Genomic blueprint of a relapsing fever pathogen in 15th century Scandinavia

Meriam Guellili1, Oliver Kersten2, Amine Namouchi1, Egil L. Bauer1, Michael Derrick3, Anne O. Jensen1, Nils C. Stenseth1, and Barbara Bramanti1,5,6

1Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, N-0316 Oslo, Norway; 2Norwegian Institute for Cultural Heritage Research, N-0155 Oslo, Norway; and 3Department of Biomedical and Specialty Surgical Sciences, Faculty of Medicine, Pharmacy and Prevention, University of Ferrara, 35-44122 Ferrara, Italy

Contributed by N. C. Stenseth, August 13, 2018 (sent for review May 4, 2018; reviewed by Sally J. Cutler and Albert R. Zink)

Louse-borne relapsing fever (LBRF), once one of many fevers ravaging Europe, has disappeared from the Western world and is now endemic to only Eastern Africa. The disease is part of a group of well-known deadly louse-borne pathogens, which specialize in vector-human transmission. Its causative pathogen is the spirochete Borrelia recurrentis, whose only known vector is the human body louse, Pediculus humanus. This sets it apart from other tick-borne relapsing fever (TBRF) pathogens, such as its most closely related strains Borrelia duttonii and Borrelia crocidurae (1, 2).

Genomes of the genus Borrelia are unique. They are composed of up to 24 circular and linear plasmids with covalently closed hairpin telomeres showcasing AT-rich genomes of very small sizes. While the chromosomes, generally around 900–930 kbp, are very conserved across species, a high level of DNA rearrangements can be found among the plasmids (3–5), with some carrying essential genes, such as the telomere resolvase gene resT on plasmid p23 of B. recurrentis and B. duttonii.

The B. recurrentis genome (1.24 Mbp) is composed of one linear chromosome and seven linear plasmids and is characterized by its low GC content (mean 28.1%). Genomic and ecological data on the pathogen are scarce, since it is challenging to cultivate and studies lack an animal model (6, 7). Research on the pathogen has therefore been limited to clinical samples, resulting thus far in the publication of one deposited reference sequence, strain A1 (5), and six additional datasets from Eastern African strains (8).

LBRF is fatal in 10–40% of untreated cases (9) and is transmitted from vector to host when the hemocoe of crushed lice comes into contact with intact mucosa or skin (9). Studies (10) have also suggested a possible transmission of LBRF via lice feces. Like all relapsing fevers (RF), it is characterized by multiple febrile episodes separated by short periods of remission. Defining symptoms of LBRF are epistaxis and jaundice (9).

The disease was first mentioned in medical texts by Hippocrates from the fourth century BCE, in which he describes a series of fevers affecting the populations of Thasos after a harsh winter (11). Further references to LBRF can be found throughout European history with prominent examples being outbreaks during the Great Irish Famine of 1846–1852 (11–13) and the post-World War I pandemic (1919–1923), which is estimated to have killed more than five million people in central Europe and Russia alone (14, 15). It has also been hypothesized that the so-called “pestis flava” or “Buidhe Chonaill” of sixth century CE Ireland was an LBRF epidemic (16). LBRF frequently emerged with typhus, which is also louse-borne and caused by the bacterium Rickettsia prowazekii (12).

Today, outbreaks of the disease can be found sporadically in Ethiopia, Eritrea, Somalia, and Sudan, where it is still endemic and until recently was the fifth most common cause of death (17). In addition to recent European reports of the disease among refugees (18), which clearly highlight how quickly a decline in hygiene and living conditions can lead to the spread of LBRF, a study by Brouqui et al. (19) detected a clear increase of individuals with IgG antibodies for LBRF in 2000 and 2002. The study therefore suggests that small, undetected outbreaks of the disease still exist to this day in European populations exposed to body lice infestation.

