



Genome Plasticity and Polymorphisms in Critical Genes Correlate with Increased Virulence of Dutch Outbreak-Related *Coxiella burnetii* Strains

Runa Kuley^{1,2}, Eric Kuijt¹, Mari A. Smits^{1,2}, Hendrik I. J. Roest³, Hilde E. Smith¹ and Alex Bossers^{1*}

¹ Department of Infection Biology, Wageningen Bioveterinary Research, Lelystad, Netherlands, ² Host Microbe Interactomics, Wageningen University and Research Centre, Wageningen, Netherlands, ³ Department of Bacteriology and Epidemiology, Wageningen Bioveterinary Research, Lelystad, Netherlands

OPEN ACCESS

Edited by:

John W. A. Rossen,
University Medical Center Groningen,
Netherlands

Reviewed by:

Sarah Jane Tozer,
The University of Queensland,
Australia
Corné Klaassen,
Erasmus Medical Center, Netherlands

*Correspondence:

Alex Bossers
alex.bossers@wur.nl

Specialty section:

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

Received: 14 March 2017

Accepted: 28 July 2017

Published: 10 August 2017

Citation:

Kuley R, Kuijt E, Smits MA, Roest HIJ,
Smith HE and Bossers A (2017)
Genome Plasticity and Polymorphisms
in Critical Genes Correlate with
Increased Virulence of Dutch
Outbreak-Related *Coxiella burnetii*
Strains. *Front. Microbiol.* 8:1526.
doi: 10.3389/fmicb.2017.01526

Coxiella burnetii is an obligate intracellular bacterium and the etiological agent of Q fever. During 2007–2010 the largest Q fever outbreak ever reported occurred in The Netherlands. It is anticipated that strains from this outbreak demonstrated an increased zoonotic potential as more than 40,000 individuals were assumed to be infected. The acquisition of novel genetic factors by these *C. burnetii* outbreak strains, such as virulence-related genes, has frequently been proposed and discussed, but is not proved yet. In the present study, the whole genome sequence of several Dutch strains (CbNL01 and CbNL12 genotypes), a few additionally selected strains from different geographical locations and publicly available genome sequences were used for a comparative bioinformatics approach. The study focuses on the identification of specific genetic differences in the outbreak related CbNL01 strains compared to other *C. burnetii* strains. In this approach we investigated the phylogenetic relationship and genomic aspects of virulence and host-specificity. Phylogenetic clustering of whole genome sequences showed a genotype-specific clustering that correlated with the clustering observed using Multiple Locus Variable-number Tandem Repeat Analysis (MLVA). Ortholog analysis on predicted genes and single nucleotide polymorphism (SNP) analysis of complete genome sequences demonstrated the presence of genotype-specific gene contents and SNP variations in *C. burnetii* strains. It also demonstrated that the currently used MLVA genotyping methods are highly discriminatory for the investigated outbreak strains. In the fully reconstructed genome sequence of the Dutch outbreak NL3262 strain of the CbNL01 genotype, a relatively large number of transposon-linked genes were identified as compared to the other published complete genome sequences of *C. burnetii*. Additionally, large numbers of SNPs in its membrane proteins and predicted virulence-associated genes were identified in all Dutch outbreak strains compared to the NM reference strain and other strains of the CbNL12 genotype. The presence of large numbers of transposable elements and mutated genes, thereof most likely

resulted in high level of genome rearrangements and genotype-specific pathogenicity of outbreak strains. Thus, the epidemic potential of Dutch outbreak strains could be linked to increased genome plasticity and mutations in critical genes involved in virulence and the evasion of the host immune system.

Keywords: *C. burnetii*, Q fever, whole genome sequencing, comparative genomics, virulence, transposons, mutations, orthologs

INTRODUCTION

Coxiella burnetii is the pathogenic agent of Q fever which is a zoonotic infectious disease (Maurin and Raoult, 1999). It is an obligate intracellular gram-negative bacterium, which thrives within the acidic parasitophorous vacuole (PV) of eukaryotic cells (Akporiaye et al., 1983). Domestic ruminants such as goats, sheep and cattle are usually the primary reservoir for the *C. burnetii* strains causing Q fever in humans (Woldehiwet, 2004). The main clinical symptoms of Q fever widely differ between hosts (Maurin and Raoult, 1999; Norlander, 2000; Woldehiwet, 2004). Abortions are usually manifested in goats (Palmer et al., 1983; Roest et al., 2011a), whereas clinical symptoms are rarely observed in cattle (To et al., 1998; Arricau-Bouvery and Rodolakis, 2005). During abortions in goats and parturition of infected does, a large number of bacteria are excreted by infected animals into the environment (Arricau Bouvery et al., 2003; Roest et al., 2012). Inhalation of pathogen-contaminated aerosols is the main route of infection in humans (Arricau Bouvery et al., 2003). Most of the infected human individuals (~60%) remain asymptomatic after exposure to the pathogen. In symptomatic patients acute infections are usually presented as flu-like illness and pneumonia. Around 1–5% of these cases can develop into chronic infections often leading to life-threatening endocarditis (Maurin and Raoult, 1999; Arricau-Bouvery and Rodolakis, 2005; Raoult et al., 2005; Mazokopakis et al., 2010).

Q fever is prevalent throughout the world, but during 2007–2010 an unprecedented Q fever outbreak occurred in The Netherlands and is referred to as the largest outbreak ever reported (Raoult et al., 2005; Roest et al., 2011b; Hammerl et al., 2015; Kuley et al., 2016). During this epidemic, several isolates from different host species were isolated and cultured. Genotyping by Multiple Locus Variable-number Tandem Repeat Analysis (MLVA) revealed the predominant presence of the so-called CbNL01 genotype and, to a lesser extent, the CbNL12 genotype among the strains isolated from The Netherlands (Roest et al., 2011a). The CbNL01 genotype was identified mostly in strains isolated from goats and specifically from human patients, indicating that goats harboring the CbNL01 genotype strain were most likely the source of the large human Q fever outbreak in The Netherlands. These genotyping studies confirmed the earlier observed epidemiological link between an increase in human Q fever cases and a high abortion rate in goats (Enserink, 2010; van der Hoek et al., 2010; Roest et al., 2011a,b; Kampschreur et al., 2013; Kuley et al., 2016). Additionally, the genotyping studies also showed that the CbNL12 genotype was majorly associated with strains originating from cattle and rarely

with strains from goat, sheep and humans (Roest et al., 2011a,b; Tilburg et al., 2012c; Mori et al., 2013).

As an intracellular pathogen, the most important factor for natural selection of *C. burnetii* could be the interactions with its specific host niches. Genome reduction and the presence of mobile genetic elements and virulence related pseudo-genes throughout the genome are predicted to be specific genome manifestations of the obligate intracellular lifestyle of this pathogen (Beare et al., 2009). Hence, a comparison of *C. burnetii* genome sequences with specific emphasize on genes involved in pathogen-host (cell) interactions or modulation thereof, may shed light on adaptation mechanisms of *C. burnetii* to various host species. Currently lipopolysaccharide (LPS) is the only biologically validated virulence factor of *C. burnetii* and relatively little is known regarding other potential virulence factors, or factors involved in pathogen-host interactions and host specificity (Beare et al., 2009; Gilk, 2012; Kuley et al., 2015a). Previous studies showed the up-regulation of several genes in a Dutch outbreak strain (602) under *in vivo* conditions as compared to *in vitro* cell-based and cell-free growth conditions. These studies suggested a role of these up-regulated genes during intracellular survival and replication of *C. burnetii* in hosts. Unfortunately, a large number of the identified *in vivo* up-regulated genes were of unknown function limiting the information derived from this study with regard to adaptation processes of the bacterium in hosts (Kuley et al., 2015a).

Coxiella burnetii strains with a different genotypic profile can infect a variable range of host species with a different efficiency. For example; the CbNL01 genotype strains are predominantly found in goats and in humans, whereas the CbNL12 genotype strains are commonly found in cattle and hardly in goats and humans (Roest et al., 2011a; Tilburg et al., 2012b,c; Mori et al., 2013). This suggests a higher susceptibility of humans and goats to *C. burnetii* of the CbNL01 genotype than to the CbNL12 genotype strains. Identification of the genomic variations between different genotypic strains by comparative analysis using bioinformatics approaches may help in the identification of the genetic factors of *C. burnetii* involved in its host and cell tropism. Till now this has not been assessed with the Dutch outbreak strains. Therefore, in the present study, we sequenced and compared whole genome sequences of several strains with different genotypic profiles and originating from various hosts to assess host-specific and/or genotype-specific genetic signatures of *C. burnetii*. Main focus during this study was on strains of the CbNL01 genotype which were predominantly found in The Netherlands during the large outbreak, in order to identify genomic signatures associated with their hyper-virulence behavior. From the performed research, we conclude

that an increased genome plasticity and an increased number of single nucleotide polymorphisms in genes, potentially involved in virulence and the evasion of bacterial recognition by the host immune system, is linked to the increased epidemic potential of the Dutch outbreak strains.

MATERIALS AND METHODS

Bacterial Strains and Isolation of Genomic DNA

A summary of the features of all the *C. burnetii* strains used in this study is indicated in **Table 1**. The strains from the Netherlands were primarily isolated from aborted placentas of goats, sheep and from heart valves of human chronic infected patients during the Q fever outbreak period (2007–2010; Roest et al., 2011a). This study also involves strains isolated from aborted placentas of goats and cattle from France. Additionally, a few strains were isolated from Q fever infected humans from elsewhere in the world. All the strains were genotyped based on MLVA (Roest et al., 2011a). The majority of the strains used for sequencing were cultured with a minimal number of sub-passages in BGM cells (Roest et al., 2012). A few strains were cultured axenically in acidified citrate cysteine medium (ACCM-2) also maintaining a minimum number of sub-passages to prevent phase variation (**Table 1**; Omsland et al., 2011). Genomic DNA was isolated using the phenol-chloroform method (Tang et al., 2013) after overnight incubation with ATL lysis buffer and proteinase K (QIAGEN, Hilden, Germany). A prior DNase treatment was performed to the bacterial pellet of cell-culture cultivated strains to eliminate most of the host derived DNA.

Genome Sequences of *C. burnetii*

Draft genome sequences of all *C. burnetii* strains (except 602) were *de novo* reconstructed from Illumina MiSeq paired-end 250 bp reads using SPAdes-3.6.2 (Bankevich et al., 2012). The draft genome sequence of the 602 strain was obtained by using Roche 454XL sequencing technology and was *de novo* assembled by Newbler 2.6. The Dutch outbreak representative strain NL3262 was fully reconstructed using a combination of sequencing technologies (Illumina MiSeq paired-end 250 bp, PacBio RS and Roche 454XL) as described previously (Kuley et al., 2016). The contigs of the draft genome sequences were scaffolded and joined into an artificial chromosome based on homology with the closest *C. burnetii* strain from the database using BLAT synteny tool (**Table 1**; Kent, 2002). Mapping the reads back on the draft genome sequences using Bowtie2 aligner (Langmead and Salzberg, 2012) confirmed its correct synteny and occasionally helped to close minor gaps. All *C. burnetii* (complete and draft) genome sequences were annotated by the NCBI Prokaryotic Genome Annotation Pipeline (http://www.ncbi.nlm.nih.gov/genome/annotation_prok). The (draft) genome sequences and plasmids of all sequenced *C. burnetii* strains have been deposited at GenBank under accession numbers indicated in **Table 1**. Published genomic sequences of several other *C. burnetii* strains used for phylogenetic studies were obtained from GenBank (**Table 2**).

Genotyping of *C. burnetii* Strains

All the sequenced strains were genotyped by MLVA using a selection of 10–12 loci described previously (Arricau-Bouvery et al., 2006; Roest et al., 2011a). As an additional verification step, genotypes of the reconstructed genome sequences as well as published genome sequences of strains from NCBI database were checked by *in silico* MLVA genotype analysis using isPCR (BLAT), MLVA primers and genotype references. *In silico* genotypes of all *C. burnetii* sequenced strains corresponded with known/measured MLVA genotypes. The *in silico* analysis of the database strains gave us information on their MLVA genotype (**Table 2**). Based on the number of repeats per genome loci a minimum spanning tree method was used to cluster the MLVA genotypes using Bionumerics (version 6.6) with default settings.

