data sharing, with the aim to enable all laboratories to contribute their genomic data to an integrated genome-based surveillance and outbreak analyses with uniform interpretation of results.

During the next project stage, the developed open-source tools will be optimized in regards to user friendliness and backward compatibility and an administration portal to manage user accounts and access rights will be established. The results of a currently ongoing real-time study will be used to validate the established workflow and will further be incorporated into training measures. It is further planned to carry out a benchmarking test with selected public health laboratories already performing WGS, to determine and compare their sequence data quality. Finally, authorities and multipliers will be surveyed, consulted and trained with the aim to make genome-based surveillance and outbreak analyses easily accessible to all authorities involved and thus to accelerate its establishment across all sectors (one-health concept).

DATA AVAILABILITY STATEMENT

The code for the AQUAMIS, Bakcharak, and chewieSnake bioinformatic pipelines is freely available from GitLab (https://gitlab.com/bfr_bioinformatics/).

AUTHOR CONTRIBUTIONS

LU and BM structured the manuscript. LU, NB, MB, UB, AD, CD, JF, AF, SH, IH, UM, JL, MP, SS, AS, SHT, IS, and BM wrote the manuscript. CD and ST developed the in-house

bioinformatic pipelines used for analysis of the sequencing data. NB, MP, and CD created the figures. All authors read and approved the manuscript.

FUNDING

The project ‘Integrated Genome-Based Surveillance of *Salmonella*’ (GenoSalmSurv) was funded by the German Ministry of Health, decision ZMVII-2518FSB709 of 26 November 2018 on the basis of a resolution of the German Bundestag by the Federal Government.

ACKNOWLEDGMENTS

BfR: We are grateful for the continuous collaboration with the German Federal State laboratories, who provided us with *Salmonella* isolates. We thank Martha Brom, Johanna Ledwolorz, Manuela Rister, and Angelina Bloch (National Reference Laboratory for the Analysis and Testing of Zoonoses [*Salmonella*]), as well as Beatrice Baumann and Katharina Thomas (Study Centre for Genome Sequencing and Analysis), for their excellent laboratory assistance.

RKI: Our thanks are given to the local and regional laboratories for continuously providing the NRC with *Salmonella* isolates from patients. Many thanks also to our skillful technicians Marita Wahnfried, Susanne Kulbe, and Monique Duwe from the *Salmonella* typing lab at the NRC. Further we wish to thank the colleagues from the RKI sequencing core unit (MF2) for constantly producing high quality sequence data as the essential basis for all subsequent analyses.

LGL: We would like to thank Stefan Hörmansdorfer, Julia Riehm, Matthias Hanczaruk, Gesine Schulze, Mareike Wenning, and Hans Bauer and their teams for providing and typing *Salmonella* isolates. Special thanks goes to the team of the Department of Bacteriology (Stefan Hörmansdorfer, Anja Berger, Katja Bengs, Bernhard Hobmaier, Turgut-Cengiz Dedeoglu, Peggy Zill, Dagmar Meitzler) for professional support in the *Salmonella* laboratory. We would also like to gratefully acknowledge Jasmin Fräßdorf, Sabine Lohrer, and Marion Linder Mayer for their excellent laboratory assistance (NGS). We further want to thank Nancy Bretschneider and Larissa Murr for testing the pipelines described in this article.

REFERENCES

- Aarestrup, F. M., Brown, E. W., Detter, C., Gerner-Smidt, P., Gilmour, M. W., Harmsen, D., et al. (2012). Integrating genome-based informatics to modernize global disease monitoring, information sharing, and response. *Emerg. Infect. Dis.* 18:e1. doi: 10.3201/eid1811.120453
- Aarestrup, F. M., and Koopmans, M. G. (2016). Sharing data for global infectious disease surveillance and outbreak detection. *Trends Microbiol.* 24, 241–245. doi: 10.1016/j.tim.2016.01.009
- Allard, M. W., Strain, E., Melka, D., Bunning, K., Musser, S. M., Brown, E. W., et al. (2016). Practical value of food pathogen traceability through building a whole-genome sequencing network and database. *J. Clin. Microbiol.* 54, 1975–1983. doi: 10.1128/JCM.00081-16
- Ashton, P. M., Nair, S., Peters, T. M., Bale, J. A., Powell, D. G., Painset, A., et al. (2016). Identification of *Salmonella* for public health surveillance using whole genome sequencing. *PeerJ* 4:e1752. doi: 10.7717/peerj.1752
- Bayer, C., Bernard, H., Prager, R., Rabsch, W., Hiller, P., Malorny, B., et al. (2014). An outbreak of *Salmonella* Newport associated with mung bean sprouts in Germany and the Netherlands, October to November 2011. *Eurosurveillance* 19:20665. doi: 10.2807/1560-7917.ES2014.19.1.20665
- Besser, J. M., Carleton, H. A., Trees, E., Stroika, S. G., Hise, K., Wise, M., et al. (2019). Interpretation of whole-genome sequencing for enteric disease surveillance and outbreak investigation. *Foodborne Pathog. Dis.* 16, 504–512. doi: 10.1089/fpd.2019.2650