While LBRF is the only known RF capable of reaching epidemic proportions, it was never possible to confirm the presence of ancient genomic | relapsing fever | vector-borne pathogen | immune evasion | aDNA

Significance

Louse-borne relapsing fever was one of the major diseases affecting Western human populations, with its last major pandemic killing millions after World War I. Despite the major roles fevers have played in epidemic events throughout history, molecular evidence for the presence of their etiological agent has been extremely scarce in historical samples worldwide. By comparing our medieval Borrelia recurrentis genome with modern representatives of the species, we offer an historical snapshot of genomic changes in an immune-evasion system and of reductive evolution in a specialized vector-borne human pathogen. This shotgun sequencing project highlights the potential for ancient DNA research to uncover pathogens which are undetectable to osteological analysis but are known to have played major roles in European health historically.

Author contributions: M.G. designed research; M.G. and O.K. collected data; M.G., O.K., and A.N. performed research; M.G. analyzed data; E.L.B., M.D., and A.O.J. provided archaeological/osteological information; and M.G., O.K., E.L.B., M.D., A.O.J., N.C.S., and B.B. wrote the paper.

Reviewers: S.J.C., University of East London; and A.R.Z., EURAC Research.

The authors declare no conflict of interest.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: Sequencing data have been deposited at the European Nucleotide Archive (accession no. PRJEB24501).

1To whom correspondence may be addressed. Email: meriam.guellili@ibv.uio.no, n.c. stenseth@ibv.uio.no, or barbara.bramanti@ibv.uio.no.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1807266115/-/DCSupplemental.

Published online September 24, 2018.
of the pathogen in archaeological samples. Previous attempts at detecting *B. recurrentis* in archaeological samples have been unsuccessful (20, 21) but have succeeded in detecting other louse-borne pathogens such as the typhus agent *R. prowazekii*. The only previously reported ancient DNA (aDNA) sequences of the genus *Borrelia* came from the Tyrolean Iceman Ötzi (22), who carried sequences matching the Lyme disease pathogen *Borrelia burgdorferi*.

The unique genomic output of these highly specialized bacteria allows high mapping specificity across the *B. recurrentis* genome (4, 5). This, in turn, allows us to infer the presence and absence of genomic regions via the level of coverage observed after mapping the raw datasets to *Borrelia* references. Additionally, the use of a shotgun dataset, as opposed to a target-enrichment sequencing strategy, did not restrict or shift the available data to known preselected modern sequences.

To ascertain that the organism represented in our metagenomic output was in fact a *B. recurrentis* strain and not one of its most closely related species, and potentially to detect signs of plasmid rearrangements, we mapped all datasets against available reference sequences for *B. duttonii* Ly and *B. crocidurae* Achema independently (Fig. 2C and SI Appendix, Fig. S4). The number of reads mapping to the chromosomal assemblies varies only slightly from one species to another, but we observe distinctly higher numbers of reads mapping with edit distance 0 to the *B. recurrentis* genome than to any of the other genomes (Fig. 2D and SI Appendix, Fig. S7 and Dataset S5).

Most *B. recurrentis* plasmids are colinear to *B. duttonii* plasmids with the exception of p16, which is colinear to a 5-kbp plasmid in *B. crocidurae*. While p16 is covered at a mean depth of coverage of 12.6x in our genome mapping, we clearly observed that noncolinear plasmids of *B. duttonii* and *B. crocidurae* are not present in our medieval strain (SI Appendix, Fig. S4).

