Microevolution and history of the plague bacillus, *Yersinia pestis*

Mark Achtman*, Giovanna Morelli*, Peixuan Zhu+, Thierry Wirth+, Ines Diehl*, Barica Kusecek*, Amy J. Vogler†, David M. Wagner†, Christopher J. Allender†, W. Ryan Easterday†, Viviane Chenal-Francisque†, Patricia Worsham*, Nicholas R. Thomson††, Julian Parkhill††, Luther E. Lindler‡‡§§, Elisabeth Carniel, and Paul Keim***

*Department of Molecular Biology, Max-Planck Institut für Infektionsbiologie, D-10117 Berlin, Germany; †Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 86011-5640; ‡Yersinia Research Unit, Institut Pasteur, 75724 Paris Cedex 15, France; ¶*U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, Frederick, MD 21702-5011; ‡‡The Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, United Kingdom; §§National Biodefense Analysis and Countermeasures Center, Frederick, MD 21703; ¶¶Department of Bacterial Diseases, Walter Reed Army Institute of Research, Silver Spring, MD 20910; and ††Translational Research Institute, Phoenix, AZ 85004

Communicated by M. S. Meselson, Harvard University, Cambridge, MA, October 28, 2004 (received for review May 7, 2004)

The association of historical plague pandemics with *Yersinia pestis* remains controversial, partly because the evolutionary history of this largely monomorphic bacterium was unknown. The microevolution of *Y. pestis* was therefore investigated by three different multilocus molecular methods, targeting genomewide synonymous SNPs, variation in number of tandem repeats, and insertion of IS100 insertion elements. Eight populations were recognized by the three methods, and we propose an evolutionary tree for these populations, rooted on *Yersinia pseudotuberculosis*. The tree invokes microevolution over millennia, during which enzootic pestoises isolates evolved. This initial phase was followed by a binary split 6,500 years ago, which led to populations that are more frequently associated with human disease. These populations do not correspond directly to classical biovars that are based on phenotypic properties. Thus, we recommend that henceforth groupings should be based on molecular signatures. The age of *Y. pestis* inferred here is compatible with the dates of historical pandemic plague. However, it is premature to infer an association between any modern molecular grouping and a particular pandemic wave that occurred before the 20th century.

**Methods**

**Bacterial Strains.** We examined 156 *Y. pestis* strains isolated from humans, fleas, and small rodents on various continents between 1946 and 1998 (Table 1, which is published as supporting information on the PNAS web site). They included isolates that had been assigned to pestoises (9 isolates) or the biovars Orientalis (94 isolates), Medievalis (27 isolates), Antiqua (25 isolates), or Microtus (1 isolate) by standard tests. *Y. pseudotuberculosis* isolates of serotypes I (8 isolates), II (2 isolates), III (1 isolate), IV (2 isolates) and V (1 isolate) also were examined.
The association of historical plague pandemics with *Yersinia pestis* remains controversial, partly because the evolutionary history of this largely monomorphic bacterium was unknown. The microevolution of *Y. pestis* was therefore investigated by three different multilocus molecular methods, targeting genomewide synonymous SNPs, variation in number of tandem repeats, and insertion of IS100 insertion elements. Eight populations were recognized by the three methods, and we propose an evolutionary tree for these populations, rooted on *Yersinia pseudotuberculosis*. The tree invokes microevolution over millennia, during which enzootic pestoides isolates evolved. This initial phase was followed by a binary split 6,500 years ago, which led to populations that are more frequently associated with human disease. These populations do not correspond directly to classical biovars that are based on phenotypic properties. Thus, we recommend that henceforth groupings should be based on molecular signatures. The age of *Y. pestis* inferred here is compatible with the dates of historical pandemic plague. However, it is premature to infer an association between any modern molecular grouping and a particular pandemic wave that occurred before the 20th century.

