Ancient *Borrelia* genomes document the evolutionary history of louse-borne relapsing fever

Pooja Swali[1,2]*, Thomas Booth[1], Cedric C.S. Tan[2], Jesse McCabe[1], Kyriaki Anastasiadou[1], Chris Barrington[3], Matteo Borrini[4], Adelle Bricking[5], Jo Buckberry[6], Lindsey Buster[7], Rea Carlin[8], Alexandre Gilardet[1], Isabelle Glocke[1], Joel Irish[9], Monica Kelly[1], Megan King[8], Fiona Petchey[10], Jessica Peto[1], Marina Silva[1], Leo Speidel[1,2], Frankie Tait[1], Adelina Teoaca[11], Satu Valoriani[9], Mia Williams[1], Richard Madgwick[12], Graham Mullan[13], Linda Wilson[13], Kevin Cootes[4], Ian Armit[14], Maximiliano G. Gutierrez[15], Lucy van Dorp[2]*, Pontus Skoglund[1]*§

*Correspondence: pooja.swali.18@ucl.ac.uk (Pooja Swali), lucy.dorp.12@ucl.ac.uk (Lucy van Dorp), pontus.skoglund@crick.ac.uk (Pontus Skoglund)

1. Ancient Genomics Laboratory, The Francis Crick Institute, London, UK
2. UCL Genetics Institute, Department of Genetics, Evolution & Environment, University College London, London, UK
3. Bioinformatics and Biostatistics, The Francis Crick Institute, London, UK
4. Faculty of Science, School of Biological and Environmental Sciences, Liverpool John Moores University, UK
5. Department of History and Archaeology, Amgueddfa Cymru/National Museum of Wales, Cardiff, UK
6. Department of Archaeology and Forensics, University of Bradford, Bradford, UK
7. School of Humanities and Educational Studies, Canterbury Christ Church University, Canterbury, UK
8. Freelance Consultant Osteoarchaeologist, Poulton Research Project, Cheshire, UK
9. School of Biological and Environmental Sciences, Liverpool John Moores University, Liverpool, UK
10. Environmental Science, University of Waikato, New Zealand
11. Canterbury Archaeological Trust, Canterbury, UK
12. School of History, Archaeology and Religion, Cardiff University, Cardiff, UK
13 University of Bristol Spelaeological Society, University of Bristol, Bristol, UK
14. Department of Archaeology, University of York, UK
15. Host-Pathogen Interactions in Tuberculosis Laboratory, The Francis Crick Institute, London, UK

§Lead Contact

Summary

Several disease-causing bacteria have transitioned from tick-borne to louse-borne transmission, a process characterised by increased virulence and genome reduction. However, the historical time frame and speed of such evolutionary transitions have not been documented with series of ancient genomes. Here, we discover four cases of *Borrelia recurrentis*, the causative agent of louse-borne relapsing fever, in Britain between ~700 and 2,300 years ago, and sequence whole

genomes up to 29-fold coverage. We estimate a recent divergence from the closest tick-borne ancestor, likely within the last ~8,000 years. We reconstruct a chronology of gene losses and acquisitions using the pan-genome of related species, and show that almost all of the reductive evolution observed in *B. recurrentis* had occurred by ~2,000 years ago, and was thus a rapid process over just a few millennia since divergence. Our observations provide a new understanding of the origins of *B. recurrentis* and documents complex reductive evolution in a specialist vector-borne pathogen.

Key words: infectious disease, genomics, ancient DNA, phylogenetics

Introduction

Several species of bacteria have undergone an evolutionary process of transitioning from tick-borne to louse-borne transmission, including the trench fever agent *Bartonella quintana*, the epidemic typhus agent *Rickettsia prowazekii*, and the agent of louse-borne relapsing fever (LBRF) *Borrelia recurrentis*. All species show a pattern of higher virulence in the louse-borne agent compared to their respective closest tick-borne relatives, and all show a striking evolutionary pattern of genome reduction[1], possibly facilitated by specialisation to the louse vector[1–3]. However, the evolutionary time frame and genomic basis of the transition from tick-borne to louse-borne transmission, and the drivers of increased virulence, remains largely unknown.

Relapsing fever, named after the recurring fevers it induces, is caused by several species of *Borrelia*, mostly spread by soft-bodied ticks, with the exception of *Borrelia miyamotoi* which is spread by hard-bodied ticks (Tick-borne relapsing fever; TBRF)[4] and *B. recurrentis. B. recurrentis* is transmitted from human to human by the bite of an infected human body louse, *Pediculus humanus humanus*[5] and is not known to have an animal reservoir. The bacteria establishes infection when the haemocoel of the infected vector is able to penetrate the mucosa and skin barrier[5]. In contrast to LBRF, TBRF is zoonotic, with multiple animal reservoir hosts, and can be found worldwide. For example, the closest related bacterial species to *B. recurrentis*, *Borrelia duttonii*, was previously found in animal reservoirs such as pigs and chickens and is today mostly found in East Africa[6,7].

The present-day *B. recurrentis* genome has an unusual genome structure, comprising a ~930 kb linear chromosome and seven linear plasmids ranging from 6-124 kb in length[8]. Though the chromosome is fairly conserved over the *Borrelia* genus, the plasmids have potential to undergo extensive rearrangements[3]. *B. recurrentis* has been estimated to have lost approximately a fifth of its genome relative to its sister species *B. duttonii*, with prominent gene loss occurring on plasmids, possibly due to the accumulation of impaired genes[3]. While it has been suggested that genome loss in the other louse-borne taxa in *Rickettsia* and *Bartonella* was primarily via elimination of inactivated genes[1,3], the exact genes involved in the mechanism for vector specification (louse or tick), and the evolutionary processes underlying genome degradation, in *Borrelia* remain unclear.

Major uncertainties surround the past and present epidemiology of *B. recurrentis* and hence the timeline over which genome reduction and vector/host specialisation have occurred. Throughout European history there have been numerous references to episodes of “epidemic fever”, and “fever lasting six or seven days, with multiple relapses”[9,10], with the earliest descriptions dating back to the 5th century BCE[10]. It has been hypothesised that LBRF was the agent of the Yellow Plague which affected Europe in 550 CE, the episodic fevers which became known as sweating sickness in northwestern Europe between 1,485-1,551CE[11,12], as well as fevers which accompanied famines in Ireland through the 17th and 18th centuries CE. However the specific agents of these historical outbreaks have not been confirmed. LBRF posed major challenges to public health during World War I and World War II, before mostly disappearing from Europe at the end of the 19th century CE. Today, LBRF remains a major cause of morbidity and mortality in Ethiopia (where it is endemic), Somalia and Sudan[13,14]. While some now consider LBRF as a neglected tropical disease (NTD), it is clear that LBRF shows the potential to re-emerge during times of overcrowding, poor access to sanitation and hygiene, and during times of conflicts and disasters[3,15].

An archaeogenetic approach to past infections represents one of the most promising tools for characterising the wider diversity and evolution of *B. recurrentis* given that there is very limited genomic data from present-day infections, and that *B. recurrentis* is challenging to grow in culture[16]. Previously, a ~550-year-old (1,430–1,465 cal. CE) 6.4-fold coverage ancient *B. recurrentis* genome was recovered from a tooth taken from a human skeleton buried in medieval Oslo, Norway (OSL9)[17]. However, we lack an understanding of the deeper genomic evolution of *B. recurrentis*, or its prevalence in Europe across time and space. Here we provide four new ancient *B. recurrentis* genomes from Britain spanning 2,300-600 years ago, from the Iron Age to the later medieval period. Leveraging these observations we confirm the contribution of *B. recurrentis* to disease in European history and document its complex evolutionary behaviour during the transition to louse-borne transmission.

Results

Detection and authentication of four ancient *B. recurrentis* genomes

We used an ancient DNA approach, including single-stranded DNA library preparation[18] which optimises retrieval of short fragments and allows cytosine deamination derived errors to be removed *in-silico*, to generate whole *B. recurrentis* genomes from four human skeletons recovered from individuals located at four archaeological sites in Britain (**Figure 1A**). We generated ~0.8 - 8.5 billion read pairs to obtain an overall X-fold genome coverage over the *B. recurrentis* A1 reference chromosome ranging from 0.8-29.4 (**Table 1**), with all libraries sequenced to either >20-fold coverage or >40% clonality (proportion of sequences identified as PCR duplicates, indicative of library saturation). These include two observations dating to the last ~2,000 years; an 11.2-fold genome (C10416) from the ‘Arras-style’ Iron Age cemetery at Wetwang Slack, East Yorkshire, contextually dated to 2,300-2,100 years ago (300-100 BCE)[19], and a 3.5-fold *B. recurrentis* genome from Fishmonger’s Swallet (C13361), a cave in the Mendip Hills, southern Gloucestershire, UK. The individual has been directly radiocarbon dated to 2,185-2,033 years ago (162 cal BCE-10 CE; 2063±28bp, BRAMS-5059)[20]. We also generated a

0.8-fold *B. recurrentis* genome from the tooth of an unprovenanced skull, from the collections of the Canterbury Archaeological Trust which is likely to be from Canterbury or the surrounding region and date to 2,000-700 years ago (1st-14th centuries CE). Finally, we generated a 29.4-fold *B. recurrentis* genome (C10976) from a rural cemetery site associated with a medieval chapel next to the village of Poulton, near Chester in Cheshire. The skeleton has been radiocarbon dated to 733-633 years ago (1,290-1,390 cal CE, 646±14bp, Wk 52986; Cootes et al. 2023). Further details of all individuals sampled in this study are available in the **Supplementary Information.**

Aligned sequences were confirmed to be authentic through assessment of evidence for cytosine deamination[21,22], distribution of number of mismatches across sequences (edit distance), even coverage across the genome and a unimodal fragment length distribution (**Figure 1B and Figure S1**)[23]. Additionally, all identified genomes were aligned to representatives of the more distantly related species *B. duttonii* Ly and *Borrelia crocidurae* DOU reference strains to confirm that all identified cases were genetically closer to the modern-day *B. recurrentis* A1 genome than other related *Borrelia* species (**Figure S2**). Mapping was also conducted to the wider *Borrelia* plasmid complement. The presence of the pl6 plasmid in these ancient British genomes, which is absent in *B. duttonii* Ly, further supports their taxonomic identification as *B. recurrentis*.

Iron Age and medieval lineages of *B. recurrentis*

To evaluate the relatedness of our ancient observations to contemporary sampled strains, we initially reconstructed a *Borrelia* phylogeny including the closest *B. duttonii* representative (Ly). Consistent with the assessment of our samples being *B. recurrentis*, all ancient genomes form a monophyletic clade together with present-day *B. recurrentis* (**Figure S3 and Figure S4**). Our medieval genome, C10976 Poulton, is positioned in a subclade with the previously published medieval genome from Norway OSL9[17]. All Iron Age genomes from Britain fall basal to this clade. Among the Iron Age genomes, C13361 Fishmongers falls on a lineage basal to C10416 Wetwang Slack despite being dated to a similar period. Although the Fishmongers genome is lower coverage (2.6-fold when aligned to the core genome), the 100% bootstrap support for this phylogenetic placement suggests the possibility of synchronic sister lineages of the species existing in Britain ~2,300-2,000 years ago. Additionally, we reconstructed a phylogeny on an alignment built using relaxed SNP filtering thresholds in order to include the lower coverage (0.8-fold) C11907 Whitefriars genome (**Methods**) (**Figure S4**). We found this genome was closely related to other Iron Age genomes, consistent with the likely chronological age range of the sample. C11907 Whitefriars was then excluded from further analysis due to its lower coverage (0.8-fold) and uncertain chronological date.

We next reconstructed a core gene alignment to assess the extent to which recombination and accessory (plasmid) gene content may influence our reconstructed relationships by identifying a set of genes shared amongst the modern sampled diversity of *B. recurrentis*. To do so, we applied the pan-genome analysis tool Panaroo[24] to all modern *B. recurrentis* (seven genomes), *B. duttonii* (two assemblies) and *B. crocidurae* (two assemblies) genomes (**Table S1**). The inferred *Borrelia* pan-genome comprised a total of 3,035 genes corresponding to a length of

2,223,831 base pairs. We observed a high degree of conservation within the *B. recurrentis* species, supporting a limited intraspecies pan-genome despite high plasmid carriage. Of these genes we identified 933 as being present in 99% of included strains, providing a core gene reference panel to which the ancient samples were aligned before phylogenetic reconstruction. Phylogenies constructed by mapping to the *B. recurrentis* A1 reference genome and core-genome alignment showed identical phylogenetic topologies, though with the latter providing a higher amount of reference diversity with improved bootstrap phylogenetic support (4,192 SNPs versus 4,200 SNPs). This suggests a limited impact of mapping bias or structural variation on our observed patterns of relatedness.

Given recombination may violate the assumptions of tree-building representations of diversity we additionally formally tested for evidence of homologous recombination using *ClonalFrameML*[25]. We observed very limited evidence of homologous recombination, with only a small fraction of the genome (<0.1%) estimated to derive from such processes Nonetheless, after pruning the alignment for the modest amount of recombination detected by ClonalFrameML, we recovered a topologically identical phylogeny (**Figure S5**).

Chronology of the divergence from the tick-borne sister species

The timeline over which *B. recurrentis* diverged from its common ancestor with *B. duttonii* and subsequently evolved vector and host specificity is uncertain. We evaluated the presence of a signal of clock-like evolution in our core-genome alignment. We formally assessed the temporal structure in our recovered phylogenies using *BactDating*[26] to test for a significant correlation between genomic diversity and sampling time using date randomisation (**Methods**). We assessed temporality including and excluding C13361 Fishmongers due to its lower coverage and obtained an R^2 value of 0.69 (p-value 0.00080) and 0.66 (p-value 0.0011) respectively, suggesting significant temporal signal across our dataset (**Figure S6, Table S2**).

We implemented formal Bayesian tip-dating calibration via *BEAST2*[27] to provide a probabilistic assessment of the divergence of sampled *B. recurrentis* from the closest sequenced relative *B. duttonii* Ly, together with an estimate of the rate of mutation over the non-recombining fraction of the core alignment (**Methods**)**.** Evaluating a suite of possible clock and demographic models, we estimate a split time of *B. recurrentis* genomes over 5-fold in coverage in our dataset from *B. duttonii* Ly ranging from 2,215-5,630 years ago. When the lower coverage C13361 Fishmongers sample is included, we estimate the split time from *B. duttonii* Ly ranges from 2,313-7,654 years ago, overlapping with the initial estimate (**Figure 2A and Figure S7, Figure S8, Table S3**). The posterior distributions of the best supported model indicate a divergence estimate of 4,724-5,630 years ago, corresponding to a rate of evolution of 5.0×10^{-8} (4.6×10^{-8} - 5.5×10^{-8} substitutions per site per year; 95% HPD values) (**Figure 2A**). Estimates from this model suggest an emergence of the Iron Age clade between 2,326-2,410 years ago with the medieval clade, including C10976 Poulton and the previously recovered OSL9 genome, dating to within the last 700 years. Our inference would also suggest a very recent emergence (46-69 years ago) of all contemporarily sampled *B. recurrentis* infections (exclusively from Africa or linked to refugee status), noting that both ancient and modern diversity are likely markedly undersampled. The divergence date from *B. duttonii Ly* reported in our temporal analysis is

consistent with a preprint study providing low-coverage observations of *B. recurrentis* dating back to the Late Neolithic ~4,600-year-old[28] (**Figure 2B**).

Patterns of pan-genome diversity and genome reduction in *B. recurrentis*

We used a pan-genome approach to establish patterns of gene content across *Borrelia* (**Methods**). Consistent with previous observations[3], we note variability in the number of genes per species, though identify that *B. recurrentis* contains ~25% less genes than *B. duttonii,* with the least variance in gene content in the pan-genome of any relapsing fever species in the genus, consistent with its hypothesised niche constraint[29] (**Figure S9, Figure S10**). This result was consistent when directly aligning to the *B. duttonii* Ly reference genome, where we estimate that during the medieval period, the high coverage C10976 Poulton genome covered ~82% of the *B. duttonii* reference (the previously published genome OSL9 covered ~75%, which may be potentially underestimated due to lower coverage) and the older Iron Age genome, C10416 Wetwang Slack, ~81%. This demonstrates that almost all of the reductionist evolution seen in *B. recurrentis* had already occurred by the Iron Age, ~2,000 years ago.

We assessed the contribution of plasmid carriage relative to *B. duttonii*, *B. crocidurae* and contemporary *B. recurrentis* (**Figure 3**). We identify three plasmids (pl26, pl27, pl28), described in *B. duttonii Ly*, that are present in the Iron Age genomes—authenticated using coverage, cytosine deamination patterns and distribution of mismatches—but are absent or at substantially lower coverage in medieval and present-day genomes (**Figure S11, Table S4**). We therefore suggest at least partial plasmid loss events, or loss of significant plasmid-borne elements ~2,000-700 years ago, between the Iron Age and the medieval lineages. Overall, by the medieval period *B. recurrentis* harboured the full suite of plasmids observed in currently sampled infections. In contrast, the Iron Age genome shows only partial coverage of *B. recurrentis* plasmid pl53, which is present in medieval and modern genomes, suggesting the complete plasmid gene complement was acquired after ~2,000 years ago (**Table 1, Figure 3**).

We next assessed gene variation, agnostic to chromosomal or plasmid affiliation, captured by our ancient *B. recurrentis* genomes given they may represent possible intermediates on the trajectory towards specialisation. To begin with, we assessed the full gene repertoire across all *B. recurrentis, B. duttonii and B. crocidurae* species identifying a total of 3035 unique genes defining the pan-genome which were then subject to presence/absence assessment, filtering and annotation (**Methods**). Given hypotheses surrounding genome reduction and plasmid stability, we particularly noted two genes in the pan-genome implicated in plasmid segregation and partitioning, *Soj* and *ParA* (annotated as *Soj_1* and *ParA_1*), which showed a temporal patterning across our dataset (**Table S5**, **Figure S12**). *Soj_1* is present in both *B. crocidurae* and *B. duttonii* and was also present in the ancient Iron age *B. recurrentis* genomes of sufficient coverage for consideration, but absent in medieval and present-day data. Assuming parsimony, this suggests gene loss on the branch leading to the descendant medieval and contemporary clades (we estimate phylogenetically, based on current data, between 2,326 - 1,115 years ago; 95% HPD). Contrastingly, *ParA*, which is an ortholog of the *Soj* gene[30] also implicated in plasmid segregation, is present solely in the medieval and modern-day genomes. A parsimonious explanation is the acquisition of this gene between the Iron Age and medieval period.