Sequence Clustering and Phylogenetics

Reconstructed as well as database obtained *C. burnetii* sequences were hierarchical clustered based on their MUMi distance (Deloger et al., 2009) and displayed as a phenogram using the BioNJ algorithm (Gascuel, 1997). The underlying MUMi distance matrix was calculated from the pair-wise non-overlapping maximal unique matches (MUMs; using Nucmer version 3.22; Kurtz et al., 2004). Relative pair-wise distances (MUMi) were obtained by dividing the sum of pair-wise MUMs by the average genome size of the two compared genomes. This MUMi distance varies between 0 and 1 where 0 indicates very similar genomes and 1 is for very distant genomes. MUMi trees were visualized in SplitsTree4 (Huson and Bryant, 2006).

For SNP detection in the sequenced draft genomes and in the complete genome of strain NL3262, reads with low quality (bases with a quality score of $\leq Q20$) were removed before SNP calling. High quality reads were used to map against reference NM genome (GenBank: NC_002971.3) using Bowtie2 short-read aligner (Langmead and Salzberg, 2012). SNP calling was performed using bcftools (Li et al., 2009) using default parameters. SNP sequence alignments were made for only those positions in NM reference strain where each of the sequenced strain had sequence coverage. Phylogenetic trees based on SNPs were created by Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) using standard settings (Sievers et al., 2011) and visualized in SplitsTree4 (Huson and Bryant, 2006). SNPs and insertions/deletions in CbNL01 and CbNL12 strains compared to NM were verified at nucleotide level of the mapped reads.

Comparison of *C. burnetii* Genome Contents

Alignment of the complete genome sequence of strain NL3262 with the sequences of the reference NM, and strains which were related to the outbreak CbNL01 genotype (Z3055, NL-Limburg) (D'Amato et al., 2014; Hammerl et al., 2015) was performed by using Mauve 2.4.0 aligner software with default parameters (Darling et al., 2004). Protein orthologs analysis of *C. burnetii* amino acid sequences obtained from GenBank files from NCBI were performed using Proteinortho software (Version 5) with default settings (Lechner et al., 2011). Orthologs and unique proteins in sequenced strains were identified from the orthologous list generated by Proteinortho. Additionally,

TABLE 1 | Summary of *C. burnetii* genome meta-data.

| Strain | Genotype | Place | Origin | Media ^b | Genome | Plasmid | Putative CDS | GC% | Accession number | | Scaffold | References |
|-------------|------------|-----------------|--------------------------|--------------------|-----------|----------------|--------------|------|------------------|----------|-------------------|---------------------------|
| | | | | | | | | | Genome | Plasmid | | |
| NL3262 | CbNLO1 | NL ^a | Goat | CF | 2,093,477 | 37,320 | 2,101 | 42.9 | CP013667 | CP013668 | - | Kuley et al., 2015b, 2016 |
| CbCVIC1 | CbNLO1 | NL | Goat | CC | 1,988,699 | 37,397 | 2,138 | 42.6 | CP014549 | CP014550 | NL3262 | This study |
| 602 | CbNLO1 | NL | Goat | CC | 1,945,399 | 37,640 | 2,078 | 42.6 | CP014836 | CP014837 | NL3262 | Kuley et al., 2015a,b |
| NLhu3345937 | CbNLO1 | NL | Human chronic Q fever | CC | 1,964,125 | 37,397 | 2,110 | 42.6 | CP014354 | CP014355 | NL3262 | Kuley et al., 2015b, 2016 |
| 42785537 | CbNLO1 | NL | Human chronic Q fever | CF | 1,997,221 | 37,447 | 2,212 | 42.6 | CP014548 | CP014547 | NL3262 | This study |
| 601 | CbNL12 | NL | Goat | CC | 1,963,598 | 37,407 | 2,079 | 42.5 | CP014551 | CP014552 | NM | Kuley et al., 2015b |
| 18430 | CbNL12 | NL | Sheep | CC | 1,964,871 | 37,407 | 2,078 | 42.6 | CP014557 | CP014558 | NM | This study |
| 701CbB1 | CbNL12 | France | Cattle | CF | 1,966,478 | 37,457 | 2,082 | 42.5 | CP014553 | CP014554 | NM | This study |
| 2574 | CbNL12 | NL | Cattle | CF | 1,967,280 | 37,457 | 2,081 | 42.5 | CP014555 | CP014556 | NM | This study |
| Henzerling | Henzerling | Italy | Human acute Q fever | CC | 1,956,892 | 37,394 | 2,064 | 42.6 | CP014559 | CP014560 | 331 ^d | Glazunova et al., 2005 |
| Heizberg | Heizberg | Greece | Human acute Q fever | CC | 1,956,650 | 37,394 | 2,062 | 42.6 | CP014561 | CP014562 | 331 | Glazunova et al., 2005 |
| Scurry | Scurry | USA | Human chronic Q fever | CC | 1,971,034 | - ^c | 2,020 | 42.5 | CP014565 | - | Q212 ^e | Glazunova et al., 2005 |
| Scherpling | Scherpling | Kyrgyzstan | Human acute Q fever | CF | 2,004,282 | 37,406 | 2,087 | 42.5 | CP014563 | KY271744 | Q154 ^f | Glazunova et al., 2005 |

^aNetherlands.^bCell-free (CF) Cell culture (CC) of *C. burnetii* strains.^cPlasmid integrated in chromosome^dCbRSA331.^eCbuG_Q212.^fCbuK_Q154.

TABLE 2 | Genome sequences of *C. burnetii* strains obtained from NCBI database for phylogenetic analysis and their *in silico* MLVA genotype.

| Strain | Accession no. | Source | <i>In silico</i> MLVA | Genome (bp) | Plasmid type | References |
|-------------------------|--------------------|----------|-----------------------|-------------|--------------|-----------------------------|
| NMRSA493 ^c | NC_002971.3 | Tick | NM | 1,995,281 | QpH1 | Beare et al., 2009 |
| Cb175 ^c | HG825990.3 | Human | NM | 1,989,565 | QpH1 | D'Amato et al., 2015 |
| Cb185 ^d | NZ_CBTH000000000.1 | – | NM | 1,991,515 | unknown | – |
| Z3055 ^c | PRJEB1438 | Sheep | * | 1,995,463 | QpH1 | D'Amato et al., 2014 |
| Cb109 ^d | AKYP000000000 | Human | CbNL01 | 2,030,000 | QpH1 | Rouli et al., 2012 |
| EV-Cb_C13 ^d | CCAM010000000 | Ruminant | CbNL01 | 2,023,172 | unknown | Sidi-Boumedine et al., 2014 |
| NL-Limburg ^d | JZWL000000000 | Human | CbNL01 | 2,214,254 | QpH1 | Hammerl et al., 2015 |
| Cb_B1 ^d | CCAH010000000 | Ruminant | CbNL12 | 2,008,014 | QpH1 | Sidi-Boumedine et al., 2014 |
| EV-Cb_BK10 ^d | CCAL010000000 | Ruminant | CbNL12 | 1,999,727 | unknown | Sidi-Boumedine et al., 2014 |
| Cb_B18 ^d | CCAI010000000 | Ruminant | CbNL12 | 2,008,445 | unknown | Sidi-Boumedine et al., 2014 |
| CbRSA331 ^c | NC_010117.1 | Human | CbRSA331 | 2,016,427 | QpH1 | – |
| Dugway ^c | NC_009727.1 | Rodents | Dugway | 2,158,758 | QpDG | Beare et al., 2009 |
| CbuG_Q212 ^c | NC_011527.1 | Human | CbuG_Q212 | 2,008,870 | integrated | Beare et al., 2009 |
| CbuK_Q154 ^c | NC_011528.1 | Human | CbuK_Q154 | 2,063,100 | QpRS | Beare et al., 2009 |
| Q321 ^d | AAYJ010000000 | Cattle | Q321 | 2,004,584 | QpDV | Beare et al., 2006 |
| Goat Q177 ^d | NZ_AAUP000000000.2 | Goat | Goat Q177 | 2,090,565 | QpRS | – |
| Cb_O184 ^d | CCAK010000000 | Ruminant | Cb_O184 | 2,168,222 | QpRS | Sidi-Boumedine et al., 2014 |
| Namibia ^d | CP007555 | Goat | Namibia | 2,101,438 | QpRS | Walter et al., 2014a |
| AuQ01 ^d | JPVV000000000 | Human | AuQ01 | 2,073,000 | QpRS | Walter et al., 2014b |

^cComplete genome sequence.

^dDraft genome sequence.

**In silico* MLVA genotype more similar to NM genotype.

–Information not available.

Integrated, plasmid sequences integrated in chromosome.

Unknown, plasmid type not known for these *C. burnetii* strains.

ortholog analysis of sequenced strains was conducted using the proteome of NM strain as a reference. Pseudo-genes were excluded from the ortholog analysis. Orthologs from different strains were plotted by means of Edwards-Venn diagrams using Venerable package (Version 2.0). For a more detailed genomic sequence comparison, Artemis Comparison Tool (ACT) (Carver et al., 2005) was used to visualize whole genome MUMmer alignments. Detailed analyses were performed on virulence related genes such as variants (mutations/deletions) in LPS encoding region, which were performed by comparing *C. burnetii* genome sequences to LPS encoding region of NM-I (GenBank: AF387640). Also, variants found in the CbNL01 and CbNL12 strains were subsequently pairwise checked against the reference NM strain using ACT and ClustalW (Thompson et al., 1994). We also performed a COG (Cluster of Orthologous Groups) enrichment-analysis on all non-synonymous genes in CbNL01 and CbNL12 strains and analyzed the proportions of different proteins for each COG category.

RESULTS

Genome Sequences of *C. burnetii* Strains

In the present study we report the new genome sequences of 11 *C. burnetii* strains. The reconstruction of the genome sequences of the NL3262 and NLhu3345937 strains was described before (Kuley et al., 2016). **Table 1** is a summary of the main features of the sequenced genomes. Except for strain NL3262,

all the other strains are reported as draft genome sequences. The assembled contigs of these sequenced strains are joined into artificial chromosomes based on scaffolding against the closest genome sequence present in the database. In *C. burnetii* genome sequences, the insertion sequence (IS) elements are dispersed all over the chromosome and are not found on the plasmid (Seshadri et al., 2003). These elements are repetitive in nature and as a consequence, the majority of contig-breaks in the draft genome assemblies mapped to the positions of an IS sequence, when analyzed against the NM reference genome sequence (GenBank: NC_002971.3). As expected, all *C. burnetii* sequenced genomes possess a single circular chromosome of around 2 Mb and a QpH1 plasmid of 37 kb size (except the Schperling and Scurry strains). The bacterial genomes of all sequenced strains are predicted to contain between 1800 and 2062 coding sequences (CDSs). Around 35–39% of these CDSs were annotated as encoding hypothetical products based on the NCBI Prokaryotic Genome Annotation Pipeline. *In silico* MLVA typing of the genome sequences corresponded with similar profiles of strains as measured by lab-based MLVA measurements (**Table 2**).

Genotype-Specific Structural Similarities between *C. burnetii* Genome Sequences

Whole (complete and draft) genome-based phylogenetic analysis was used to infer the relationship between sequenced *C. burnetii* strains of different genotypes. The strains were initially isolated from different hosts such as goats, sheep, cattle and human

patients (both acute and chronically infected; **Table 1**). The analysis also included whole genome sequences of *C. burnetii* strains that have become available in the NCBI database (**Table 2**). Genome sequences of some of the strains from the database, present as draft sequences (several contigs), were scaffolded similarly to our reconstructed genomes and joined into artificial chromosomes based on the reference NM strain. The phylogenetic analysis of these genome sequences were based on MUMi distance. This approach takes into account the number of maximum unique and exact matches (MUMs) of a given minimal length shared by the two genome sequences being compared (Kurtz et al., 2004; Deloger et al., 2009). The phylogenetic analysis was conducted for genomes excluding the plasmids, as presence or absence of plasmids did not result in any difference in hierarchical clustering of *C. burnetii* strains.