### Subject Terms

*Yersinia pestis*, plague, insertion element, molecular clock, history, evolution, pandemic, SNP, VNTR
**napA.** The entire napA gene was PCR-amplified from *Y. pseudotuberculosis* strain IP32953 (primers: AGTGCAAGCTT-TCAGGCCACTACCCGTTCAG and CATCACGGATC-CATGAACACTCAGCCGGGAGCG), digested with BamHI plus HindIII, ligated into the corresponding multicloning site at 146/207 of expression vector pQE30 (Qiagen, Valencia, CA), and cloned into *Escherichia coli* SCS1. One resulting recombinant plasmid (pBE696), which contains the expected insert size according to sequencing, was used for complementation of the inability to reduce nitrate.

To screen for the napA613 mutation, a 430-bp product was PCR-amplified (primers: GTACGACCCTAATCTGGATG and GATGTTGTTGCGTGAAGCCA) (annealing temperature: 54°), followed by sequencing of the internal 155-bp product (napA positions 562–716) at 58° from both strands (primers: TTGTATGGCCTCTCGGTG and TTGGTAAGTGGAGAGACGG). *The napA613 mutation results in a unique MboII site that also can be used for rapid screening.*

**MLVA.** A total of 43 loci were screened for size variation of fluorescently labeled PCR amplicons, as described (26). Fragments of common sizes were inferred to represent homologous alleles, and the inability to amplify a PCR product was scored as missing data.

**IS100 Typing.** A total of 31 genomic loci that contain IS100 were identified by BLAST searches of the genome of strain CO92 (27) (molecular group 1.ORI). Eight additional locations where IS100 is integrated into the chromosome of strains IP554 (1.ANT) and IP564 (2.MED) but not that of CO92 were identified by inverse PCR as follows. Chromosomal DNA was ligated after digestion with eight endonucleases lacking target sequences in IS100 (BamHI, ClaI, HindIII, SstI, BflI, DraI, KpnI, or NcoI). Fragments flanking IS100 were PCR-amplified by using oligonucleotide primers within IS100 (CTACTCTTCCTGCTGGCA and TAGCAGAAGCTTCTCAGG) and cloned into vector pCR2.1 (Invitrogen) in *E. coli* INVaF™. PCR amplification using M13 reverse and T7 promoter universal primers identified 125 inserts of unique sizes among 1,375 transformants, whose sequences then were compared to the genome of CO92.

Oligonucleotide primers that flank each of the 39 insertion sites by ~100 bp were used for PCRs. Sizing of the PCR amplicons by agarose gel electrophoresis indicated whether an IS100 insertion was present (~2,200 bp) or absent (~200 bp), and the inability to amplify a PCR product was scored as missing data. Data on 11 locations are presented here (Fig. 5 and Table 2, which are published as supporting information on the PNAS web site); the other locations were excluded because they yielded similar results to the 11 locations or were characterized by high frequencies of missing data or homoplasies. Note that the inability to amplify Y45 in *Y. pseudotuberculosis* reflects the absence of an IS1541 insertion that contains the target site for that particular IS100 insertion.

**Genomic Analyses.** Reciprocal-best FASTA hits with >40% predicted amino acid identity over >80% of the protein length were used to identify 3,283 potential orthologous coding sequences (CDSs) from pairwise comparisons of the genomes of 91001 (34), CO92 (1.ORI) (27), and KIM (2.MED) (29). These CDSs were then screened for sSNPs. We excluded sSNPs in 30 CDSs that were within regions of low sequence complexity, within CDSs with multiple paralogs, or where the CDS was lacking in *Y. pseudotuberculosis* (CI32953 (GenBank accession no. NC_006155) according to pairwise BLAST analyses. Three more putative sSNPs in the CO92 genome and one within the KIM genome were excluded because they reflected sequencing errors, leaving 76 sSNPs in 3,250 orthologous CDSs (Tables 3–5, which are published as supporting information on the PNAS web site).