Six newly published de novo genomes by Marosevic et al. (8) stemming from Eastern African refugees treated for relapsing fever seem to have reinforced previously raised doubts (25) regarding the length of plasmids p1124 and p16 (NC_011263.1). The new strains all exhibit a ca. 40-kbp extension at the 5′ end of the previously reported p1124 and a ca. 1-kbp reduction at the 3′ end of p16, much like their respective colinear plasmids p1165 in *B. duttonii* and a 5-kbp-long plasmid in *B. crocidurae*. Interestingly, our ancient *B. recurrentis* genome also displays signs of a longer p1124 and a shorter p16 (SI Appendix, Figs. S4 and S10) compared with the reference strain A1. Our mapping to p16 was completely missing 1 kbp at the 3′ end of the plasmid. Similarly, while aligning our data to p1165 of *B. duttonii*, we observed uniform mapping at a mean depth of coverage of 3x across the entire length of the plasmid, including the missing 40-kbp extension at the 5′ end. It is highly likely that these discrepancies can be linked to the deposition of an incomplete genome assembly of the original reference genome strain A1, since both our medieval European strain and the modern African strains show the same variations. This was further validated by pulse-field gel electrophoresis and the amplification of a gene ortholog to *cliC* in *B. duttonii* at 27,465–28,535 bp (26).

Upon further inspection of the plasmids, we found that three identical 582-bp-long variable short proteins (vsp) genes are missing at the 3′ end of plasmids p133, p137, and p153 (Fig. 3). Strikingly, these genes mark the end of mapping for plasmid p133 and p137, while p153 retains a hypothetical protein at its 3′ end. The first ca. 130 bp at the 5′ end of the vsp gene are covered by nonunique reads restricted to the gene across all three plasmids. This indicates that p133, p137, and p153 are shorter in strain OSL9 and that we could be seeing signs of undetected plasmid rearrangements involving a similar vsp gene. The genes missing after the 582-bp vsp genes are copies of five variable long proteins (vlp) genes and multiple hypothetical proteins. Identical copies of a 785-bp-long vlp pseudogene are missing across all three plasmids, with the copy on pl53 being only 475 bp long. While the deposited assembly of this plasmid is known to be missing parts of its telomeric sequence, it might also be missing the end of the 785-bp vlp pseudogene and one or more of the four additional identical vlp pseudogenes missing across pl37 and pl53 (Fig. 3 and Dataset S2).

Additionally, plasmids p137 and p153 have a decreased coverage in the intervals 23,822–28,471 and 39,440–44,087 bp, respectively.
The sequences are both 4,648 bp long and share 99% identity over their entire length. These intervals distinguish themselves by a patchy and low coverage, with most vlp and vsp genes in the region being covered ca. 50–80% (mean coverage 75.2%), while the covered vlp and vsp genes outside these intervals show a mean coverage of 96.9%, hinting at potential sequence degradation or plasmid rearrangements. Overall, 11 vlp pseudogenes and three vsp genes are missing, with one additional vsp pseudogene and three vlp genes showing signs of potential degradation.

We investigated the presence of known frameshift and stop-gain mutations throughout our ancient genome and found that, with the exception of oppA-1, all pseudogenes, which B. recurrentis A1 acquired during its reductive evolution and divergence from B. duttonii Ly, were present in the medieval strain. Our alignments to both B. recurrentis A1 and B. duttonii Ly support (depth of coverage 5x) that the oppA-1 gene could still be active in the OSL9 strain by retaining its ancestral glutamine (residue 59) (SI Appendix, Fig. S8).

After aligning all OSL9A-B reads to the chromosome of the reference strain, B. recurrentis A1, we detected 321 SNPs (Dataset S3). These SNPs were then combined with all other SNPs found in RF Borrelia strains included in this study to build a phylogenetic tree using the maximum likelihood method. Before building the phylogeny, we checked for the presence of recombination using a phi test and ClonalFrameML software but could not detect any sign of recombination, as these approaches yielded a P value of 0.8287 and a likelihood of −612.47, respectively. The midpoint-rooted tree was generated using PhyML. The phylogeny clearly shows two clusters represented by LBFR and TBRF species (Fig. 4), and, as previously reported, the African RF isolates cluster together (8). Interestingly, the ancient genome recovered in Oslo is clustered between B. duttonii and the African B. recurrentis strains in the phylogeny. Compared with all other Borrelia in this study, 164 SNPs are specific to the OSL9 strain. These SNPs are distributed as 77 nonsynonymous SNPs, 77 synonymous SNPs, and 10 intergenic SNPs.