Exploration of the genomic neighbourhoods of both the *ParA_1* and *Soj_1* genes (**Methods**), further supports that these genes are today localised on different backgrounds, with *Soj_1* found on the *B. duttonii Ly* chromosome and *ParA_1* on the pl53 *B. recurrentis* plasmid.

Temporal variation in functional genes

Borrelia relapsing fevers use a suite of plasmid-encoded antigenic phase variation as a mechanism of immune evasion. This is mediated by variable large proteins (vlp) and variable short proteins (vsp), together known as the variable major proteins (vmp). Antigenic variation of these surface-exposed lipoproteins likely plays an important role in evading host-acquired immunity, and allowing for the bacteria to persist within its host population, though it is unclear how stable this mechanism has been through evolutionary time[31–33]. We found that the medieval lineage, comprising C10976 Poulton and OSL9, both had similar vmp profiles, with three of the four chromosomal borne vmps characteristic of present-day *B. recurrenti*s gained by medieval times (**Figure 3**). However, two of these *vlp* genes have been identified as pseudogenes in modern *B. recurrentis* genomes, and so despite being gained in *B. recurrentis* these are believed to be functionally redundant in modern genomes.

When evaluating vmp profiles over the *B. recurrentis* plasmids, we noted the absence of a number of *vmp* genes, many of which are pseudogenes, in the medieval genomes, in particular at the 3' ends of the pl33, pl37 and pl53 plasmids (**Figure 3**); an observation also observed in the previously published OSL9 genome from the same time period[17]. Interestingly, these regions (5' pl35, pl37 and pl53) are mostly present in the Iron Age sample C10416 Wetwang Slack but absent in our basal Iron Age sample C13361 Fishmongers. Though the exact gene complement of plasmid-borne vmp genes is difficult to formally assess in this case due to the lower overall coverage of this genome, this observation is consistent with the potential for interspecies variability in vlp profiles, as is seen in tick-borne relatives (*B. crocidurae* and *B. duttonii*).

Evaluating other hallmarks of infective behaviour[3], we note that BDU 1, a p35-like antigen implicated in fibronectin binding in *B. burgdorferi* (Lyme disease)[3,34], and an outer membrane protein, and which is absent in *B. recurrentis*, is also absent across our Iron Age and medieval observations supporting an early loss following the split with *B. duttonii Ly (***Figure S12)**. A similar pattern was seen for several further outer membrane proteins, including those involved in host complement system inactivation (BDU 2, BDU 3, BDU 5)[32], which are also absent in *B. recurrentis* by our earliest observation ~2,000 years ago. We do however observe some temporal variation suggesting an ongoing process of genome adaptation, for example the aforementioned loss of *Soj* (annotated as BDU 429) and truncation of an uncharacterised protein BDU 430 in the modern *B. recurrentis* genome, which are also absent/truncated in the medieval genomes but present in the Iron Age genomes.

Finally, we assessed SNPs and indels of functional relevance (**Figure S13**)[3]. For example, the *oppA-1* gene is a pseudogene in *B. recurrentis* A1 due to an in-frame stop mutation, but in *Borrelia burgdorferi*, it is shown to play an essential role in metabolic function as well as survival in different environments[35]. As with the previously reported medieval genome from OSL9[17], this gene is found in its ancestral form in all of our ancient samples. This suggests the inactivation of

oppA-1 occurred relatively recently, we estimate within the last ~1,115 years (**Figure 2**). Similarly, the *smf* gene and the *mutS* gene show an in-frame stop mutation and a frameshift mutation in present-day *B. recurrentis*. All ancient genomes with data at this locus show the in-frame stop mutation in the *smf* gene, with the higher coverage genomes (C10416 Wetwang Slack and C10976 Poulton) also supporting the presence of the *mutS* frameshift mutation. Conversely, the *recA* gene is still functional in the high coverage Iron Age genomes. Hence, the true recombination efficiency is unknown for these genomes even if we find little detectable signal of recombination in our genomics analysis (**Figure S5**).

Discussion

Here we reconstruct the complex evolutionary history of *B. recurrentis* by retrieving and analysing four ancient *Borrelia* genomes from Britain across a ~1,500-year time span. Our work confirms the presence of the pathogen in Europe in both the Iron Age and later medieval periods, while extending *B. recurrentis* whole genome observations by ~1,600 years, adding markedly to existing data from the species, with only eight genomes (seven contemporary and one medieval) available prior to this study. While it is unclear whether there is any link between our detected infections and historically attested outbreaks in Britain, we note that the high coverage recovery of genomes achieved here suggests that the individuals studied likely died from acute infections with high levels of bacteremia.

In the context of published datasets, our work confirms the existence of a closely related medieval phylogenetic clade which existed from at least 600 years ago and spanned Britain and the Scandinavian peninsula. In addition, we recover previously unknown ~2,000-year-old basal lineages from Iron Age Britain. The phylogenetic placements of the Iron Age genomes may suggest that multiple lineages of *B. recurrentis* existed at this time. Harnessing this temporal structure we find support for a relatively recent divergence of sampled *B. recurrentis* strains from the closest relative *B. duttonii* Ly, to within the last ~8,000 years (across models and datasets). While we caution that a more complete time series could improve these calibrations and divergence estimates, they estimates are consistent with a late Neolithic/Early Bronze Age emergence of the agent of LBRF. This could for example have been linked to changes in human lifestyles which were more favourable to the human body louse vector, for instance gradually increasing levels of sedentism during the development of agriculture, as well as the emergence of densely-occupied mega-settlements in regions of Eastern Europe specifically[36].

To date, no *B. recurrentis* whole genomes have been identified earlier than these estimates. Sikora *et al*[28] report in an unpublished preprint 31 observations of *B. recurrentis* in the ancient skeletal record based on the assignment of sequencing reads from human skeletons, mostly from Eurasia (**Figure 2B and Figure S14**), all of which are lower than 0.17-fold genome coverage. The earliest recovered observation is from Denmark dating to ~4,600BP (RISE61; 0.003X)[28,37]. Due to the low coverage of these *B. recurrentis* observations through time, it was not possible to incorporate these observations into our phylogenetics analysis, though further sequencing of these or other ancient samples may prove vital in establishing the spatio-temporal timeline of infections as well as potentially uncovering previously unseen genomic diversity. Contemporary infections are also undersampled, with limited available

genomic data mostly linked to cases in East Africa or associated with migrants on their journey to Europe[38]. As such it is challenging to disentangle many of our ancient observations from the expectation of population structure between African and European strains. Further data, from different time periods and geographic regions of the world, may result in our temporal estimates being amended and would be required to assess any connection between the population structure of the bacteria and human host movements.

Our estimates, over the non-recombining fraction of the genome, nonetheless indicate a fairly recent timeline over which presently sampled *B. recurrentis* diverged from its closest sequenced relative, *B. duttonii Ly*. We note that some have suggested that *B. recurrentis* is a degraded form of *B. duttonii* and hence the lines between species demarcations may well have been blurred in deeper history. Aside from undersampling, we can also not exclude the possibility of rate variation in the history of *B. recurrentis*. While a relaxed clock model was less well supported by our Bayesian phylogenetic analyses, it is plausible that ecological influences on mutation rate; as observed in other bacterial pathogens, may have played a complex role difficult to capture by our temporal reconstruction approach[39,40].

Nonetheless, our work indicates the need for rapid genome decay in order for *B. recurrentis* to exhibit ~20% reduction in its genome relative to *B. duttonii Ly* within ~8,000 years; an evolutionary behaviour which has been linked to specialism to the human body louse vector and potentially resulting in enhanced pathogenicity[41]. Strikingly, we observe that much of this decay had already occurred by the time of our oldest sampled infection ~2,000 years ago. It has been suggested that this accelerated evolution, particularly in light of the slow global mutation rate, may have been supported by plasticity in the wider accessory genome of *B. recurrentis*. It remains unclear precisely to which extent the genome size difference between the two species differs due to the reductive evolution in B. recurrentis, as we also observe some unique gains in the *B. duttonii* Ly genome. However, aided by the high coverage Iron Age and medieval *B. recurrentis* genomes, we can demonstrate that at least some of the decay towards the extant *B. recurrentis* genome was ongoing between these periods. In particular, we identify the partial loss of three plasmids (or extensive plasmid-borne elements) between *B. duttonii* and our Iron Age samples and later medieval and contemporary *B. recurrentis* strains; an event we estimate as most likely to have occurred between 2,326 and 1,115 years ago.

The extent to which such loss events may still be ongoing is unclear, though it has been suggested that disruption in plasmid partitioning genes relative to *B. duttonii* homologs may indicate a degree of ongoing reductive process[3]. Indeed, we note an interesting temporal patterning of major chromosomal and plasmid partitioning genes *Soj* and *ParA*, best described for their role in *Bacillus subtilis* and *Escherichia coli*[42]. While it has previously been reported that *B. recurrentis* lacks a chromosomal *Soj* homologue[3], using a pan-genomics approach we identify that *Soj* was retained until at least the Iron Age period before being lost by the time of the earliest of our two medieval observations. Intriguingly we simultaneously reconstruct the gain of *ParA*, a distant homolog of *Soj*. Such observations highlight the fluidity of the process of genome reduction, with the suggestion of necessary acquisition of some functionally relevant genes as a likely outcome of large-scale plasmid loss events. Another plausible contributor to

the pattern of genome decay is the loss of DNA repair mechanisms, with the inactivation of genes such as *recA* resulting in the bacteria becoming dependent on its human/vector hosts[8]. Our data suggests that the loss of *recA*, as a DNA repair mechanism, may also be a reasonably recent event, given we find the *recA* gene is still functional in our Iron Age observation. We however note that both *smf* and *mutS*, also implicated in DNA repair, are disrupted across our ancient samples, supporting the importance of this mechanism in the wider propensity for genome loss.

The transition to host specialised pathogen from a more generalist ancestor will also have exerted a selective pressure on the bacteria. Within *Borrelia*, the *vmp* and *vmp*-like genes offer an important mechanism to allow persistence and resurgence of relapsing fevers, with antigenic variation during infection achieved through a process of silent *vmp* genes being transferred to the expression locus leading to the generation of new surface protein variants[31–33]. Our work supports some temporal variability within *B. recurrentis*, particularly in the *vmp* genes located at the 3' end of the pl33, pl37 and pl53 plasmids, which we observe as absent in medieval samples. While this may represent a change in antigenic behaviour, it is also plausible that the pl33, pl37 and pl53 plasmids are shorter or subject to genomic rearrangements in these strains, an observation also suggested by Guellil and colleagues who detected a similar patterning in the only other ancient *B. recurrentis* full genome published prior to this study[17]. We also note far richer diversity in vmp profiles in non-*recurrentis* species indicating the extent of antigenic plasticity may have been very different prior to host specialisation.

Together we highlight how ancient microbial DNA can be used to enhance our understanding of the age and diversity of significant but understudied pathogens. Our work highlights the value of temporal data in pinpointing the timing and patterning of the process of host/vector specialisation; supporting a prevailing background of accelerated genome reduction notwithstanding more recent key instances of gene gains and losses. We can however, only speculate on whether these ancient bacteria from Britain were adapted to be louse-borne or tick-borne form of relapsing fever. Additional work is required to build a mechanistic understanding of the genomic basis for each vector niche.

Acknowledgements
We thank the Advanced sequencing facility at the Francis Crick Institute for technical support. This work was supported by the Vallee Foundation, the Wellcome Trust (217223/Z/19/Z), and Francis Crick Institute core funding (FC001595) from Cancer Research UK, the UK Medical Research Council, and the Wellcome Trust. P. Skoglund. was supported by the European Molecular Biology Organisation, and the European Research Council (grant no. 852558). LvD was supported by a UKRI Future Leaders Fellowship MR/X034828/1 and European Union END-VoC consortium grant (agreement no. 101046314). P. Swali was supported by the Francis Crick Institute core funding (FC001595 to P. S.) and a UKRI Future Leaders Fellowship grant (MR/X034828/1 to L. vD.). [Insert all coauthor grants with initials of coauthor]. We are grateful to the University of Bristol Spelaeological Society, the University of Bradford, Hull and East Riding Museum and Canterbury Archaeological Trust and the Poulton Archaeological Trust for allowing

access to collections and facilitating sampling [Please add any other relevant funding]. We would also like to thank Meriam Guellil and Marina Escalera for help and advice in the analysis.

Author contributions

Conceptualisation: P. Swali, L.vD, P. Skoglund, Software: P. Skoglund, Data processing and curation: P. Swali, C.B. A.G., Formal analysis: P. Swali, L.vD, Visualisation: P. Swali, Investigation: P. Swali., T.B., C.T., J.M., K.A., C.B., A.G., I.G., M. Kelly., J.P., M.S., L.S., F.T.,M.W., L.vD., P. Skoglund, Resources: M.B., A.B., J.B., L.B., R.C., G.M., R.M., J.I., M.King., F.P., J.P., A.T., S.V L.W., K.C., I.A, Supervision: L.vD., P. Skoglund. Writing—original draft: P. Swali, L.vD, P. Skoglund. Writing – review & editing: P. Swali, T.B, C.T, L.vD, P. Skoglund

Data Availability

All sequence data will be available in the European Nucleotide Archive upon publication.

Code availability

Code to reconstruct a consensus sequence using strand-aware removal of cytosine-deamination-derived errors is available at (https://github.com/pontussk/mpileup2consensus.py/).

Declaration of Interests

The authors declare no competing interests.

FIGURE AND TABLES (Main Text)

Table 1. Sequencing metrics for the four individuals recovered in this study when mapped to the *B. recurrentis* A1 reference genome (chromosome and plasmids) requiring a mapping quality of MQ1.

Individual	C10416 Wetwang Slack	C10976 Poulton	C13361 Fishmongers	C11907 Whitefriars

Archaeological dates		2250-2050 years ago (context)	733-633 years ago (C14 dating)	2185-2033 years ago (C14 dating)	2000-700 years ago (context)
Individual		C10416 Wetwang Slack	C10976 Poulton	C13361 Fishmongers	C11907 Whitefriars
Total sequences		8,453,668,864	841,565,272	2,615,551,750	2,040,000,513
Sequences aligned to *B. recurrentis* A1 (Chromosome and accessory genome)		526,415	888,736	139,860	164,451
Sequences aligned after duplicate removal		264,871	715,574	82,053	18,372
Clonality (%)		49.7	19.5	41.3	88.8
Proportion *B. recurrentis* (%)		0.0062	0.1056	0.0053	0.0081
X-fold coverage when aligned to *B. recurrentis* A1 chromosome and plasmids (overall Q1)		10.2	28.6	3.1	0.7
Breadth of coverage >1x (%)		95.6	98.7	85.2	37.1
Breadth of coverage >2x (%) [all]		93.1	98.6	65.9	13.0
Breadth of coverage >3x (%) [all]		89.3	98.4	44.5	4.6
Chromosome (NC_011244.1)	**X-fold coverage**	11.2	29.4	3.5	0.8
	Breadth of coverage >1x (%)	99.6	100.0	90.4	40.3
Plasmid pl6 (NC_011263.1)	**X-fold coverage**	18.8	57.4	6.5	2.1
	Breadth of coverage >1x (%)	85.9	86.1	85.4	65.0
Plasmid pl23 (NC_011252.1)	**X-fold coverage**	13.7	48.2	4.2	1.0
	Breadth of coverage >1x (%)	90.6	99.7	87.6	52.7
Plasmid pl33 (NC_011253.1)	**X-fold coverage**	6.6	16.5	1.7	0.6
	Breadth of coverage >1x (%)	95.1	94.4	69.3	31.9
Plasmid pl35 (NC_011255.1)	**X-fold coverage**	13.1	40.9	3.7	1.1
	Breadth of coverage >1x (%)	96.9	100.0	92.5	55.2
Plasmid pl37 (NC_011258.1)	**X-fold coverage**	3.1	14.8	0.7	0.2
	Breadth of coverage >1x (%)	61.1	80.4	34.8	14.7

Plasmid pl53 (NC_011260.1)	**X-fold coverage**	2.9	22.5	0.9	0.2
	Breadth of coverage >1x (%)	45.3	86.1	34.1	9.0
Plasmid pl124 (NC_011246.1)	**X-fold coverage**	6.5	23.4	2.0	0.3
	Breadth of coverage >1x (%)	98.0	100.0	82.9	22.8
To Core Genome Alignment	**X-fold coverage**	9.4	27.4	2.6	0.5
	Breadth of coverage >1x (%	96.9	98.0	87.2	36.7

Figure 1. A. Geographic location of the four ancient *B. recurrentis* genomes sequenced in this study together with OSL9 previously published by Guellil and colleagues. B. Circos plot with the coverage of ancient genomes across the *B. recurrentis* chromosome and plasmids when

aligned to the *B. recurrentis* A1 reference genome (GCF_000019705.1). A window size of 100bp for the chromosome and 10bp for the plasmids was used to provide the normalised coverage per window plotted. To allow for visualisation, the coverage for each genome was scaled by the maximum coverage per genome (C10416 Wetwang Slack, 70; C13361 Fishmongers, 20; C11907 Whitefriars, 10; C10976 Poulton, 170; OSL9, 40).

