bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402362; this version posted February 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

CoxBase: an online platform for epidemiological surveillance, visualization, analysis and typing of *Coxiella burnetii* genomic sequences.

Akinyemi. M. Fasemore [1,2,4], Andrea Helbich [2], Mathias. C. Walter [2], Thomas Dandekar [5], Gilles Vergnaud [6], Konrad U. Förstner [4*], Dimitrios Frangoulidis [2,3*]

[1]University of Würzburg, Würzburg, Germany. [2]Bundeswehr Institute of Microbiology, Munich, Germany. [3]Bundeswehr Medical Service Headquarters VI-2, Medical Intelligence & Information, Munich, Germany. [4]ZB MED - Information Centre for Life Science, Cologne, Germany. [5]Department of Bioinformatics, Biocenter, Am Hubland, University of Würzburg, Würzburg, Germany. [6] Institute for Integrative Biology of the Cell (I2BC), Université Paris-Saclay, 91198, Gif-sur-Yvette cedex, France.

ABSTRACT

Q (query) fever is an infectious zoonotic disease caused by the Gram-negative bacteria *Coxiella burnetii*. Although the disease has been studied since decades, it still represents a threat due to sporadic outbreaks across farms in Europe. The absence of a central platform for *Coxiella* typing data management in an important epidemiological gap which is relevant in the case of an outbreak. To fill this gap, we have designed and implemented an online, open-source, and, web-based platform called CoxBase (https://coxbase.q-gaps.de). This platform includes a database that holds genotyping information of more than 400 *Coxiella* isolates alongside metadata that annotates them. We have also implemented features for *in silico* genotyping of completely or minimally assembled *Coxiella* sequences using five different typing methods, querying existing isolates, visualization of isolate's geodata via aggregation on a world map and submission of new isolates. We tested our *in silico* typing method on 50 *Coxiella* genomes downloaded from the RefSeq database and we successfully genotyped all except for cases where the sequence quality was poor. We identified new spacer sequences using our implementation of the MST *in silico* typing method, and established adaA gene phenotypes for all 50 genomes as well as their plasmid types.

INTRODUCTION

Q (query) fever is an infectious zoonotic disease that affects humans and small ruminants like sheep, goat and cattle. It was first described among abattoir workers in Queensland, Australia, with symptoms of febrile illness in 1937 (1). The causative agent is a Gram-negative, pleomorphic, obligate intracellular bacterium called *Coxiella burnetii*. It has a world wide distribution and persists in biological and environmental reservoirs like milk, hay and dust, which can act as sources for sporadic outbreaks in livestock (2).

Since its first description as a febrile illness in Australia, the pathology of Q fever is now more understood and has been described as usually sub-clinical in ruminants but may manifest in form of late term abortion in pregnant ruminant females (3). In humans, the disease can be observed in two different forms. The first form is the acute disease, which is usually self-limiting, might occur alongside symptoms such as febrile illness, fever and severe headaches. It has been shown to happen in 40% of the primary Q fever cases. The second form is the chronic form, usually long lasting, characterized by endocarditis and can be severe and in dire cases fatal. It occurs in 1-5% of primary cases, the remaining cases are usually subclinical/asymptomatic and are also defined as acute disease (4) (2).

The epidemiology of this disease has been linked to the interplay of several dynamic factors including but not limited to, vector diversity, reservoir type and worldwide distribution of the disease (5). Another important point for disease control is the absence of a central platform that connects the different ends of the large and growing field of *Coxiella* research.

As a result, data from *Coxiella* research are dispersed over the academic space and if collected at a point is usually specific to a single method. The implication of these is that speed of research flow is significantly impeded especially in urgent cases of outbreaks where strain comparison and discrimination is vital to the control of the aetiological agent.