Based on the clustering analysis, the strains could be divided into 4 main clades (indicated as 1a, 1b, 2 and 3 in **Figure 1**). The majority of the *C. burnetii* sequenced strains segregated into two clusters [cluster 1 (1a and 1b) and cluster 2, **Figure 1**]. These two clusters correspond to the two major genotypes (CbNL01 and CbNL12) of strains isolated from The Netherlands, respectively. Strain NL3262 and the database strain NL-Limburg clustered closely together (Cluster 1a, **Figure 1**). The genome of strain NL3262 is a complete reconstructed genome derived and curated using different sequencing technologies (Kuley et al., 2016). The

NL-Limburg genome is a draft genome generated with a single sequencing technology (PacBio) and its assembly constitutes 4 contigs representing the chromosome (Hammerl et al., 2015). The contigs of NL-Limburg cover the complete genome of NL3262 without any gaps (as visualized by ACT; data not shown). The ends of the contigs in NL-Limburg contain overlapping sequences of around 50,000 bp. Due to these additional sequences (which are probably the result of assembly error) the genome of strain NL-Limburg (represented as ** in **Figure 1**) is clustered at some distance away from strain NL3262 (MUMi distance 0.025). Upon removal of these repetitive sequences the MUMi distance between strains NL-Limburg (indicated in cluster 1a without **) and NL3262 decreases to 0.01, indicating nearly identical genomes with differences based on 8 SNPs. Draft genome sequences of the database strains (Cb109 and EV-Cb_C13) belonging to the CbNL01 genotype (**Figure 2** and **Table 2**) clustered with the CbNL01 strains (602, CbCVIC1, 3345937, 42785537) originating from The Netherlands in cluster 1b (**Figure 1**). Strains NL3262 and NL-Limburg belong to the CbNL01 genotype strains as well (**Figure 2** and **Table 2**; Hammerl et al., 2015; Kuley et al., 2016). The major differences between the complete and draft genome sequences of the strains belonging to the clusters 1a and 1b are the genomic locations of highly repetitive transposon sequences. The genome sequences of strains NL3262 and NL-Limburg contain 121 and

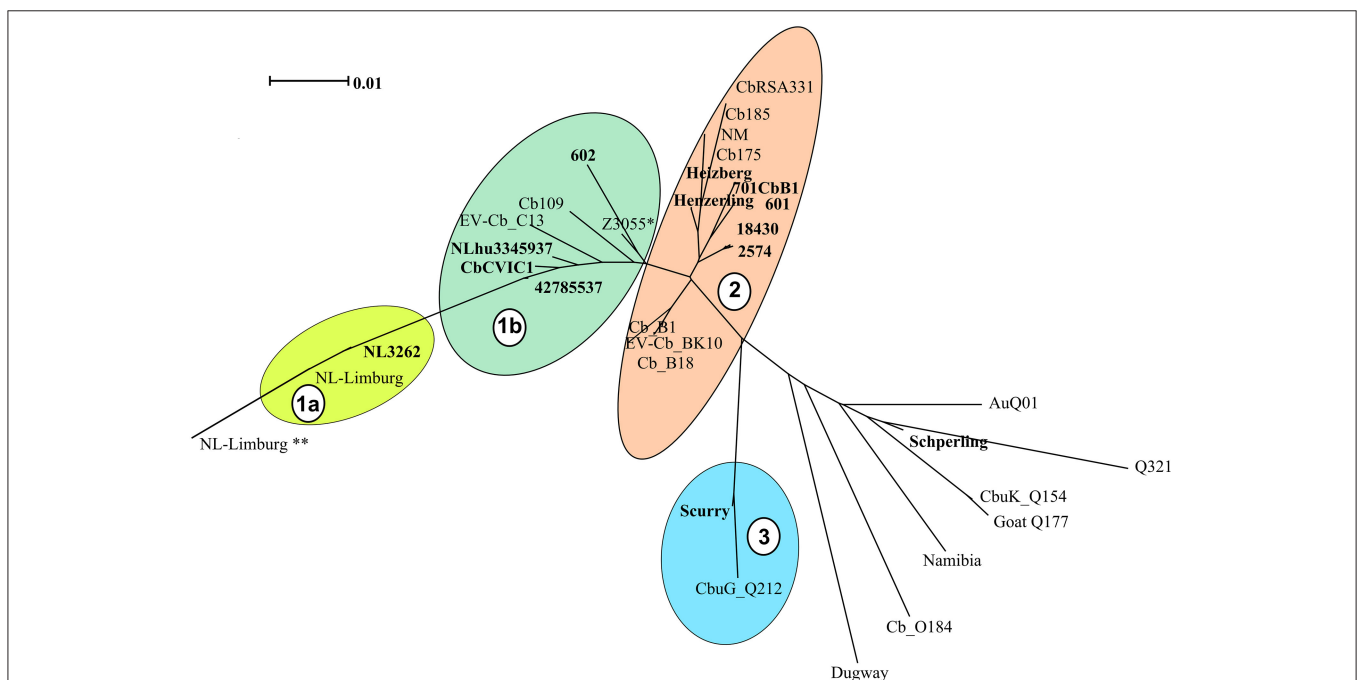
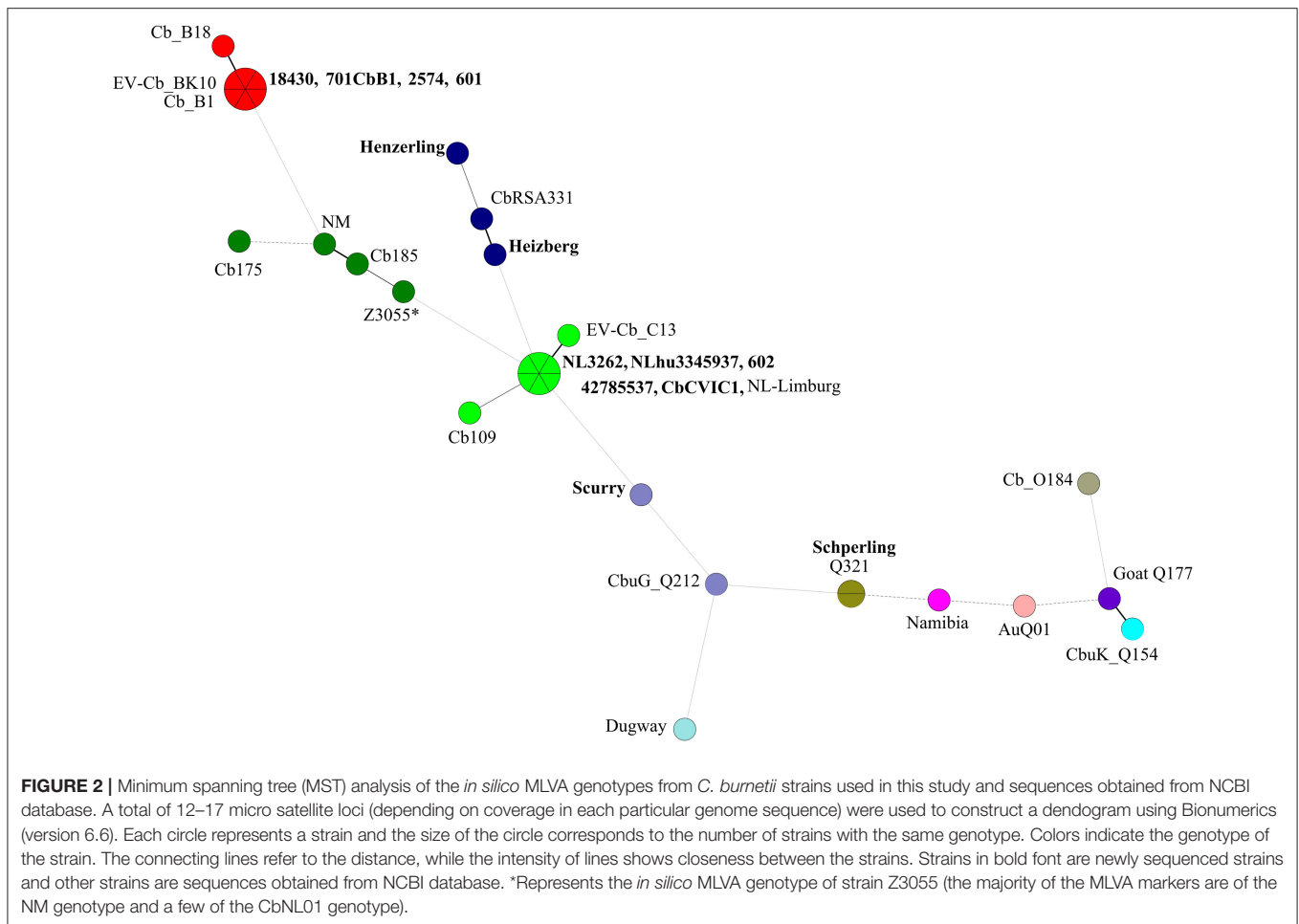


FIGURE 1 | Phylogenetic relationships among *C. burnetii* genome sequences. *C. burnetii* whole (complete and draft) genomes from newly sequenced strains and strains obtained from NCBI database were hierarchical clustered and are displayed as a phenogram using the BioNJ algorithm. Strains in bold font are newly sequenced strains and other strains are genome sequences obtained from NCBI database. Scale bar indicates 1% of genome variability. Clades representing strains from the same MLVA genotype are indicated by colored clouds. Clades contain strains with distance ranging from 0.010 to 0.021. Clade 1a and 1b: CbNL01, Clade 2: CbNL12 and NM like genotype, Clade 3: Scurry genotype (Plasmid-less strains), non-clustered strains are each of different genotypes (except for the sequenced Schperling strain and NCBI obtained Q321 strain, which are of same genotype). *Represents the genome of strain Z3055 which is equidistant from cluster 1 and 2, ** represent the genome of strain NL-Limburg with duplicated regions.



128 IS elements respectively scattered around the chromosome. These IS elements are missing in draft genome sequences, most probably due to their presence at contig-break regions, as indicated above. Thus, due to these structural differences, the draft genome sequences cluster very closely to each other (Cluster 1b) relative to the complete genome sequences (Cluster 1a). The complete sequence of NL3262 and the nearly complete genome sequence of NL-Limburg (belonging to genotype CbNL01) differ with a MUMi distance of around 0.05–0.06 from the draft genome sequences in cluster 1b which also belong to CbNL01 genotype.

Cluster 2 includes the draft genome sequences of strains Henzerling, Heinzberg and the sequences of the CbNL12 genotype group of strains 601, 701CbB1, 2574, and 18430 (Figure 1). Apart from these, cluster 2 contains the database strains belonging to the CbNL12 genotype (draft genome sequences: CbB1, Cb_B18, and EV-Cb_BK10) as well as the reference NM like genotype strains (NM, CbRSA331, Cb175, and Cb185) (Figure 2 and Table 2). The complete genome sequence of strain Z3055 was placed at an equidistance (MUMi distance 0.02) from clusters 1 (CbNL01) and 2 (CbNL12) (Figure 1). An *in silico* PCR analysis of strain Z3055 showed few MLVA markers similar to CbNL01 genotype but the majority indicative for the NM genotype.

This resulted in the presence of Z3055 in-between the nodes of the NM and CbNL01 genotype strains in an MST analysis (Figure 2). Based on the MST and phylogenome analysis, we assessed the genome sequences of strains Z3055 and NM to be very closely related. Cluster 3 included strain Scurry and the database strain CbuG_Q212, both belonging to the same genotype (Figures 1, 2). These strains clustered separately as they contain plasmid-homologous regions integrated in their genomes unlike other strains (Figure 1 and Table 1; Willems et al., 1997). From the phylogenetic analysis, the Schperling strain was found to be the most distant strain as compared to the other *de novo* sequenced strains used in this study (MUMi distance 0.05; Figure 1). The genome sequences of strains Q321, CbuK_Q154, and Goat Q177 were most similar to the strain Schperling with a MUMi distance of around 0.025–0.03. However, except for strain Q321, these strains belonged to different genotypes (Figure 2). Hence, a phylogenetic clustering of the strains based on an MLVA classification rather than on the host-origin background indicated a high-level of genome sequence similarity of genotype-specific *C. burnetii* strains.