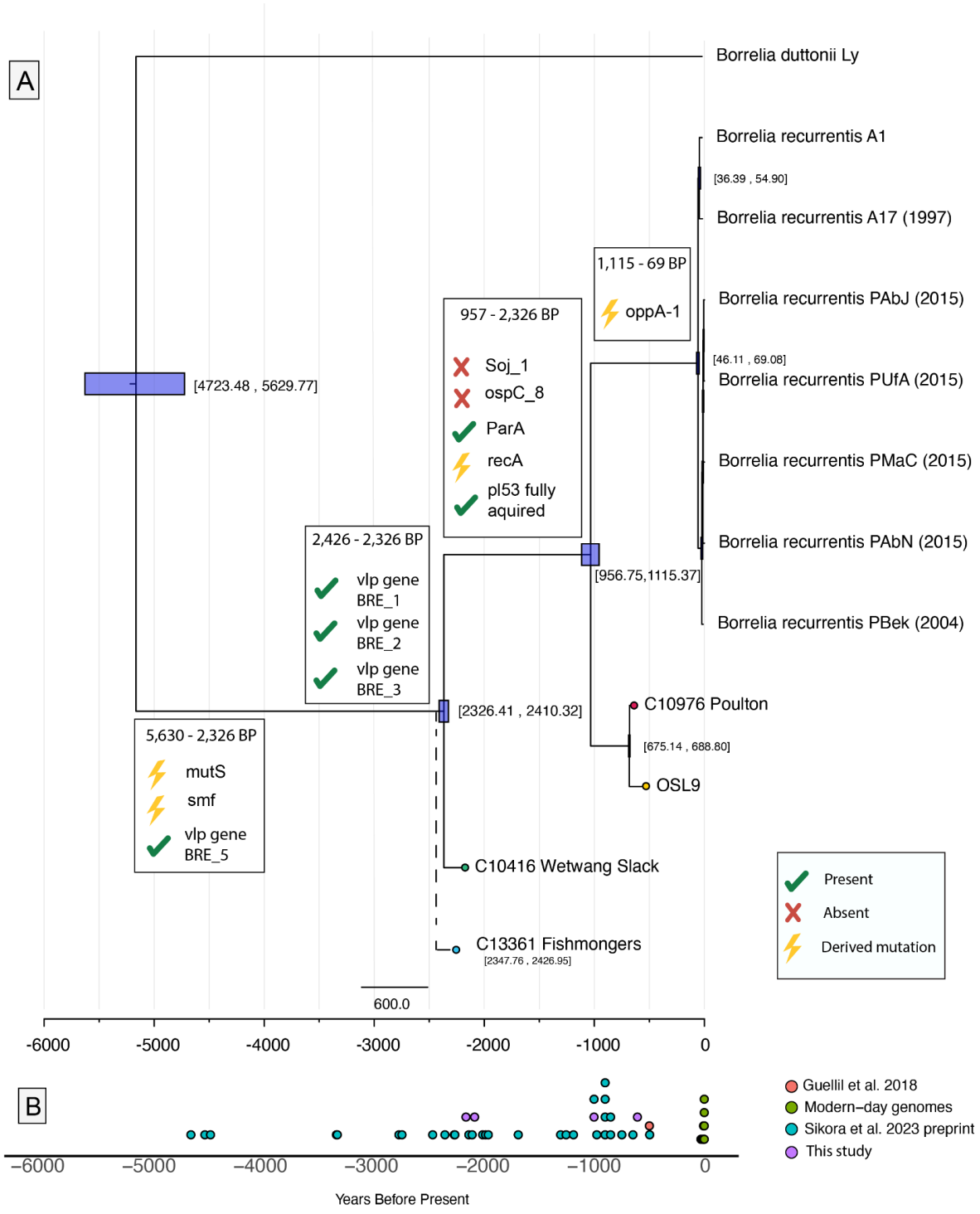


Figure 2. A. Bayesian tip-calibrated maximum clade credibility time tree from Beast2, providing the best supported model following path-sampling.The 95% higher posterior density is indicated with purple boxes and within brackets. The placement of Fishmongers is indicative following a relaxed tip-calibration analysis. Ancient samples are highlighted by coloured tips. Key gene loss and gain events described in text are highlighted at the relevant phylogenetic nodes. B. Timeline

providing the estimated age of *B. recurrentis* observations recovered from ancient DNA reported in this and other studies.

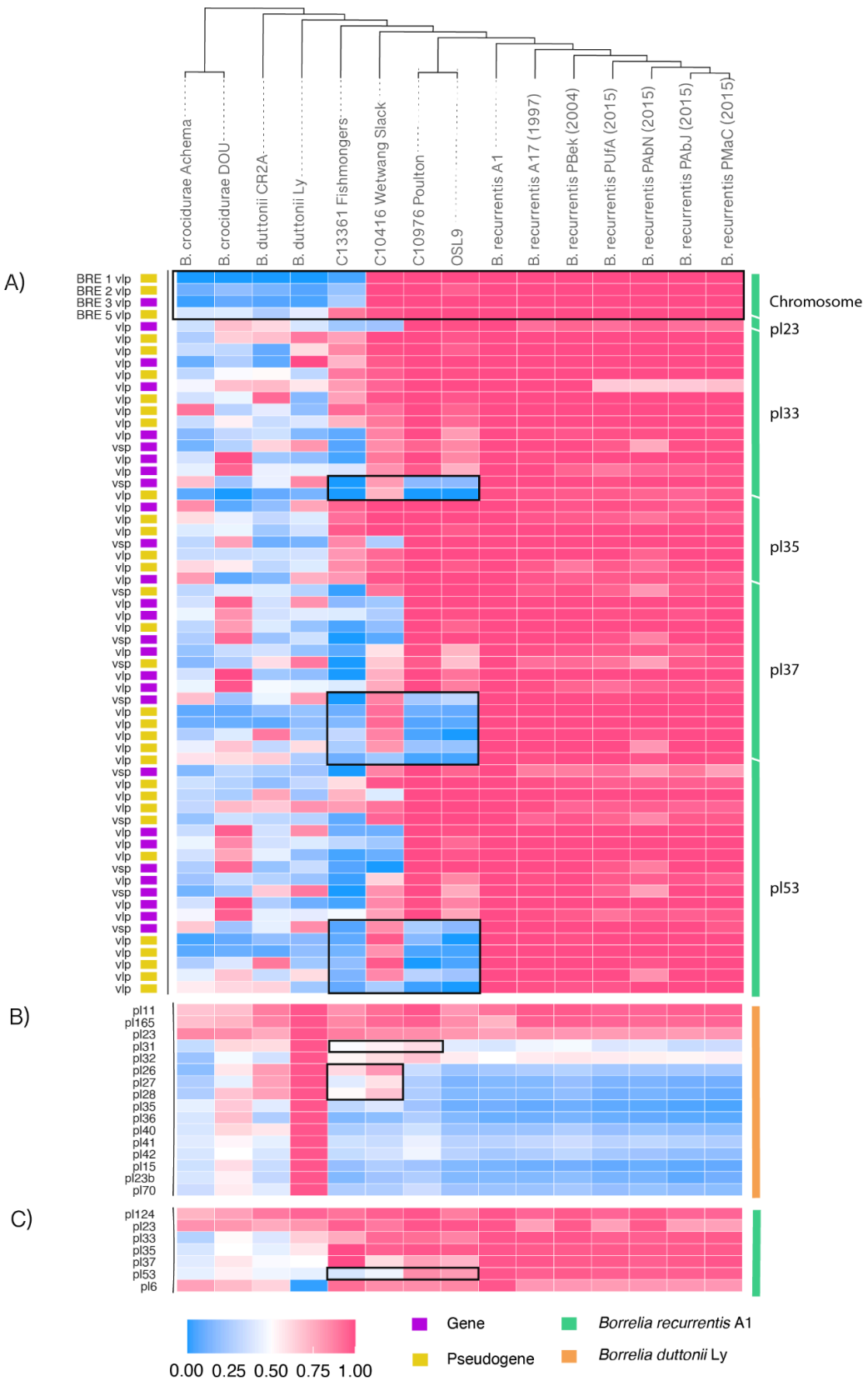

B. crocidurae Achema
B. crocidurae DOU
B. duttonii CR2A
B. duttonii Ly
C13361 Fishmongers
C10416 Wetwang Slack
C10976 Poulton
OSL9
B. recurrentis A1
B. recurrentis A17 (1997)
B. recurrentis PBek (2004)
B. recurrentis PUfA (2015)
B. recurrentis PAbN (2015)
B. recurrentis PAbJ (2015)
B. recurrentis PMaC (2015)
A)
BRE 1 vlp
BRE 2 vlp
BRE 3 vlp
BRE 5 vlp
Chromosome
pl23
pl33
pl35
pl37
pl53
B)
pl11
pl165
pl23
pl31
pl32
pl26
pl27
pl28
pl35
pl36
pl40
pl41
pl42
pl15
pl23b
pl70
C)
pl124
pl23
pl33
pl35
pl37
pl53
pl6
0.00 0.25 0.50 0.75 1.00
Gene
Pseudogene
Borrelia recurrentis A1
Borrelia duttonii Ly

Figure 3.

Gene losses/gains across variable major proteins, *B. duttonii* Ly and *B. recurrentis A1* plasmids. Ancient and modern genomes were aligned to single reference *B. recurrentis* (green) or *B. duttonii* (orange) (**Methods**). Regions of interest highlighted in text are outlined with a black box. Cladogram provides the relationship between different genomes based on a SNP phylogeny . **A)** Normalised coverage across the variable major proteins on the *B. recurrentis* A1 chromosome and plasmids (pl), using *BEDTools* v2.29.2. Positions of the genomes and whether they are classified as genes (yellow) or pseudogenes (purple) were provided in Guellil et al[17] using previously annotated genomes from the NCBI database. **B)** Breadth of coverage for *B. duttonii Ly* plasmids using SAMTools v1.3.1 with a mapping quality of Q1. **C)** Breadth of coverage for *B. recurrentis* A1 plasmids using *SAMTools v1.3.1* with a mapping quality of Q1.

Methods

Archaeological Context:

Wetwang Slack is an Iron Age Arras-style cemetery in East Yorkshire dating to 2300-2100 years ago (300-100 BCE). Sample C10416 was taken from a left mandibular third molar from Burial 240. Fishmonger's Swallet is a cave site in the Mendip Hills in southern Gloucestershire. Sample C13361 was taken from a left first molar in a disarticulated human mandible (A320) that has been radiocarbon dated to 2185-2033 years ago (162 cal. BCE - 10 cal. CE; 2063±28bp, BRAMS-5059[20]). Sample C11907 was taken from a maxillary left 3rd molar from an ancient cranium (CW29) held by the Canterbury Archaeological Trust (CAT). The provenance of CW29 is uncertain but based on the collections held at CAT the cranium must be from a site in Canterbury or the surrounding region and date to 2000-700 years ago (0-1300 CE). Sample C10976 was taken from a maxillary canine of an adult male (Sk 435) buried in a cemetery associated with a medieval chapel at Poulton, Cheshire. Radiocarbon dates from Sk 435 and other parts of the cemetery indicate that Sk 435 dates to 733-633 years ago (1290-1390 cal. CE, 646±14bp, Wk 52986[43]). For more information on the archaeological context see **Supplementary Material**.

Sampling, DNA extraction and library preparation

One tooth from each individual included in this study was sampled and processed in a dedicated cleanroom facility at the Francis Crick Institute. An EV410-230 EMAX Evolution Dentistry drill was used to clean the surface of the tooth and both the cementum and multiple fractions of the dentine were sampled, resulting in ~11- 35 mg of powder from the dentine. 300 µl (<10 mg of powder), 600 µl (10-25 mg) or 1000 µl (>25 mg of powder) of extraction buffer (0.5 EDTA pH 8.0, 0.05% Tween-20, 0.25 mg/ml Proteinase K[44]) was added to the dentine powders and incubated for 24 hours at 37°C. They were then centrifuged for 2 minutes at 13,200 rpm in a table centrifuge and 140 µl of the supernatant was transferred into LVL tubes for automated extraction on an Agilent Bravo Workstation[45]. Extracts were turned into single-stranded DNA libraries[46], then double-indexed[47] and underwent paired-end sequencing with a 2×100 paired-end read configuration on the Illumina HiSeq4000 or NextSeq500 and Novaseq platform (**Table 1** for sequencing effort per library). All samples were processed alongside negative extraction controls as well as positive and negative library controls.

One library, C10416 from Wetwang Slack, underwent size selection to remove fragments shorter than 35 bp and longer than 150 bp, as in Gansauge et al. 2020[46]. Specifically, 100 ng of the initial library was biotinylated and streptavidin beads were used to isolate the non-biotinylated strand and obtain a single-stranded library. This sample was then loaded on a denaturing polyacrylamide gel along with 35 bp and 150 bp insert markers, and fragments within the desired sequence length were physically excised and eluted from the gel, after overnight incubation. The resulting size-selected libraries were further amplified and sequenced on the Illumina NovaSeq.

Bioinformatic Processing, Metagenomic Screening and Authentication

Samples were initially processed via the nf-core/eager v2 pipeline(Fellows Yates et al. 2021). First, adapters were removed, paired-end reads were merged and bases with a quality below 20 were trimmed using AdapterRemoval v2[49] with –trimns –trimqualities –collapse –minadapteroverlap 1 and –preserve5p. Merged reads with a minimum length of 35 bp were mapped to the hs37d5 human reference genome with Burrows-Wheeler Aligner (BWA-0.7.17 aln)[50] using the following parameters "*-l 16500 -n 0.01*"[18,51]. We then analysed sequences that did not align successfully to the human genome using Kraken2[52] and identified individuals as putatively positive for *B. recurrentis* by assessing an excess number of observed *k*-mers (sequence matches).

These libraries were subsequently aligned to the *B. recurrentis* A1 reference genome (chromosome and plasmids; GCF_000019705.1) using BWA-0.7.17 aln[50] parameters "*-l 16500 -n 0.01 -o 2*". Duplicates were removed by keeping only the first sequence out of any set of sequences with the same start position and length (https://github.com/pontussk/samremovedup). We assessed the authenticity of the final set of sequences using the following criteria[23]: i) the observation of postmortem damage, ii) the number of sequences being negatively correlated with edit distance from the reference genome, and iii) an unimodal fragment length distribution via DamageProfiler[53] iv) even breadth of coverage across the *B. recurrentis* A1 reference genome using SAMTools v1.3.1 *depth*[54]. Additionally, these libraries were aligned to the *B. duttonii Ly* and the *B. crocidurae DOU* reference genome (NC_011229.1 and NZ_CP004267.1 respectively) and their edit distance distributions were compared (**Figure S2**). Screening libraries that passed these authentication criteria, were taken forward for further shotgun sequencing. For the final BAM files, we merged shotgun BAM files using SAMTools *merge* resulting in a final chromosome coverage of 28.6X, 9.6X, 2.3X and 0.8X coverage for C10976 Poulton, C10416 Wetwang Slack, C13361 Fishmongers and C11907 Whitefriars, respectively when aligned to *B. recurrentis* A1 (**Table 1, Figure S1**).

Dataset Curation

All published genomes used in this study are listed in **Table S1**. This includes all modern genome assemblies available from NCBI (accessed April 2024). In addition, we *de-novo* assembled six modern genomes available on the SRA linked to BioProject PRJNA378726 using UniCycler v0.50 in short-read only mode, recovering high quality assemblies (N50>700,00 in all cases). The previously sequenced ancient *B. recurrentis* genome OSL9[17] was downloaded from ENA following which adapters were removed and reads were merged using AdapterRemoval2 and processed identically to the described method for individuals in this study (Methods, Bioinformatic Processing).

Alignment Approaches and Phylogenetic Reconstruction

Two approaches were used to construct an alignment for phylogenetic inference. The first used a reference mapped approach to the *B. recurrentis* A1 reference genome (as described above). The second was to construct a core gene reference alignment, built using an assessment of gene content in modern strains. For the latter, we initially applied Panaroo v1.1.2 on the 11

(seven *B. recurrentis,* two *B. duttonii* and two *B. crocidurae)* modern RF genomes and assemblies available specifying the *–core* flag in relaxed *-mode* to obtain a core alignment of genes featuring in 99% of considered genes. We then aligned all genomes (ancient and modern) to the core gene sequence using BWA-0.7.17 *aln* parameters "*-l 16500 -n 0.01 -o 2*", and processed to remove duplicates by keeping only the first sequence in case multiple sequences had the same start and end positions (https://github.com/pontussk/samremovedup). For both approaches, for published assemblies, where short-reads were not available, we applied SeqKit to generate pseudo-reads for which the mapping pipeline could be applied.

In both cases, modern genomes were converted to fasta sequences keeping all bases with a coverage of 1 using HTSBOX[55]. For the ancient genomes, we computed Base Alignment Qualities using SAMTools "*mpileup -E*", restricted to a minimum phred-scaled mapping quality of 30 and base quality 30 using SAMTools v1.3.1. The MD field was modified to record mismatches to the reference using SAMtools *calmd*. Relative to the A1 reference genome and core genome alignment, C to T transition mutations on the forward strand and G to A transition mutations on the reverse strand were masked to correct for the possible effects of cytosine deamination in the single-stranded ancient DNA sequences (https://github.com/pontussk/mpileup2consensus.py/blob/main/mpileup2consensusfasta.py). Additionally, using this tool we filtered out all heterozygous base calls and only kept sites with a minimum coverage of 3 calls per site, resulting in a filtered fasta for each ancient genome.

Ancient and modern Fastas were then concatenated and polymorphic positions were identified for the initial maximum likelihood phylogeny. Using https://github.com/pontussk/fasta_nomissing.py, polymorphic positions with a threshold of maximum missingness of 20% per site and a maximum missingness of 20% per genome were implemented resulting in 4,200 sites when aligned to the core-genome (**Figure S3**) and taken forward for maximum likelihood phylogenetic reconstruction in IQ-TREE v.1.6.12. Due to these thresholds, C11907 Whitefriars was excluded from most downstream analysis. We implemented model testing using the ModelFinder in IQ-TREE[56], which suggested a TIM+F+ASC as the best-fit model according to the Bayesian information Criterion. We implemented 1000 rapid bootstrap replicates and rooted the maximum likelihood phylogeny in FigTree using *B. duttonii Ly* as an outgroup.

Recombination and Temporal analysis

Using the initial SNP tree and the whole core-genome alignment, ClonalFrameML v1.13 was applied to detect homoplasies and putative recombination tracts. Identified recombinant tracts were masked from the alignment **(Figure S5)**.

The recombination pruned core genome alignments and corresponding phylogenies were inspected for signatures of temporal evaluation using the *roototip()* function implemented in BactDating[26], evaluating empirical significance following 10,000 randomisations of the sampling dates. Temporality was assessed for a dataset including and excluding the C13361 Fishmongers sample due to its low coverage, and using phylogenies built on alignments with and without the inclusion of transitions. In all cases we obtained a highly significant temporal regression (**Figure S6, Table S2**).