Four additional sSNPs and 11 nonsynonymous changes were identified during our screening procedures (Tables 6 and 7, which are published as supporting information on the PNAS web site).

**sSNP Screening.** PCR products spanning sSNPs were amplified over 25 cycles in 25-μl volumes, containing 5 ng of DNA from each of 1–4 test strains plus a reference strain (CO92, IP520, or 91001), polymerase (1.25 units, Optimase, Transgenomic, Omaha, NE), as well as specific primers (Table 8, which is published as supporting information on the PNAS web site). PCR products were analyzed by denaturing HPLC with a DNA-SepK Cartridge, (Wave® Nucleic Acid Fragment Analysis System, Transgenomic) at the temperatures indicated in Table 8.

**Phylogenetic Methods.** Data were stored as numerical character sets in BIOPHINERICS 4.0 (Applied Maths, Sint-Martens-Latem, Belgium), which was also used to calculate Hamming distance matrices of the number of shared alleles between isolates. PAUP*4.0 (30) was used for parsimony analysis, and MEGA 2.0 (31) was used for neighbor joining.

**Results**

**Pestoides and Microtus Belong to *Y. pestis.* Because of their ability to ferment melibiose and rhamnose, it was unclear whether pestoides were more closely related to *Y. pseudotuberculosis* or *Y. pestis* (32). We therefore sequenced six housekeeping gene fragments from nine pestoides isolates. These fragments are identical among the classical *Y. pestis* biovars but variable in *Y. pseudotuberculosis* (3). The pestoides sequences were identical to those from *Y. pestis.* Similarly, in *silico* analyses of the genome (28) of biovar Microtus strain 91001 also yielded sequences identical to those from *Y. pestis,* except for a homopolymeric stretch of seven adenines in *manB,* which contains only six adenines in *other pestis* isolates. Thus, despite phenotypic differences, pestoides and Microtus belong to *Y. pestis.*

**Genomic Branch Order and Age.** Pairwise comparisons of the three genomic sequences from *Y. pestis* that are currently available (27–29) revealed 76 conservative sSNPs within 3,250 orthologous CDSs. For each sSNP, the ancestral nucleotide was deduced (27–29) revealed 76 conservative sSNPs within 3,250 orthologous CDSs. For each sSNP, the ancestral nucleotide was deduced (27–29) revealed 76 conservative sSNPs within 3,250 orthologous CDSs. For each sSNP, the ancestral nucleotide was deduced. For each sSNP, the ancestral nucleotide was deduced. For each sSNP, the ancestral nucleotide was deduced.
have accumulated. The correct synonymous mutation rate between E. coli and Typhimurium is the synonymous distance between them (0.94) (35) divided by twice the time since these organisms separated (140 million years) (36), or 3.4 × 10⁻⁹ per year. The frequency of sSNPs per potential sSNP divided by that rate then yields the age estimates for Y. pestis that are shown in Fig. 1. We estimate that 13,000 years of evolutionary history separate CO92 and KIM and that the time since 91001 separated from branch 0 is longer (10,000 years) than since CO92 or KIM diverged from their common ancestor (average of 6,500 years).

Molecular Groupings. sSNPs could be useful for epidemiological or forensic purposes as molecular markers for specific populations within Y. pestis. Therefore, 40 sSNPs in 38 gene fragments (total length of 11.2 kb) that marked branches 0, 1, or 2 (Tables 3 and 4) were screened among 105 diverse isolates of Y. pestis by dHPLC (Fig. 6, which is published as supporting information on the PNAS web site). Four additional sSNPs were identified by these procedures (Table 6), for a total of 44. The nucleotides at these 44 positions are identical among Orientalis isolates, except that sSNP s34 is specific to CO92 and s36 is specific for a different Orientalis isolate. However, although most (Medievalis) isolates that cannot reduce nitrate were indistinguishable from KIM (Fig. 6), others were very different.