To highlight this challenge, there are up to five known genotyping methods for discriminating *Coxiella* species namely Multiple Locus Variable-number Tandem Repeat Analysis (MLVA) (6) (7), Multispacer typing (MST) (8), IS1111 typing (9), AdaA gene typing (10) and Plasmid typing (11) (12). MLVA and IS1111 typing require the measurement of PCR amplification products. MST requires the sequencing of intergenic regions whereas AdaA typing is based upon the

*To whom correspondence should be addressed. Tel: +4922147878909; Email: foerstner@zbmed.de, dimitriosfrangoulidis@bundeswehr.org

Keywords: Coxiella burnetii, Q Fever, Web service
https://coxbase.q-gaps.de
© 2021

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402362; this version posted February 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

sequencing of one coding sequence. All methods allow to detect a correlation between geographic origin and genotype and are useful for typing strains in endemic regions as well as clinical entities (5) (10). MLVA, MST and IS1111 methods offer a higher resolution compared to the other two methods (5).

A researcher interested in typing a new *Coxiella* strain is likely to employ more than a single method to obtain quality proof or at least to employ the methods accessible in his particular setting. Access to a database resource with strain information and metadata will be necessary for comparison purpose.

Presently there are two of such resources that house *Coxiella* genotyping data. The first is the MLVA databank (http://mlva.i2bc.paris-saclay.fr/mlvav4/genotyping/) and the second is the MST database (https://ifr48.timone.univ-mrs.fr/mst/coxiella_burnetii/), for the other genotyping methods there are no available database resources.

First we sought to overcome the lack of additional genotyping resource, then we sought to consolidate on the existing resources via introduction of new features such as visualization of allelic reference for MST typing, aggregation of MLVA groups and introduction of MLVA genotypes for better comparison. To this end, we have developed an online, open, web-based platform called CoxBase, which caters for vital aspects of internet based *Coxiella* research. This platform also includes a database that contains over 400 *C. burnetii* isolates from different countries. It has been implemented with a user interface for quick retrieval of isolate information as well as a submission channel to add to the growing body of new *Coxiella* isolates.

Also, we sought to unify all *Coxiella* typing systems under a single platform, alongside all the published details of *Coxiella* genotyping, including primers for genotyping protocols, as well as phenotypes, for the purpose of strain discovery and comparison. We implemented an *in silico* genotyping option for all major genotyping systems of *C.burnetii* based on whole genomic sequences.

Finally, we included visualization systems to quickly summarize all metadata on country level, maps for enhanced geographic localization of isolates and a worldwide distribution map of all *C. burnetii* isolates in our database. Here, we present our platform, its current scope, usage and capabilities.

MATERIALS AND METHODS

Webserver components

The server is run by Apache HTTP server (version 2.4.29), on a machine hosted by de.NBI Cloud services. The server components can be grouped under 2 main sections, the front end and the back end.

The front end. The main component of the front end is the web user interface, this is designed to accept user queries as well as submissions, send data to the back end and present data back to the user. Styling was achieved through an assortment of the cascading style sheet (CSS) Bootstrap framework (https://getbootstrap.com/), jQuery UI library and custom CSS style scripts. The validation of forms and events processing is achieved with JavaScript. The user interface accepts two kinds of data input: FASTA formatted whole genomic sequence (contigs or complete assembly) for typing purposes, and typing profiles via multi locus variable number tandem repeat analysis (MLVA) (7) and Multilocus sequence typing (MST) (8) for isolate comparison and discovery.

The back end. The back end handles user requests and uses a MySQL database to store data. Requests are handled via an Apache server (https://httpd.apache.org/) which then communicates via the Web Server Gateway Interface (WSGI) to a python pyramid framework application (https://trypyramid.com/). The application processes the request and communicates via the SQLAlchemy library (https://www.sqlalchemy.org) to the MySQL storage.

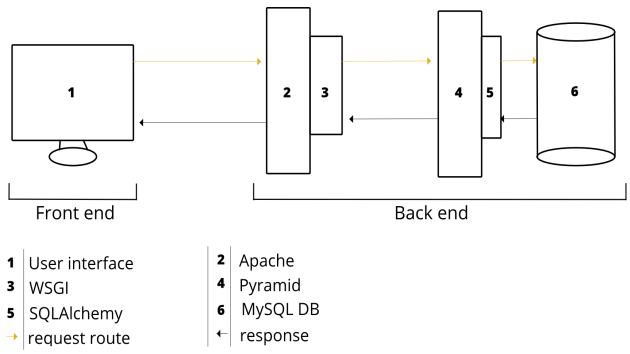

Figure 1. CoxBase Server Architecture.