Resulting alignments were subsequently taken forward for formal Bayesian tip-dating calibration implemented in the BEAST2 workflow[27]. In all cases, variant positions were considered (with corresponding correction for the base composition of invariant sites) with the prior on the tip dates corresponding to the date of sample collection, or, where a range of dates was given (either in contemporary samples or corresponding to radiocarbon dates) the mean estimate was used as an initial prior. Following evaluation of possible substitution models in BModelTest, in all cases a GTR substitution model was best supported and selected for further analysis. BEAST2 was run assuming either of a strict or relaxed (exponential distribution) molecular clock exploring three distinct priors for the demographic model: coalescent constant, coalescent exponential and coalescent bayesian skyline, specifying 50 million chains sampling every 1000. Resulting chains were inspected for convergence in TRACER, requiring an effective sampling space (ESS) of >200, with resulting posterior estimates extracted following discarding the first 10% of chains as burn-in. All runs exhibited a significant difference between the posteriors obtained when sampling from the prior (data absent model). Finally, model fit was assessed using the path-sampling model, requiring 100 steps over 250,000 chains, to establish marginal likelihoods for the models and corresponding Bayes Factors. All results are provided in **Table S3** with posterior distributions available in **Figures S7** and **Figures S8** .

Pangenome Analysis and Evaluation of Gene Content

We constructed a full pan-genome given diversity in modern *B. recurrentis*, *B. duttonii* and *B. crocidurae* genomes (**Table S1**). To do so we initially applied Prokka v1.12(Seemann 2014) to generate gene annotations for each modern genome. Panaroo v1.1.2 with the *-pan* flag in relaxed mode and otherwise default thresholds was implemented and then used to identify gene clusters and create a list of genes present in all of the given modern genomes. We identified 14,475 genes and gene clusters across all modern relapsing fever *Borrelia* species, which were subsequently filtered for pseudogenes or genes of unusual lengths. Among the modern *B. recurrentis* genomes, Panaroo identified ~1,100 of these genes as present.

To identify the presence and absence of genes in the pan-genome we aligned simulated short-read modern FASTQs using BWA-0.7.17 *mem* with the parameters "*-B 40 -O 60 -E 10 -L100*" to the pangenome and the ancient FASTQs using BWA *aln* with the parameters "*-l 16500*

-n 0.01 -o 2". We then used the pangenome gene clusters. This allowed us to set an arbitrary threshold to identify genes as present required a minimum threshold of 70% coverage and below 30% coverage for absence. In addition we only kept genes which showed phylogenetically congruent patterning, resulting in consideration of 71 well resolved genes of interest **(Table S5**). We performed *in silico* functional predictions for these genes (**Methods**) using the EggNOGG-Mapper web server (http://eggnog-mapper.embl.de/) and InterproScan[57], both with default parameters. Putative gene functions were inferred from the annotations provided by both tools **(Table S6).** The gene neighbourhood of identified hits, including *Soj_1* and *ParA*, were assessed using the Panaroo pan-genome network graphs, visualised in Cytoscape v3.10.2. In this case, gene homology was obtained by translating the nucleotide sequence of *Soj_1* and *ParA* to amino acids before aligning in MUSCLE and calculating pairwise identity.

Visual inspection of SNPs and Indels

An additional reference based analysis was also carried out, aligning all genomes to *B. recurrentis* A1 and *B. duttonii* Ly. Here, all ancient and simulated short-read modern genomes were aligned to each of the aforementioned references using the described BWA aln pipeline for ancient samples and BWA *mem* for the modern genomes (specifying the same parameters as when aligning to the pangenome). After the removal of duplicates, we assessed the coverage across both chromosomes, plasmids, and previously reported genes, and manually inspected previously reported SNP mutations using Intergrated Genome Viewer (IGV) (**Figure S13**). Additionally, we assessed regions of missingness by identifying regions with a gap in coverage larger than 500bp in the ancient genomes using BEDTools(Quinlan and Hall 2010) when aligned to each of the *B. recurrentis A1* reference genome and the *B. duttonii* Ly reference genome (**Figure S16**, **Table S7**). We noted some marked drops in coverage consistent with genomic deletions or rearrangements. As an example, despite the low coverage of C11907 Whitefriars, we could confirm a unique 26.1kb deletion when aligned to both colinear plasmid pl124 in *B. recurrentis* A1 and pl165 in *B. duttonii* Ly (**Figure 1B, Figure S15, Table S7**). The pl124 plasmid in *B. recurrentis* A1 has been previously observed to exhibit a ~40kb deletion in comparison to the *B. duttonii* plasmid[3]. Marosevic et al[8] identified this 40 kb region to also be present in eight modern-day *B. recurrentis* genomes from East Africa, suggesting this deletion may be a recent event, or potentially the product of an incomplete assembly of the A1 reference genome plasmid. We find this 40kb region is also present in all of our ancient *B. recurrentis* observations, consistent with these proposed scenarios.

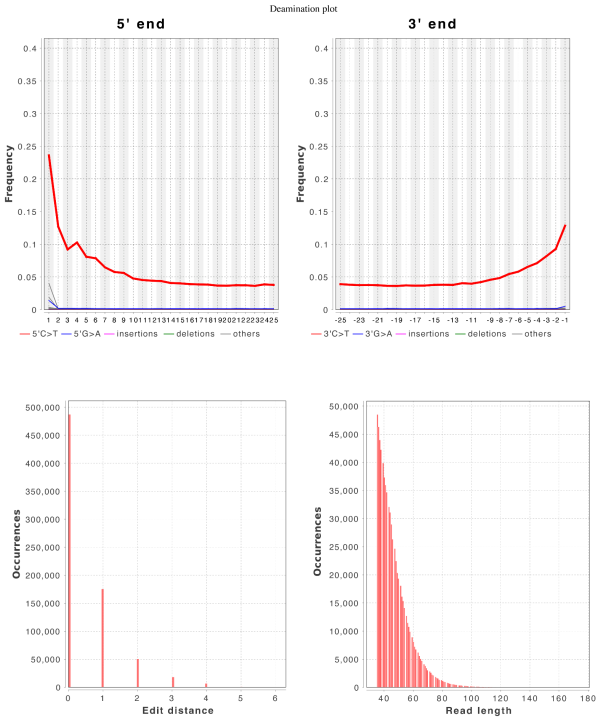
Supplemental Information

Supplementary Figures

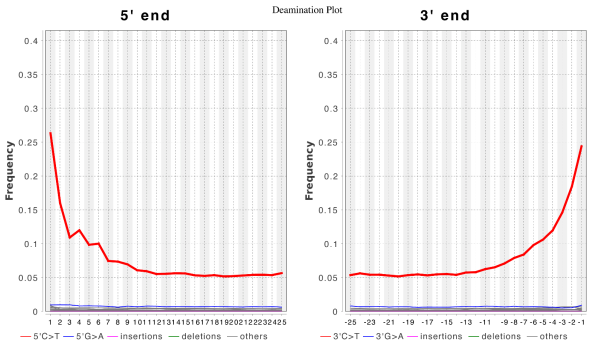
Figure S1	Ancient pathogen authentication of four ancient genomes from this study when aligned to *B. recurrentis* A1 reference.

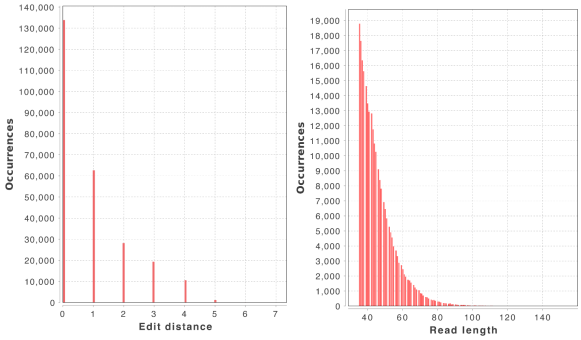
Figure S2	Edit distance of four ancient genomes from this study when aligned to *B. recurrentis* A1, *B. duttonii* Ly, and *B. crocidurae* DOU reference.
Figure S3	Maximum Likelihood SNP phylogeny of ancient and modern *B. recurrentis* genomes and *B. duttonii* outgroup when aligned to Panaroo core-genome.
Figure S4	Maximum Likelihood SNP phylogeny of ancient and modern *B. recurrentis* genomes and *B. duttonii Ly* outgroup when aligned to *B. recurrentis* A1 reference genome with relaxed filtering thresholds.
Figure S5	ClonalFrameML homoplasy recombination analysis of core genome.
Figure S6	BactDating, Root-to-tip regression with and without inclusion of C13361 Fishmongers.
Figure S7	Beast2 tip-calibrated time tree when C13361 Fishmongers genome is included.
Figure S8	Posterior Distributions from Bayesian tip-dating analysis.
Figure S9	Per species distribution of the total number of genes identified across the *Borrelia* wide pan-genome after filtering for truncated genes.
Figure S10	Pan-genome diversity across the *Borrelia* (RF) genus, comprising 14,475 genes
Figure S11	Plasmid authentication of B. duttonii plasmids pl26, pl27, pl28 in C10416 Wetwang Slack and C13361 Fishmongers.
Figure S12	Presence/absence across pan-genome genes and gene clusters.
Figure S13	Integrated Genome Viewer of SNPs and INDELs
Figure S14	Map and timeline showing the distribution of ancient and modern *B. recurrentis* genome through space and time.
Figure S15	Gap size distribution plot to identify deletions over 500bp.
Figure S16	Circos plots providing the per sample coverage across the *B. duttonii Ly* reference genome.

a. C10976 Poulton
Number of used reads: 737,914 (100.0% of all input reads)

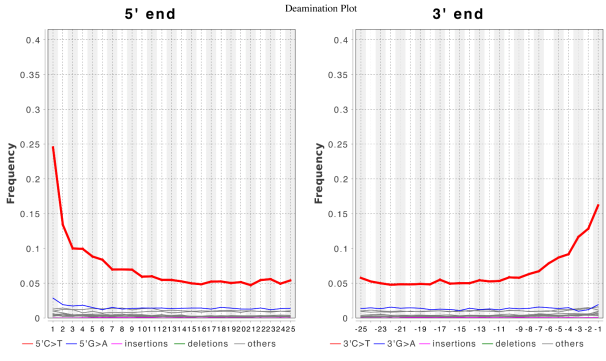


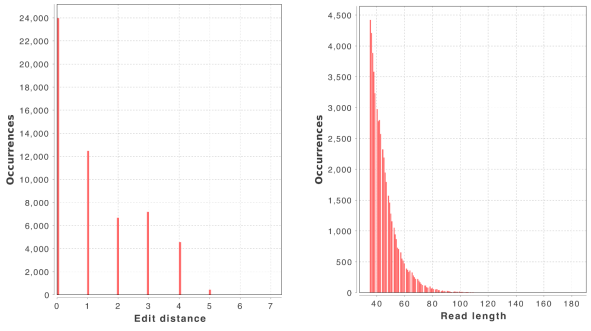
b. C10416 Wetwang Slack
Number of used reads: 255,646 (100.0% of all input reads)



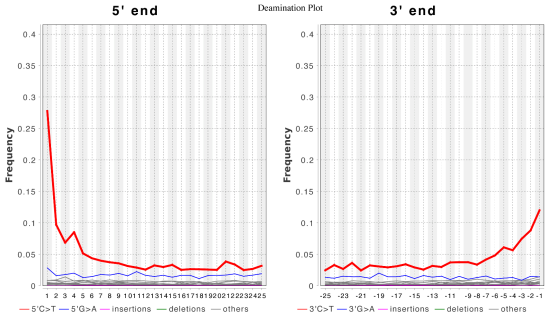


c. C13361 Fishmongers
Number of used reads: 55,328 (100.0% of all input reads)





d. C11907 Whitefriars
Number of used reads: 19,555 (100.0% of all input reads)



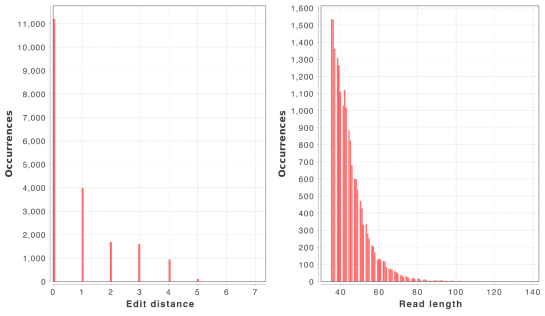


Figure S1. Ancient pathogen authentication of four ancient genomes from this study when aligned to *B. recurrentis* A1 reference. Edit distance, Damage and fragment length distribution for the concatenated genomes for each sample when aligned to *B. recurrentis* chromosome and plasmids (a1 reference genome) via DamageProfiler[53]. **a)** C10796 from Poulton **b)** C10416 from Wetwang Slack **c)** C13361 from Fishmongers **d)** C11907 from Whitefriars.

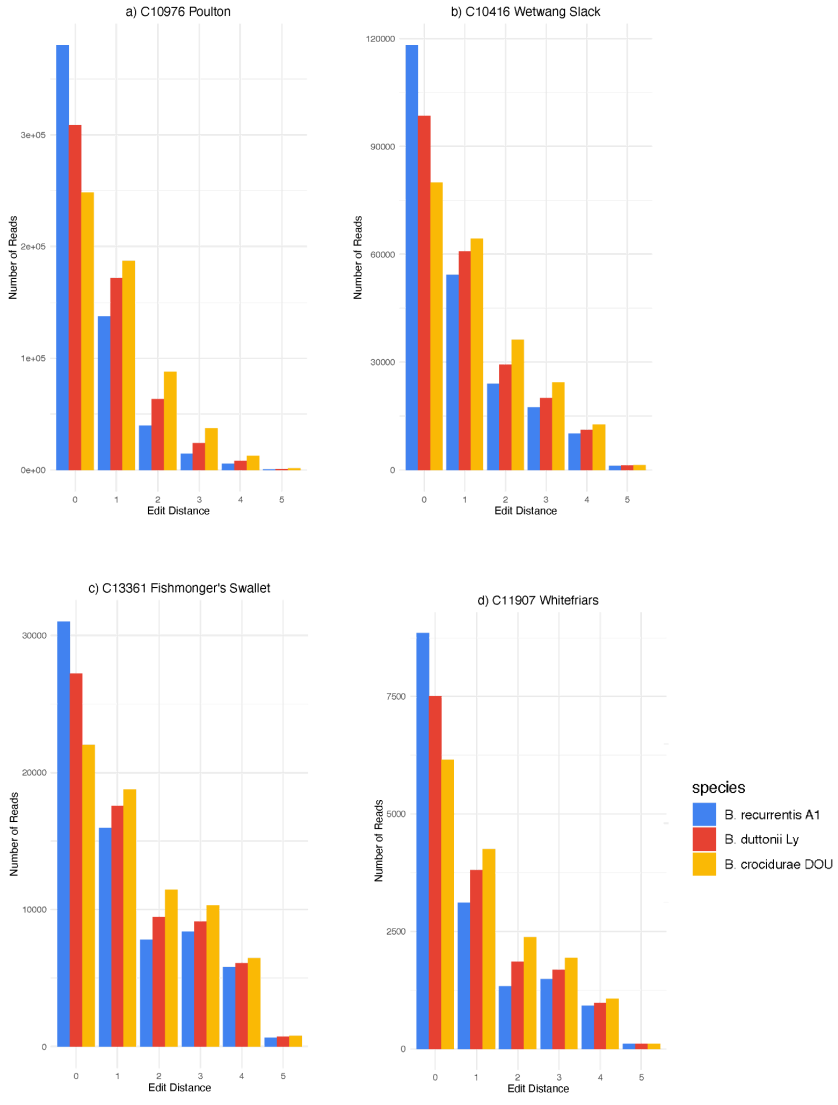


Figure S2. Edit distance of four ancient genomes from this study when aligned to *B. recurrentis* A1, *B. duttonii* Ly, and *B. crocidurae* DOU reference. Edit distance (x-axis) against number of reads (y-axis) of all newly considered ancient genomes from this study, when aligned to *B. recurrentis* A1 (blue), *B. duttonii* Ly (red) and *B. crocidurae* DOU (yellow) reference genomes. **a)** C10796 from Poulton **b)** C10416 from Wetwang Slack **c)** C13361 from Fishmongers **d)** C11907 from Whitefriars

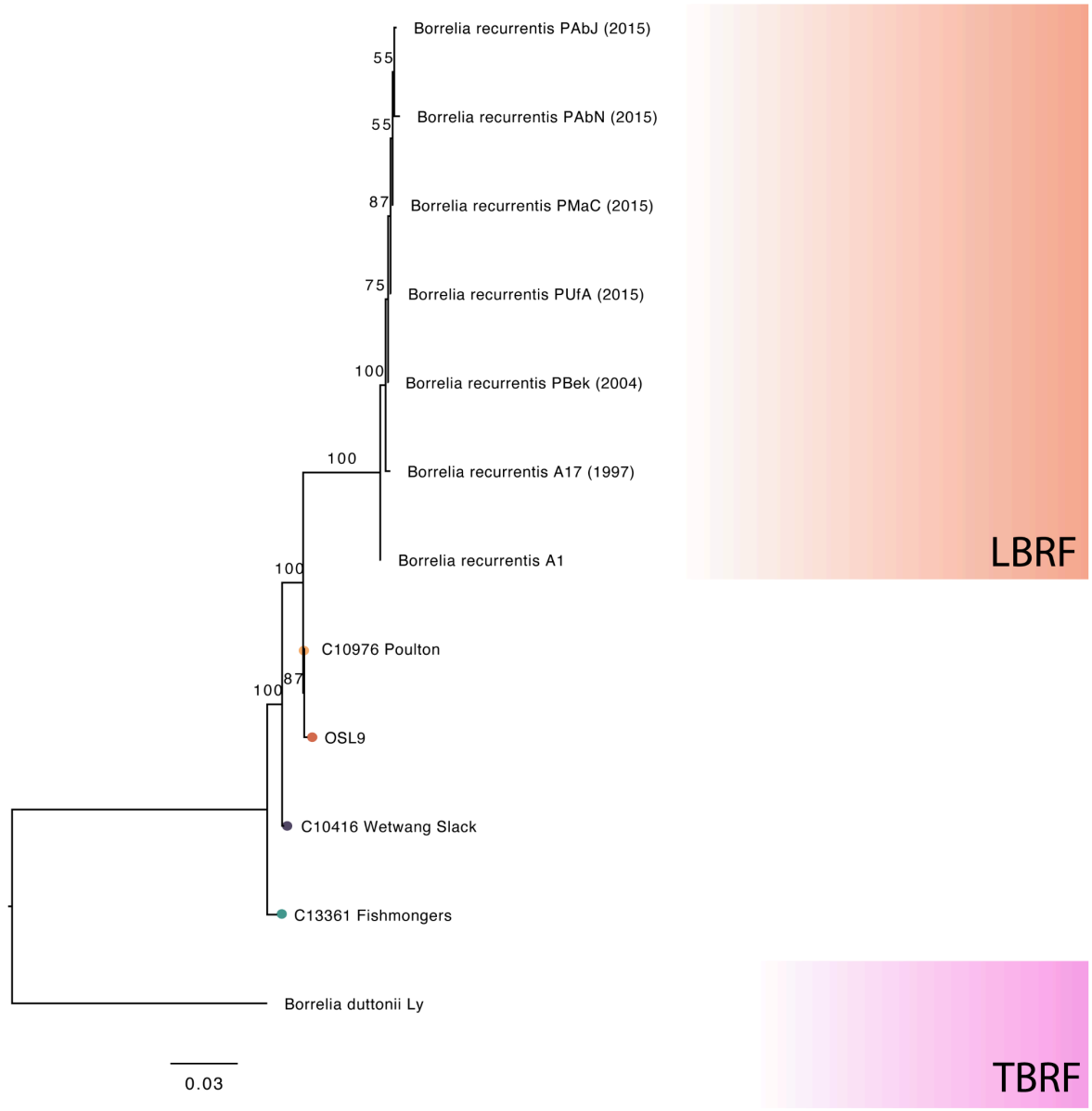


Figure S3. Maximum Likelihood SNP phylogeny of ancient and modern *B. recurrentis* genomes and *B. duttonii* outgroup when aligned to Panaroo core-genome. Maximum likelihood phylogenetic tree (GTR+F+ASC according to AIC) constructed on variability over the core genome SNP alignment of 2,006 sites filtering for 20% missing per site and per genome including transitions.