These and other discrepancies (see below) between classical biobar designations and molecular groupings stimulated us to devise a nomenclature that is based on molecular relatedness but includes mnemonic biobar designations to facilitate the transition. The group of bacteria related to Orientalis is referred to as 1.ORI to reflect the association of the Orientalis phenotype with branch 1 and classical Medievalis isolates are referred to as 2.MED (Figs. 2 and 3). Antiqua isolates split into distinct groups on each of branches 1 and 2, designated 1.ANT and 2.ANT, which were isolated in Africa and East Asia, respectively. Branch 0 includes almost all pestoides isolates (groups 0.PE1, 0.PE2, and 0.PE3) as well as the Micrurus isolate, 91001 (0.PE4).

A strong discovery bias affects the particular sSNPs that were used for screening because they were defined by a comparison between only three genomes (0.PE4, 1.ORI, and 2.MED). As a result, the current set of sSNPs can indicate the branch order and time of separation for molecular groups from which genome sequences are not (yet) available (0.PE1–0.PE3, 1.ANT, and 2.ANT), but is not particularly informative about their genetic diversity and age (37). Therefore, we screened Y. pestis by an independent approach, MLVA, which should yield neutral estimates of the pairwise genetic distances between all isolates. MLVA of 43 variable number of tandem repeats detected 102 unique patterns among 104 isolates of Y. pestis and Y. pseudotuberculosis. After phylogenetic clustering, the patterns clustered together in

![Fig. 1. Age of Y. pestis. sSNPs were identified by pairwise genome comparisons between 91001 (0.PE4), CO92 (1.ORI), and KIM (2.MED). For each sSNP, one of the alternative nucleotides is present at the corresponding position within the genome of Y. pseudotuberculosis strain IP32953. sSNPs on branch 0 (Table 4) were identical in IP32953 and 91001 and also identical in KIM and CO92, but differed between these pairs. Other sSNPs were unique to the branches, as indicated. To calculate ages, the number of sSNPs was divided by the 777,520 potential sSNPs within the 3,250 homologous gene pairs, and that distance was then divided by the molecular clock rate of 3.4 × 10⁻⁹ per year.](Image 1)

![Fig. 2. Evolutionary branch order within Y. pestis. (a–d) Simplified branch order of the major groups as indicated by sSNPs (a), MLVA (b), and IS100 insertions (c and d), based on data in Figs. 3, 6, and 7. The primary inconsistencies between a and b–d are indicated in orange and purple. The differences in branch order between c and d reflect different interpretation of insertion events (green text). Nodes along branches are indicated by circles, the sizes of which indicate the number of isolates. (e) Consensus evolutionary order of IS100 insertions (Yxx) and synonymous mutations (sxo). The diagram also indicates the inferred order of phenotypic changes (Rha⁻, Mel⁻, and Nit⁻) and nutritional mutations (glpD, napA316), except for the Nit⁻ isolates in 2.ANT, which are not indicated. Sources of isolates according to grouping: 0.PE1, former Soviet Union (4 isolates); 0.PE2, former Soviet Union (3 isolates); 0.PE3, Africa (1 isolate); 0.PE4, China (1 isolate); 1.ANT, Africa (21 isolates); 1.ORI, global (95 isolates); 2.ANT, East Asia (5 isolates); and 2.MED, Kurdistan (26 isolates).](Image 2)
molecular groups that were consistent with those found by sSNP analysis (Fig. 3), except that all branch lengths were relatively long. The branch order of a neighbor-joining dendrogram indicated that 2.MED and 2.ANT represent sister clades, as do 0.PE1, 0.PE2, and 0.PE3, consistent with the sSNP data (Fig. 3). However, unlike the three branch structure described above, 1.ANT was more distinct from 1.ORI than are 2.MED/2.ANT, and 0.PE4 did not cluster together with 0.PE1–0.PE3 (Figs. 2b and 3). Similar results were obtained when the MLVA data were analyzed with other clustering algorithms (data not shown).