Genome typing

We have implemented five different *in silico* typing methods for *Coxiella* sequences on the server. The MLVA typing method (6), the MST method (8), the adaA gene typing method (10), the plasmid typing method (12) and the 1S1111 typing method (9). The typing programs were implemented in the Python web application.

Establishing the typing features.

MLVA typing: The MLVA typing feature accepts as input, genomic sequences either as contigs or complete assembly in FASTA format. The lengths of fourteen MLVA amplicons (when present) are extracted *in silico* with the e-PCR tool (13) using primers from Frangoulidis *et al.*, 2014 (Table 1) (7) updated in http://mlva.i2bc.paris-saclay.fr/MLVAnet/spip.php?rubrique50. The repeat number is calculated with the formula below :

$$\mathrm{RN} = \frac{(\mathrm{AL} - \mathrm{FL})}{\mathrm{RS}} \quad \textbf{(1)}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402362; this version posted February 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

where RN = repeat number, AL = amplicon length, FL = flanking length and RS = repeat size.

Table 1. Table of used MLVA markers and their primer sequences

Marker name	Primer sense	Primer sequence
ms01	forward	GCCCTTGTCATCTTGCGG
	reverse	TCAAGTATTAATGAGCGTCG
ms03	forward	TGTCGATAAATCGGGAAACTT
	reverse	ACTGGGAAAAGGAGAAAAAGA
ms20	forward	CTGAAACCAGTCTTCCCTCAAC
	reverse	CTTTATCTTGGCCTCGCCCTTC
ms21	forward	AGCATCTGCCTTCTCAAGTTTC
	reverse	TGGGAGGTAGAAGAAAAGATGG
ms22	forward	GGGGTTTGAACATAGCAATACC
	reverse	CAATATCTCTTTCTCCCGCATT
ms23	forward	GGACAAAAATCAATAGCCCGTA
	reverse	GAAAACAGAGTTGTGTGGCTTC
ms24	forward	ATGAAGAAAGGATGGAGGGACT
	reverse	GATAGCCTGGACAGAGGACAGT
ms26	forward	GCAATCCAGTTGGAAAGAA
	reverse	ATTGAAGTAATCCATCGATGATT
ms27	forward	TTTTGAGTAAAGGCAACCCAAT
	reverse	CAAACGTCGCACTAACTCTACG
ms28	forward	AATGGAGTTTGTTAGCAAAGAAA
	reverse	AAAGACAAGCAAAACGATAAAAA
ms30	forward	ATTTCCTCGACATCAACGTCTT
	reverse	AGTCGATTTGGAAACGGATAAA
ms31	forward	ACAGGCCGGTATTCTAACC
	reverse	CCTCAGCACCCATTCAG
ms33	forward	TAGGCAGAGGACAGAGGACAGT
	reverse	ATGGATTTAGCCAGCGATAAAA
ms34	forward	TGACTATCAGCGACTCGAAGAA
	reverse	TCGTGCGTTAGTGTGCTTATCT

For every submitted job, a unique identifier is generated that can be used to retrieve the results historically from the database within three weeks after the date of submission. The results of the MLVA typing are presented in form of a table with all the calculated parameters. A feature to search the database for closely related MLVA profiles is also provided.

MST typing: The *in silico* MST method accepts genomic sequence in FASTA format. The first step is amplicon detection via USEARCH (14). This is done using the MST primers from Glazunova *et al.*, 2005 (8). The allele type is determined by aligning detected amplicon sequence globally with known alleles in the MST library (https://ifr48.timone.univ-mrs.fr/mst/coxiella_burnetii/spacers.html). Novel sequences with no match are also reported. The detected MST profile can be used as a query to the database to find the corresponding MST group.

IS1111 typing: The IS1111 typing is based on detection of localizations adjacent to IS1111 elements (9). This is a binary detection method meaning the discrimination is based on the absence or presence of an amplicon from a given location. For the *in silico* detection, we employed the e-PCR tool (13) to detect amplicons based on the primers described by (9) and extended by Bleichert & Hanczaruk 2012 (unpublished). The presence or absence is highlighted with a green (+) and red (-) respectively as shown in Figure 2.