- BFR (2020). *Salmonellen-Bekämpfungsprogramm – Ergebnisse für das Jahr 2019*. Available online at: <https://www.bfr.bund.de/cm/343/salmonellen-bekaempfungsprogramm-ergebnisse-fuer-2019.pdf> (accessed September–December, 2020).
- BVL (2016a). *Berichte zur Lebensmittelsicherheit – Zoonosen-Monitoring 2015*. Berlin: Bundesamt für Verbraucherschutz und Lebensmittelsicherheit. Available online at: https://bvl.bund.de/SharedDocs/Downloads/01_Lebensmittel/04_Zoonosen_Monitoring/Zoonosen_Monitoring_Bericht_2015.pdf?__blob=publicationFile&v=6 (accessed September–December, 2020).
- BVL (2016b). *Zoonoses Monitoring 2015 – Summary of Findings and Conclusions*. Available online at: https://www.bvl.bund.de/SharedDocs/Downloads/01_Lebensmittel/04_Zoonosen_Monitoring/Zoonosen_Monitoring_Bericht_2015_en-summary.pdf?__blob=publicationFile&v=6 (accessed September–December, 2020).
- BVL (2018a). *Berichte zur Lebensmittelsicherheit – Zoonosen-Monitoring 2017*. Berlin: Bundesamt für Verbraucherschutz und Lebensmittelsicherheit. Available online at: https://www.bvl.bund.de/SharedDocs/Downloads/01_Lebensmittel/04_Zoonosen_Monitoring/Zoonosen_Monitoring_Bericht_2017.pdf?__blob=publicationFile&v=2 (accessed September–December, 2020).
- BVL (2018b). *Zoonoses Monitoring 2017 – Summary of Findings and Conclusions*. Available online at: https://www.bvl.bund.de/SharedDocs/Downloads/01_Lebensmittel/04_Zoonosen_Monitoring/Zoonosen_Monitoring_Bericht_2017_en-summary.pdf?__blob=publicationFile&v=6 (accessed September–December, 2020).
- BVL (2019). *Berichte zur Lebensmittelsicherheit – Zoonosen-Monitoring 2018*. Berlin: Bundesamt für Verbraucherschutz und Lebensmittelsicherheit. Available online at: https://www.bvl.bund.de/SharedDocs/Downloads/01_Lebensmittel/04_Zoonosen_Monitoring/Zoonosen_Monitoring_Bericht_2018.pdf?__blob=publicationFile&v=7 (accessed September–December, 2020).
- Carattoli, A., Zankari, E., García-Fernández, A., Voldby Larsen, M., Lund, O., Villa, L., et al. (2014). *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* 58, 3895–3903. doi: 10.1128/AAC.02412-14
- Chen, L., Yang, J., Yu, J., Yao, Z., Sun, L., Shen, Y., et al. (2004). VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* 33, D325–D328. doi: 10.1093/nar/gki008
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560
- Chiu, C.-H., Su, L.-H., and Chu, C. (2004). *Salmonella enterica* serotype Choleraesuis: epidemiology, pathogenesis, clinical disease, and treatment. *Clin. Microbiol. Rev.* 17, 311–322. doi: 10.1128/CMR.17.2.311-322.2004
- Dallman, T., Inns, T., Jombart, T., Ashton, P., Loman, N., Chatt, C., et al. (2016). Phylogenetic structure of European *Salmonella* Enteritidis outbreak correlates with national and international egg distribution network. *Microb. Genomics* 2:70. doi: 10.1099/mgen.0.000070
- Dangel, A., Berger, A., Messelhäuser, U., Konrad, R., Hörmansdorfer, S., Ackermann, N., et al. (2019). Genetic diversity and delineation of *Salmonella* Agona outbreak strains by next generation sequencing, Bavaria, Germany, 1993 to 2018. *Eurosurveillance* 24:303. doi: 10.2807/1560-7917.ES.2019.24.18.1800303
- Deng, X., den Bakker, H. C., and Hendriksen, R. S. (2016). Genomic epidemiology: Whole-genome-sequencing-powered surveillance and outbreak investigation of foodborne bacterial pathogens. *Annu. Rev. Food Sci. Technol.* 7, 353–374. doi: 10.1146/annurev-food-041715-033259
- Deng, X., Desai, P. T., den Bakker, H. C., Mikoleit, M., Tolar, B., Trees, E., et al. (2014). Genomic epidemiology of *Salmonella enterica* serotype enteritidis based on population structure of prevalent lineages. *Emerg. Infect. Dis.* 20, 1481–1489. doi: 10.3201/eid2009.131095
- ECDC (2016). *Expert Opinion on Whole Genome Sequencing for Public Health Surveillance Strategy to Harness Whole Genome Sequencing to Strengthen EU Outbreak Investigations and Public Health Surveillance*. Stockholm. Available online at: <https://www.ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/whole-genome-sequencing-for-public-health-surveillance.pdf> (accessed September–December, 2020).
- EFSA, and ECDC (2015). The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2013. *EFSA J.* 13:3991. doi: 10.2903/j.efsa.2015.3991