To resolve differences between discrepant branch orders, we applied still a third molecular grouping method, namely the presence or absence of the IS100 insertion element at 11 distinct genomic locations (Fig. 5 and Fig. 7, which is published as supporting information on the PNAS web site). Except for 0.PE1, 0.PE2, and 0.PE4, which were not distinguished by this method, the same molecular groups were found within 131 isolates as with the other two methods. The IS100 results confirmed the split between branches 1 and 2 (Fig. 2) and revealed minor subdivisions within 1.ANT (1.ANT.a and 1.ANT.b) and 2.ANT (2.ANT.a and 2.ANT.b) that were consistent with the results from MLVA. However, branch 0 was lacking in the most parsimonious interpretation (Fig. 2d) and first reappeared in a less parsimonious interpretation involving one more step (Fig. 2c). According to the latter interpretation, an insertion of IS100 at Y23 predated the separation of all Y. pestis molecular groups but was subsequently lost by excision during the evolution of branch 2. We conclude that the molecular groupings represent major populations and that the patterns of descent within Y. pestis correspond to a three branch structure. Characteristic sSNPs and changes in IS100 patterns are summarized in a consensus tree containing eight populations and six subpopulations that is shown in Fig. 2e.

A Signature Mutation in napA. According to the data presented here and by others (8, 10), the inability to reduce nitrate is common to distantly related organisms in 2.MED, 0.PE1, 0.PE4, and 2.ANT (3/5 isolates). We found that the sequence of the entire nap operon is identical between strains IP564 (2.MED), IP554 (1.ANT), and CO92 (1.ORI), except for a premature stop codon in IP564 (Fig. 4A) within the napA gene, which encodes a periplasmic nitrate reductase. This stop codon, which we designated napA613, prevents IP564 from reducing nitrate because nitrate reduction was restored by complementation with an intact napA gene from Y. pseudotuberculosis strain IP32953 (Fig. 4B).

The napA613 mutation is a diagnostic marker for 2.MED, and an inability to reduce nitrate by some isolates from other groups has a different genetic basis. For example, 2.ANT.b strain IP546 (Nepal) was originally classified as Medievalis because it is impaired in nitrate reduction. However, IP546 possesses a WT napA sequence and, upon reexamination, we found that IP546 does reduce nitrate weakly on extended cultivation (Fig. 3). In contrast, modern stocks of 1.ANT strain IP566 do not reduce nitrate because of a deletion, acquired in the laboratory, which encompasses the napA operon. However, IP566 did reduce nitrate originally, as expected for 1.ANT strains, and older DNA preparations yielded a weak napA PCR product. Finally, one 2.MED isolate, pestoides J, has been designated pestoides because it ferments melibiose (but not glycerol). In this study, we found napA613 in 24 2.MED isolates (Table 1), including pestoides J, but not in 98 other strains, including seven from 0.PE1, 0.PE4, or 2.ANT that do not reduce nitrate. Similar results have recently been published by other investigators (8, 10).

Discussion

Populations Versus Biovars. We propose that Y. pestis should be subdivided into populations based on molecular groupings, eight of which are defined here, rather than biovar. The same eight molecular groupings were detected among 156 isolates by three independent methods, except that 0.PE1, 0.PE2, and 0.PE4 were not distinguished by IS100 typing. Assignments to these groupings were unambiguous and consistent for the 60 isolates that were tested by all three methods (Table 1), with only minor exceptions (Supporting Text, which is published as supporting information on the PNAS web site). We infer that these molecular grouping represent distinct bacterial populations. Independent support for the existence of these populations also can be deduced from other molecular analyses, which have examined subsets of the diversity examined here (3–8).