**Discussion**

LBFR is characterized by multiple relapses of fever, which are believed to be caused by bacterial immune evasion systems. RF Borreliae use plasmid-encoded antigenic phase variation within a clonal population as a mechanism of immune evasion. It utilizes the sequential expression of different variants of an antigenic surface protein to evade the host immunity, prolong the infection, and promote transmission. Hence, each new wave of infection, in this case febrile relapse, is characterized by a new serotype (5, 6, 27–29). In B. recurrentis, these proteins are the surface lipoproteins vlp and vsp, which are also known to be its main proinflammatory proteins (30). The genes expressing these proteins are arranged across plasmids in silent and expressed copies, and the interaction between the host immunity and the phase-variation mechanism encoded in the pathogen’s genome are believed to be responsible for the febrile relapses of LBFR (29). The number of phase-variation loci or loci families dictates the number of possible serotypes and, thus, the number of genome variants theoretically available to evade the host immunity,
with some estimating up to $2^{10}$ different states for every 20 phase-variation sites (28, 29, 31). As a result, although LBRF can be treated with antibiotics, no vaccine against the disease exists today. The only effective antibodies against LBRF seem to be directed at vlp and vsp genes and thus only can be manufactured to fight one randomly appearing serotype at a time (32). We observed that a number of vlp and vsp genes are absent or potentially degraded in our medieval strain (Fig. 3 and Dataset S2), but are present in modern-day African strains and have orthologs in other RF genomes. Some are frameshifted or incomplete in A1 and might have been pseudogenes, but others, like the vsp copies found at the beginning of each missing interval, were fully functional genes. We can hypothesize, based on the data at hand, that the observed loss of antigenic phase-variation loci could have led to a difference in phenotype compared with modern strains and hypothetically could have influenced the number of febrile relapses, i.e., the number of serotypes that the medieval strain of the pathogen might have been capable of generating. While experimental infections by LBRF have yielded up to 10 relapses (9, 32), most sources cite up to five relapses for the pathogen (33). Similarly, most untreated non-European cases recorded during the World War I/World War II pandemics reached only up to five relapses (11, 34). Historically, Creighton (12) recorded varying numbers of observed relapses across known LBRF outbreaks. An outbreak in 18th century Dublin cites that patients were prone to relapses “even sometimes to the third” (35). Overall, most outbreaks, for which the number of observed relapses was supplied in historical texts, saw one or two or “one or more” relapses. However, it should be noted that data on modern untreated cases of LBRF are rare, and most available historical sources only refer to “multiple relapses” or “prone to relapse” (12, 36), considerably reducing the amount of data to be evaluated (36).

Compared with B. duttonii Ly, the modern reference strain B. recurrentis A1 has lost a large number of intact vlps and vsps and causes fewer relapses in human patients (5, 33), but this difference is even more pronounced in the medieval strain. The missing genes make up all archival copies of six phase-variation loci. Overall, this translates to a genome reduction of 1.2% of the pan-genome and 5.1–21% of the affected plasmids, which, in combination with a decrease in GC content (Dataset S4), is in line with observed reductive genome evolution of specialized and highly pathogenic bacteria (37, 38). The presence of additional unknown plasmids seems unlikely, as the number of plasmids tends to follow the same reductive trend as the overall genome (37). Instead, we can detect additional gene loss at the 3' end of some plasmids.

This reductive evolution characterizes B. recurrentis' epidemic potential and its increased virulence compared with other RFs. Similar to Bartonella quintana and Rickettsia prowazekii (38, 39), other well-known body lice-to-human transmission specialists, B. recurrentis shows accelerated rates of genome degradation caused by adaptation to host-restricted vectors and functional trade-offs, resulting in a degraded genome, reduced genome size, low coding content, and increased virulence (5, 31, 37, 38).