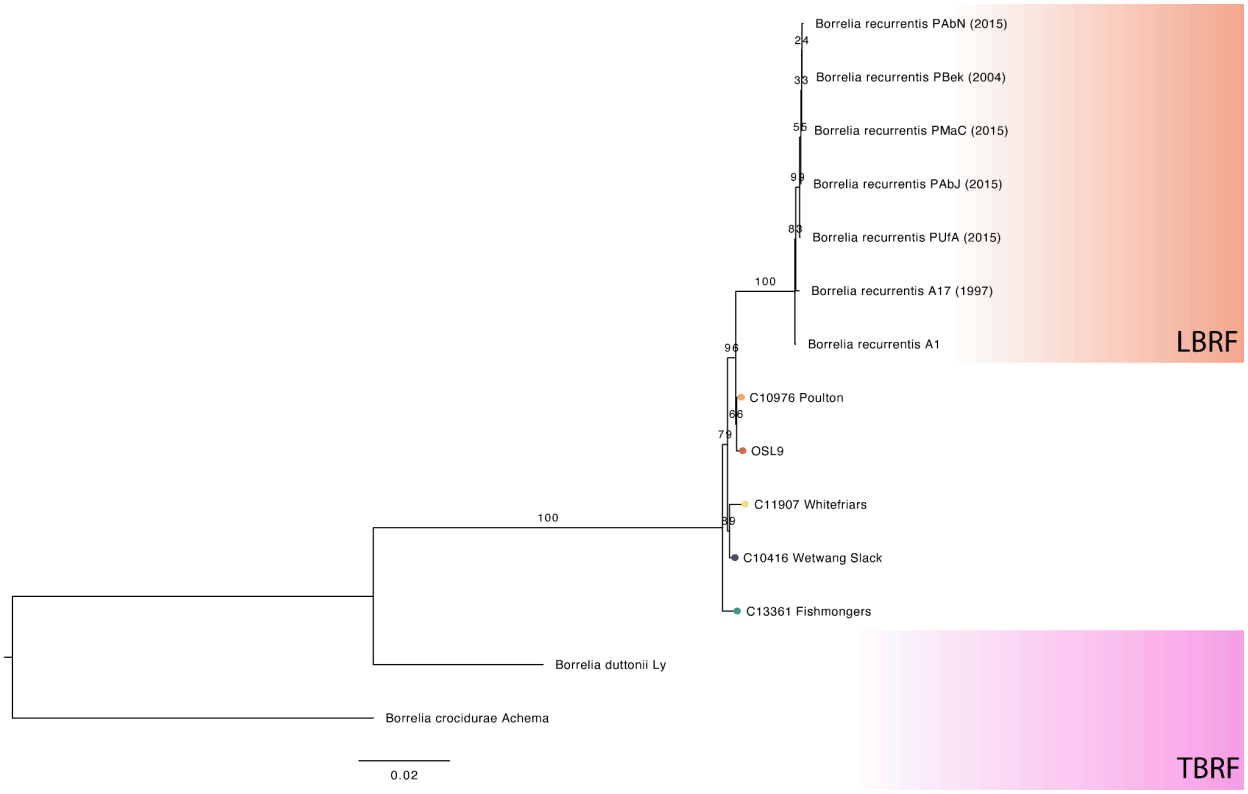


Figure S4. Maximum Likelihood SNP phylogeny of ancient and modern *B. recurrentis* genomes and *B. duttonii Ly* outgroup when aligned to *B. recurrentis* A1 reference genome with relaxed filtering thresholds. Maximum likelihood phylogenetic tree (K3Pu+F+ASC model implemented with rapid bootstrapping) constructed on variability over a reference based alignment to the *B. recurrentis* A1 reference genome (3,354 sites, 20% missingness per site, maximum missingess across genome 90%). All ancient genomes were filtered for a minimum base fold coverage of 3 except for C11907 Whitefriars which was filtered for a minimum coverage of 2.

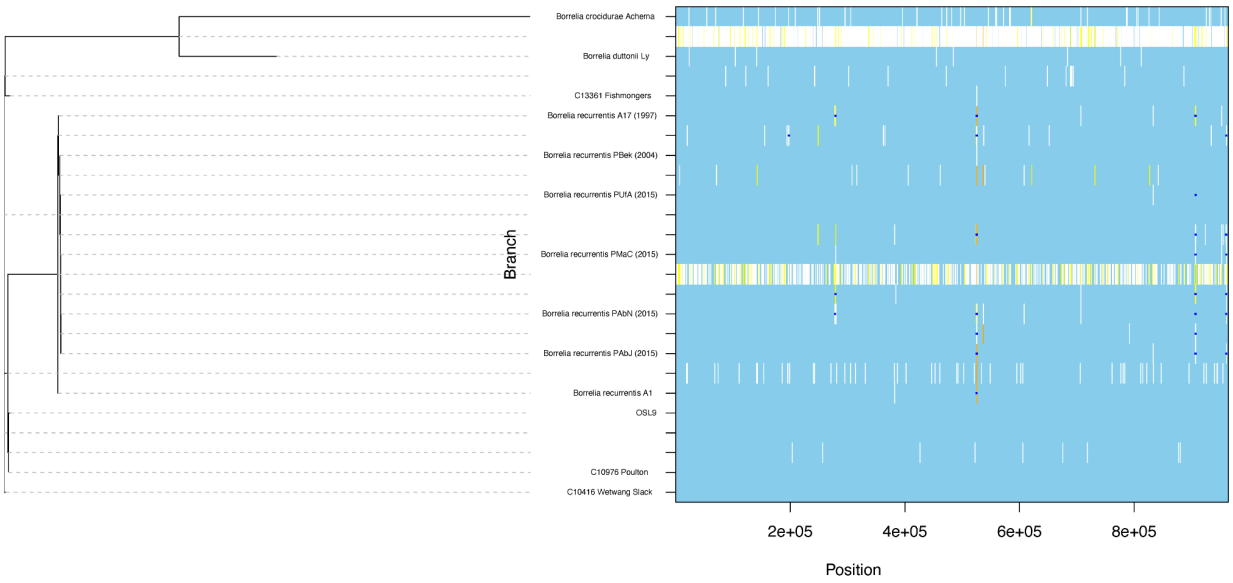


Figure S5. ClonalFrameML homoplasy recombination analysis of core genome. Reconstructed substitutions (white vertical bars) are shown for each branch of the maximum likelihood tree. Dark blue horizontal bars indicate recombination events detected by the analysis with yellow providing homoplasic positions.

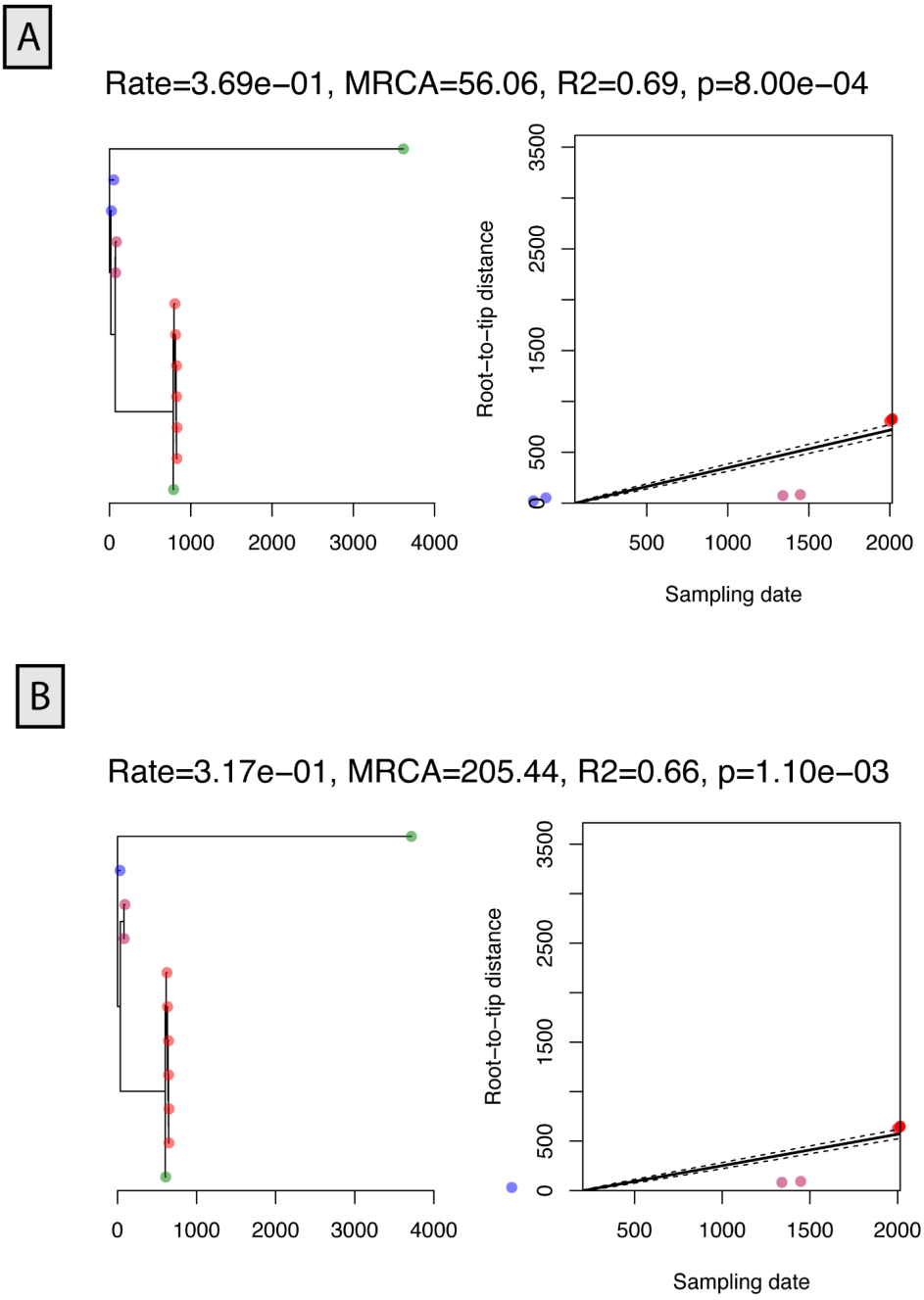


Figure S6. BactDating, Root-to-tip regression with and without inclusion of C13361 Fishmongers. Linear regression of root-to-tip (y-axis) versus sampling date (x-axis) across the phylogeny including (A) and excluding (B) C13361 Fishmongers, implemented via the roototip() function available in BactDating. P-values at top provide the empirical p-value following 10,000 steps of the BactDating date randomisation procedure.

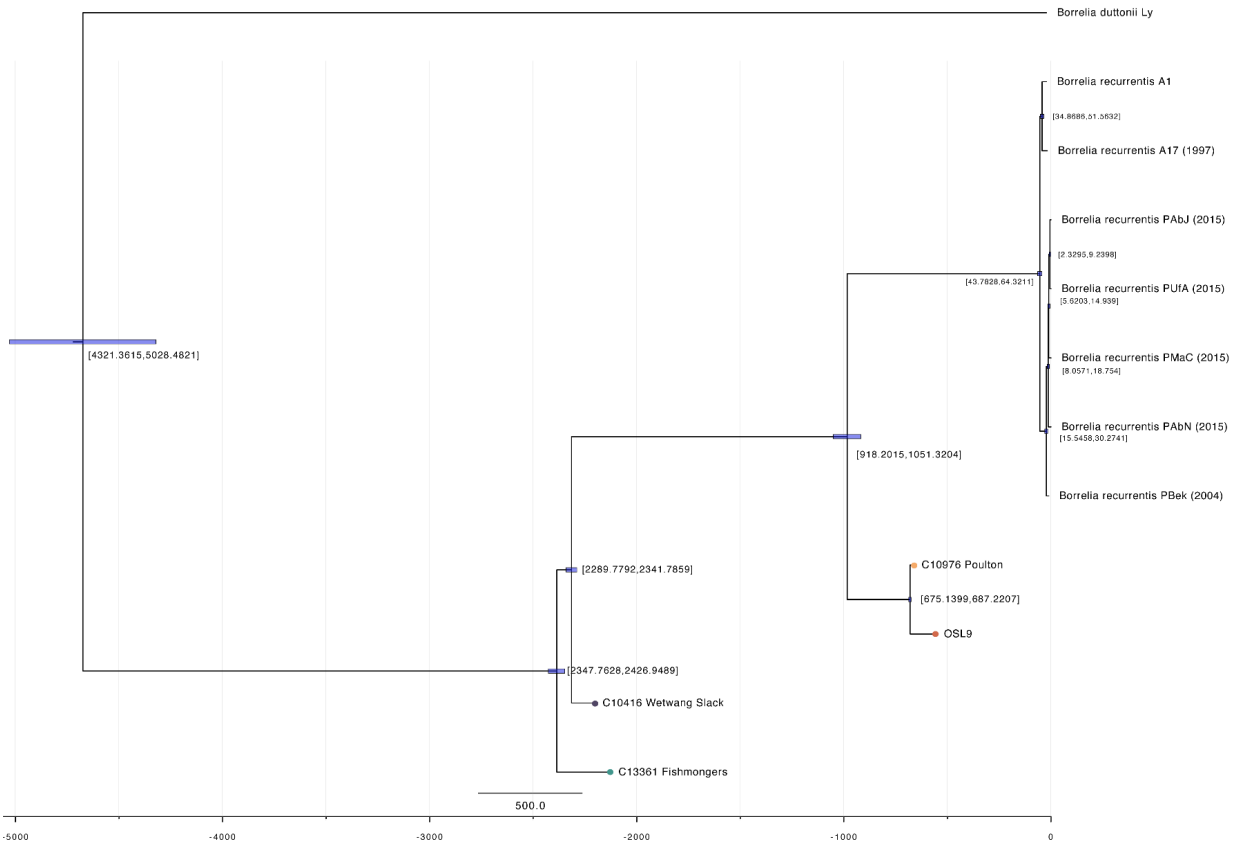


Figure S7. Beast2 tip-calibrated time tree when C13361 Fishmongers genome is included. Bayesian tip-calibrated maximum clade credibility time tree from Beast2, providing the best supported model following path-sampling when C13361 Fishmongers is included in the alignment. Confidence intervals around nodes provided the 95% higher posterior density. Ancient samples are highlighted by coloured tips.