The populations are only partially compatible with the classical phenotypic categories designated as biovars. An inability to reduce nitrate, the hallmark of biovar Medievalis, is found among isolates from groups 2.MED, 2.ANT, 0.PE1, and 0.PE4, probably because of multiple, independent molecular events. Similarly, biovar Antiqia includes unrelated organisms from 1.ANT and 2.ANT that can ferment glycerol and reduce nitrate. Finally, the designation pestoides for organisms that can ferment melibiose and rhamnose combines a variety of diverse organisms from 0.PE1, 0.PE2, and 0.PE3. Thus, biovars are not necessarily monophyletic and should not be used for evolutionary or taxonomic purposes.

Molecular groupings also are not necessarily a reliable indicator of phenotype. One 2.MED isolate (pestoides J) was unable to ferment glycerol and, concordant with other results (4), 1.ORI includes one isolate (Nich51) that can ferment glycerol. Similarly, some 2.ANT isolates can reduce nitrate, whereas others cannot. The multilocus molecular markers that are defined here provide the basis for a common language for classifying the diversity and relatedness among isolates from distinct geographical areas, such as enzootic isolates from the former Soviet Union and China. These isolates manifest extensive phenotypic diversity but their genetic
relationships remain unresolved (7, 9). For example, molecular tests could be used to determine whether Central Asian isolates that were previously designated as *altaica* and *hissarica* (9) belong to the same population (0.PE4) as Microtus strain 91001 from China (10), with which they share phenotypic properties. Many Central and East Asian isolates probably will fall into the populations described here, whereas others may quite possibly define new groupings.

**Detecting Phylogenetic Structure in a Highly Monomorphic Species.** Each of the three screening methods used here has distinct advantages and disadvantages for deducing the phylogenetic structure of *Y. pestis*. MLVA was the most discriminatory but the boundaries of population groupings were somewhat ambiguous. Furthermore, the high mutation rate of variable number of tandem repeat loci resulted in very long branch lengths, with corresponding problems for tree reconstruction. As a result, MLVA did not correctly detect the binary split between branches 1 and 2. We hoped that IS100 analyses would combine adequate discrimination with reliable classification. However, the most parsimonious tree was partially wrong because of hotspots for genomic rearrangements and excision events at the Y23 and Y36 loci (data not shown), and the IS100 analysis also suffered from a higher proportion of missing data (0.04 versus 0.02 for either sSNPs or MLVA). Although it is conceivable that screening additional genomic locations would have resulted in more reliable conclusions, our unpublished data do not support this possibility. Four additional locations that we analyzed in detail were difficult to interpret because of high homoplasies and still other locations could not be reliably amplified from numerous isolates (data not shown). Thus, IS100 analyses are probably not ideal for classification and phylogeny of *Y. pestis*.

Of the three methods, sSNP analyses are the easiest to interpret from an evolutionary viewpoint. No homoplasies were detected, and most branches were supported by multiple, independent sSNPs. However, *Y. pestis* is so monomorphic that three complete genome sequences of 4.5 MB differed by only 76 conservative sSNPs, most of which were specific for the 1.ORI, 2.MED, and 0.PE4 populations represented by the three genomes. A definitive sSNP-based classification will probably only be possible after at least one genome has been sequenced from each of the other five populations. For the moment, the sSNP-based resolution within branch 0, 1.ANT, and 2.ANT is scanty, and the best current estimates of genetic diversity within these populations are given by the MLVA and IS100 data. As a result, the evolutionary branch order along branch 0 should be considered as a working hypothesis for subsequent investigations.

With time, as additional genomes are sequenced, sSNP analysis may become the method of choice for determining the evolutionary branch structure and molecular groupings within highly uniform species. Genotyping of bacteria might then be efficiently performed by a hierarchical approach (38) in which molecular markers for the branch structure are used to group bacteria into populations before using more variable methods with higher resolution, such as high-throughput SNP typing, whole gene microarrays (6, 7), or MLVA, for subdivision into genotypes. Although multiple nonsynonymous polymorphisms were found here (Table 7), the frequency of nonsynonymous SNPs was only slightly higher than the frequency of SNPs within our pairwise genome comparisons. Similarly, only 14–16 genotypes were detected by whole gene microarrays (6, 7). In contrast, MLVA might be particularly suitable for genotyping within a hierarchical approach because it distinguished 102 patterns among 104 isolates and correlated strongly with geographical source within 1.ORI (Fig. 8, which is published as supporting information on the PNAS web site).