- EFSA, and ECDC (2017). *The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-Borne Outbreaks in 2017*. Available online at: <https://efsa.onlinelibrary.wiley.com/doi/full/10.2903/j.efsa.2018.5500> (accessed September–December, 2020).
- EFSA, and ECDC (2018). *The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-Borne Outbreaks in 2017*. Available online at: <https://efsa.onlinelibrary.wiley.com/doi/full/10.2903/j.efsa.2018.5500> (accessed September–December, 2020).
- EFSA, and ECDC (2019). The European Union one health 2018 zoonoses report. *EFSA J.* 17:e05926. doi: 10.2903/j.efsa.2019.5926
- EFSA Panel on Biological Hazards (2019). Whole genome sequencing and metagenomics for outbreak investigation, source attribution and risk assessment of food-borne microorganisms. *EFSA J.* 17:e05898. doi: 10.2903/j.efsa.2019.5898
- Egli, A., Blanc, D. S., Greub, G., Keller, P. M., Lazarevic, V., Lebrand, A., et al. (2018). Improving the quality and workflow of bacterial genome sequencing and analysis: paving the way for a Switzerland-wide molecular epidemiological surveillance platform. *Swiss Med. Wkly.* 18:14693. doi: 10.4414/smw.2018.14693
- Enkelmann, J., von Laer, A., Simon, S., Fruth, A., Lachmann, R., Michaelis, K., et al. (2020). Disentangling outbreaks using whole-genome sequencing: concurrent multistate outbreaks of *Salmonella* Kottbus in Germany, 2017. *Epidemiol. Infect.* 148:e51. doi: 10.1017/S0950268820000394
- Faensen, D., Claus, H., Benzler, J., Ammon, A., Pfoch, T., Breuer, T., et al. (2006). SurvNet@RKI – a multistate electronic reporting system for communicable diseases. *Eurosurveillance* 11, 7–8. doi: 10.2807/esm.11.04.00614-en
- FAO (2016). *Applications of Whole Genome Sequencing (WGS) in Food Safety Management*. Available online at: <http://www.fao.org/3/a-i5619e.pdf> (accessed September–December, 2020).
- Feldgarden, M., Brover, V., Haft, D. H., Prasad, A. B., Slotta, D. J., Tolstoy, I., et al. (2019). Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob. Agents Chemother.* 63:e00483-19.
- Grüning, B., Dale, R., Sjödin, A., Chapman, B. A., Rowe, J., Tomkins-Tinch, C. H., et al. (2018). Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nat. Methods* 15, 475–476. doi: 10.1038/s41592-018-0046-7
- Hawkey, J., Edwards, D. J., Dimovski, K., Hiley, L., Billman-Jacobe, H., Hogg, G., et al. (2013). Evidence of microevolution of *Salmonella* Typhimurium during a series of egg-associated outbreaks linked to a single chicken farm. *BMC Genom.* 14:800. doi: 10.1186/1471-2164-14-800
- Healy, J., and Bruce, B. (2017). “Salmonellosis (nontyphoidal),” in *CDC Yellow Book 2020: Health Information for International Travel*, eds G. W. Brunette and J. B. Nemhauser (New York, NY: Oxford University Press). Available online at: <https://wwwnc.cdc.gov/travel/page/2020-yellow-book-about>
- Hoffmann, M., Miller, J., Melka, D., Allard, M. W., Brown, E. W., and Pettengill, J. B. (2020). Temporal dynamics of *Salmonella enterica* subsp. *enterica* serovar Agona isolates from a recurrent multistate outbreak. *Front. Microbiol.* 11:478. doi: 10.3389/fmicb.2020.00478
- Jagadeesan, B., Gerner-Smidt, P., Allard, M. W., Leuillet, S., Winkler, A., Xiao, Y., et al. (2019). The use of next generation sequencing for improving food safety: translation into practice. *Food Microbiol.* 79, 96–115. doi: 10.1016/j.fm.2018.11.005
- Jain, S., Mukhopadhyay, K., and Thomassin, P. J. (2019). An economic analysis of *Salmonella* detection in fresh produce, poultry, and eggs using whole genome sequencing technology in Canada. *Food Res. Int.* 116, 802–809. doi: 10.1016/j.foodres.2018.09.014
- Jansen, A., Frank, C., Prager, R., Oppermann, H., and Stark, K. (2005). Bundesweiter Ausbruch durch *Salmonella* Give in Deutschland im Jahr 2004. *Z. Für Gastroenterol.* 43, 707–713. doi: 10.1055/s-2005-858256
- Jayeola, V., McClelland, M., Porwollik, S., Chu, W., Farber, J., and Kathariou, S. (2020). Identification of novel genes mediating survival of *Salmonella* on low-moisture foods via transposon sequencing analysis. *Front. Microbiol.* 11:726. doi: 10.3389/fmicb.2020.00726
- Jolley, K. A., and Maiden, M. C. (2010). BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11:595. doi: 10.1186/1471-2105-11-595