However, this type of evolution usually involves the loss of regulatory genes (37), and while this is also the case for our medieval strain, there is one exception. The oppA operon, which encodes for an ABC transporter, is significantly involved in the uptake of oligopeptides in many bacterial species (40, 41). OppA-1, which likely plays a critical role in host environment adaptation and essential metabolic functions (42), is a pseudogene in B. recurrentis A1 due to an in-frame stop codon. However, our medieval strain retains the ancestral glutamine much like the TBRF pathogen B. duttonii. While we can only speculate about the effect of this mutation on the ecological life cycle of the disease, it is interesting to note that the inactivation of oppA-1 seems to be much more recent than the rest of the pseudogenes in B. recurrentis. These pseudogenes are degraded to the same extent in both B. recurrentis lineages, with the exception of some antigenic phase-variation loci, which the medieval strain seems to have lost altogether. Therefore, we could hypothesize that, while both lineages have continued their reductive evolution, they have done so in different ways.

The full ecological dynamics of LBRF remain unclear (43). The European lineage presented in this study probably evolved in a distinctly different environment than known modern African strains. The extent to which anthropogenic pressure on pathogen and vector impacted this evolution is difficult to assess based on a single representative in the medieval lineage. As the only epidemic and louse-borne RF, LBRF has been assumed to be responsible for all RF epidemics recorded throughout European

Fig. 3. Alignment of OSL9A-B reads to B. recurrentis plasmids p33 (A), p37 (B), and p53 (C). Across all plasmids a section at the 3' end of the linear plasmids is missing starting with 80% of a vsp gene. The regions boxed in red share identities of 99–100%. Tiled reads seen within the boxed regions are nonunique reads mapping to sequences conserved across multiple plasmids.

Fig. 4. RL phylogeny based on chromosomal alignments (Datasets 53 and 59). The European LBRF strain (this study) is shown in red, and African LBRF strains are shown in blue. The legend displays the branch length.
history. It was known to have reached epidemic proportions in susceptible populations around key events, such as wars and famines. LBFR is a disease that probably would only infrequently have resulted in the type of mass mortality that would warrant the planning of mass graves, and, historically, scattered cases of LBFR have been recorded between epidemics, usually among the poor (12). Although at the time of the burial of individual OSL9/SZ50522 the town was still affected by the economic decline caused by the Black Death in the mid-1300s (44, 45), which probably left parts of the population vulnerable to disease and malnourishment, the results reported in this study represent an isolated case of the disease. Given the available data, we cannot speculate on the presence or size of an outbreak of LBFR in 15th century Oslo, especially since the number of studied samples from the same time period is limited (Dataset S1). However, our results can tentatively corroborate the involvement of *B. recurrentis* in the European epidemics reported in historical text because of its being the only known epidemic RF, the specificity of the reported symptoms, the identification of the spirochete during epidemics in 1868, and the discovery of its relation to lice in 1907 (33, 46). Finally, the studied individual probably died at the height of bacteremia, allowing the detection of the pathogen in relatively high quantities within aDNA shotgun datasets.

Previously assembled ancient pathogens have generally been limited to diseases that could leave visible marks on the skeletons or mummified tissues of affected individuals (e.g., leprosy, tuberculosis, smallpox) (47–49). However, recent studies on pathogens that are invisible in the archaeological record have started to emerge (48, 50, 51). These studies are of particular relevance, as they clearly illustrate the osteological paradox (52, 53). It is hoped that, as one of the few epidemic fevers believed to have played a major role in the disease landscape of historical Europe, the addition of LBFR will lead the way for the detection of diseases which might not have much relevance to Western health today and therefore are less represented in the genomic databases and literature. The unique genome of *B. recurrentis* has provided a rare opportunity in aDNA to study the genomic make-up of an ancient pathogen and catch a glimpse of the evolutionary process that accompanies the environmental adaption and pathogenesis of specialized human pathogens. Furthermore, the results detailed in our study illustrate the importance of human body lice as a vector throughout European history, which has recently also been suggested for plague (54).

