Yersinia pestis and the Plague of Justinian 541–543 AD: a genomic analysis


Summary

Background Yersinia pestis has caused at least three human plague pandemics. The second (Black Death, 14–17th centuries) and third (19–20th centuries) have been genetically characterised, but there is only a limited understanding of the first pandemic, the Plague of Justinian (6–8th centuries). To address this gap, we sequenced and analysed draft genomes of Y pestis obtained from two individuals who died in the first pandemic.

Methods Teeth were removed from two individuals (known as A120 and A76) from the early medieval Aschheim-Bajuwarenring cemetery (Aschheim, Bavaria, Germany). We isolated DNA from the teeth using a modified phenol-chloroform method. We screened DNA extracts for the presence of the Y pestis-specific pla gene on the pPCP1 plasmid using primers and standards from an established assay, enriched the DNA, and then sequenced it. We reconstructed draft genomes of the infectious Y pestis strains, compared them with a database of genomes from 131 Y pestis strains from the second and third pandemics, and constructed a maximum likelihood phylogenetic tree.

Findings Radiocarbon dating of both individuals (A120 to 533 AD [plus or minus 98 years]; A76 to 504 AD [plus or minus 61 years]) places them in the timeframe of the first pandemic. Our phylogeny contains a novel branch (100% bootstrap at all relevant nodes) leading to the two Justinian samples. This branch has no known contemporary representatives, and thus is either extinct or unsampled in wild rodent reservoirs. The Justinian branch is interleaved between two extant groups, 0.ANT1 and 0.ANT2, and is distant from strains associated with the second and third pandemics.

Interpretation We conclude that the Y pestis lineages that caused the Plague of Justinian and the Black Death 800 years later were independent emergences from rodents into human beings. These results show that rodent species worldwide represent important reservoirs for the repeated emergence of diverse lineages of Y pestis into human populations.


Introduction Between 541 and 543 AD, the Plague of Justinian, traditionally regarded as the first of three human plague pandemics, spread from either central Asia or Africa across the Mediterranean basin into Europe, killing an estimated 100 million people according to the contemporary scholar Procopius (although this history is disputed), contributing to the end of the Roman empire, and marking the transition from the classical to the Medieval period. Subsequent outbreaks of this disease occurred in 8–12 year cycles for two centuries after the initial epidemic with estimated mortality of 15–40%. Although suspected to have been caused by Yersinia pestis, uncertainties in historical epidemiology have led some researchers to propose other pathogens as the causative agent of the Plague of Justinian, such as an influenza virus. However, the symptoms described by Procopius are very similar to those later reported during the second pandemic of the 14–17th centuries (ie, the Black Death), from which Y pestis has been genomically characterised, implicating this bacterial species as the infectious agent of the first pandemic.

Although the presence of Y pestis in victims of the Plague of Justinian has been confirmed by the sequencing of Y pestis-specific genomic regions from skeletal material of individuals buried in Sens, France, and Bavaria, Germany, a detailed genomic characterisation of the strains that caused the first pandemic and their relation to subsequent plague pandemics is absent. In particular, it is unclear whether the Justinian strains were direct ancestors of the Black Death strains and later plague outbreaks, or form a novel lineage of Y pestis. Similarly, it is unclear whether the second and third pandemics were the result of continued re-emergences of one pandemic strain, or were novel emergences from diverse rodent reservoirs. To address these questions we report and analyse draft genomes of Y pestis obtained from individuals who died in the first pandemic.

Methods DNA extraction, screening, preparation, and indexing Teeth were removed from two individuals (known as A120 and A76) from the Aschheim-Bajuwarenring...
Figure 1: Source of materials
(A) Aerial photograph showing the location of the grave of individual A120 in the Aschheim–Bajuwarenning cemetery, Bavaria, Germany. (B) Shared grave of A120 and two other individuals. (C) Skeleton of A120. (D) Sampled tooth from A120. (E) Goods obtained from the grave of A120 that were used to estimate the age of this burial to 525–550 AD. Parts (A) and (B) reproduced with permission of Hans Peter Volpert. Part (E) reproduced with permission of Doris Gutsmiedl-Schumann.

cemetery, an early medieval graveyard in Aschheim, Bavaria, Germany, which contained 438 individuals buried in single and a few multiple graves. All laboratory procedures were done at the McMaster Ancient DNA Centre (McMaster University, Hamilton, ON, Canada).