- Jourdan-da Silva, N., Fabre, L., Robinson, E., Fournet, N., Nisavanh, A., Bruyand, M., et al. (2018). Ongoing nationwide outbreak of *Salmonella* Agona associated with internationally distributed infant milk products, France, December 2017. *Eurosurveillance* 23:852. doi: 10.2807/1560-7917.ES.2018.23.2.17-00852
- Knoblauch, A., Bratschi, M., Zuske, M., Althaus, D., Stephan, R., Hächler, H., et al. (2015). Cross-border outbreak of *Salmonella enterica* ssp. *enterica* serovar Bovismorbificans: multiple approaches for an outbreak investigation in Germany and Switzerland. *Swiss Med. Wkly.* 15:14182. doi: 10.4414/sm.w.2015.14182
- Koch, J., Schrauder, A., Alpers, K., Werber, D., Frank, C., Prager, R., et al. (2005). *Salmonella* Agona outbreak from contaminated Aniseed, Germany. *Emerg. Infect. Dis.* 11, 1124–1127. doi: 10.3201/eid1107.041022
- Koster, J., and Rahmann, S. (2012). Snakemake – a scalable bioinformatics workflow engine. *Bioinformatics* 28, 2520–2522. doi: 10.1093/bioinformatics/bts480
- Kozyreva, V. K., Truong, C.-L., Greninger, A. L., Crandall, J., Mukhopadhyay, R., and Chaturvedi, V. (2017). Validation and implementation of clinical laboratory improvements act-compliant whole-genome sequencing in the public health microbiology laboratory. *J. Clin. Microbiol.* 55, 2502–2520. doi: 10.1128/JCM.00361-17
- Leekitcharoenphon, P., Sørensen, G., Löfström, C., Battisti, A., Szabo, I., Wasyl, D., et al. (2019). Cross-border transmission of *Salmonella* Choleraesuis var. Kunzendorf in European pigs and wild boar: Infection, genetics, and evolution. *Front. Microbiol.* 10:179. doi: 10.3389/fmicb.2019.00179
- Llarena, A., Ribeiro-Gonçalves, B. F., Nuno Silva, D., Halkilahti, J., Machado, M. P., Da Silva, M. S., et al. (2018). INNUENDO: A cross-sectoral platform for the integration of genomics in the surveillance of food-borne pathogens. *EFSA Support. Publ.* 15:1498. doi: 10.2903/sp.efsa.2018.EN-1498
- Mandilara, G., Mellou, K., Karadimas, K., Georgalis, L., Polemis, M., Georgakopoulou, T., et al. (2016). An outbreak of a possibly new *Salmonella enterica* subspecies *enterica* serovar with the antigenic formula 11:z41:e,n,z15, Greece, March to May 2016: preliminary results. *Eurosurveillance* 21:30265. doi: 10.2807/1560-7917.ES.2016.21.25.30265
- Meinen, A., Simon, S., Banerji, S., Szabo, I., Malorny, B., Borowiak, M., et al. (2019). Salmonellosis outbreak with novel *Salmonella enterica* subspecies *enterica* serotype (11:z41:e,n,z15) attributable to sesame products in five European countries, 2016 to 2017. *Eurosurveillance* 24:543. doi: 10.2807/1560-7917.ES.2019.24.36.1800543
- Methner, U. (2019). *Salmonellose der Rinder – Tiergesundheitsjahresberichte 2010–2019*. Greifswald: Friedrich-Loeffler-Institute.