IS element	state	IS element	state	IS element	state	IS element	state	IS element	state
IS1111-1	+	IS1111-2	+	IS1111-3	+	IS1111-4	+	IS1111-5	+
IS1111-6	+	IS1111-7	+	IS1111-8	+	IS1111-9	+	IS1111-10	+
IS1111-11	+	IS1111-12	+	IS1111-13	+	IS1111-14	+	IS1111-15	+
IS1111-16	+	IS1111-17	+	IS1111-18	+	IS1111-19	+	IS1111-20	+
IS1111-21	-	IS1111-22	-	IS1111-23	-	IS1111-24	-	IS1111-25	-
IS1111-26	-	IS1111-27	-	IS1111-28	-	IS1111-29	-	IS1111-30	-
IS1111-30	-	IS1111-31	-	IS1111-32	-	IS1111-34	-	IS1111-35	-
IS1111-36	-	IS1111-37	-	IS1111-38	-	IS1111-39	-	IS1111-40	-
IS1111-41	-	IS1111-42	-	IS1111-43	-	IS1111-44	-	IS1111-45	-
IS1111-46	-	IS1111-47	-	IS1111-48	-	IS1111-49	-	IS1111-50	-
IS1111-51	-	IS1111-53	-	IS1111-54	-	IS1111-55	-	IS1111-56	-
IS1111-57	-	IS1111-58	-	IS1111-59	-	IS1111-60	-	IS1111-61	-
IS1111-84	-								

Figure 2. IS1111 typing result of RSA 439 as calculated on the CoxBase platform.

adaA and Plasmid typing: adaA phenotype has been earlier reported to correlate with plasmid type (10), therefore we combined these two typing methods together. Five different variants of the adaA gene have been reported, three single nucleotide variants (wildtype, A431T SNP, and repeat) and two deletion variants (Q154 deletion and Q212 deletion) (10). In our implementation, we first try to detect if the coding sequence of the adaA gene exists withing the genome to be typed. For this, we used the USEARCH tool (14) and the primer sequence for the detection of the entire adaA open reading frame (684 bases) (10). If an amplicon exits, we subsequently evaluate its length. If the length is longer than 684 bases then we assign it the adaA insertion genotype, and if it is shorter we assign it the incomplete adaA genotype. If it is exactly 684 bases, then we evaluate type of SNP at position 431 of the amplicon sequence. For the detection of plasmid type, we employed 4 primers that have been used for direct identification of *C. burnetii* plasmids via laboratory PCR methods (12) (11) (15).

Isolate discovery and comparison

The CoxBase platform offers features for the discovery and comparison of *Coxiella* strains through several approaches. One approach is to query the database based on metadata and genotype features like country, host type, plasmid type, year of isolation, MLVA genotype and MST group. The advantage of this approach is that it's fine grained and the fields can be aggregated to build more specific queries. Another approach utilizes a faceted search, this approach is more suitable for refining queries based on reviewed criteria. Other approaches rely on making queries based on known typing profiles via MLVA or MST typing schema. This is implemented as follows:

For a user who wishes to discover isolates with a specific isolate profile (MST or MLVA). They need to provide a complete or partial profile (MLVA or MST) of the isolate which they are interested in. Usually one marker is enough for a search but for a more defined and reliable results, at least 6 markers should be provided for the MLVA query and

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402362; this version posted February 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

10 for the MST. For ease of comparison, isolates with similar profiles are pooled together in a single row in the query results. Profile entries can then be expanded with the click of a button called View profile entries in the final column of the result table. A list of all isolates with that profile is provided with metadata. Geographical information is visualized through a Leaflet Map (https://leafletjs.com/). The aim is to provide a geographic orientation that can be used to estimate physical proximity of the isolates. Figure 3 shows a single isolate from the A14 MLVA group that was isolated from a sheep in Hannover, Germany. The last approach relies on grouping based on geographical location. A user interested in isolates from a particular country will approach the distribution map. He can access a comprehensive table of all isolates from the country of interest after clicking on the country marker located on the map.