We extracted DNA using a modified phenol-chloroform method as previously described. We screened DNA extracts for the presence of the *Y. pestis*-specific *pla* gene on the pPCP1 plasmid using primers and standards from an established assay with minor modifications of the cycling conditions (appendix 1 p 2). We followed a previously published protocol with recommended modifications for library preparation of all our DNA extracts (appendix 1 pp 2–3).

DNA enrichment
We designed a 1 million Agilent SureSelect DNA capture array (Agilent, Santa Clara, USA) using the program PanArray (version 1.0) as per the manufacturer’s instructions. We aimed to target and capture sequences from the core *Y. pestis* chromosome (strain CO92, NC_003143), three primary plasmids (pPCP1, pMT1, and pCD1), and genes from 155 other *Y. pestis* strains not present in the core chromosomal genome (all targets are listed in appendix 1 pp 18–21). Although this array contained probes representing all unique genomic components from 156 *Y. pestis* strains it was not capable of capturing unique genomic sequences that might have been present in the Justinian lineage. 2 million 60-mer probes were tiled every five bases and distributed in genomically contiguous sections across two 1 million arrays. In-solution enrichment was done according to the manufacturer’s protocol (Myroarray, Ann Arbor, MI, USA). We designed three in-solution bait sets for the complete human mitochondrial genome, the complete *Y. pestis* CO92 genome, and all 2298 previously confirmed polymorphic sites plus all novel single-nucleotide polymorphisms (SNPs) discovered in our novel strains. We did 46 serial targeted enrichments from 23 uracil DNA glycosylase (UDG)-treated DNA libraries derived from four teeth from individual A120 and one tooth from A76.

We sequenced the enriched libraries on the Illumina HiSeq 1500 platform (Illumina, San Diego, CA, USA). Adapter sequences 3’ of the inserts were trimmed with cutadapt (version 1.0). We then merged the reads using FLASH (version 1.2.4). Finally, we mapped the trimmed
merged reads using BWA (version 0.7.5; appendix 1 p 5). All read files have been deposited at the National Centre for Biotechnology Information Sequence Read Archive (accession number SRP033879).

To ensure a minimum of five-times coverage of all novel SNPs discovered in the Justinian strains in addition to known SNPs within Y pestis, we designed 7083 80-mer in-solution baits and serially enriched and sequenced 14 A120 and A76 UDG-treated DNA libraries.

We identified SNPs specific to the Justinian samples using Geneious (version 6.1.2) at five-times read coverage and 90% variant frequency and Freebayes (version 0.9.9) at five-times read coverage and 90% variant frequency with the parameters: use mapping quality, exclude unobserved genotypes, alternate allele frequency, ploidy of one, minimum mapping quality of 20, minimum base quality of 20, exclude multiple nucleotide variants, no complex events, no indels, and disabled population priors. SNPs that were called by only one of the two programs were visually examined for authenticity and were only included in subsequent phylogenetic analysis when total variant calls were at least 90% (final SNPs are in appendix 2).

A maximum likelihood phylogenetic tree was estimated for 133 strains of Y pestis with an additional strain of Y pseudotuberculosis (IP32953) used as an outgroup (appendix 1 p 17). We did analysis with the PhyML program (version 3) using the generalised time-reversible GTR+ , model of nucleotide substitution with substitution rates (GTR) and among-site rate variation (Δ) estimated from the empirical data. To infer the rate and timescale of Y pestis evolution we used a Bayesian Markov chain Monte Carlo approach (BEAST program, version 1.7.5).

Role of the funding source
The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to all the data in the study. DMW, ECH, and HP made the final decision to submit for publication.

Results
Radiocarbon dating of both individuals (A120 to 533 AD [plus or minus 98 years]; A76 to 504 AD [plus or minus 61 years]), and the dating of associated grave goods (figure 1b), places them in the timeframe of the first pandemic. To assess the authenticity of the DNA isolated from the remains, we used in-solution enrichment targeting the human mitochondrial genome and sequenced samples to 218-times (A120) and 1600-times (A76) coverage. The consensus sequences obtained were distinct haplotypes matching closest to J1c3 (A76) and H8c (A120; appendix 1 pp 22–23), both of

![Figure 2: Coverage plots (log scale, outer ring) and GC percentage content (orange inner ring) for the draft genome of Justinian sample A120+A76](image-url)
which are identified at high frequencies in contemporary European and Middle Eastern populations. Damage profiles (appendix 1 pp 12–15) of libraries not treated with UDG were consistent with ancient damaged DNA and support the authenticity of our samples.