A phylogenetic tree was also reconstructed using the identified SNPs among the sequenced strains with respect to the reference

strain NM using bcftools (Figure 3; Li et al., 2009). The SNP tree was constructed using Clustal Omega (Sievers et al., 2011). At SNP level, strains belonging to similar genotypes clustered together with clade 1 containing the CbNL01 genotype strains; clade 2 containing the strains Henzerling and Heizberg and clade 3 containing the CbNL12 genotype strains. This indicates that the presence of similar SNPs in the genomes of the same genotype strains relative to the reference strain NM. Strains Schperling and Scurry clustered separately, indicating that these strains contain SNPs distinct from the SNPs present in other sequenced strains compared to the reference strain NM. The data corresponds to the presence of distinct genotypes of Schperling and Scurry compared to other sequenced strains (Figure 2). This SNP based clustering of genotype-specific strains further confirms the highly similar genome sequences of strains belonging to the same genotype.

Conserved and unique Gene Content in *C. burnetii* Sequenced Genomes

A comparison of the genome sequences of *C. burnetii* shows that all sequenced strains are highly conserved in terms of size, number of coding sequences and in nucleotide composition (Table 1). The degree of conservation is illustrated by the comparison of the coding capacity of these genomes: 1447 predicted proteins are orthologs in all sequenced 13 *C. burnetii* strains in this study (with a total CDS count ranging from 1800 to 2062) among which 1409 of the predicted proteins are orthologs with reference NM. Remaining 38 proteins not shared with NM are majorly hypothetical protein products mostly occurring due to modifications (such as point mutations,

small deletions/ insertions of <50 bp) in existing NM pseudo-genes (Supplementary Table 1). The 1409 orthologs among all strains constitute the coding core genome. All sequenced *C. burnetii* strains contain the QpH1 plasmids, except for strain Schperling containing QpRS plasmid and the plasmidless strain Scurry. The QpH1 plasmids from sequenced strains were all very similar to the reference NM QpH1 plasmid (Accession No. NC_002118.1) with differences on an average of 58 point mutations. The draft plasmid sequence of Schperling was similar to the published sequence of plasmid QpRS (Accession No. Y15898.1) with differences in 63 SNP and a gap of 1810 bp (data not shown).

Ortholog analysis of the proteome of the sequenced *C. burnetii* strains showed the presence of a few unique proteins annotated as hypothetical protein in CbNL01 and CbNL12 strains (data not shown). Further analysis of genes encoding these unique proteins showed that these genes were present at the beginning or end of contigs in assembled draft genome sequences. As contig ends of draft genome sequences are highly repetitive, genes annotated in these regions were not considered as unique genes, because these genes were truncated transposase genes. Taking this factor into account and other manual annotation corrections, we found no unique complete genes encoded in CbNL01, CbNL12, Henzerling, and Heizberg strains. Instead of unique genes for each strain we did find orthologs that were genotype-specific. These orthologs were shared between strains of the same genotype only and were annotated as hypothetical proteins. When we compared these genotype-specific genes at the nucleotide level to the genome of the reference strain NM, mutations or partial ORF

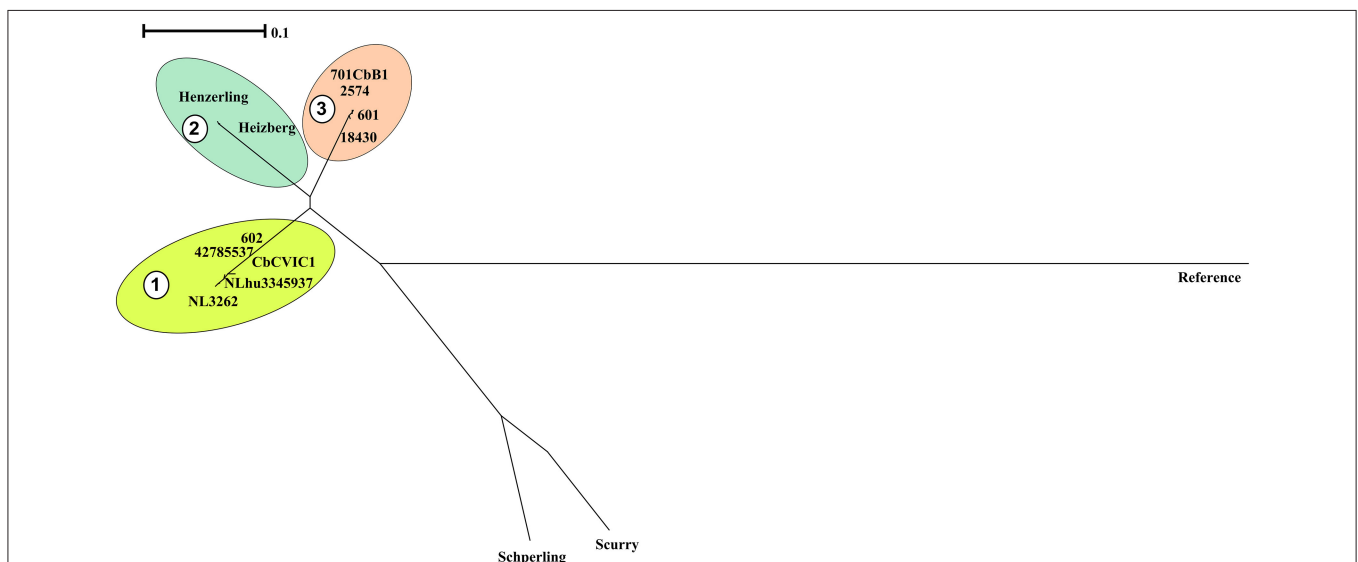


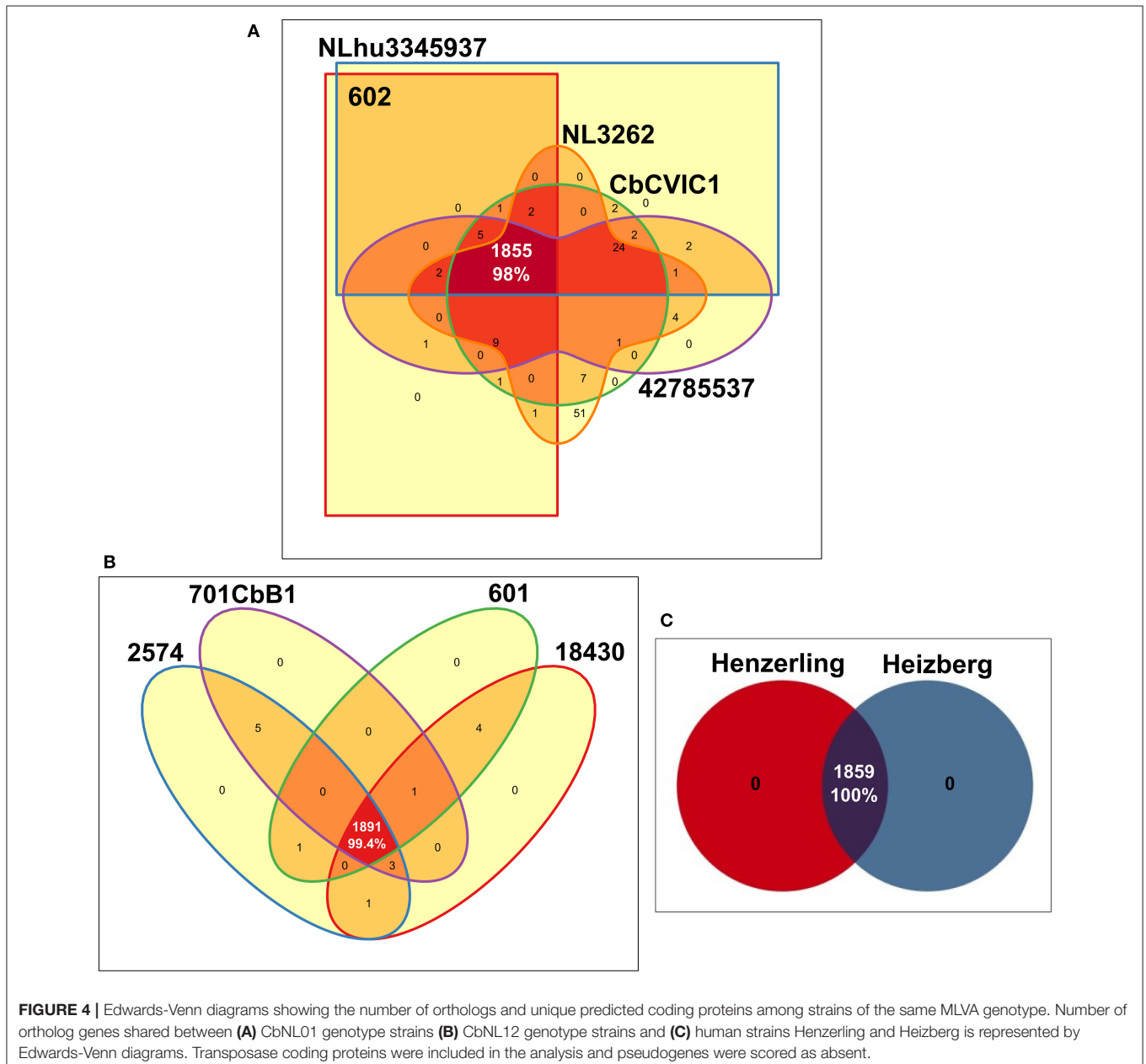
FIGURE 3 | Phylogenetic relationships among sequenced *C. burnetii* genomes based on SNP analysis. The phylogenetic tree is based on SNPs among genomes of the sequenced strains with respect to the genome of reference strain NM using bcf tools. The SNP tree was constructed using Clustal Omega and the clustering is displayed as a phenogram using the BioNJ algorithm. The scale bar indicates the number of substitutions per site. Clades representing strains from the same genotype are indicated by colored clouds. Clade 1:CbNL01, Clade 2: Henzerling and Heizberg (similar genotype), Clade 3: CbNL12. The Schperling and Scurry strains are present separately due to the presence of distinct SNPs relative to other sequenced strain genomes. NM indicates the polymorphic regions in sequenced strain genomes relative to the NM reference genome.

deletion in existing NM pseudo-genes were observed for which altered functionality or regulation cannot be directly inferred (Supplementary Table 2).

By comparing the proteome of the 13 sequenced strains (**Table 1**) 50 and 98 unique protein encoding genes were identified in strains Scurry and Schperling, respectively. Out of these, 3 and 17 genes are exclusively found in Scurry and Schperling, whereas the other genes were orthologs with the closely related strains CbuG_Q212 and CbuK_Q154, respectively (**Figure 1**). These unique genes in Scurry and Schperling strains were annotated as hypothetical proteins and result from point mutations in existing pseudo-genes of CbuG_Q212 and CbuG_Q154 (data not shown).