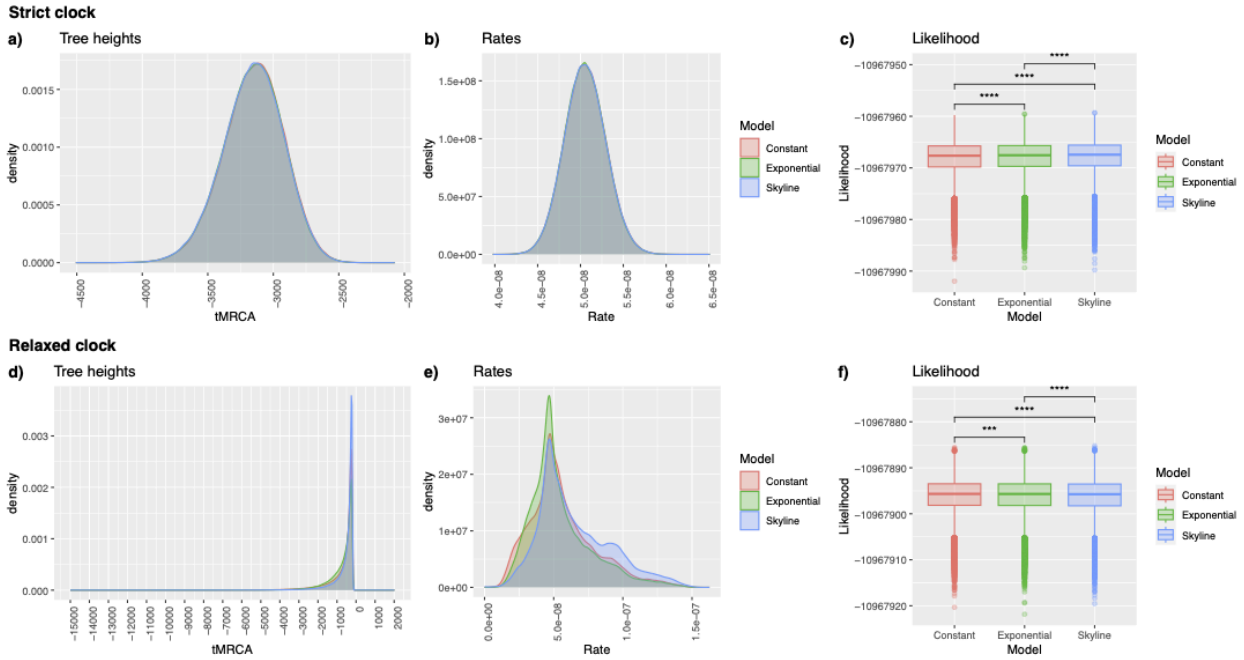


Figure S8. Posterior Distributions from Bayesian tip-dating analysis. Posterior distributions estimated under a strict clock (top row) and relaxed clock (bottom row) allowing three possible specifications of demographic models (red - coalescent constant, green - coalescent exponential, blue - coalescent skyline). Estimated tree heights are given in the left-hand panels (a,d), estimated clock rates are provided in the right-hand panels (b,e). Boxplots provide the posterior distribution of likelihood values under a strict (c) and relaxed (f) clock model.

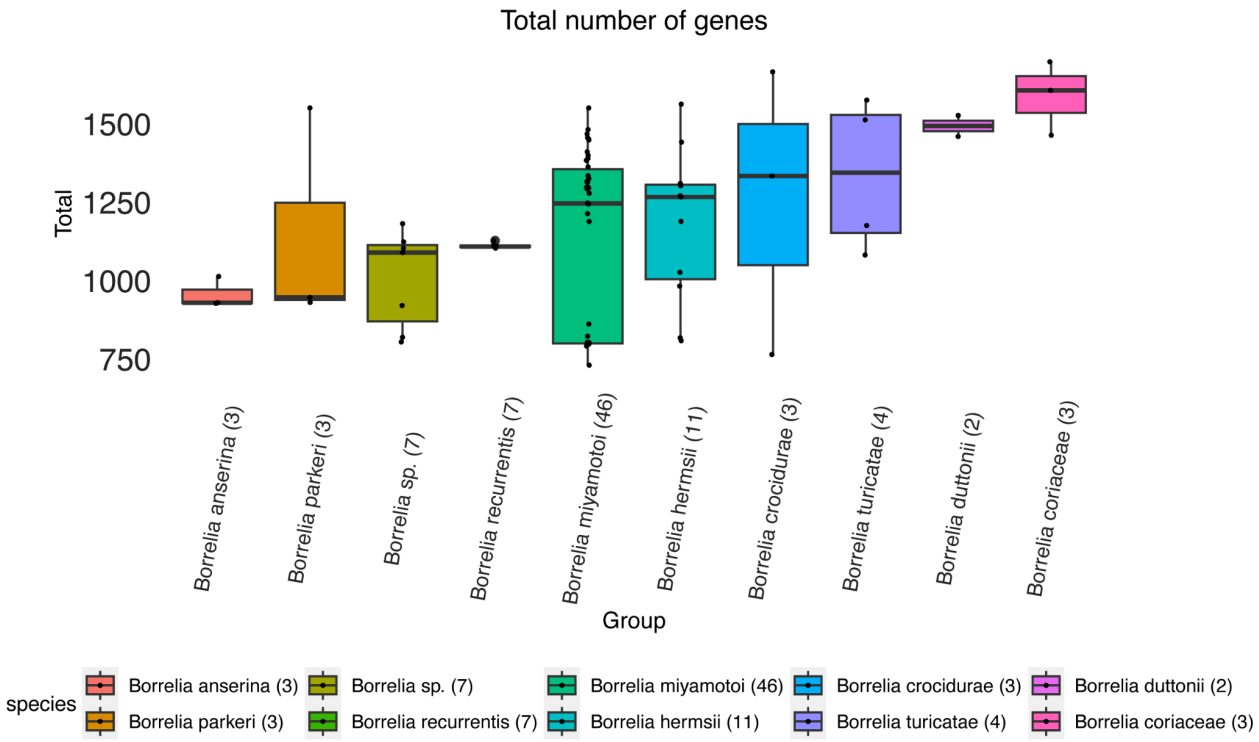


Figure S9. Per species distribution of the total number of genes identified across the *Borrelia* wide pan-genome after filtering for truncated genes. The count of the number of genomes considered per species is provided in parentheses. Genes were filtered for truncated genes using Panaroo *–filter* parameter. Species where only one isolate was available was excluded from this plot.

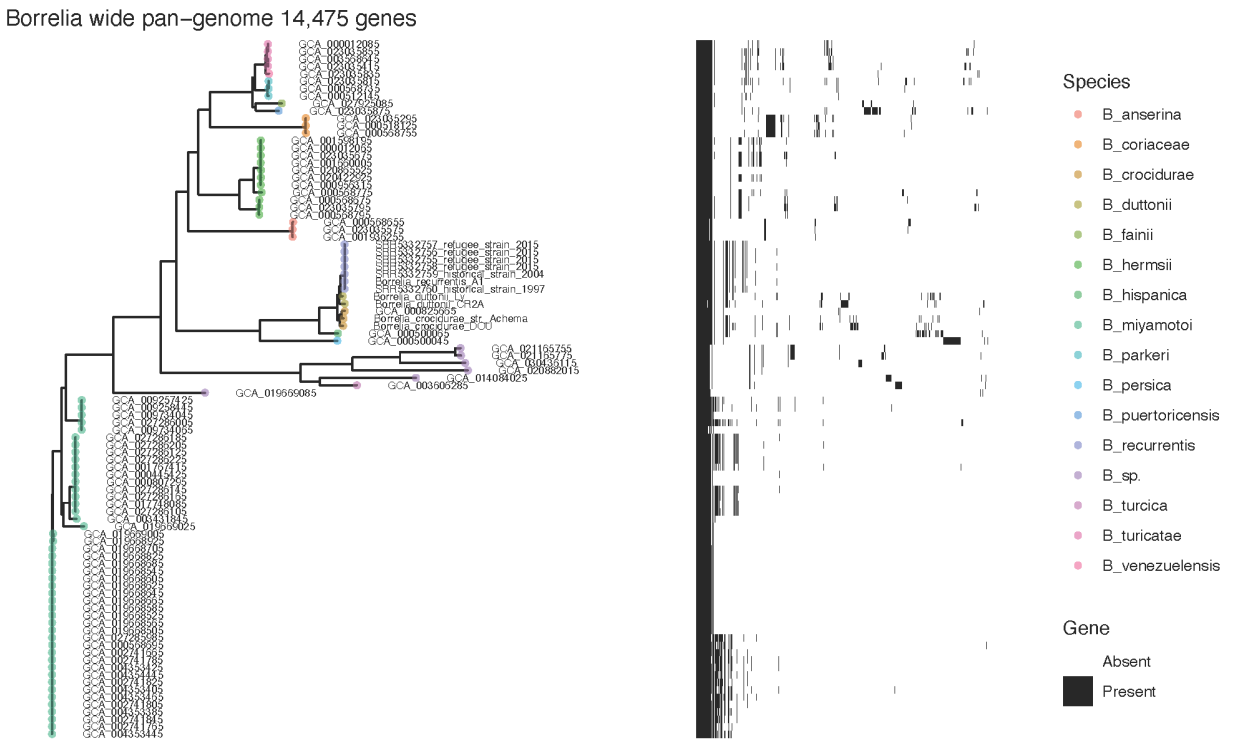


Figure S10. Pan-genome diversity across the *Borrelia* (RF) genus, comprising 14,475 genes. The phylogeny at right provides a neighbour-joining representation based on the gene presence and absence over the 14,475 pan-genome. Heatmap at right provides the presence (black) and absence (white) across the pan-genome, with each column providing an individual gene.

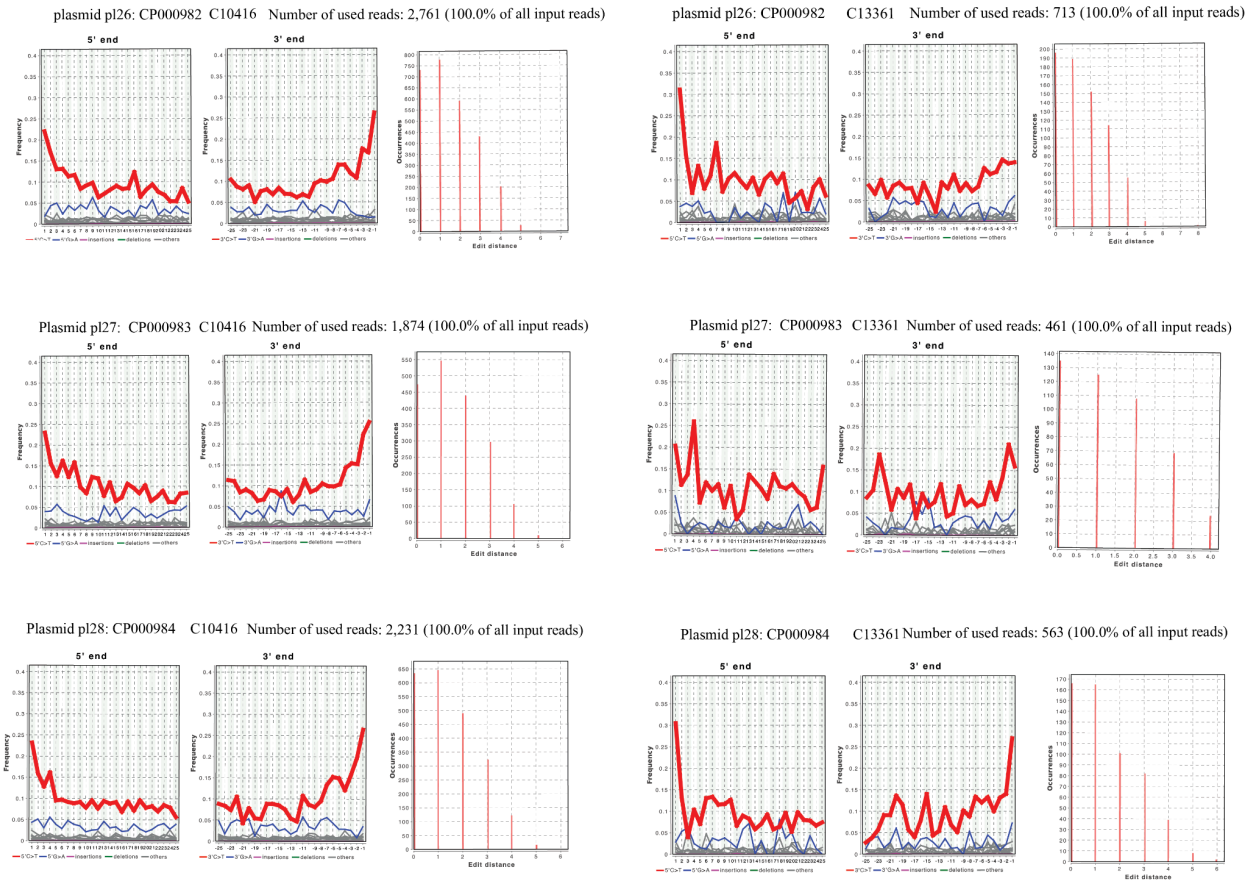


Figure S11. Plasmid authentication of *B. duttonii* plasmids pl26, pl27, pl28 in C10416 Wetwang Slack and C13361 Fishmongers. Damage (deamination) pattern and edit distance of C10416 Wetwang Slack and C13361 Fishmongers when aligned to *B. duttonii* plasmids pl26, pl27 and pl28 which are shown to be higher coverage than all other *B. recurrentis* genomes.

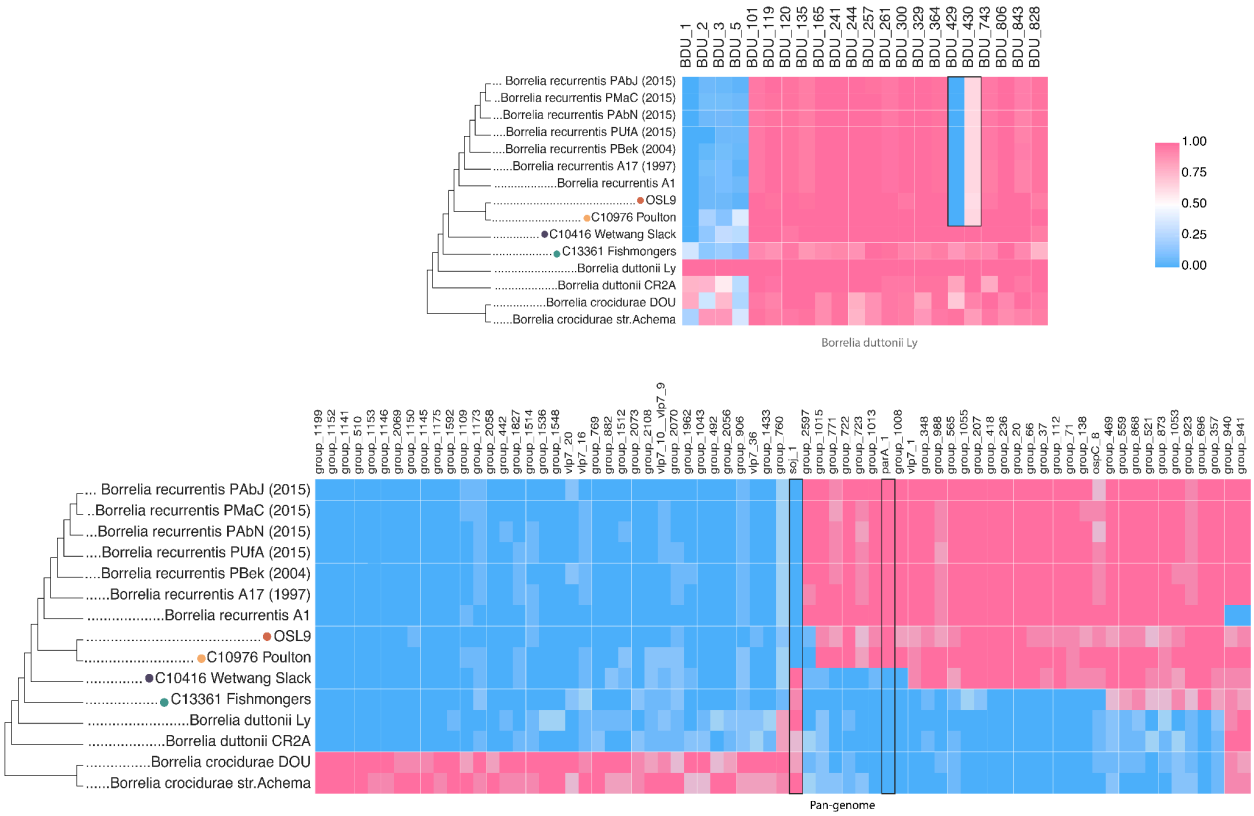


Figure S12. Presence/absence across pan-genome genes and gene clusters. Ancient and modern genomes were aligned to the pan-genome reference built on all modern *B. recurrentis*, *B. duttonii* and *B. Crocidurae* genomes (**Methods**). . Normalised coverage across the gene and gene clusters was calculated using *BEDtools* v2.29.2. Genes that had a coverage between 0.3 and 0.7 were filtered out and genes that has a consensus coverage across the clade modern *Borrelia*, medieval clade was kept. Additionally, genes that are in the same state (either all absent or all present) in the B. duttonii Ly and ancient and modern B. recurrentis were also filtered out. This resulted in 71 genes out of the 3035 genes identified in the pan-genome output. Regions of interest highlighted in text are outlined with a black box. Cladogram provides the relationship between different genomes based on a SNP phylogeny

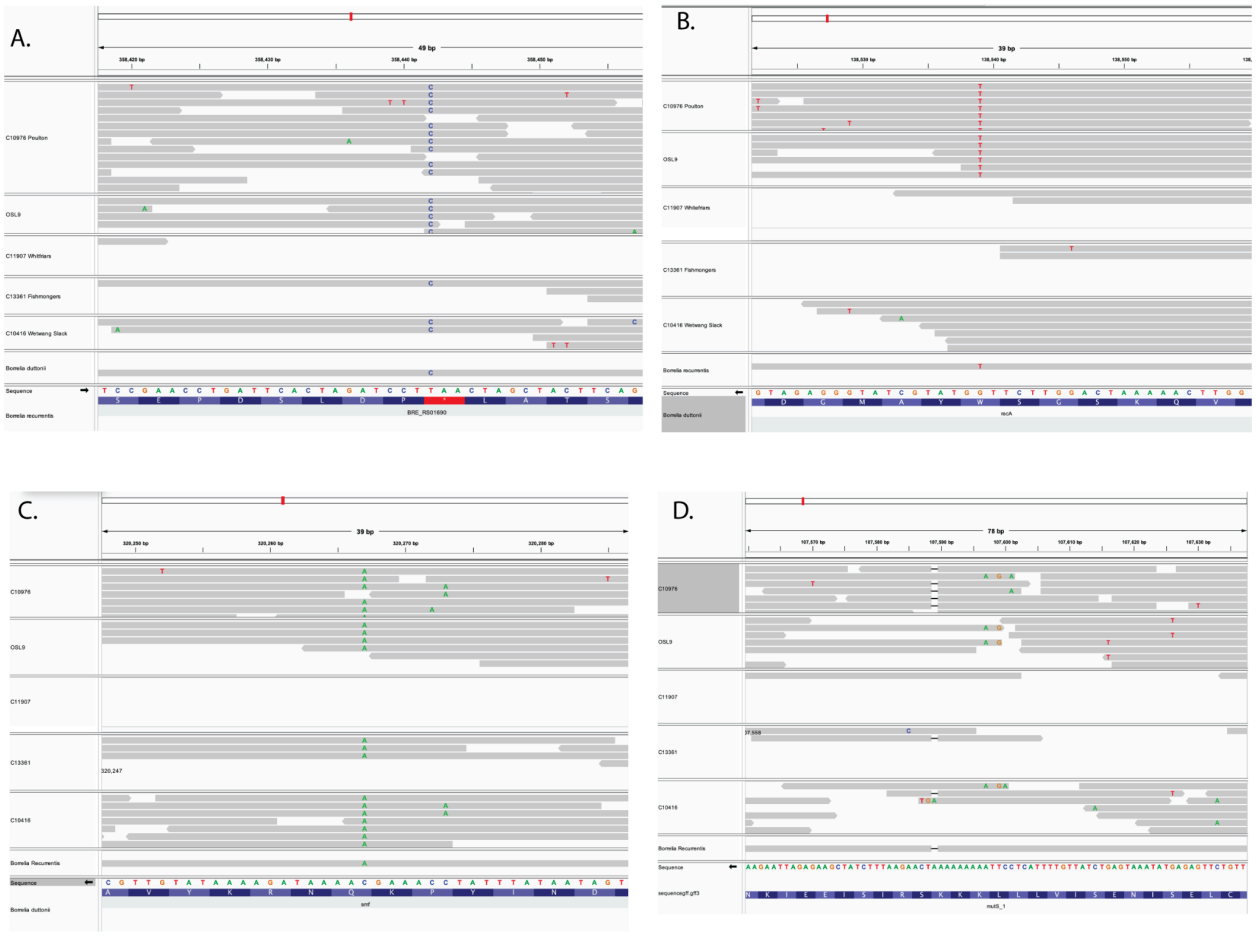


Figure S13. Integrated Genome Viewer of SNPs and INDELs. Previously identified SNPs and INDELS shown to be in different states in *B. recurrentis* A1 and *B. duttonii Ly*, visualised with IGV 2.17.4.