DNA enrichment and sequencing produced 5,118,424 bp of sequence. After trimming and merging, raw mapped reads with identical 5’ and 3’ end coordinates and strand origin (direction) were collapsed to establish unique counts. This procedure resulted in 683,191 reads of an average length of 51-8 bp that mapped to the CO92 reference, with an average of 7-6 times coverage for 91-5% of the chromosome, 99-9% for pCD1, 96-7% for pMT1, and 100-0% for pCP1 (figure 2, appendix 1 p 22). Because the overall quality of the sequences and the total coverage obtained for strain A76 was much lower than that for strain A120, we analysed data from A76 and A120 together as well as A120 individually.

We used two approaches to infer the placement of Justinian strains (A120 and A120+A76) in the Y pestis phylogeny. First, we queried the allele state in A120 and A120+A76 at 2,298 SNPs identified in a large-scale phylogenetic analysis of Y pestis. Of these SNPs, 1,509 positions from A120 and 1,673 positions from A120+A76 were covered at more than four-times depth and were present in at least 90% of the sequences obtained (appendix 1 pp 25–27), yielding identical and unambiguous placement of the Justinian samples along branch N04-N05 in the phylogeny of Cui and colleagues. Second, we inferred a maximum likelihood phylogenetic tree (figure 3) using 2,268 polymorphic positions in 130 contemporary Y pestis strains, the draft genome from the second pandemic strain, and the two Justinian samples (appendix 3). One (representative) Y pseudotuberculosis genome was used as an outgroup.

Our phylogeny is much the same as that obtained previously but contains a novel branch (100% bootstrap at all relevant nodes) leading to the two Justinian samples. Notably, this branch has no known contemporary representatives, and so is either extinct or unsampled in wild rodent reservoirs. The Justinian branch is interleafed between two extant groups, 0.ANT1 and 0.ANT2 (first pandemic; Justinian’s Plague—540 AD; CEF).
and 0.ANT2 (figure 3), and distant from strains associated with the second and third pandemics. These extant groups are identified in long-term plague foci in the Xinjiang region of China where they are ecologically established within rodent and flea populations. These findings are compatible with the theory that the first pandemic emerged from rodents in long-term plague foci in China, not Africa as was originally proposed by Procopius, and spread east to Europe along established trade routes (eg, the Silk Road; figure 4).

To further assess whether the Y pestis strains associated with the Justinian plague represented a novel emergence event with no later sampled descendants we compared the maximum likelihood tree with a second phylogeny in which the Justinian strains were constrained to be the direct ancestor of those associated with the second and third pandemics. The likelihood of the maximum likelihood tree (figure 3) was significantly better than the competing phylogeny (p<0.001, Shimodaira-Hasegawa test), strongly supporting the idea that the Plague of Justinian was an independent emergence that did not give rise to later human pandemics. Thus, the Black Death is the result of a separate emergence of Y pestis, some 800 years later, from rodents into human beings that led to pandemic plague.

The contrasting phylogenetic positions of the Plague of Justinian and the Black Death strains are noteworthy (figure 3). The Black Death strain sits near the base of branch 1, strongly suggesting that the second pandemic gave rise to all later Y pestis strains along branch 1, including those directly responsible for the third pandemic (1.ORI group; figure 3). These descendants include the 1.ANT group, which is currently ecologically established in rodent populations in central Africa; the 1.IN1, 1.IN2, and 1.IN3 groups, currently identified in rodent populations in western, central, and southern China, respectively; and the 1.ORI group, which arose in southeastern China and spread from Hong Kong worldwide during the third pandemic, leading to the establishment of new rodent foci in Africa, Asia, North America, and South America that persist to this day. A viable explanation for these phylogeographic patterns is that Y pestis spread from Asia to Europe during the second pandemic and then from Europe or northern Africa to central Africa establishing the 1.ANT group. Under this scenario, Y pestis also spread back from China to Europe (or northern Africa) along the same established trade routes giving rise to the 1.IN populations, which later gave rise to the 1.ORI group that caused the third pandemic (figure 4). Historical records are consistent with this scenario because the 1.IN3 group is identified in the Yunnan province of China, where the third pandemic originated. Additionally, some 1.IN3 strains exhibit the derived orientalis phenotype, which is conserved in all 1.ORI strains but not identified in any other groups.

