Infection, Genetics and Evolution 26 (2014) 172–179

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

# Genetic variations of live attenuated plague vaccine strains (*Yersinia pestis* EV76 lineage) during laboratory passages in different countries

Yujun Cui [a,1], Xianwei Yang [a,1], Xiao Xiao [a], Andrey P. Anisimov [b], Dongfang Li [c], Yanfeng Yan [a], Dongsheng Zhou [a], Minoarisoa Rajerison [d], Elisabeth Carniel [e], Mark Achtman [f,g], Ruifu Yang [a,*], Yajun Song [a,f,*]

[a] *State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing 100071, China*
[b] *State Research Center for Applied Microbiology and Biotechnology, Obolensk, Moscow Region, Russia*
[c] *BGI-Shenzhen, Shenzhen 518083, China*
[d] *Institut Pasteur de Madagascar, Antananarivo, Madagascar*
[e] *Yersinia Research Unit, National Reference Laboratory, Institut Pasteur, Paris, France*
[f] *Environmental Research Institute, University College Cork, Cork, Ireland, United Kingdom*
[g] *Warwick Medical School, University of Warwick, Coventry, United Kingdom*

ARTICLE INFO

*Article history:*
Received 8 February 2014
Received in revised form 13 May 2014
Accepted 21 May 2014
Available online 4 June 2014

*Keywords:*
*Yersinia pestis*
Live vaccine
EV76
Genetic variation

ABSTRACT

Plague, one of the most devastating infectious diseases in human history, is caused by the bacterial species *Yersinia pestis*. A live attenuated *Y. pestis* strain (EV76) has been widely used as a plague vaccine in various countries around the world. Here we compared the whole genome sequence of an EV76 strain used in China (EV76-CN) with the genomes of *Y. pestis* wild isolates to identify genetic variations specific to the EV76 lineage. We identified 6 SNPs and 6 Indels (insertions and deletions) differentiating EV76-CN from its counterparts. Then, we screened these polymorphic sites in 28 other strains of EV76 lineage that were stored in different countries. Based on the profiles of SNPs and Indels, we reconstructed the parsimonious dissemination history of EV76 lineage. This analysis revealed that there have been at least three independent imports of EV76 strains into China. Additionally, we observed that the *pyrE* gene is a mutation hotspot in EV76 lineages. The fine comparison results based on whole genome sequence in this study provide better understanding of the effects of laboratory passages on the accumulation of genetic polymorphisms in plague vaccine strains. These variations identified here will also be helpful in discriminating different EV76 derivatives.



## 1. Introduction

Plague, caused by *Yersinia pestis*, has claimed millions of human lives during the three major pandemics in human history, and it is still an epidemic and endemic disease in some areas in Africa, Asia, and the Americas (Anisimov et al., 2004; Stenseth et al., 2008; WHO, 2009). *Y. pestis* is a gram-negative bacterium that can be transmitted to humans and susceptible animals from natural rodent reservoirs through bites of infected fleas, contacts with infected animals, or inhalation of infectious droplets. Because of its high mortality rate and respiratory transmission route, *Y. pestis* remains a major public health concern and one of the most dangerous bioterrorism agents (Greenfield et al., 2002).

Various types of plague vaccines have been developed, based on killed whole cells, live attenuated strains, subunit antigens, recombinant *Salmonella* or virus derivatives, enteropathogenic *Yersinia* et al. (Alvarez and Cardineau, 2010; Derbise et al., 2012; Smiley, 2008; Titball and Williamson, 2004). However, only the first two types of vaccines have been widely used in humans (Meyer, 1970; Meyer et al., 1974; Wang et al., 2013). EV76 is the most widely used live plague vaccine during the campaigns of human plague vaccination (Sun and Curtiss, 2014; Sun et al., 2011).