Highly Similar *C. burnetii* Gene Content within the Same Genotype-Groups

A comparison of the coding capacity of the genomes of strains belonging to the same genotype showed 98, 99.4 and 100% orthologs in predicted proteins of CbNL01, CbNL12, and the human strains (same genotype Henzerling and Heizberg strains), respectively (**Figure 4**). Fifty-one unique genes (non-orthologs) of NL3262 encode for transposase associated proteins (**Figure 4A**). These genes are not properly annotated in 602, CbCVIC1, NLhu3345937, and 42785537, because they fall within contig-break regions of the draft genome sequences. The high level of similarities of the proteomes between these strains indicates that the gene content is highly conserved within

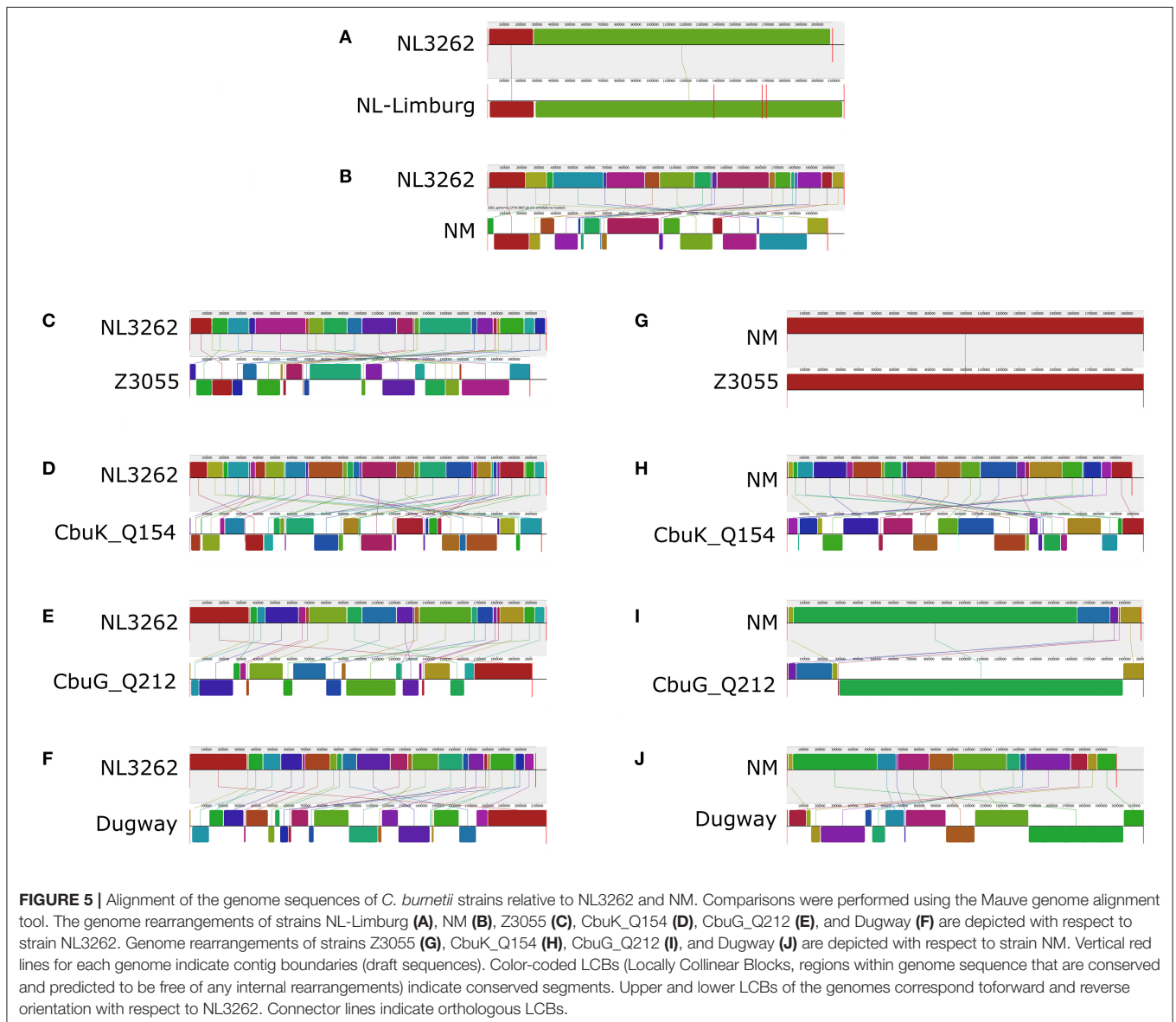


strains of the same genotype. As all the 13 *C. burnetii* strains included in this study were initially isolated from different hosts, a comparison was made to assess the percentage of proteins shared by strains derived from each host-species. These comparisons resulted in the presence of 94.6, 100, 87.9, and 87% orthologs in predicted proteins of goat, cattle, human-acute, and human-chronic isolated strains respectively (Supplementary Figure 1). The lower percentage of orthologs shared by strains obtained from various host species is due to the presence of different genotypes within each host-species, except for the cattle-derived strains, as both cattle strains were of same CbNL12 genotype. For example; the strains isolated from goat include both CbNL01 (NL3262, 602, CbCVIC1) and CbNL12 (601) genotypes (Supplementary Figure 1B). Further, the strains isolated from acute and chronic human Q fever patients contain more divergent strains (Schperling and Scurry respectively).

Because of this, the protein orthologs of acute and chronic human strains are comparatively less (~87%, Supplementary Figures 1C,D). Taken together, the percentage of orthologs in strains obtained from similar host-species is less conserved than the strains obtained from similar genotype. This clearly indicates the presence of genotype-specific functional characteristics of *C. burnetii* strains.

Architecture of the NL3262 Complete Genome Sequence

To assess the genome architecture of strain NL3262 and to study its putative genome rearrangements, the complete genome sequence of strain NL3262 was aligned with the genome sequence of strains NL-Limburg, NM, Z3055, Cbuk_Q154, CbuG_Q212 and Dugway using the Mauve genome aligner (Figure 5). Among the different *C. burnetii* strains, the CbNL01



genotype strains NL3262 and NL-Limburg encode the highest number of transposase genes (106-112 IS110 family transposase, 3 ISAs1 family transposase and 12-13 remnant transposase genes, respectively). Other completely sequenced *C. burnetii* genomes generally contain only 30-60 transposase genes (Beare et al., 2009; D'Amato et al., 2014). Compared to the genome sequence of NL3262, chromosomal rearrangements in strains NL-Limburg, NM, Z3055, Cbuk_Q154, CbuG_Q212, and Dugway strains have resulted in 2, 19, 21, 32, 21, and 23 Locally Collinear Blocks (LCBs) of the same gene content and gene order respectively (Figures 5A–F). NM and Z3055 strains share almost all LCBs indicating a similar genome structure (Supplementary Table 3). Cumulatively, the LCBs of NM, Z3055, Cbuk_Q154, CbuG_Q212, and Dugway represent 116 genomic breakpoints relative to the genome of strain NL3262 (Figures 5B–F). In the genome sequence of strain NL3262, 102 (88%) of the breakpoints occurred at intact or a remnant transposase sequences. These transposase mostly belonged to IS110 family and very rarely to IS30 and ISAs1 family (Supplementary Table 3). The presence of transposase genes at almost all genome breakpoints suggests an important role of homologous recombination in order to establish the observed genome rearrangements of the Dutch outbreak strains. Unlike other strains, the gene order in strains NL3262 and NL-Limburg are nearly identical, yielding only 2 LCBs. The absence of rearrangements between NL3262 and NL-Limburg, indicates a similar genome structure of these CbNL01 genotype Dutch outbreak strains isolated from goat and human patient (Figure 5A).

Increased Number of Transposons in Dutch Outbreak CbNL01 MLVA Genotype Strains

Genome sequencing of a number of Dutch outbreak strains belonging to the CbNL01 MLVA genotype resulted in 132, 135, and 123 contigs for strains NLhu3345937, 42785537, and CbCVIC1, respectively. In contrast, sequencing of the CbNL12 strains 601, 2574, 18430, and 701CbB1 resulted in 43, 40, 66, and 42 contigs, respectively. As described above, the majority of the contig-break points contained repeat sequences and corresponded mostly to transposase genes when mapped against the NM reference genome. Based on this observation, we anticipate that the observed number of contigs roughly corresponds to the number of transposase genes encoded by the strain. Based on this assumption, the CbNL01 and CbNL12 strains contain around 130 and 47 contigs, respectively. The assumed high numbers of transposons in the draft genome sequences of CbNL01 strains also matches with the number of annotated transposase genes in strains NL3262 (121 intact and remnant transposons) and NL-Limburg strains (128 intact and remnant transposons) belonging to the same genotype group. Taken together, the largest number of transposons are encoded by the CbNL01 Dutch outbreak strains as these strains contain two to three fold more transposase genes compared to other strains of *C. burnetii* sequenced so far (Beare et al., 2009; D'Amato et al., 2014).

Differences in Dutch Strains at SNP Level

To further assess the differences in the genome sequences of the Dutch strains from different genotypic groups, we performed a detailed NM reference-centric variant analysis. For this analysis we used the complete genome sequences of the NL3262 strain (Kuley et al., 2016) and the draft genome sequence of the CbNL01 and CbNL12 strains. We found an average of 2514 SNPs between the CbNL01 strains and NM. Among these, 417 SNPs were present in intergenic regions, 1041 were synonymous SNPs in coding regions and 1056 were non-synonymous SNP which corresponded to 702 mutated genes. Between the CbNL12 strains and NM an average of 2078 SNPs were found, with 340 SNPs in intergenic regions, 889 SNPs in coding regions (synonymous SNP) and 849 non-synonymous SNPs which corresponded to 581 mutated genes. With respect to NM, these mutated genes in CbNL01 and CbNL12 genotype strains majorly encoded hypothetical proteins, membrane proteins, transporter proteins, DNA repair proteins, translation-related proteins and a few virulence-associated proteins (explained in detail below). Compared to the genome of NM, the genomes of the CbNL01 strains showed a deletion of 3,600 bp (containing ankyrin-repeat containing protein; CBU_0072), a deletion of 2,300 bp (containing several hypothetical proteins; CBU_0877-CBU_0880) and a larger deletion of 9,700 bp (containing the peptidoglycan catabolism proteins; CBU_1101-CBU_1112). A deletion of around 2,250 bp (containing several hypothetical proteins; CBU_0016-CBU_0019) was observed in the CbNL01 and CbNL12 genome sequences, relative to the NM genome sequence (Supplementary Figure 2).

With respect to mutations in membrane protein encoding genes, a total of 64 and 48 genes were mutated in the CbNL01 and CbNL12 strains out of the 104 membrane protein encoding genes in the NM strain. Thirty eight of these genes contained the same mutation in both genotype strains. Additional mutations in 26 hypothetical membrane proteins were specific for all the CbNL01 strains. Although, the exact functions of these hypothetical protein products are not known, changes in these cell surface proteins could potentially contribute to an altered antigenic profile of the CbNL01 strains. Large numbers of mutations were also seen in transporter genes, where a total of 41 and 25 genes were mutated in the CbNL01 and CbNL12 strains (from a total of 75 transporters in NM). Among these, 20 genes contained identical mutations in both genotype strains, whereas 21 genes contained mutations specific for the CbNL01 strains. These transporter genes belonged mostly to the ABC (ATP-binding cassette transporters) and MFS (major facilitator superfamily) family transporters. Further COG analysis of the mutated transporter genes showed that these genes could be primarily involved in transportation of amino acid, carbohydrate and ions (COG categories E, G, and P respectively). Moreover, we also identified mutations in several translation-related proteins in the outbreak CbNL01 strains. Around 29 of these mutated genes were also found in the CbNL12 strains, whereas 15 mutated genes were specifically found in the CbNL01 strains (from a total of 121 translation genes in NM; Supplementary Figure 2).

Recognized virulence factors of *C. burnetii* are genes involved in primary defense mechanisms against oxidative stress. These include several DNA repair genes involved in base excision repair, nucleotide excision repair, mismatch repair, addAB mediated recombinational repair systems, and oxidative stress enzymes (Mertens and Samuel, 2012). Our variant analysis showed high number of mutations in the DNA repair genes; wherein 13 mutated genes were shared in both genotype strains and mutations in 7 genes were specific to the CbNL01 strains (from a total of 26 DNA repair genes in NM). On the contrary, we did not find any mutations in genes encoding for oxidative stress enzymes among the various strains indicating that these genes are highly conserved. The secondary defense mechanisms of *C. burnetii* includes the manipulation of host cell processes (Mertens and Samuel, 2012). A Dot/Icm type IV secretion system (T4SS) is predicted to be encoded by the genome of *C. burnetii* homologous to the *L. pneumophila* Dot/Icm system. This is an important transfer system associated with the delivery of effector proteins from the bacterial cell to the host cytosol mediating PV formation and other cellular events required for bacterial maintenance and survival (Seshadri et al., 2003; van Schaik et al., 2013; Kuley et al., 2015a). Among the T4SS genes, 4 genes were mutated in both genotype strains and an additional 4 genes were mutated specifically in the CbNL01 strains. Furthermore, 12 effector proteins characterized by the presence of eukaryotic like protein domains were also mutated in the CbNL01 and CbNL12 strains relative to NM. These genes are predicted to be virulence-related as the encoded proteins contain eukaryotic-like domains which may mimic host proteins and modulate host responses required for successful bacterial infection (Supplementary Figure 2). Taken together, compared to the reference strain NM, genotype-specific mutations of genes were present in the Dutch strains. Furthermore, a relatively larger number of non-synonymous mutations were present in critical genes of the CbNL01 strains compared to the CbNL12 strains.