- Moura, A., Tourdijman, M., Leclercq, A., Hamelin, E., Laurent, E., Fredriksen, N., et al. (2017). Real-time whole-genome sequencing for surveillance of *Listeria monocytogenes*, France. *Emerg. Infect. Dis.* 23, 1462–1470. doi: 10.3201/eid2309.170336
- Müller, L., Kjelsø, C., Frank, C., Jensen, T., Torpdahl, M., Søborg, B., et al. (2016). Outbreak of *Salmonella* Strathcona caused by datterino tomatoes, Denmark, 2011. *Epidemiol. Infect.* 144, 2802–2811. doi: 10.1017/S0950268816000121
- Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., et al. (2016). Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 17:132. doi: 10.1186/s13059-016-0997-x
- Phillips, A., Sotomayor, C., Wang, Q., Holmes, N., Furlong, C., Ward, K., et al. (2016). Whole genome sequencing of *Salmonella* Typhimurium illuminates distinct outbreaks caused by an endemic multi-locus variable number tandem repeat analysis type in Australia, 2014. *BMC Microbiol.* 16:211. doi: 10.1186/s12866-016-0831-3
- Pightling, A. W., Pettengill, J. B., Luo, Y., Baugher, J. D., Rand, H., and Strain, E. (2018). Interpreting whole-genome sequence analyses of foodborne bacteria for regulatory applications and outbreak investigations. *Front. Microbiol.* 9:1482. doi: 10.3389/fmicb.2018.01482
- Pisani, E., and AbouZahr, C. (2010). Sharing health data: good intentions are not enough. *Bull. World Health Organ.* 88, 462–466. doi: 10.2471/BLT.09.074393
- RKI (2015). *Infektionsepidemiologisches Jahrbuch Meldepflichtiger Krankheiten für 2014*. Available online at: https://www.rki.de/DE/Content/Infekt/Jahrbuch/Jahrbuch_2014.pdf?__blob=publicationFile (accessed September–December, 2020).
- RKI (2016). *Infektionsepidemiologisches Jahrbuch Meldepflichtiger Krankheiten für 2015*. Available online at: https://www.rki.de/DE/Content/Infekt/Jahrbuch/Jahrbuch_2015.pdf?__blob=publicationFile (accessed September–December, 2020).
- RKI (2017). *Infektionsepidemiologisches Jahrbuch Meldepflichtiger Krankheiten für 2016*. Available online at: <http://edoc.rki.de/docviews/abstract.php?lang=ger&id=5220> (accessed September–December, 2020).
- RKI (2018). *Infektionsepidemiologisches Jahrbuch Meldepflichtiger Krankheiten für 2017*. Available online at: <https://edoc.rki.de/handle/176904/5820> (accessed September–December, 2020).
- RKI (2019). *Infektionsepidemiologische Jahrbuch Meldepflichtiger Krankheiten für 2018*. Berlin: RKI, doi: 10.25646/5978
- Roche, D. G., Lanfear, R., Binning, S. A., Haff, T. M., Schwanz, L. E., Cain, K. E., et al. (2014). Troubleshooting public data archiving: suggestions to increase participation. *PLoS Biol.* 12:e1001779. doi: 10.1371/journal.pbio.1001779
- Schielke, A., Rabsch, W., Prager, R., Simon, S., Fruth, A., Helling, R., et al. (2017). Two consecutive large outbreaks of *Salmonella* Muenchen linked to pig farming in Germany, 2013 to 2014: is something missing in our regulatory framework? *Eurosurveillance* 22:30528. doi: 10.2807/1560-7917.ES.2017.22.18.30528