A) Ancestral state of a polymorphism in the *oppA-1* gene in ancient samples from this study when aligned to *B. recurrentis A1.* The derived allele, which is present in *B recurrentis,* results in an inframe stop codon. **B)** In-frame stop codon in the *recA* gene in *B. recurrentis* and medieval genomes, from a Tryptophan TGG (W) in *B. duttonii Ly* to a TAG stop codon (seen as a T on the forward strand), in the *B. recurrentis* A1 genome and the two medieval genomes included in this study. C10416 Wetwang Slack confidently shows the functional form of the gene, as does C11907 Whitefriarsbut with the support of only 1 read. **C)** In-frame stop codon in the *smf* gene. IGV plot showing A on the forward strand in *B. recurrentis* A1 and all ancient genomes, resulting in a TAA in the reverse strand leading to an in-frame stop codon. The *B. duttonii* sequence is shown in the direction of protein synthesis. **D)** A frameshift mutation in the *mutS* gene, which is present in *B. recurrentis*, is represented by a deletion. Here we show the presence of reads aligning to the *mutS* gene with the same frameshift mutation present in *B recurrentis A1*, also present in C10416 Wetwang Slack, C13361 Fishmongers, and C10976 Poulton. Due to the low number of reads aligning to this region and the potential misalignment of the reads, it is unclear as to whether the frameshift mutation is absent or present in the other ancient individuals, or whether their presence is an artefact caused by misalignment.

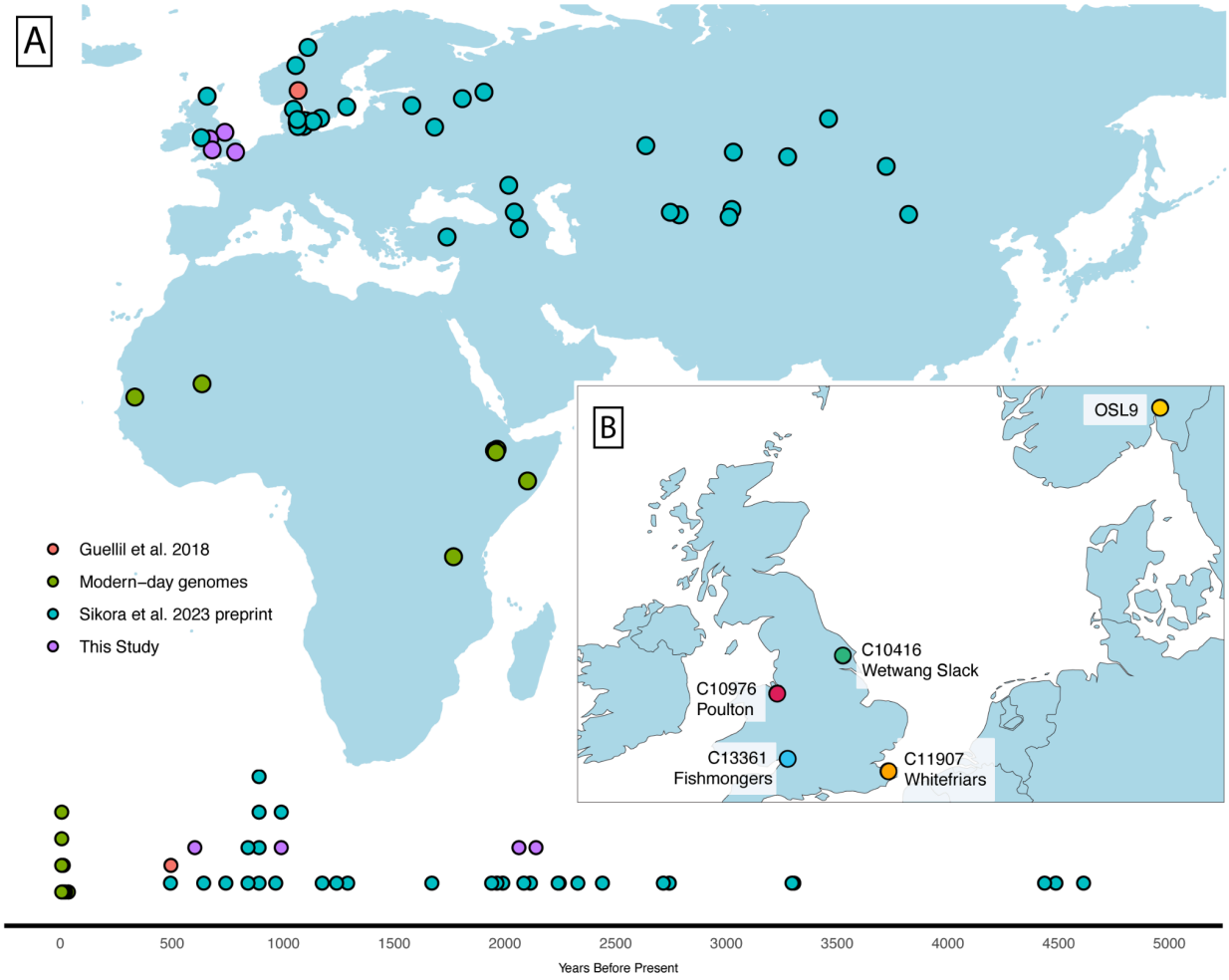


Figure S14. Map and timeline showing the distribution of ancient and modern *B. recurrentis* genome through space and time. Map providing the geographic location and timeline of *B. recurrentis* observations to date. Note that observations from Sikora et al. 2023 preprint comprise hits assessed from partial recovery of sequencing reads as opposed to whole genome observations.

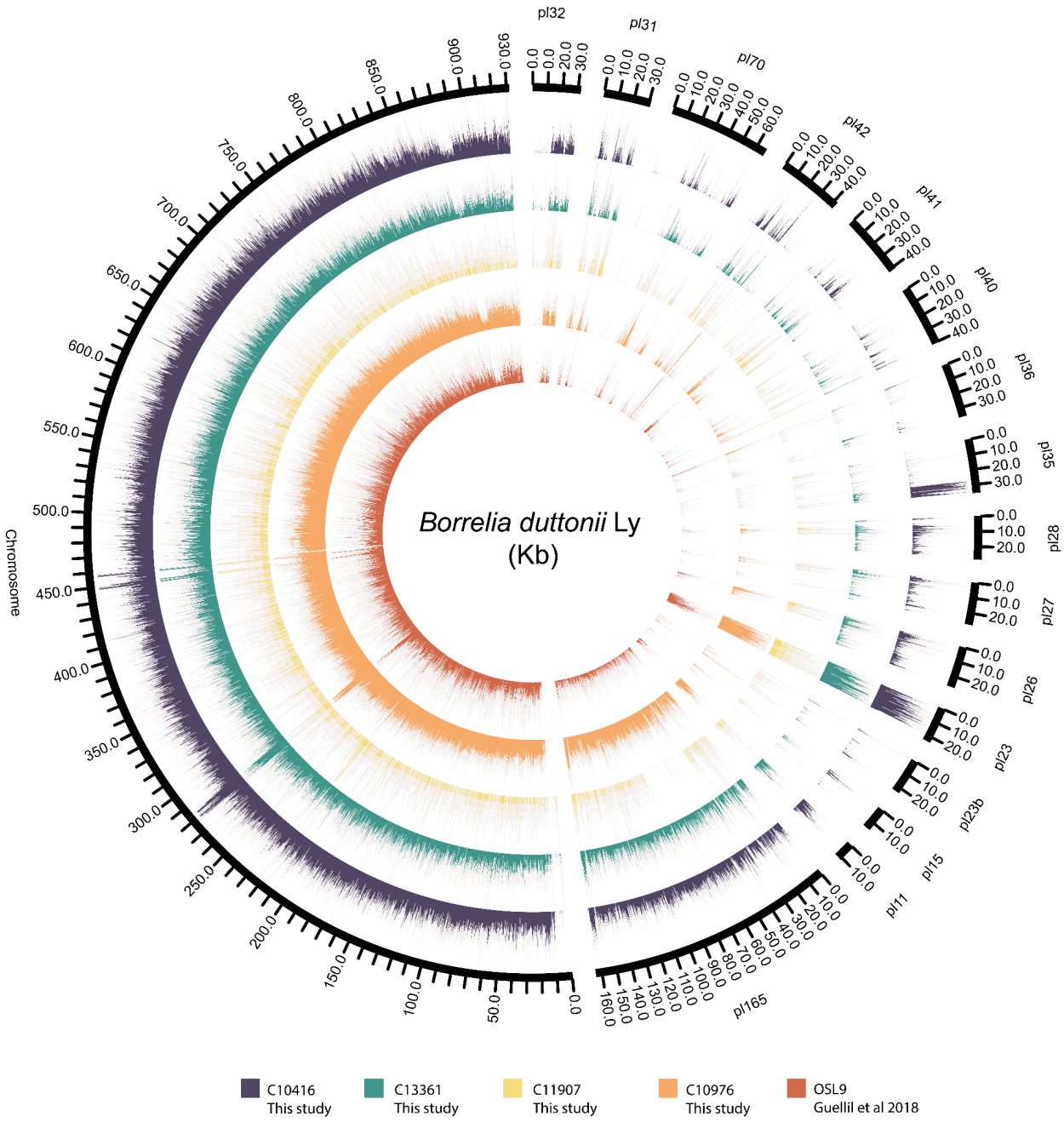
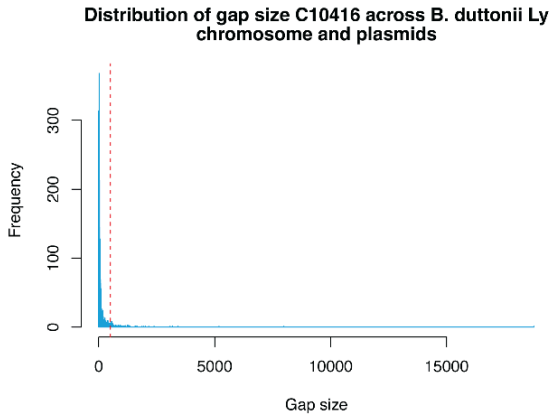
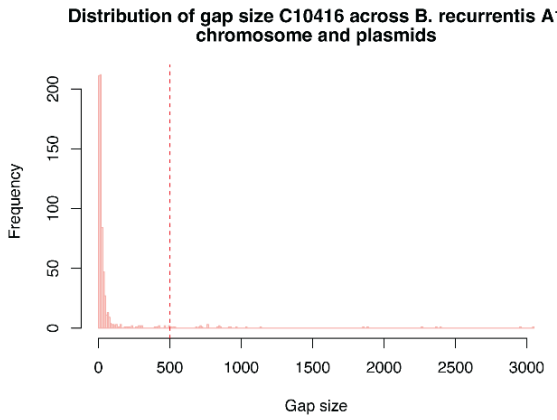
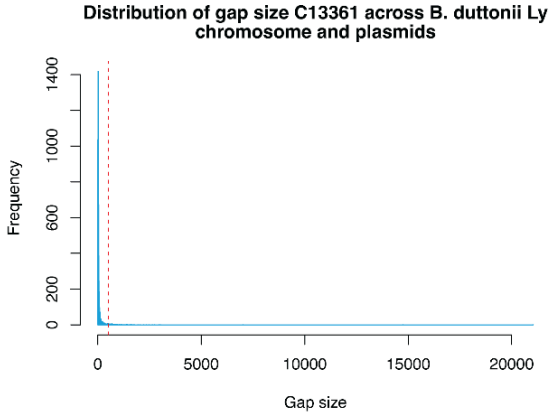
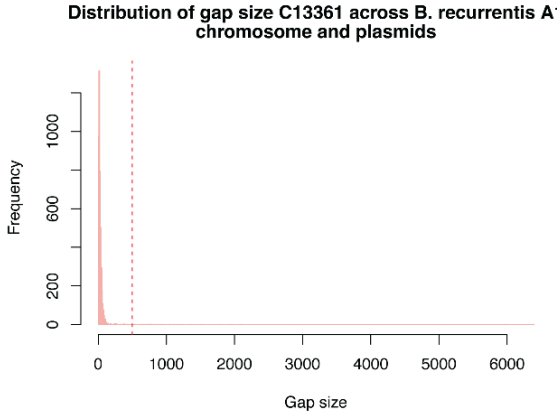


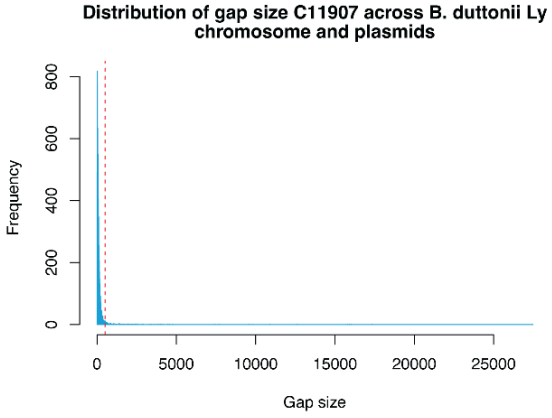
Figure S15. Circos plots providing the per sample coverage across the *B. duttonii Ly* reference genome. Circos plot providing the coverage of ancient genomes across the *B. duttonii* chromosome and plasmids when aligned to the *B. duttonii Ly* reference genome (GCF_000019685.1). A window size of 100bp for the chromosome and 10bp for the plasmids was used to provide the normalised coverage per window plotted.

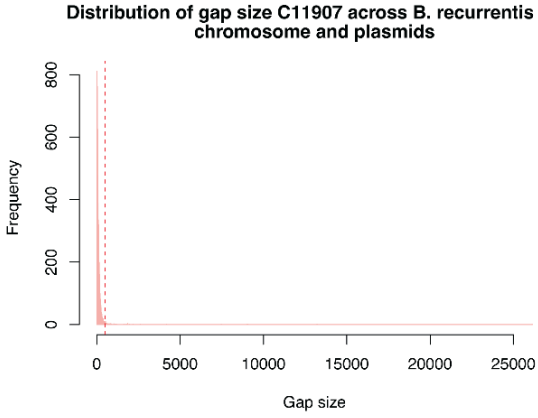

Distribution of gap size C10416 across B. duttonii Ly chromosome and plasmids
Frequency
0
100
200
300
0
5000
10000
15000
Gap size

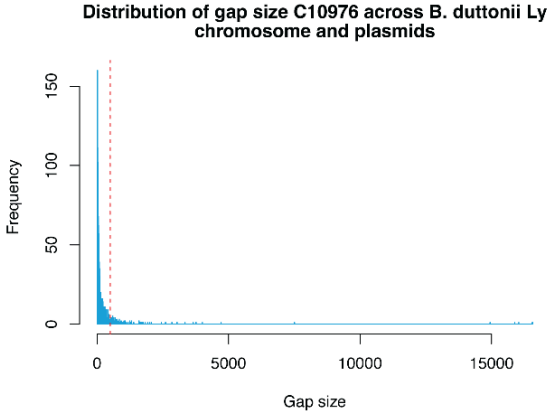

Distribution of gap size C10416 across B. recurrentis A1 chromosome and plasmids
Frequency
0
50
100
150
200
0
500
1000
1500
2000
2500
3000
Gap size


Distribution of gap size C13361 across B. duttonii Ly chromosome and plasmids
Frequency
0
200
600
1000
1400
0
5000
10000
15000
20000
Gap size


Distribution of gap size C13361 across B. recurrentis A1 chromosome and plasmids
Frequency
0
200
600
1000
0
1000
2000
3000
4000
5000
6000
Gap size


Distribution of gap size C11907 across B. duttonii Ly chromosome and plasmids
Frequency
0
200
400
600
800
0
5000
10000
15000
20000
25000
Gap size


Distribution of gap size C11907 across B. recurrentis A1 chromosome and plasmids
Frequency
0
200
400
600
800
0
5000
10000
15000
20000
25000
Gap size


Distribution of gap size C10976 across B. duttonii Ly chromosome and plasmids
Frequency
0
50
100
150
0
5000
10000
15000
Gap size

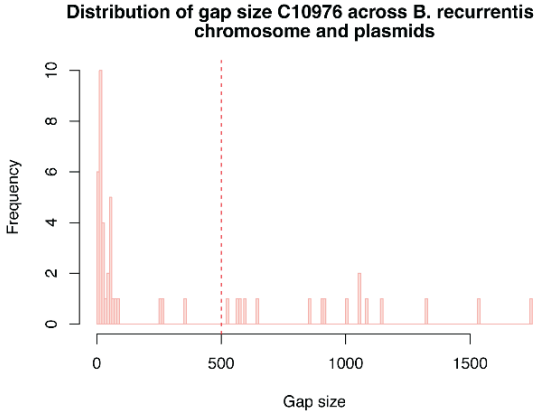

Distribution of gap size C10976 across B. recurrentis A1 chromosome and plasmids
Frequency
0
2
4
6
8
10
0
500
1000
1500
Gap size

Figure S16. Gap size distribution plot to identify deletions over 500bp. Missingness distribution across ancient individuals from this study when aligned to the *B. duttonii* Ly chromosome (blue) and *B. recurrentis* A1 chromosome and plasmids (pink). Positions of missingness window above 500bp (red dashed line) are reported in **Table S7**.