Changes in LPS Encoding Genes and Its Effect on Phase Variation

Repeated *in vitro* passages of *C. burnetii* strains have been shown to induce antigenic variation and loss of virulence characteristics due to the transition into a truncated LPS structure. This truncation of LPS is associated with chromosomal deletions of O-antigen coding genes, located in a 38 kb region in the genome resulting in antigenic phase variation from phase I to phase II (Hoover et al., 2002; Denison et al., 2007; Kuley et al., 2015b). The genome sequences established in this study, were compared with the LPS encoding region (CBU0676–CBU0706) of NM-phase I (GenBank: AF387640) to visualize deletions in the LPS encoding region. The human strains Henzerling, Heizberg, Schperling, and Scurry contained no deletions in this region (Table 3). However, a deletion of 201 bp (NM-phase I LPS encoding region: 14604–14805) was consistently observed in this region in all sequenced (601, 2574, 701CbB1, 18430) and database strains (CbB1, CbB18, EV-Cb_BK10) of the CbNL12 genotype (Table 3). This deletion corresponded to a portion of CBU_0686 gene (old gene name: JB153-5), a paralog predicted

TABLE 3 | Deletion of LPS O-antigen encoding genes of *C. burnetii* strains compared to NM-phase I sequence (GenBank: AF387640).

| Strain | Genotype | Deletion | No. of bp | Gene containing deletion | Remarks |
|-------------------------|------------|---------------|-----------|--------------------------|---------------------|
| NL3262 | CbNL01 | 20,476–20,830 | 354 | CBU_0691 | Partly ^b |
| CbCVIC1 | CbNL01 | 20,476–20,830 | 354 | CBU_0691 | Partly |
| 602 | CbNL01 | 20,476–20,830 | 354 | CBU_0691 | Partly |
| 42785537 | CbNL01 | 20,476–20,830 | 354 | CBU_0691 | Partly |
| NLhu3345937 | CbNL01 | 20,476–20,830 | 354 | CBU_0691 | Partly |
| NL-Limburg ^a | CbNL01 | 561–7,902 | 7341 | CBU_0676, CBU_0683 | Partly |
| | | | | CBU_0677 | Full ^c |
| | | | | to CBU_0682 | |
| | | 20,476–20,830 | 354 | CBU_0691 | Partly |
| Cb109 ^a | CbNL01 | 20,476–20,830 | 354 | CBU_0691 | Partly |
| EV-Cb_C13 ^a | CbNL01 | 20,476–20,830 | 354 | CBU_0691 | Partly |
| 601 | CbNL12 | 14,604–14,805 | 201 | CBU_0686 | Partly |
| 18430 | CbNL12 | 14,604–14,805 | 201 | CBU_0686 | Partly |
| 701CbB1 | CbNL12 | 14,604–14,805 | 201 | CBU_0686 | Partly |
| 2574 | CbNL12 | 14,604–14,805 | 201 | CBU_0686 | Partly |
| CbB1 ^a | CbNL12 | 14,604–14,805 | 201 | CBU_0686 | Partly |
| CbB18 ^a | CbNL12 | 14,604–14,805 | 201 | CBU_0686 | Partly |
| EV-Cb_BK10 ^a | CbNL12 | 14,604–14,805 | 201 | CBU_0686 | Partly |
| Henzerling | Henzerling | None | – | – | – |
| Heizberg | Heizberg | None | – | – | – |
| Scurry | Scurry | None | – | – | – |
| Schperling | Schperling | None | – | – | – |

^aDatabase strains.

^bDeletions of a portion of gene.

^cDeletions of complete genes.

to encode a 2-Oxoacid dehydrogenase (Hoover et al., 2002). As a paralogous gene, deletion of a portion of this gene might not affect the LPS synthesis as the function might be compensated by the other gene in the genome. Similarly, a deletion of 354 bp (NM-phase I LPS encoding region: 20476–20830) was consistently observed in this region in all sequenced (NL3262, 602, CbCVIC1, NLhu3345937, 42785537) and database strains (NL-Limburg, Cb109, EV-Cb_C13) of the CbNL01 genotype. This deletion corresponded to a portion of CBU_0691 gene (old gene name: JB153-10, Table 3) and is predicted to be involved in virenose synthesis which is one of the unusual sugars only present in the *C. burnetii* LPS-phase I structure (Toman et al., 1998). Due to the deletion, CBU_0691 gene is frame-shifted and might consequently not encode for enzymes involved in virenose synthesis. We hypothesize that this deletion of 355 bp may induce the start of phase shifting in CbNL01 strains. Since the genome sequences of 602 (CbNL01 genotype) was obtained from *in vitro* grown cell cultures, we looked back into our previous functional analysis data of the 602 strain cultured in cells maintaining a low passage number. Our transcriptome measurements showed that the CBU_0691 gene was actively transcribed in cells (Kuley et al., 2015b). Although the deletion

of a portion of this gene results in frame shifting of CBU_0691, it is still possible that gene CBU_0691 is transcribed but that no protein product is formed. In addition, a complete gene deletion (corresponding to CBU_0677 to CBU_0682) and a deletion of part of the genes (corresponding to CBU_0676 and CBU_0683) were observed in the NLhu3345937 strain of the CbNL01 genotype. The NLhu3345937 strain was passaged in cells for only 11 times and already showed deletion of several O-antigen genes, indicating a high rate of phase shifting in this strain under *in vitro* culture conditions (Table 3). Apart from these deletions, non-synonymous mutations in CBU_0688 and CBU_0698 were present in all CbNL01 strains relative to NM. Additionally, mutations in 12 O-antigen genes were observed in all CbNL01 and CbNL12 strains, compared to NM (Supplementary Figure 2). Taken together, we observed genotype-specific deletions and mutations in the LPS encoding region at a higher rate in CbNL01 strains than in CbNL12 strains.

DISCUSSION

In this study we used genome comparisons to detail the extent of genomic diversity of several *C. burnetii* strains. Focus was given to goat and human strains of the Dutch predominant CbNL01 genotype to identify genomic aspects resulting in its anticipated increased virulence characteristics. In all the strains, we firstly performed phylogenome analysis to infer the relationships between the sequenced strains. Further, we analyzed orthologs and mutations among all sequenced strains with respect to the NM reference strain to assess differences between strains at the gene and nucleotide level.

Regardless of whether the sequence data was derived from complete or draft genome sequences, the phylogenetic clustering of the strains was based on the MLVA genotype rather than on host origin. Clustering of the genome sequences was similar to the clustering based on *in silico* MLVA genotyping. Moreover, a similar genotype-specific clustering of strains was observed based on SNP analysis. The inferred topology of *C. burnetii* strains is consistent with different analysis methods and also with previous phylogenetic studies based on MLVA and MST genotypic data (Roest et al., 2011a; Tilburg et al., 2012b,c). Based on these findings, a close relationship of genotype-specific *C. burnetii* strains was clearly observed. The phylogenetic analysis, as described here, is an improved genotyping tool, besides the MLVA and MST methods, as it uses whole genome sequences for comparisons of strains rather than a few genomic loci. Nevertheless, similar results were obtained with the three techniques. Thus the results of our study validate the traditional genotyping methodologies as sensitive and less-time consuming tools used for outbreak investigations and molecular epidemiology of pathogens.

Among the different Dutch outbreak strains, the NL3262 strain has now a complete genome sequence. The NL-Limburg database strain has a draft genome sequence, although with contigs covering the complete genome of NL3262. Both strains are clonal as the difference between these strains is limited to 8 point mutations. Compared to each other, both strains NL3262

and NL-Limburg have only one DNA rearrangement in their genomes (Figure 5) and contain around 121-130 transposase encoding genes. This is the highest number of transposase genes encoded by any other *C. burnetii* strain (Beare et al., 2009). Upon phylogenetic genome analysis, strains NL3262 and NL-Limburg cluster together in cluster 1a. These strains differ from other CbNL01 strains in cluster 1b by point mutations (an average of 12 SNPs) and annotation of transposons, which are the major contig break regions in these draft genome sequences (Figure 1). Thus, based on phylogenetic and SNP analysis, the outbreak goat (NL3262, 602, CbCVIC1) and human strains (NLhu3345937, 42785537, NL-Limburg) of the CbNL01 genotype cluster together and possess almost identical genomes. Furthermore, these data confirm that goat are the source of the Dutch human Q fever outbreak as suggested previously by genotyping (Figures 1, 2; Roest et al., 2011a).

Both CbNL01 and CbNL12 strains were isolated in The Netherlands during the outbreak period from 2007 to 2010. The CbNL12 genotype strains were the second most prevalent strains after the CbNL01 strains and were only isolated from cattle, goat and sheep, but not from humans. The CbNL12 genome sequence was shown to be similar to the genome sequence of the reference strain NM (Figures 1, 2). Compared to CbNL01 strains, the CbNL12 strains clustered separately based on phylogenomics (cluster 1 and 2, Figure 1) but the distance between these clusters was very small (MUMi distance 0.022–0.03) indicating closely related genomes. Although the CbNL12 and CbNL01 strains belong to distinct MLVA genotypes, the major differences were based on point mutations (average of 2400 SNPs) and the number of transposase encoding genes. The cattle strains 2574 and 701CbB1 used in this study are of the CbNL12 genotype and from different geographical locations. The 701CbB1 strain is an archival sample isolated in France and a similar genotype strain (2574) was recently isolated during the Dutch Q fever outbreak period. The cattle strains were clustered together phylogenetically, indicating a close relationship between these strains. In addition, the published genome sequences of 3 strains (Cb_B1, Cb_B18, EV-Cb_BK10) of the CbNL12 genotype, which were isolated from ruminants in France, were also present in the same cluster (Cluster 2, Figure 1). The presence of similar genome strains obtained from various geographical locations may imply a clonal spread of CbNL12 strains over different European countries. This idea is strengthened by the fact that strains of the CbNL12 genotype were identified in consumer cow milk products obtained from various European countries (Tilburg et al., 2012a). As the CbNL12 strains are not detected in human samples but are spread over many European countries, it could be hypothesized that humans are less susceptible to CbNL12 genotype strains, decreasing its risk for causing an outbreak in humans.

The core genome of sequenced *C. burnetii* strains constitutes up to 75% of the complete gene repertoire of the sequenced *C. burnetii* strains, indicating a conservative nature of evolution. Gene ortholog analysis showed a 98% overlap between strains of the same genotype which is higher than the overlap between strains of the same host-origin. This indicates that

strains of the same genotype are of clonal origin with highly conserved gene content (**Figure 4**). Unique genes were not identified in the Dutch outbreak strains and other closely related strains (CbNL12, Henzerling, and Heizberg), although genotype specific genes were identified (Supplementary Table 2). A search of host-specific genes and signatures (based on point mutation) didn't yield any results, indicating that host-specific genes and mutation were absent. Thus, genotype-specific changes were clearly observed in strains, whereas host-specific changes in the strains could not be observed. Based on these findings, the observed phenotypic differences in clinical manifestations of Q fever in natural hosts, goats (abortion), cattle (rarely clinical symptoms), human acute infections (asymptomatic) and human chronic infections (endocarditis), is most probably due to host-specific (immune) responses to an infection with *C. burnetii* or other complex pathogen-related factors (such as host-tropism) which are not known yet.

The difference between the genome sequences of the Dutch outbreak strains and the reference strain NM were based on point mutations, small deletions in parts of genes and a few deletions of complete genes in the CbNL01 strains (Supplementary Figure 2 and **Table 3**). Novel genes, potentially associated with increased bacterial virulence, were absent in the CbNL01 strains. This suggests that modifications of existing genes have given rise to the high-virulent features of the outbreak strains rather than the acquisition of novel genetic factors by *C. burnetii*. The present findings support previous observations that genomes of *C. burnetii* lack broad genetic variance, due to absence of DNA exchange machinery and the intracellular lifestyle, which limits opportunities for genetic exchange (Seshadri et al., 2003; Beare et al., 2006; D'Amato et al., 2014). Upon assessing the variants in the Dutch strains, relative to NM, genotype-specific mutations in the CbNL01 strains were predominantly found in hypothetical membrane proteins, transporter proteins, translation proteins, DNA repair protein encoding genes and T4SS genes (Supplementary Figure 2). Our study showed a large number of mutations in membrane protein encoding genes of CbNL01 strains, relative to the CbNL12 strains. Amino acid changes in these membrane proteins may lead to a distinct antigenic profile of the CbNL01 strains, which may result in a change in immune evasion efficiencies. Based on these findings, we speculate that a shift in the bacterial antigenic repertoire has resulted in the increased zoonotic potential of the outbreak strains. Furthermore, mutations in transporter genes involved majorly in metabolite transportation, could be crucial for efficiently assessing host-derived nutrients to adapt in the host hostile environments. Additionally, a large number of mutations in the protein synthesis machinery could be a novel mechanism for adaptation of CbNL01 strains to the host environment. These mutations may result in altered levels of protein synthesis, thereby influencing the growth rate of the bacterium in the host niche (D'Amato et al., 2014). Increased numbers of mutations in DNA-repair genes of the CbNL01 strains could be related to an altered/increased capability of these strains to fight against oxidative stress. This feature might aid in efficient maintenance

of the bacterial cell integrity in harsh environments of host cell PV. Finally, the mutations in genes encoding virulence related effector proteins and the T4SS proteins delivering system could contribute to an increased/altered virulence potential of CbNL01 strains by an enhanced/altered protein export. This could be beneficial for *C. burnetii* in effective manipulation of host components or host intracellular processes to enhance successful replication and persistent infections in hosts (Beare et al., 2009).