As in other Y pestis phylogenies, our maximum likelihood tree is notable for a substantial variation in branch length that has no correlation with sampling time (eg, the long branch leading to the 0.PE2 group; figure 3). In particular, the lineage of the Justinian strains is roughly the same distance from the root of the tree as the strain from the Black Death, despite being about 800 years older.

Figure 4: Hypothetical scenario for the geographic spread of Yersinia pestis
The lineage associated with the first pandemic might have arisen in Asia (China outlined) and then travelled along established trade routes, such as the Silk Road (purple lines), to Europe. The lineage associated with the second pandemic might also have spread to Europe from Asia along trade routes and from Europe via trade routes along the Nile river to central Africa. It might also have travelled back to China to establish the 1.IN populations, which later gave rise to the 1.ORI group involved in the third pandemic. The orange square shows the location of the Justinian samples, the green square in the UK shows the location of the East Smithfield Black Death samples, the green square in central Africa shows the location of the extant 1.ANT group, and the blue square in China shows Hong Kong, from which the third pandemic was disseminated globally.
virulence. Of the 66 unique chromosomal and plasmid SNPs identified in A120 and A76 combined, 23 were non-synonymous, two of which are identified in well-characterised virulence factors—ail and yopJ (appendix 1 pp 27–29). Ail is an outer membrane protein that promotes bacterial resistance to complement-mediated killing and adhesion to host cells. A novel Arg→Gly aminoacid substitution occurs in an extracellular region (loop 3) of the Ail protein from the Justinian strains, which could potentially increase resistance to complement-mediated killing or adherence, or both.34,35 YopJ is an effector protein with acetyltransf erase activity that is secreted into host cells. Compared with CO92, the Lys→Glu substitution in YopJ of the Justinian (and Black Death) strains increases the antihost activity of this effector.6 Additionally, using our new pan-array we were able to identify a roughly 15 kb genomic island in the Justinian strains and also in the Black Death strain that we previously sequenced (appendix 1 p 16).1 This genomic island, referred to as DFR 4 (different region 4),16 was lost in some Y pestis strains, such as CO92. DFR 4 contains several genes with potential roles in virulence, including ccm2a, which encodes a multidrug ABC transporter ATPase. DFR 4 is currently identified in Y pseudotuberculosis and some Y pestis strains (in 0.PE4 and other groups on branch 2). This island is an example of the decay common to the highly plastic Y pestis genome.16 What part DFR 4 and novel non-synonymous SNPs might have played in virulence is unknown. However, additional sequences from ancient plaque pandemics will allow us to assess the virulence capacity of these mutations. In the present report we have focused exclusively on possible genomic differences within our ancient Y pestis strains that might have accounted for the high mortality associated with the first plague pandemic. However, host and environmental factors also probably had a role in determining mortality.

Discussion

Our genomic analysis shows that the strains of Y pestis involved in the Plague of Justinian form a novel branch on the Y pestis phylogeny. This branch has no known contemporary representatives, and so is either extinct or unsampled in wild rodent reservoirs. The Justinian branch is interleaved between two extant groups, O.ANT1 and 0.ANT2, and is distant from strains associated with the second and third pandemics. Our phylogeny showed a substantial variation in branch length that had no correlation with sampling time. We identified 23 non-synonymous SNPs in the Justinian strains, two of which occurred in well-characterised virulence factors—ail and yopJ. The Justinian strains also include the genomic island known as DFR 4, which contains several genes with potential roles in virulence. These genomic components, combined with host and environmental factors, might have contributed to the high (up to 40%) mortality associated with the Plague of Justinian.