Supplementary Tables

Table S1	All modern and ancient genomes used in this analysis.

Table S2	Assessment of temporal signal following the BactDating root-to-tip evaluation procedure.
Table S3	BEAST2 posterior probability estimates following assessment of the results over three demographic models tested using both strict and relaxed clock priors.
Table S4	Coverage across plasmids when aligned to B. duttonii Ly reference genome
Table S5	Normalised coverage after filtering for the 71 genes post filtering showing temporal patterning across our dataset
Table S6	Identified gene ontology annotations in the 71 filtered genes and gene clusters from EggNoGG and interproScan
Table S7	Identified regions of missingness greater than 500bp when aligned to B. recurrentis A1 and B. duttonii Ly chromosome

Supplementary Note S1

Archaeological Context:
Wetwang Slack (Ian Armit):
The individual with LBRF is Burial 240 at Wetwang Slack, an Arras-style square-ditched barrow cemetery from East Yorkshire. The human remains are kept at the University of Bradford on loan from the Hull and East Riding Museum. This individual, identified osteologically as an adult female, was buried without any accompanying grave goods under a square barrow. The lack of grave goods is not especially unusual at Wetwang Slack, even for primary barrow burials such as this. This individual has been identified as the mother or daughter (order currently unknown) of another adult female (Burial 303), and a second-degree relative of a third adult female (Burial 270), both also primary burials under square barrows. This all suggests that she was a locally-born member of the community and of reasonably high status (in the sense that she was a primary rather than secondary burial - but that is true of around half the burials at the site, so not high status in any major hierarchical sense).

Poulton (Kevin Cootes): Sk 435 was recovered during the excavation of a medieval burial ground and associated chapel in rural Poulton, Cheshire, England. The site was used for interment of individuals from a small farming community, with historical documents and radiocarbon dates indicating use between the 13th and 16th centuries CE[43,58]. The remains are curated at Liverpool John Moores University as part of their teaching and research collection. Sk 435 was identified osteologically as a mature adult of undetermined sex, who was approximately 40-44 years old at the time of death. The body was laid out in typical medieval fashion, extended in a supine position with west–east orientation. Neither grave goods nor a burial marker were present. Continual re-use of the cemetery resulted in the individual's legs

and portions of the pelvis being removed by later interments. Radiocarbon dating of the remains provided a date of 1290-1390 cal. CE at 2 δ (Wk52986. 646 ± 14BP).

Fishmongers (Adelle Bricking, Graham Mullan, Linda Wilson): Fishmonger's Swallet is a small stream sink located near the village of Alveston in the Vale of Berkeley, South Gloucestershire. Initial clearing of the swallet was undertaken by Clive Grace, a local fishmonger, from whom the site derives its name. Subsequently, David Hardwick and the Hades Caving Club assumed control of the dig, revealing a substantial quantity of disarticulated human and animal bones, particularly canids, concentrated within an area designated the Bone Idle Chamber (Hardwick 2022). This discovery attracted the attention of the Channel 4 programme Time Team and a three-day archaeological investigation conducted by the team uncovered additional human and animal bone fragments, along with evidence of postholes on the surface near the swallet (Horton 2022).

Post-excavation analysis of the human remains by Cox and Loe (2022) showed a minimum number of five individuals, but possibly six adult females and five adult males represented by the assemblage. The degree of fragmentation and environmental staining is limiting for macroscopic observations, nevertheless some evidence of pathology (Paget's disease, degenerative joint disease and abscesses) and perimortem trauma was seen on some elements (Cox and Loe 2022).

It is important to note that a small creek flows into the swallet from the surface, carrying modern debris down a short vertical shaft into the Bone Idle Chamber. The co-mingling of modern and archaeological material means it is impossible to interpret the depositional history of the site stratigraphically, so radiocarbon dating is essential. A total of sixteen radiocarbon dates have been obtained from skeletal material and recent programme comprising seven human (BRAMS-5057, BRAMS-5058, BRAMS-5059, BRAMS-5060) and six canid (BRAMS-5061, BRAMS-5062, BRAMS-5063, BRAMS-6671, BRAMS-6672, BRAMS-6673) bone have produced dates centred on the later Iron Age (Bricking et al. 2022, Bricking and Mullan forthcoming).

References:

Bricking, A., Peto, J., Horton, M. and Mullan, G. 2022. Fishmonger's Swallet, Alveston, Gloucestershire, radiocarbon dating. Proceedings of the University of Bristol Spelaeological Society 29(1), 29-32.

Cox, M. and Loe, L. 2022. The human skeletal remains from Fishmonger's Swallet, Alveston, Gloucestershire: Evidence for anthropogenic modification. Proceedings of the University of Bristol Spelaeological Society 29(1), 33-66.

Hardwick, D. 2022. Fishmonger's Swallet, near Alveston, Gloucestershire: Description and history. Proceedings of the University of Bristol Spelaeological Society 29(1), 7-18.

Horton, M. 2022. Archaeology and Television at Fishmonger's Swallet. Proceedings of the University of Bristol Spelaeological Society 29(1), 23-28.

Canterbury (Adelina Teoaca): Sample C11907 comprised dentine powder taken from a left maxillary third molar from a complete cranium (CW 29) held in the archives of Canterbury Archaeological Trust (CAT) in Canterbury, Kent. The provenance and date of cranium are uncertain but the scope of CATs archive both today and historically mean that it must come from eastern Kent and date to 2000-700 years ago (0-1300 cal. CE) .

References

1. Andersson, J.O., and Andersson, S.G. (1999). Genome degradation is an ongoing process in Rickettsia. Mol. Biol. Evol. *16*, 1178–1191.

2. Alsmark, C.M., Frank, A.C., Karlberg, E.O., Legault, B.-A., Ardell, D.H., Canbäck, B., Eriksson, A.-S., Näslund, A.K., Handley, S.A., Huvet, M., et al. (2004). The louse-borne human pathogen Bartonella quintana is a genomic derivative of the zoonotic agent Bartonella henselae. Proc. Natl. Acad. Sci. U. S. A. *101*, 9716–9721.

3. Lescot, M., Audic, S., Robert, C., Nguyen, T.T., Blanc, G., Cutler, S.J., Wincker, P., Couloux, A., Claverie, J.-M., Raoult, D., et al. (2008). The genome of Borrelia recurrentis, the agent of deadly louse-borne relapsing fever, is a degraded subset of tick-borne Borrelia duttonii. PLoS Genet. *4*, e1000185.

4. Talagrand-Reboul, E., Boyer, P.H., Bergström, S., Vial, L., and Boulanger, N. (2018).

Relapsing Fevers: Neglected Tick-Borne Diseases. Front. Cell. Infect. Microbiol. *8*, 98.

5. ECDC (2017). European Centre for Disease Prevention and Control. https://www.ecdc.europa.eu/en/louse-borne-relapsing-fever/facts.

6. McCall, P.J., Hume, J.C.C., Motshegwa, K., Pignatelli, P., Talbert, A., and Kisinza, W. (2007). Does tick-borne relapsing fever have an animal reservoir in East Africa? Vector Borne Zoonotic Dis. *7*, 659–666.

7. Drancourt, M. (2020). 80 - Relapsing Fever and Borrelioses. In Hunter's Tropical Medicine and Emerging Infectious Diseases (Tenth Edition), E. T. Ryan, D. R. Hill, T. Solomon, N. E. Aronson, and T. P. Endy, eds. (Elsevier), pp. 641–645.

8. Marosevic, D., Margos, G., Wallich, R., Wieser, A., Sing, A., and Fingerle, V. (2017). First insights in the variability of Borrelia recurrentis genomes. PLoS Negl. Trop. Dis. *11*, e0005865.

9. Rutty, J. (1770). A Chronological History of the Weather and Seasons, and of the Prevailing Diseases in Dublin: With the Various Periods, Successions, and Revolutions, During the Space of Forty Years : with a Comparative View of the Difference of the Irish Climate and Diseases, and Those of England and Other Countries (Robinson and Roberts).

10. Trevisan, G., Cinco, M., Trevisini, S., di Meo, N., Ruscio, M., Forgione, P., and Bonin, S. (2021). Borreliae Part 2: Borrelia Relapsing Fever Group and Unclassified Borrelia. Biology *10*. 10.3390/biology10111117.

11. Shaw, M.B. (1933). A Short History of the Sweating Sickness. Ann. Med. Hist. *5*, 246–274.

12. Heyman, P., Simons, L., and Cochez, C. (2014). Were the English sweating sickness and the Picardy sweat caused by hantaviruses? Viruses *6*, 151–171.

13. Yimer, M., Abera, B., Mulu, W., Bezabih, B., and Mohammed, J. (2014). Prevalence and risk factors of louse-borne relapsing fever in high risk populations in Bahir Dar city Northwest, Ethiopia. BMC Res. Notes *7*, 615.

14. Warrell, D.A. (2019). Louse-borne relapsing fever (Borrelia recurrentis infection). Epidemiol. Infect. *147*, e106.

15. Pavia, C.S. (2020). Chapter 3 - Immunologic detection of Lyme disease and the related borrelioses. In Methods in Microbiology, C. S. Pavia and V. Gurtler, eds. (Academic Press), pp. 41–74.

16. Cutler, S.J. (2015). Relapsing Fever Borreliae: A Global Review. Clin. Lab. Med. *35*, 847–865.

17. Guellil, M., Kersten, O., Namouchi, A., Bauer, E.L., Derrick, M., Jensen, A.Ø., Stenseth, N.C., and Bramanti, B. (2018). Genomic blueprint of a relapsing fever pathogen in 15th

century Scandinavia. Proc. Natl. Acad. Sci. U. S. A. *115*, 10422–10427.

18. Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., de Filippo, C., et al. (2012). A High-Coverage Genome Sequence from an Archaic Denisovan Individual. Science *338*, 222–226.

19. Jay, M., Haselgrove, C., Hamilton, D., Hill, J.D., and Dent, J. (2012). Chariots and context: New radiocarbon dates from wetwang and the chronology of iron age burials and brooches in east Yorkshire. Oxf. J. Archaeol. *31*, 161–189.

20. Bricking, A., Hayes, A., and Madgwick, R. (2022). An interim report on histological analysis of human bones from Fishmonger's Swallet, Gloucestershire. Proceedings of the University of Bristol Speleological Society *29*, 67–86.

21. Brotherton, P., Endicott, P., Sanchez, J.J., Beaumont, M., Barnett, R., Austin, J., and Cooper, A. (2007). Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. Nucleic Acids Res., gkm588.

22. Briggs, A.W., Stenzel, U., Johnson, P.L.F., Green, R.E., Kelso, J., PrÃ¼fer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., et al. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. Proceedings of the National Academy of Sciences *104*, 14616–14621.

23. Key, F.M., Posth, C., Krause, J., Herbig, A., and Bos, K.I. (2017). Mining Metagenomic Data Sets for Ancient DNA: Recommended Protocols for Authentication. Trends Genet. *33*, 508–520.

24. Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J.A., Gladstone, R.A., Lo, S., Beaudoin, C., Floto, R.A., et al. (2020). Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biol. *21*, 180.

25. Didelot, X., and Wilson, D.J. (2015). ClonalFrameML: efficient inference of recombination in whole bacterial genomes. PLoS Comput. Biol. *11*, e1004041.

26. Didelot, X., Croucher, N.J., Bentley, S.D., Harris, S.R., and Wilson, D.J. (2018). Bayesian inference of ancestral dates on bacterial phylogenetic trees. Nucleic Acids Res. *46*, e134.

27. Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., et al. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. *15*, e1006650.

28. Sikora, M., Canteri, E., Fernandez-Guerra, A., Oskolkov, N., Ågren, R., Hansson, L., Irving-Pease, E.K., Mühlemann, B., Nielsen, S.H., Scorrano, G., et al. (2023). The landscape of ancient human pathogens in Eurasia from the Stone Age to historical times. bioRxiv, 2023.10.06.561165. 10.1101/2023.10.06.561165.

29. Ahmed, N., Dobrindt, U., Hacker, J., and Hasnain, S.E. (2008). Genomic fluidity and pathogenic bacteria: applications in diagnostics, epidemiology and intervention. Nat. Rev. Microbiol. *6*, 387–394.

30. Murray, H., and Errington, J. (2008). Dynamic control of the DNA replication initiation protein DnaA by Soj/ParA. Cell *135*, 74–84.

31. Vink, C., Rudenko, G., and Seifert, H.S. (2012). Microbial antigenic variation mediated by homologous DNA recombination. FEMS Microbiol. Rev. *36*, 917–948.

32. Röttgerding, F., and Kraiczy, P. (2020). Immune Evasion Strategies of Relapsing Fever Spirochetes. Front. Immunol. *11*, 1560.

33. Gilmore, R.D., Armstrong, B.A., Brandt, K.S., Van Gundy, T.J., Hojgaard, A., Lopez, J.E., and Kneubehl, A.R. (2023). Analysis of variable major protein antigenic variation in the relapsing fever spirochete, Borrelia miyamotoi, in response to polyclonal antibody selection pressure. PLoS One *18*, e0281942.

34. Fischer, J.R., LeBlanc, K.T., and Leong, J.M. (2006). Fibronectin binding protein BBK32 of the Lyme disease spirochete promotes bacterial attachment to glycosaminoglycans. Infect. Immun. *74*, 435–441.

35. Wang, X.-G., Kidder, J.M., Scagliotti, J.P., Klempner, M.S., Noring, R., and Hu, L.T. (2004). Analysis of differences in the functional properties of the substrate binding proteins of the Borrelia burgdorferi oligopeptide permease (Opp) operon. J. Bacteriol. *186*, 51–60.

36. Rascovan, N., Sjögren, K.-G., Kristiansen, K., Nielsen, R., Willerslev, E., Desnues, C., and Rasmussen, S. (2019). Emergence and Spread of Basal Lineages of Yersinia pestis during the Neolithic Decline. Cell *176*, 295–305.e10.

37. Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. Nature *522*, 167–172.

38. ECDC (2015). Rapid Risk Assessment: Louse-borne relapsing fever in the EU, 19 November 2015.

39. Matic, I. (2019). Mutation Rate Heterogeneity Increases Odds of Survival in Unpredictable Environments. Mol. Cell *75*, 421–425.

40. Murray, G.G.R., Balmer, A.J., Herbert, J., Hadjirin, N.F., Kemp, C.L., Matuszewska, M., Bruchmann, S., Hossain, A.S.M.M., Gottschalk, M., Tucker, A.W., et al. (2021). Mutation rate dynamics reflect ecological change in an emerging zoonotic pathogen. PLoS Genet. *17*, e1009864.

41. Murray, G.G.R., Charlesworth, J., Miller, E.L., Casey, M.J., Lloyd, C.T., Gottschalk, M., Tucker, A.W.D., Welch, J.J., and Weinert, L.A. (2021). Genome Reduction Is Associated

with Bacterial Pathogenicity across Different Scales of Temporal and Ecological Divergence. Mol. Biol. Evol. *38*, 1570–1579.

42. Zückert, W.R., and Meyer, J. (1996). Circular and linear plasmids of Lyme disease spirochetes have extensive homology: characterization of a repeated DNA element. J. Bacteriol. *178*, 2287–2298.

43. Cootes, K., Axworthy, J., Borrini, M., Carlin, R., Irish, J., Jordan, D., King, M., Russ, H., Swallow, R., Thomas, M., et al. (2023). Poulton, Cheshire: The investigation of a rural chapel in an evolving medieval landscape. Church Archaeology *23*, 43–53.

44. Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., et al. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl. Acad. Sci. U. S. A. *110*, 15758–15763.

45. Rohland, N., Glocke, I., Aximu-Petri, A., and Meyer, M. (2018). Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. Nat. Protoc. *13*, 2447–2461.

46. Gansauge, M.-T., Aximu-Petri, A., Nagel, S., and Meyer, M. (2020). Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. Nat. Protoc. *15*, 2279–2300.

47. Kircher, M., Sawyer, S., and Meyer, M. (2011). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res., gkr771.

48. Yates, J.A.F., Fellows Yates, J.A., Lamnidis, T.C., Borry, M., Valtueña, A.A., Fagernäs, Z., Clayton, S., Garcia, M.U., Neukamm, J., and Peltzer, A. Reproducible, portable, and efficient ancient genome reconstruction with nf-core/eager. Preprint, 10.1101/2020.06.11.145615 10.1101/2020.06.11.145615.

49. Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes *9*, 88.

50. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760.

51. Poullet, M., and Orlando, L. (2020). Assessing DNA Sequence Alignment Methods for Characterizing Ancient Genomes and Methylomes. Frontiers in Ecology and Evolution *8*, 105.

52. Wood, D.E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome Biol. *20*, 257.

53. Neukamm, J., Peltzer, A., and Nieselt, K. (2021). DamageProfiler: Fast damage pattern calculation for ancient DNA. Bioinformatics. 10.1093/bioinformatics/btab190.

54. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

55. Li, H. htsbox (Github).

56. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589.

57. Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236–1240.

58. Emery, M.M., Gibbins, D.J.L., and Matthews, K.J. (1996). The Archaeology of an Ecclesiastical Landscape: Chapel House Farm, Poulton (Cheshire) 1995 (Chester City Council, Department of Development and Leisure Services).