Our study shows that non-synonymous SNPs in a number of membrane protein encoding genes, transporter genes, genes encoding virulence-related effector proteins and O-antigen genes of CbNL01 strains resulted in gene frame-shifts (Supplementary Figure 2). These frame shifts result in the gain of pseudo-genes, which could be a mechanism of "genome reduction." It could be speculated that lost gene functions might be compensated by other genes. Alternatively, the bacteria might not require this function anymore due to its intimate relationship with host cells (Merhej et al., 2009). The presence of frame-shift mutations might increase the likelihood of the removal of whole genes by DNA deletion processes (Kuo and Ochman, 2010), as evidenced by the observed deletion of an ankyrin repeat protein ortholog to CBU_0072 in all CbNL01 strains (Supplementary Figure 2). It could be speculated that loss of genes can be a crucial mechanism for specialization of the bacteria to survive in a specific niche as seen previously for other bacterial species (Georgiades and Raoult, 2011; Rolain et al., 2013). On the whole, our findings favor the reductive evolution of genomes in the CbNL01 strains, where increased gene loss could have led to bacterial specialization in a host niche (Georgiades and Raoult, 2011).

In this study we focused on the O-antigen encoding region, the most prominently known virulence-related genomic region of *C. burnetii*, which is prone to deletions resulting in phenotypic phase shifting (Hoover et al., 2002). The sequenced strains were cultivated both in cell-based and in cell-free systems and the passage numbers were kept at the minimum to avoid phenotypic phase shifting. Although we maintained low passage numbers (passages between 4 and 14), the CbNL01 and CbNL12 genotype strains showed partial gene deletions in this region. Such deletions were not observed in the remaining human strains used in this study (Schperling, Scurry, Henzerling, and Heizberg) (**Table 3**). Deletion of portion of the CBU_0691 and CBU_0686 genes in CbNL01 and CbNL12 strains, respectively were identified. We anticipate that these deletions could have a significant effect on the structure and function of LPS as these genes are primarily involved in LPS biosynthesis. Although, these genes are polymorphic and frameshifted, the Dutch strains (CbNL01: NL3262, 602 and CbNL12: 601) were shown to be virulent in a mice virulence bioassay (Kuley et al., 2015b). Hence, based on these observations the effect of genotype-specific deletions in the LPS encoding region and its effect on virulence of the Dutch strains are not clearly known and should be the further investigated.

Transposable elements are present in *C. burnetii* genomes and contribute to genome plasticity (Beare et al., 2009). The

Dutch outbreak strains (NL3262, NL-Limburg) encode more than 100 transposase genes which is the highest number of transposase genes found in *C. burnetii* (30 in NM, Z3055, Dugway; 59 in Cbuk_Q154 and 40 in CbuG_Q212). Although we cannot infer the exact number of transposons in draft genome sequences, we anticipate that the higher number of contigs in the sequenced CbNL01 strains (average of 130) corresponds to the higher number of transposons, as contig ends often map to transposase genes relative to NM genome. The CbNL01 and NM genomes are similar in the majority of their gene content, but differ significantly in genome structure due to a large number of genome rearrangements. Moreover, increased genome arrangements were also seen in NL3262 compared to complete genomes of Z3055, Cbuk_Q154, CbuG_Q212, and Dugway strains (Figure 5). Previous comparative studies of *C. burnetii* complete genomes (Beare et al., 2009) as well as our current analysis showed genome rearrangements at around 0, 21, 6, and 13 chromosomal locations in NM compared to Z3055, Cbuk_Q154, CbuG_Q212, and Dugway strains respectively (Figures 5G–J). Our analysis also showed genome rearrangements of NL3262 at 21, 31, 21, and 23 locations compared to Z3055, Cbuk_Q154, CbuG_Q212, and Dugway strains respectively (Figures 5C–F, Supplementary Table 3). Taken together, comparison of Cbuk_Q154, CbuG_Q212 and Dugway strains showed 40 chromosomal breakpoints relative to NM chromosome and 96 chromosomal breakpoints relative to NL3262 chromosome. This clearly shows high level of rearrangements specific for the outbreak genotype than other *C. burnetii* strains known to date. Around 88% of these chromosomal breakpoints in NL3262 occurred due to homologous recombination of transposons (Figure 5, Supplementary Table 3).

The high number of transposase genes (IS elements) is a unique feature of *C. burnetii* genomes, compared to other obligate intracellular pathogens (*Rickettsia*, *Chlamydia*, *Mycobacterium leprae*) and are assumed to be associated with adaptation of the pathogen to different intracellular niches (Seshadri et al., 2003). These IS elements associated with genome rearrangements provide evidence for the notion that the high genome plasticity of the outbreak strains aid in an improved acclimatization of the bacteria in its intracellular niche. However, there is no direct evidence that such large-scale genomic rearrangements have impact on the functionalities of *C. burnetii*. On the other hand, genetic transposition events have been previously suggested to be involved in the generation of genetic variation and adaptation of other bacteria species in various environments (Hall, 1991). Based on the current findings and the literature, it can be hypothesized that the large number of transposons in the outbreak strains might resulted in extensive genome rearrangements in response to environmental challenges, including those in the host. Such rearrangements can result in DNA insertions/deletions, generation of polymorphisms, pseudo-genes (genes disrupted by IS elements) or modulate the gene expression by gene

order rearrangements favoring its growth and survival in different niches (Hall, 1991; Seshadri et al., 2003; Beare et al., 2006, 2009; Pallen and Wren, 2007; Rohmer et al., 2007).

In conclusion, the present study provides the first complete study on phylogenetic analysis of Dutch outbreak strains, a few additionally selected strains, and all publicly available *C. burnetii* genome sequences. Hierarchical clustering based on genome content, the *in silico* MLVA genotypes and SNP analysis, showed a genotype-specific clustering of strains, showing highly similar genomes of *C. burnetii* within each genotype. The differences between the Dutch outbreak strains relative to the reference NM and CbNL12 strains were mainly based on single nucleotide polymorphisms. We hypothesize that high numbers of transposons, genotype specific polymorphisms in membrane protein encoding genes and virulence related genes could have resulted in a highly flexible genome, an altered antigenic profile and specialized virulence mechanisms. The observed subtle genetic differences form the basis for the increased epidemic potential and the Q fever outbreak in The Netherlands. Finally, the comparative analysis of several *C. burnetii* strains from this study has the potential to provide information on cross-protective vaccine candidates. Additionally, the consistent genotype-specific genetic differences; such as SNPs and deletions could be used as additional markers for better characterization of *C. burnetii* isolate.

AUTHOR CONTRIBUTIONS

RK carried out design of study, sample collection, genome sequencing, analysis, interpretation of data and drafting the manuscript. EK was involved in data analysis. HS and HR were involved in interpretation of data and critically revising the manuscript. MS was involved in revising the manuscript. AB carried out conception and design of study, data interpretation and helped to draft the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank Frank Harders, Albert de Boer and Annemieke Dinkla for technical assistance. We also thank Dimitrios Frangoulidis from Bundeswehr institute of Microbiology, Munich, Germany and A. Rodolakis from INRA, France for kindly providing us with Henzerling, Heizberg, Scurry, Schperling and 701CbB1 *C. burnetii* strains. This work was supported by the Dutch Ministry of Economic Affairs WOT-01-002-005.06, WOT-01-002-005.05, KB-12-005.01-019, and KB-12-005.01-005.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2017.01526/full#supplementary-material>