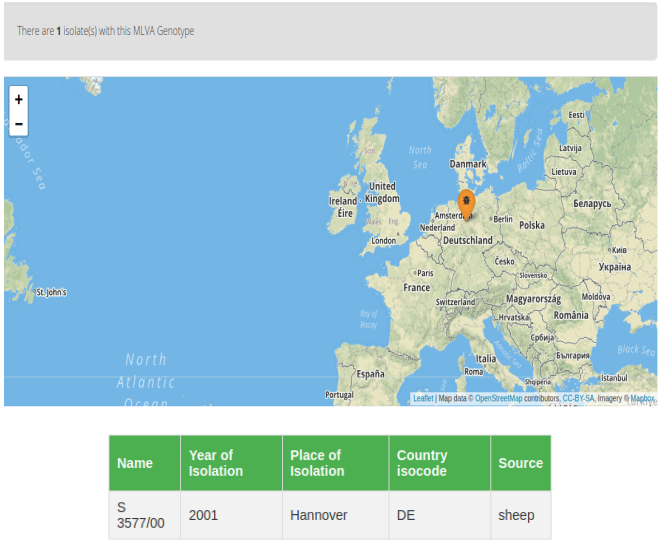

Figure 3. A strain from the A14 MLVA group that was isolated from a sheep in Hannover, Germany.

Visualization

We implemented an interactive visualization feature based on the Chart.js (https://www.chartjs.org/) JavaScript visualization library. This can be accessed through the dashboard link on country markers in the distribution map. Distribution plots for metadata categories such as host type, year of isolation, place of isolation as well as genotype could help answer questions such as the most predominant host type in a particular location as illustrated in Figure 4.

Data collection sets and Testing

We tested our implementation on fifty *Coxiella* whole genomic sequences from RefSeq. (Table of genomes in supplementary table). We had eleven complete chromosome assembly, thirteen chromosome assembly, fifteen contigs and eleven scaffolds. The average genome size was 2.01Mb. Forty of these were of unknown genotypes (not published if known), so we sought to type these genomes as well. The genome sequences in FASTA format were downloaded

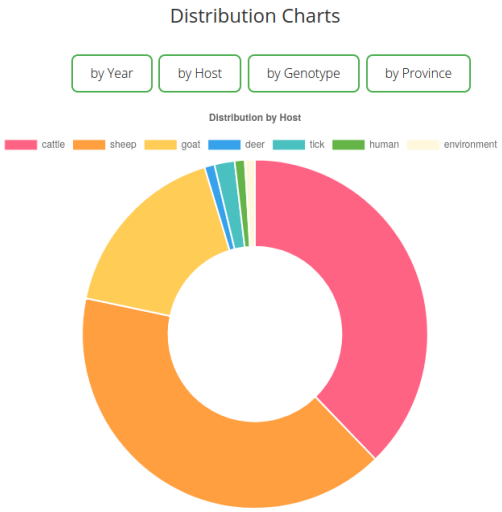

Figure 4. Donut plot of host data from Germany and it shows that most common hosts are sheep and cattle.

from the RefSeq database and stored without any form of modification. The genomes were typed individually using the different typing methods on our platform. After which the results were compared to known strains in our database. The results are discussed below

RESULTS: APPLICATION EXAMPLES

MLVA typing

Based on the *in silico* typing scheme we could separate the fifty genomes under thirteen MLVA groups. We observed that some markers were often not detected *in silico*. We computed the effectiveness for each marker (The probability of the marker producing an amplicon *in silico* as a percentage) based on the 50 *Coxiella* genomes. We observed that six of the fourteen markers had a marker effectiveness of one hundred percent, six markers between ninety-eight to seventy-two percent effectiveness and two markers with an effectiveness of fifty-six and twenty percent (Table of Marker effectiveness in Supplementary material). For genomes with known MLVA types (RSA 493 and Dugway), we were able to classify them into the correct MLVA group as described already in Frangoulidis *et al*., (2014) (7).

MST typing

We were able to classify the fifty genomes into twelve MST groups, except for two genomes whose profile had no matching MST group and one with an undetermined spacer, therefore it could not be placed in an appropriate MST group. We were able to confirm the MST group of genomes with published MST genotypes (Ohio, Henzerling, and Nine mile), however, our results differ from the published MST genotype of the Dugway strain. The spacer profile indicated it belonged

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402362; this version posted February 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to the MST type 60 and not MST 20 as published. Nearly half of the genomes (n=24/50) are either MST 16 or MST 61. All the MST primers had an amplicon *in silico*. We also detected six new spacer alleles in seven genomes. The MST markers were reliable. However, we observed that detection of a new spacer allele might be a false positive due to incomplete sequences in the alignment. Therefore, we implemented an alignment visualization feature for the detection of new spacer sequences based on the BlasterJS library (16).