In our Bayesian Markov chain Monte Carlo analysis, use of various methods to account for rate variation did not produce a robust estimate for either the rate of nucleotide substitution or times to common ancestry (figure 3; appendix 1 pp 6–7). An absence of temporal structure, which complicates all molecular clock dating, was confirmed through regressions of date (ie, year) of sampling against root-to-tip genetic distance on the maximum likelihood phylogeny for all strains (correlation coefficient 0.25), contemporary strains only (0.27), and the Black Death strain and its descendants (0.20). We therefore conclude, as have others,7 that the evolution of Y pestis is characterised by extensive rate variation, probably attributable to a combination of fluctuations in the number of bacterial replication cycles (generations) per unit time, such as between endemic and epidemic cycles, or changes in selection pressures through time, including localised adaptive evolution, or combinations of these factors. Whatever the cause of this rate variation, these data suggest that previous molecular-clock-derived estimates of the timescale of Y pestis evolution,5,11 including the date of its divergence from Y pseudotuberculosis, might be erroneous. Such extensive rate variation contrasts with studies of other bacteria such as Mycobacterium leprae in which samples collected over about 1000 years show substantial temporal structure.11 Importantly, without a firm timescale for Y pestis evolution we cannot exclude that earlier dated epidemics (Plague of Athens of 430 BC) or pandemics (Antonine Plague 165–180 AD) of unknown cause are potentially derived from repeated, novel emergences of Y pestis into human populations as was the case with later pandemics.

Estimated mortality from the Justinian Plague was up to 40%,1 which might mean that the strains of Y pestis associated with this pandemic were of intrinsically high virulence. Of the 66 unique chromosomal and plasmid SNPs identified in A120 and A76 combined, 23 were non-synonymous, two of which are identified in well-characterised virulence factors—ail and yopJ (appendix 1 pp 27–29). Ail is an outer membrane protein that promotes bacterial resistance to complement-mediated killing and adhesion to host cells. A novel Arg→Gly aminoacid substitution occurs in an extracellular region (loop 3) of the Ail protein from the Justinian strains, which could potentially increase resistance to complement-mediated killing or adherence, or both.34,35 YopJ is an effector protein with acetyltransf erase activity that is secreted into host cells. Compared with CO92, the Lys→Glu substitution in YopJ of the Justinian (and Black Death) strains increases the antihost activity of this effector.6 Additionally, using our new pan-array we were able to identify a roughly 15 kb genomic island in the Justinian strains and also in the Black Death strain that we previously sequenced (appendix 1 p 16).1 This genomic island, referred to as DFR 4 (different region 4),16 was lost in some Y pestis strains, such as CO92. DFR 4 contains several genes with potential roles in virulence, including ccm2a, which encodes a multidrug ABC transporter ATPase. DFR 4 is currently identified in Y pseudotuberculosis and some Y pestis strains (in 0.PE4 and other groups on branch 2). This island is an example of the decay common to the highly plastic Y pestis genome.16 What part DFR 4 and novel non-synonymous SNPs might have played in virulence is unknown. However, additional sequences from ancient plague pandemics will allow us to assess the virulence capacity of these mutations. In the present report we have focused exclusively on possible genomic differences within our ancient Y pestis strains that might have accounted for the high mortality associated with the first plague pandemic. However, host and environmental factors also probably had a role in determining mortality.

Discussion

Our genomic analysis shows that the strains of Y pestis involved in the Plague of Justinian form a novel branch on the Y pestis phylogeny. This branch has no known contemporary representatives, and so is either extinct or unsampled in wild rodent reservoirs. The Justinian branch is interleaved between two extant groups, 0.ANT1 and 0.ANT2, and is distant from strains associated with the second and third pandemics. Our phylogeny showed a substantial variation in branch length that had no correlation with sampling time. We identified 23 non-synonymous SNPs in the Justinian strains, two of which occurred in well-characterised virulence factors—ail and yopJ. The Justinian strains also include the genomic island known as DFR 4, which contains several genes with potential roles in virulence. These genomic components, combined with host and environmental factors, might have contributed to the high (up to 40%) mortality associated with the Plague of Justinian.

In our Bayesian Markov chain Monte Carlo analysis, use of various methods to account for rate variation did not produce a robust estimate for either the rate of nucleotide substitution or times to common ancestry (figure 3; appendix 1 pp 6–7). An absence of temporal structure, which complicates all molecular clock dating, was confirmed through regressions of date (ie, year) of sampling against root-to-tip genetic distance on the maximum likelihood phylogeny for all strains (correlation coefficient 0.25), contemporary strains only (0.27), and the Black Death strain and its descendants (0.20). We therefore conclude, as have others,7 that the evolution of Y pestis is characterised by extensive rate variation, probably attributable to a combination of fluctuations in the number of bacterial replication cycles (generations) per unit time, such as between endemic and epidemic cycles, or changes in selection pressures through time, including localised adaptive evolution, or combinations of these factors. Whatever the cause of this rate variation, these data suggest that previous molecular-clock-derived estimates of the timescale of Y pestis evolution,5,11 including the date of its divergence from Y pseudotuberculosis, might be erroneous. Such extensive rate variation contrasts with studies of other bacteria such as Mycobacterium leprae in which samples collected over about 1000 years show substantial temporal structure.11 Importantly, without a firm timescale for Y pestis evolution we cannot exclude that earlier dated epidemics (Plague of Athens of 430 BC) or pandemics (Antonine Plague 165–180 AD) of unknown cause are potentially derived from repeated, novel emergences of Y pestis into human populations as was the case with later pandemics.