REFERENCES

- Akporiaye, E. T., Rowatt, J. D., Aragon, A. A., and Baca, O. G. (1983). Lysosomal response of a murine macrophage-like cell line persistently infected with *Coxiella burnetii*. *Infect. Immun.* 40, 1155–1162.
- Arricau Bouvery, N., Souriau, A., Lechopier, P., and Rodolakis, A. (2003). Experimental *Coxiella burnetii* infection in pregnant goats: excretion routes. *Vet. Res.* 34, 423–433. doi: 10.1051/vetres:2003017
- Arricau-Bouvery, N., and Rodolakis, A. (2005). Is Q fever an emerging or re-emerging zoonosis? *Vet. Res.* 36, 327–349. doi: 10.1051/vetres:2005010
- Arricau-Bouvery, N., Hauck, Y., Bejaoui, A., Frangoulidis, D., Bodier, C. C., Souriau, A., et al. (2006). Molecular characterization of *Coxiella burnetii* isolates by infrequent restriction site-PCR and MLVA typing. *BMC Microbiol.* 6:38. doi: 10.1186/1471-2180-6-38
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Beare, P. A., Samuel, J. E., Howe, D., Virtaneva, K., Porcella, S. F., and Heinzen, R. A. (2006). Genetic diversity of the Q fever agent, *Coxiella burnetii*, assessed by microarray-based whole-genome comparisons. *J. Bacteriol.* 188, 2309–2324. doi: 10.1128/JB.188.7.2309-2324.2006
- Beare, P. A., Unsworth, N., Andoh, M., Voth, D. E., Omsland, A., Gilk, S. D., et al. (2009). Comparative genomics reveal extensive transposon-mediated genomic plasticity and diversity among potential effector proteins within the Genus *Coxiella*. *Infect. Immun.* 77, 642–656. doi: 10.1128/IAI.01141-08
- Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M.-A., Barrell, B. G., and Parkhill, J. (2005). ACT: the Artemis comparison tool. *Bioinformatics* 21, 3422–3423. doi: 10.1093/bioinformatics/bti553
- D'Amato, F., Eldin, C., Georgiades, K., Edouard, S., Delerce, J., Labas, N., et al. (2015). Loss of TSS1 in hypervirulent *Coxiella burnetii* 175, the causative agent of Q fever in French Guiana. *Comp. Immunol. Microbiol. Infect. Dis.* 41, 35–41. doi: 10.1016/j.cimid.2015.04.003
- D'Amato, F., Rouli, L., Edouard, S., Tyczka, J., Million, M., Robert, C., et al. (2014). The genome of *Coxiella burnetii* Z3055, a clone linked to the Netherlands Q fever outbreaks, provides evidence for the role of drift in the emergence of epidemic clones. *Comp. Immunol. Microbiol. Infect. Dis.* 37, 281–288. doi: 10.1016/j.cimid.2014.08.003
- Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704
- Deloger, M., El Karoui, M., and Petit, M.-A. (2009). A genomic distance based on MUM indicates discontinuity between most bacterial species and genera. *J. Bacteriol.* 191, 91–99. doi: 10.1128/JB.01202-08
- Denison, A. M., Massung, R. F., and Thompson, H. A. (2007). Analysis of the O-antigen biosynthesis regions of phase II isolates of *Coxiella burnetii*. *FEMS Microbiol. Lett.* 267, 102–107. doi: 10.1111/j.1574-6968.2006.00544.x
- Eenserink, M. (2010). Questions abound in Q-Fever explosion in the Netherlands. *Science* 327, 266–267. doi: 10.1126/science.327.5963.266-a
- Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* 14, 685–695.
- Georgiades, K., and Raoult, D. (2011). Genomes of the most dangerous epidemic bacteria have a virulence repertoire characterized by fewer genes but more toxin-antitoxin modules. *PLOS ONE* 6:e17962. doi: 10.1371/journal.pone.0017962
- Gilk, S. D. (2012). “Role of Lipids in *Coxiella burnetii* Infection,” in *Coxiella burnetii: Recent Advances and New Perspectives in Research of the Q Fever Bacterium Advances in Experimental Medicine and Biology*, eds R. Toman, R. A. Heinzen, J. E. Samuel, and J.-L. Mege (Springer), 199–213. Available online at: http://link.springer.com/chapter/10.1007/978-94-007-4315-1_10 (Accessed on: October 25, 2014)
- Glazunova, O., Roux, V., Freylikman, O., Sekeyova, Z., Fournous, G., Tyczka, J., et al. (2005). *Coxiella burnetii* genotyping. *Emerg. Infect. Dis.* 11, 1211–1217. doi: 10.3201/eid1108.041354
- Hall, B. G. (1991). Is the occurrence of some spontaneous mutations directed by environmental challenges? *New Biol.* 3, 729–733.
- Hammerl, J. A., Mertens, K., Sprague, L. D., Hackert, V. H., Buijs, J., Hoebe, C. J., et al. (2015). First draft genome sequence of a human *Coxiella burnetii* isolate, originating from the largest Q fever outbreak ever reported, the Netherlands, 2007 to 2010. *Genome Announc.* 3:e00445-15. doi: 10.1128/genomeA.00445-15
- Hoover, T. A., Culp, D. W., Vodkin, M. H., Williams, J. C., and Thompson, H. A. (2002). Chromosomal DNA deletions explain phenotypic characteristics of two antigenic variants, phase II and RSA 514 (crazy), of the *Coxiella burnetii* nine mile strain. *Infect. Immun.* 70, 6726–6733. doi: 10.1128/IAI.70.12.6726-2733.2002
- Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267. doi: 10.1093/molbev/msj030
- Kampschreur, L. M., Hagenaars, J. C. J. P., Wielders, C. C. H., Elsmann, P., Lestrade, P. J., Koning, O. H. J., et al. (2013). Screening for *Coxiella burnetii* seroprevalence in chronic Q fever high-risk groups reveals the magnitude of the Dutch Q fever outbreak. *Epidemiol. Infect.* 141, 847–851. doi: 10.1017/S0950268812001203
- Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome Res.* 12, 656–664. doi: 10.1101/gr.229202
- Kuley, R., Bossers-deVries, R., Smith, H. E., Smits, M. A., Roest, H. I. J., and Bossers, A. (2015a). Major differential gene regulation in *Coxiella burnetii* between *in vivo* and *in vitro* cultivation models. *BMC Genomics* 16:953. doi: 10.1186/s12864-015-2143-7
- Kuley, R., Smith, H. E., Frangoulidis, D., Smits, M. A., Jan Roest, H. I., and Bossers, A. (2015b). Cell-Free Propagation of *Coxiella burnetii* does not affect its relative virulence. *PLoS ONE* 10:121661. doi: 10.1371/journal.pone.0121661
- Kuley, R., Smith, H. E., Janse, I., Harders, F. L., Baas, F., Schijlen, E., et al. (2016). First complete genome sequence of the Dutch veterinary *Coxiella burnetii* strain NL3262, originating from the largest global Q fever outbreak, and draft genome sequence of its epidemiologically linked chronic human isolate NLhu3345937. *Genome Announc.* 4:e00245-16. doi: 10.1128/genomeA.00245-16
- Kuo, C.-H., and Ochman, H. (2010). The extinction dynamics of bacterial pseudogenes. *PLoS Genet* 6:e1001050. doi: 10.1371/journal.pgen.1001050
- Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12. doi: 10.1186/gb-2004-5-2-r12
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Lechner, M., Findeiss, S., Steiner, L., Marz, M., Stadler, P. F., and Prohaska, S. J. (2011). Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* 12:124. doi: 10.1186/1471-2105-12-124
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinform. Oxf. Engl.* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Maurin, M., and Raoult, D. (1999). Q fever. *Clin. Microbiol. Rev.* 12, 518–553.
- Mazokopakis, E. E., Karefilakis, C. M., and Starakis, I. K. (2010). Q fever endocarditis. *Infect. Disord. Drug Targets* 10, 27–31. doi: 10.2174/187152610790410918
- Merhej, V., Royer-Carenzi, M., Pontarotti, P., and Raoult, D. (2009). Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. *Biol. Direct* 4:13. doi: 10.1186/1745-6150-4-13
- Mertens, K., and Samuel, J. E. (2012). Defense mechanisms against oxidative stress in *Coxiella burnetii*: adaptation to a unique intracellular niche. *Adv. Exp. Med. Biol.* 984, 39–63. doi: 10.1007/978-94-007-4315-1_3
- Mori, M., Boarbi, S., Michel, P., Bakinahe, R., Rits, K., Wattiau, P., et al. (2013). *In vitro* and *in vivo* infectious potential of *Coxiella burnetii*: a study on belgian livestock isolates. *PLoS ONE* 8:e67622. doi: 10.1371/journal.pone.0067622
- Norlander, L. (2000). Q fever epidemiology and pathogenesis. *Microb. Infect. Inst. Pasteur.* 2, 417–424. doi: 10.1016/S1286-4579(00)00325-7
- Omsland, A., Beare, P. A., Hill, J., Cockrell, D. C., Howe, D., Hansen, B., et al. (2011). Isolation from animal tissue and genetic transformation of *Coxiella burnetii* are facilitated by an improved axenic growth medium. *Appl. Environ. Microbiol.* 77, 3720–3725. doi: 10.1128/AEM.02826-10
- Pallen, M. J., and Wren, B. W. (2007). Bacterial pathogenomics. *Nature* 449, 835–842. doi: 10.1038/nature06248
- Palmer, N. C., Kierstead, M., Key, D. W., Williams, J. C., Peacock, M. G., and Vellend, H. (1983). Placentitis and abortion in goats and sheep in Ontario caused by *Coxiella burnetii*. *Can. Vet. J.* 24, 60–61.
- Raoult, D., Marrie, T., and Mege, J. (2005). Natural history and pathophysiology of Q fever. *Lancet Infect. Dis.* 5, 219–226. doi: 10.1016/S1473-3099(05)70052-9

- Roest, H. I. J., Ruuls, R. C., Tilburg, J. J. H. C., Nabuurs-Franssen, M. H., Klaassen, C. H. W., Vellema, P., et al. (2011a). Molecular epidemiology of *Coxiella burnetii* from ruminants in Q fever outbreak, the Netherlands. *Emerg. Infect. Dis.* 17, 668–675. doi: 10.3201/eid1704.101562.
- Roest, H. I. J., Tilburg, J. J. H. C., van der Hoek, W., Vellema, P., van Zijderveld, F. G., Klaassen, C. H. W., et al. (2011b). The Q fever epidemic in The Netherlands: history, onset, response and reflection. *Epidemiol. Infect.* 139, 1–12. doi: 10.1017/S0950268810002268
- Roest, H.-J., van Gelderen, B., Dinkla, A., Frangoulidis, D., van Zijderveld, F., Rebel, J., et al. (2012). Q Fever in pregnant goats: pathogenesis and excretion of *Coxiella burnetii*. *PLoS ONE* 7:e48949. doi: 10.1371/journal.pone.0048949
- Rohmer, L., Fong, C., Abmayr, S., Wasnick, M., Larson Freeman, T. J., Radey, M., et al. (2007). Comparison of *Francisella tularensis* genomes reveals evolutionary events associated with the emergence of human pathogenic strains. *Genome Biol.* 8:R102. doi: 10.1186/gb-2007-8-6-r102
- Rolain, J. M., Vayssier-Taussat, M., Saisongkroh, W., Merhej, V., Gimenez, G., Robert, C., et al. (2013). Partial disruption of translational and posttranslational machinery reshapes growth rates of *Bartonella birtlesii*. *mBio* 4:e00115-13. doi: 10.1128/mBio.00115-13
- Rouli, L., Rolain, J.-M., El Filali, A., Robert, C., and Raoult, D. (2012). Genome sequence of *Coxiella burnetii* 109, a doxycycline-resistant clinical isolate. *J. Bacteriol.* 194:6939. doi: 10.1128/JB.01856-12
- Seshadri, R., Paulsen, I. T., Eisen, J. A., Read, T. D., Nelson, K. E., Nelson, W. C., et al. (2003). Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5455–5460. doi: 10.1073/pnas.0931379100
- Sidi-Boumedine, K., Ellis, R. J., Adam, G., Prigent, M., Angen, Ø., Aspán, A., et al. (2014). Draft genome sequences of six ruminant *Coxiella burnetii* isolates of European origin. *Genome Announc.* 2, e00285–e00214. doi: 10.1128/genomeA.00285-14
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75
- Tang, F., Bossers, A., Harders, F., Lu, C., and Smith, H. (2013). Comparative genomic analysis of twelve *Streptococcus suis* (pro)phages. *Genomics* 101, 336–344. doi: 10.1016/j.ygeno.2013.04.005
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- Tilburg, J. J. H. C., Roest, H. I. J., Nabuurs-Franssen, M. H., Horrevorts, A. M., and Klaassen, C. H. W. (2012a). Genotyping reveals the presence of a predominant genotype of *Coxiella burnetii* in consumer milk products. *J. Clin. Microbiol.* 50, 2156–2158. doi: 10.1128/JCM.06831-11
- Tilburg, J. J. H. C., Roest, H.-J. I. J., Buffet, S., Nabuurs-Franssen, M. H., Horrevorts, A. M., Raoult, D., et al. (2012b). Epidemic Genotype of *Coxiella burnetii* among Goats, Sheep, and Humans in the Netherlands. *Emerg. Infect. Dis.* 18, 887–889. doi: 10.3201/eid1805.111907
- Tilburg, J. J. H. C., Rossen, J. W. A., van Hannen, E. J., Melchers, W. J. G., Hermans, M. H. A., van de Bovenkamp, J., et al. (2012c). Genotypic diversity of *Coxiella burnetii* in the 2007–2010 Q fever outbreak episodes in The Netherlands. *J. Clin. Microbiol.* 50, 1076–1078. doi: 10.1128/JCM.05497-11
- To, H., Htwe, K. K., Kako, N., Kim, H. J., Yamaguchi, T., Fukushi, H., et al. (1998). Prevalence of *Coxiella burnetii* infection in dairy cattle with reproductive disorders. *J. Vet. Med. Sci.* 60, 859–861.
- Toman, R., Škultéty, L., Ftáček, P., and Hricová, M. (1998). NMR study of virenose and dihydrohydroxystreptose isolated from *Coxiella burnetii* phase I lipopolysaccharide. *Carbohydr. Res.* 306, 291–296. doi: 10.1016/S0008-6215(97)10037-4
- van der Hoek, W., Dijkstra, F., Schimmer, B., Schneeberger, P. M., Vellema, P., Wijkman, C., et al. (2010). Q fever in the Netherlands: an update on the epidemiology and control measures. *Euro. Surveill.* 15:19520. Available online at: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19520>
- van Schaik, E. J., Chen, C., Mertens, K., Weber, M. M., and Samuel, J. E. (2013). Molecular pathogenesis of the obligate intracellular bacterium *Coxiella burnetii*. *Nat. Rev. Microbiol.* 11, 561–573. doi: 10.1038/nrmicro3049
- Walter, M. C., Öhrman, C., Myrtenäs, K., Sjödin, A., Byström, M., Larsson, P., et al. (2014a). Genome sequence of *Coxiella burnetii* strain Namibia. *Stand. Genomic Sci.* 9, 22. doi: 10.1186/1944-3277-9-22
- Walter, M. C., Vincent, G. A., Stenos, J., Graves, S., and Frangoulidis, D. (2014b). Genome sequence of *Coxiella burnetii* Strain AuQ01 (Arandale) from an Australian patient with Acute Q Fever. *Genome Announc.* 2:e00964-14. doi: 10.1128/genomeA.00964-14
- Willems, H., Ritter, M., Jäger, C., and Thiele, D. (1997). Plasmid-homologous sequences in the chromosome of plasmidless *Coxiella burnetii* Scurry Q217. *J. Bacteriol.* 179, 3293–3297.
- Woldehiwet, Z. (2004). Q fever (coxiellosis): epidemiology and pathogenesis. *Res. Vet. Sci.* 77, 93–100. doi: 10.1016/j.rvsc.2003.09.001

Conflict of Interest Statement: 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 © 2017 Kuley, Kuijt, Smits, Roest, Smith and Bossers. 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) or licensor 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.