- Schroeder, S., Harries, M., Prager, R., Höfig, A., Ahrens, B., Hoffmann, L., et al. (2016). A prolonged outbreak of *Salmonella* infantis associated with pork products in central Germany, April–October 2013. *Epidemiol. Infect.* 144, 1429–1439. doi: 10.1017/S0950268815002629
- Schürch, A. C., Arredondo-Alonso, S., Willems, R. J. L., and Goering, R. V. (2018). Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. *Clin. Microbiol. Infect.* 24, 350–354. doi: 10.1016/j.cmi.2017.12.016
- Silva, M., Machado, M. P., Silva, D. N., Rossi, M., Moran-Gilad, J., Santos, S., et al. (2018). chewBBACA: a complete suite for gene-by-gene schema creation and strain identification. *Microb. Genomics* 4:166. doi: 10.1099/mgen.0.000166
- Simon, S., Trost, E., Bender, J., Fuchs, S., Malorny, B., Rabsch, W., et al. (2018). Evaluation of WGS based approaches for investigating a food-borne outbreak caused by *Salmonella enterica* serovar Derby in Germany. *Food Microbiol.* 71, 46–54. doi: 10.1016/j.fm.2017.08.017
- Stöcker, P., Rosner, B., Werber, D., Kirchner, M., Reinecke, A., Wichmann-Schauer, H., et al. (2011). Outbreak of *Salmonella* montevideo associated with a dietary food supplement flagged in the Rapid Alert System for Food and Feed (RASFF) in Germany, 2010. *Euro Surveill. Bull. Eur. Sur Mal. Transm. Eur. Commun. Dis. Bull.* 16:20040.
- Timme, R. E., Rand, H., Sanchez Leon, M., Hoffmann, M., Strain, E., Allard, M., et al. (2018). GenomeTrakr proficiency testing for foodborne pathogen surveillance: an exercise from 2015. *Microb. Genomics* 4:185. doi: 10.1099/mgen.0.000185
- Uelze, L., Grütze, J., Borowiak, M., Hammerl, J. A., Juraschek, K., Deneke, C., et al. (2020). Typing methods based on whole genome sequencing data. *One Health Outlook* 2:3. doi: 10.1186/s42522-020-0010-1
- Werber, D., Dreesman, J., Feil, F., van Treeck, U., Fell, G., Ethelberg, S., et al. (2005). International outbreak of *Salmonella* Oranienburg due to German chocolate. *BMC Infect. Dis.* 5:7. doi: 10.1186/1471-2334-5-7
- Yoshida, C. E., Kruczkiwicz, P., Laing, C. R., Lingohr, E. J., Gannon, V. P. J., Nash, J. H. E., et al. (2016). The *Salmonella in silico* typing resource (SISTR): an open web-accessible tool for rapidly typing and subtyping draft *Salmonella* genome assemblies. *PLoS One* 11:e0147101. doi: 10.1371/journal.pone.0147101
- Zhou, Z., Alikhan, N.-F., Sergeant, M. J., Luhmann, N., Vaz, C., Francisco, A. P., et al. (2018). GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res.* 28, 1395–1404. doi: 10.1101/gr.232397.117

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Uelze, Becker, Borowiak, Busch, Dangel, Deneke, Fischer, Flieger, Hepner, Huber, Methner, Linde, Pietsch, Simon, Sing, Tausch, Szabo and Malorny. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.