**History of Pandemics.** We previously suggested that *Y. pestis* may have evolved in Africa shortly before Justinian’s plague of 541...
Domini (3). Instead, >10,000 years has elapsed since 0.PE4 split from branch 0 (Fig. 1) and Y. pestis probably spread globally long before Justinian’s plague, as indicated by the isolation of representatives of branch 0 from the former Soviet Union (0.PE1 and 0.PE2, China (H11022), and Africa (0.PE3)). Furthermore, it is possible that Y. pestis arose in Asia where all three branches (0.PE1, 0.PE2, 0.PE4, 1.ORI, and 2.ANT) are found, rather than Africa, from which branch 2 has not been isolated. High diversity is often a good indicator of the geographical source of microbes.

Devignat (2) suggested on the basis of geographical sources and epidemiological observations that each of the three biovars was responsible for an independent pandemic wave. The age estimates presented here confirm that Y. pestis is old enough to have caused historical pandemics of plague. And the epidemiological data supporting an association of pandemic plague since the mid-1890s with Orientalis clearly implicate 1.ORI as the cause of the third pandemic. However, a putative association of older pandemics with unique biovars is not interpretable, especially because biovars Medievalis and Antiqua are polyphyletic, and because Y. pestis contains eight populations, many more than are needed to account for three pandemic waves.

One could attempt to refine Devignat’s hypothesis by associating Justinian’s plague with the Black Death with specific populations, such as 1.ANT and 2.MED, respectively. The following considerations argue against such a refinement. The frequent current isolation of 1.ANT from Africa does not necessarily indicate that it existed there 1500 years ago. Even if 1.ANT did exist in Africa at the time, other Y. pestis groupings may have caused Justinian’s plague, particularly because 0.PE3 strain Angola was also isolated from Africa. The Black Death did begin in Central Asia, and 2.MED isolates have been collected in “Kurdistan” (Table 1) (corresponding to areas in Iran, Iraq, and Turkey) and China (10).

However, Central Asia also includes parts of the former Soviet Union where 0.PE1 and 0.PE2 were isolated. Also, 2.MED is possibly too young to have caused the Black Death, because it is as uniform as 1.ORI, whose lack of diversity probably reflects clonal expansion over only 100 years. Thus, Devignat’s hypothesis is no longer convincing, and we can only hope for direct data from ancient DNA. (16, 17). The molecular signatures described here might facilitate such studies and were indeed originally designed for that purpose.

The history of plague and the population structure of Y. pestis is difficult to elucidate, because most cases of human disease occurred before the introduction of microbiology, modern disease is most frequent in areas that are remote from centers of molecular biology, and the causative organism is so unusually monomorphic. The results presented here provide a foundation for historical analyses, as well as a precise terminology based on molecular signatures that can be used for future epidemiological investigations. We also have readdressed the association between historical disease and modern isolates while providing technology that can hopefully supply a solid basis for future investigations of that association.

Note Added in Proof: Independent amplification of Y. pestis-specific DNA from Justinian’s plague has now been reported (39). Y. pestis-specific DNA from Justinian’s plague and the Black Death has been shown to most closely resemble biovar Orientalis (40).

We gratefully acknowledge technical support by Ying Liu, Matt Van Ert, and Tatum Simonson, extensive discussions with Edward C. Holmes (Oxford University, Oxford), the generous gift of DNA strain 91001 by Ruoding Yuan (Institute of Microbiology and Epidemiology, Beijing), and financial support from the Deutsche Forschungsgemeinschaft (Grant Ae 36/9-3) and the National Institutes of Health-National Institute of General Medical Sciences (Grant R01-GM060795).