Estimated mortality from the Justinian Plague was up to 40%,1 which might mean that the strains of Y pestis associated with this pandemic were of intrinsically high
Our study provides a new perspective on one of the most devastating disease outbreaks in human history and the dynamics of plague emergences in general (panel). We have shown that the strain associated with the Plague of Justinian is distinct from those associated with later human plague pandemics. Thus, we infer that Y. pestis has emerged from rodent reservoirs at several timepoints in history to cause pandemics in human beings. In the case of the lineage associated with the first pandemic, this emergence seems to be an evolutionary dead-end, because no extant descendants of this lineage have been identified so far. By contrast, the lineage associated with the second pandemic was more successful in that it gave rise to new rodent foci, one of which was later responsible for the third pandemic, in turn establishing new endemic rodent foci worldwide. Thus, each successive pandemic is associated with increasingly successful spatial dissemination of Y. pestis; the first seems to have died out, the second spread back and forth between Asia, Europe, and Africa, whereas the third became established worldwide (figure 4).

Why the Y. pestis lineage associated with the Plague of Justinian eventually died out is unclear. That it probably caused human epidemics for several centuries\(^6\) suggests that it was well adapted to human transmission. As a consequence, several viable explanations for its extinction might be a scarcity of susceptible hosts (people, or rodents, or both), insufficient numbers of susceptible hosts in a background of widespread population immunity, or mutations that arose and spread in the human genome that conferred resistance to this particular plague strain. The success of the lineage (or lineages) associated with the two most recent pandemics is probably attributable partly to human mobilisation,\(^8\) since increasing trade between countries and continents is known to have moved people and rodents with Y. pestis infection around the world.\(^2\)

The Plague of Justinian and, indeed, the emergence of all three plague pandemics, might be tightly linked to climatic instability; all were preceded by periods of exceptional rainfall\(^10\)–\(^11\) and ended during periods of climatic stability (around 700–1000 AD in the case of the Plague of Justinian).\(^9\) Irrespective of the effect of climate, the epidemiological pattern that we propose suggests that several Y. pestis lineages, which are currently ecologically established in rodent foci worldwide, remain capable of emerging and igniting epidemics of plague in human beings, as they have repeatedly in the past.

Acknowledgments
We thank Doris Gutsmiedl-Schümann, Anja Pütz, the Bavarian State Office for the Preservation of Monuments, Charlotte Yates, Gerry Wright, Lisa Seifert, and Ann Hering for materials, advice, and assistance. This work was supported by the Michael G DeGroote Institute of Infectious Disease Research, Social Sciences and Humanities Research Council Canada, McMaster University, a Canada Research Chair to HP, the Cowden Endowment at Northern Arizona University, and the US Department of Homeland Security (2010-ST-081-000015; HSHQDC-10-C-00119). JBB is supported by the Northeast Biodefense Center US4-AI057158-Lipkin. ECH is supported by an NHMRC Australia Fellowship.

References
1 Zietz BP, Dunkelberg H. The history of the plague and the research on the causative agent Yersinia pestis. Int J Hyg Env Health 2004; 207: 165–78.
20 Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBolet J 2011; 17: 10–12.

Contributors
DMW, HCS, MH, ECH, PSK, J-MR, and HP designed the study. JK, MH, DNB, AD, JE, MK, CI, IW, J-MR, and GG gathered the data. JK, JWS, MF, AD, BG, NW, JE, MK, JBB, HCS, SDW, TP, DMW, ECH, and HP analysed the data. DP, DJDE, ECH, DMW, and HP wrote the initial draft and all authors contributed to subsequent drafts. DMW, ECH, and HP funded the work.

Conflicts of interest
J-MR has financial interest in Mycnaaray (Ann Arbor, MI, USA) who provided in-solution capture kits baits used in this work. All other authors declare that they have no conflicts of interest.