AdaA and Plasmid typing

The plasmid of the twenty-three genomes out of the fifty genomes in the dataset were already known. We were able to confirm the known plasmid type and extended the information with their adaA gene phenotypes, and for genomes whose plasmid type was not known we were able to detect the plasmid type as well as their adaA gene phenotypes. We determined 43 out of the 50 genomes carried a plasmid. QpH1 was present in 34 instances. Nine out of the fifty genomes had an adaA deletion phenotype and the remaining were adaA wildtype. Out of the 41 positive strains, seven had a repeat insertion in there adaA gene sequence, while four genomes showed a SNP at position 431 of the adaA gene sequences. We observed that six out of the seven genomes with an adaA repetition phenotype belonged to the same MST profile.

Phylogenetic analysis

We implemented two types of visualization for phylogenetic trees. The first tree is a GrapeTree (17) implementation that can be used to visualize genomic relationships of grouped data based on their MLVA profiles. The resulting tree can be color coded based on metadata, editable and can also be exported into several image formats. The second tree is implemented using the PhyD3 visualization library (18). This is especially useful for locating MLVA profile in the MLVA genotype tree, thereby associating a strain with a new MLVA profile with its closest MLVA genotype.

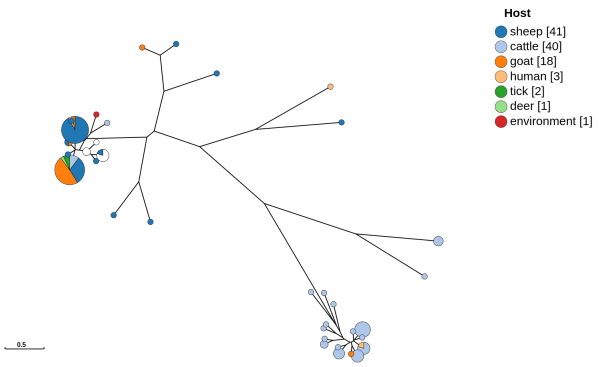

Figure 5. GrapeTree visualization of *C. burnetii* isolates from Germany on CoxBase based on MLVA genotyping. Distinctive clusters based on metadata such as host type can be inferred from such a tree.

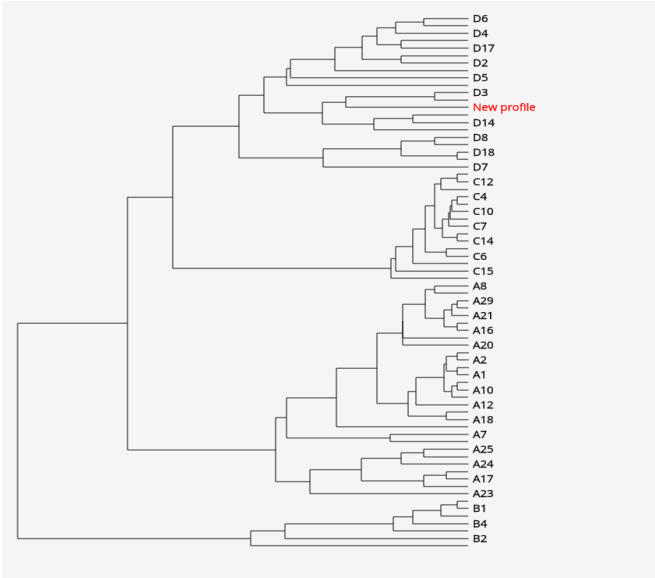

Figure 6. Unrooted phylogenetic tree of all MLVA genotypes. Highlighted node shows the position of *C. burnetii* strain Q321 that was isolated from a cow's milk in Russia. MLVA typing was done via CoxBase.

DISCUSSION

Here we present a platform that was built with the aim to overcome the lack of a centralized genomic data resource for *Coxiella burnetii*.