Future research into the ecological dynamics of LBFR and surveillance of populations affected by lice infestation in industrialized countries is needed to better understand the persistence mechanism of the disease and the dangers of spread in susceptible Western communities. Furthermore, more insights into potential outbreaks of the disease in lice-infested Western populations, as seen in Brouqui et al. (19), and genomic sequences from affected patients would allow us to determine if LBFR persisted in Europe and thus is phylogenetically related to the lineage recovered in this study or was imported on one or more occasions from outside Europe. Additional genomes, European or otherwise, may further elucidate the evolution and pathogenicity of LBFR.

**Methods**

**Samples.** We sampled two well-preserved molars from nine individuals (Dataset S1) recovered from the site of St. Clement’s St. Nicolai’s church graveyard in Oslo, under clean conditions at the Norwegian Institute for Cultural Heritage Research, Oslo. One exception was tooth OSL68, which was an incisor. The individuals all stem from different periods of the graveyard, which spans the 11th to the 15th century (Dataset S1). Upon discovery of *B. recurrentis* reads in OSL9, all sampled individuals were screened for the presence of the pathogen via qPCR but were negative (SI Appendix, Fig. S3), including the only potentially contemporary individual OSL9/SZ50522. No hits for *B. recurrentis* were found in any other samples or blank controls.

**RF Borreliae Mappings.** The merged reads were mapped noncompetitively to the *Borrelia recurrentis* A1 strain (ASM1970v1), *Borrelia duttonii* Ly strain (ASM1968v1), and *Borrelia crocidurae* Achama strain (ASM25934v1) reference sequences using BWA-ALN (-n 0 -1 -16500) and BWA Samse (61). The generated data were converted to bam format and sorted using SAMtools (62, 63). Duplicates were marked using Picard MarkDuplicates module (64). Indels were realigned using GATK Realigner and IndelRealigner modules (65, 66). Using mapDamage2.0 (67), we quantified aDNA damage patterns for each dataset and recalibrated the quality scores of likely damaged bases. Libraries from tooth B yielded significantly more sequences mapping to *B. recurrentis* than libraries from tooth A, with an average 11,346 reads per dataset mapping to the reference compared with an average of 3,402 reads per dataset for tooth A.

**Phylogeny.** Species for which only contigs were available were assembled with Multi-CAR (68) using the reference sequences for *B. recurrentis* A1 and *Borrelia hermsii* HS1. For six recently published *B. recurrentis* strains (8), we mapped available Illumina reads to the *B. recurrentis* A1 chromosome assembly using BWA-MEM, sorted the data with SAMtools (62), marked duplicates using Picard (64), and realigned around indels with GATK (65, 66). For all BAM files, SNPs were called with SAMtools mpileup (-r -ugf -B) and BCFTools call (-v) and filter (-s LOWQUAL -i %QUAL > 19). SNPs were annotated using snpToolkit (69). Complete genomes were compared using Parsnp (70), and polymorphic sites were extracted using ging. Two methods, Phi test (71) and ClonalFrameML (72), were used to assess the presence of recombination in our dataset. The midpoint-rooted tree was generated using phyML.

**ACKNOWLEDGMENTS.** Data analysis was performed on the Abel Cluster owned by the University of Oslo and the Norwegian Metacenter for High Performance Computing and operated by the Department for Research Computing at USIT, the University of Oslo information technology department. Skeletal material was provided courtesy of the Museum of Cultural History (University of Oslo) after application to the museum and the Norwegian National Committee for Research Ethics on Human Remains. This project was funded by the European Research Council under FP7-IDEAS-ERC Program Grant 324249 MedPlag.