This is the first genotyping platform that combines all the disparate typing systems of *Coxiella burnetii*. Similar platforms exist for other bacteria species such as PubMLST albeit usually focused on a single typing system.

Several features are particularly novel and unique: We combined five typing methods to enable rapid identification of *Coxiella* strains as well as the visualization of the metadata coupled to the geographical distribution. The latter format is particularly useful to study and control outbreaks, the major shortcoming for which our platform was constructed.

We have also included several features that could assist researchers to understand the variability within the genomes of *C. burnetii* in an epidemiological context. We have leveraged on technologies such as NGS, cloud computing and database to create an open web resource that can be used to genotype draft or completely assembled *C. burnetii* genomic sequences as well as compare them to existing strains. Our approach also brought together different aspects of *Coxiella* research including epidemiological surveillance, sequence analyses and phylogeny under a single platform. The strength of *in silico* typing methods rely on, to a significant degree, the quality of the input sequence. We observed that certain MLVA markers are also not as effective as others, however, a combination of multiple markers of different typing schemes should improve the averaging effect. Nevertheless, our implementations suggest that *in silico* typing can be an indispensable tool for rapid genotyping of *Coxiella* genomic sequences. We tested the implementation on 50 *C. burnetii* genomes from NCBI and we were able to type all except for cases where the sequence quality was not good enough. We

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402362; this version posted February 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

could observe perfect corroboration with known genotypes when we used our implementation to type these sequences except for one case where we argue that the published profile might not be correct as the observed spacer profile differed in all alleles compared to the published profile. One limitation of our method is in the adaA gene typing. Although we can distinguish between the different adaA gene positive variants, we are yet to implement a feature to differentiate between the deletion variants (if it is a Q212 deletion or Q514 deletion). For now we only report if the adaA gene deletion exists in a given sequence and not the variant of the deletion type. We implemented a retrieval feature on CoxBase that will enable researchers to access the results of their typing analyses up to three weeks after their submission date. This would ease collaboration efforts on typing projects and reduce the complexity of information sharing. We have also implemented a genome browser for sequence visualization to accompany sequence typing investigations most especially primer analysis. Finally, we implemented a submission feature for researchers who wish to share new MLVA or MST profiles. We hope this platform will provide researchers with the opportunity to investigate the variability between *C. burnetii* genomes as well as help to better understand the epidemiology of the Q fever disease in terms of genotype correlations with metadata like host specificity and geographical information. We will update the platform periodically to keep the data current and curated.

ACKNOWLEDGEMENTS

This research has been funded by the Federal Ministry of Education and Research of Germany (BMBF, project number 01KI1726E). Furthermore, the work was supported by the BMBF-funded de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI) (031A537B, 031A533A, 031A538A, 031A533B, 031A535A, 031A537C, 031A534A, 031A532B).

Conflict of interest statement. None declared.

REFERENCES

1. Derrick, E. H. (July, 1983) Q Fever, a New Fever Entity: Clinical Features, Diagnosis and Laboratory Investigation. *Reviews of Infectious Diseases,* **5**(4), 790–800.
2. Mertens, K., Gerlach, C., Neubauer, H., and Henning, K. (March, 2017) Q fever An Update. *Current Clinical Microbiology Reports,* **4**(1), 61–70.
3. Arricau Bouvery, N., Souriau, A., Lechopier, P., and Rodolakis, A. (August, 2003) Experimental *Coxiella burnetii* infection in pregnant goats: excretion routes. *Veterinary Research,* **34**(4), 423–433.
4. Kampschreur, L. M., Delsing, C. E., Groenwold, R. H. H., Wegdam-Blans, M. C. A., Bleeker-Rovers, C. P., de Jager-Leclercq, M. G. L., Hoepelman, A. I. M., van Kasteren, M. E., Buijs, J., Renders, N. H. M., Nabuurs-Franssen, M. H., Oosterheert, J. J., and Wever, P. C. (May, 2014) Chronic Q Fever in the Netherlands 5 Years after the Start of the Q Fever Epidemic: Results from the Dutch Chronic Q Fever Database. *Journal of Clinical Microbiology,* **52**(5), 1637–1643.
5. Massung, R. F., Cutler, S. J., and Frangoulidis, D. (2012) Molecular Typing of *Coxiella burnetii* (Q Fever). In Toman, R., Heinzen, R. A., Samuel, J. E., and Mege, J.-L., (eds.), *Coxiella burnetii: Recent Advances and New Perspectives in Research of the Q Fever Bacterium*, Vol. 984, pp. 381–396 Springer Netherlands Dordrecht.
6. Arricau-Bouvery, N., Hauck, Y., Bejaoui, A., Frangoulidis, D., Bodier, C. C., Souriau, A., Meyer, H., Neubauer, H., Rodolakis, A., and Vergnaud, G. (2006) Molecular characterization of *Coxiella burnetii* isolates by infrequent restriction site-PCR and MLVA typing.. *BMC Microbiology,* p. 14.
7. Frangoulidis, D., Walter, M. C., Antwerpen, M., Zimmermann, P., Janowetz, B., Alex, M., Böttcher, J., Henning, K., Hilbert, A., Ganter, M., Runge, M., Münsterkötter, M., Splettstoesser, W. D., and Hanczaruk, M. (October, 2014) Molecular analysis of *Coxiella burnetii* in Germany reveals evolution of unique clonal clusters. *International journal of medical microbiology: IJMM,* **304**(7), 868–876.
8. Glazunova, O., Roux, V., Freylikman, O., Sekeyova, Z., Fournous, G., Tyczka, J., Tokarevich, N., Kovacova, E., Marrie, T. J., and Raoult, D. (August, 2005) *Coxiella burnetii* Genotyping. *Emerging Infectious Diseases,* **11**(8), 1211–1217.
9. Denison, A. M., Thompson, H. A., and Massung, R. F. (October, 2007) IS1111 insertion sequences of *Coxiella burnetii*: characterization and use for repetitive element PCR-based differentiation of Coxiella burnetii isolates. *BMC microbiology,* **7**, 91.
10. Frangoulidis, D., Splettstoesser, W. D., Landt, O., Dehnhardt, J., Henning, K., Hilbert, A., Bauer, T., Antwerpen, M., Meyer, H., Walter, M. C., and Knobloch, J. K.-M. (January, 2013) Microevolution of the Chromosomal Region of Acute Disease Antigen A (adaA) in the Query (Q) Fever Agent *Coxiella burnetii*. *PLOS ONE,* **8**(1), e53440.
11. Lautenschläger, S., Willems, H., Jäger, C., and Baljer, G. (July, 2000) Sequencing and characterization of the cryptic plasmid QpRS from *Coxiella burnetii*.. *Plasmid,* **44**(1), 85–88.
12. Willems, H., Thiele, D., and Krauss, H. (July, 1993) Plasmid based differentiation and detection of *Coxiella burnetii* in clinical samples. *European Journal of Epidemiology,* **9**(4), 411–418.
13. Schuler, G. D. (May, 1997) Sequence Mapping by Electronic PCR. *Genome Research,* **7**(5), 541–550.
14. Edgar, R. C. (October, 2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics,* **26**(19), 2460–2461.
15. Zhang, G. Q., Hotta, A., Mizutani, M., Ho, T., Yamaguchi, T., Fukushi, H., and Hirai, K. (1998) Direct Identification of Coxiella burnetii Plasmids in Human Sera by Nested PCR. *Journal of Clinical Microbiology,* **36**(8), 2210–2213.
16. Blanco-Mguez, A., Fdez-Riverola, F., Snchez, B., and Loureno, A. (October, 2018) BlasterJS: A novel interactive JavaScript visualisation component for BLAST alignment results. *PLOS ONE,* **13**(10), e0205286.
17. Zhou, Z., Alikhan, N.-F., Sergeant, M. J., Luhmann, N., Vaz, C., Francisco, A. P., Carrio, J. A., and Achtman, M. (September, 2018) GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Research,* **28**(9), 1395–1404 Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.
18. Kreft, L., Botzki, A., Coppens, F., Vandepoele, K., and Van Bel, M. (05, 2017) PhyD3: a phylogenetic tree viewer with extended phyloXML support for functional genomics data visualization. *Bioinformatics,* **33**(18), 2946–2947.