EV76 was derived from a strain isolated from a child who died of bubonic plague in Madagascar in 1926 (named as EV following

* Corresponding authors. Address: State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing 100071, China. Tel.: +86 13611182355 (Y. Song).
*E-mail addresses:* ruifuyang@gmail.com (R. Yang), songyajun88@gmail.com (Y. Song).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.meegid.2014.05.023
1567-1348/

the first two letters of the patient's name). After 76 serial subcultures on nutritive agar over a 6 year period at the Institut Pasteur in Antananarivo (Madagascar), the EV76 strain became spontaneously strongly attenuated while conferring protection against plague in different animal models (Girard and Robic, 1942). EV76 was first used as a human vaccine in Madagascar in the year of 1934, and was subsequently distributed to other countries such as Senegal, Argentina, Brazil, Tunisia, former USSR, former Belgian Congo, Vietnam, China, etc., where it was utilized for human vaccination in endemic areas (Gallut and Girard, 1961; Girard, 1963). A subculture of the EV76 vaccine strain was sent in 1936 by the Institut Pasteur in Antananarivo to the Russian Anti-Plague Research Institute "Microbe" (Saratov, Russia) and to NIIEG (Russian abbreviation of the Scientific-Research Institute for Epidemiology and Hygiene, Kirov, Russia) (Girard, 1963).

Comparative studies performed in 1942–1947 and 1959–1960 on the protective efficacy of various EV76 cultures stored in different Russian microbiological institutes indicated that, only the one provided directly by the Institut Pasteur in Antananarivo and kept under vacuum dried conditions at the NIIEG with a minimal number of reseedings, preserved its initial level of protection against plague (Saltykova and Faibich, 1975). Based on this observation, the less immunogenic EV76 variants were removed and replaced by the NIIEG EV76 strain for human vaccination in the former USSR and Mongolia, and overall at least 10 million individuals were vaccinated by the year of 1974 (Saltykova and Faibich, 1975). China also used EV76 as vaccine for a long time mainly for the people at risk (plague surveillance workers, marmot hunters, etc.).

As mentioned above, differences in both protective efficacy and residual side effects were observed among various EV76 derivatives kept in different institutions, and it was hypothesized that these differences resulted from the accumulation of genetic mutations during serial subcultures *in vitro* (Saltykova and Faibich, 1975). It had been reported that, due to the presence of numerous insertion sequence elements, the genome of *Y. pestis* is prone to frequent rearrangements (Guiyoule et al., 1994; Parkhill et al., 2001), which may lead to the loss of genetic material and therefore to a decrease in virulence and protective potential. Indeed, the attenuation of EV76 upon *in vitro* subcultures has been shown to be (at least in part) due to the disruption of a large chromosomal region, the *pgm* locus (Kutyrev et al., 1989; Podladchikova et al., 2002). DNA microarray analysis of various vaccine and wild type *Y. pestis* strains also identified additional gene losses that may also participate to the variability in the level of pathogenicity and immunogenicity of diverse isolates (Zhou et al., 2004).

*Y. pestis* is a "young" pathogen which evolved from *Yersinia pseudotuberculosis* 2600–28,600 years ago (Morelli et al., 2010). The relatively short evolutionary history of *Y. pestis* accounts for its limited phenotypic and genetic diversity, and *Y. pestis* is termed as a genetically monomorphic pathogen (Achtman, 2008). The intraspecies diversity of *Y. pestis* had been described in details elsewhere (Anisimov et al., 2004). Our previous work identified a total of 2298 single nucleotide polymorphisms (SNPs) amongst 133 *Y. pestis* genomes. A fully parsimonious phylogenetic tree based on these polymorphisms provided a frame for robust population structure analysis (Cui et al., 2013). Among the draft genomes decoded in our previous study, EV76-CN is an EV76 lineage strain kept in China which served as the seed strain for vaccine manufacturing in Lanzhou Institute of Biological Products.

Here we compared the draft genome of EV76-CN with two available genomes of *Y. pestis* wild isolates from Madagascar to identify EV76 lineage-specific SNPs and Indels (insertion and deletion shorter than 10 bp). These polymorphic sites were then screened in other EV76 derivatives from various institutions in different countries to infer the EV76 evolutionary scenario following its transfer to various laboratories and during manufacturing.

## 2. Materials and methods

### 2.1. Bacterial strains

A total of 36 *Y. pestis* strains were used in this study: seven wild strains isolated in Madagascar, and 29 known or suspected EV76 lineage strains from different laboratories in six countries including China, Russia, France, Finland, Madagascar, India and former Belgian Congo (Table 1). All strains were cultivated in nutrient agar at 28 °C for 48 h, and then the genomic DNAs was extracted by using conventional SDS lysis and phenol-chloroform extraction method as previously described (Li et al., 2008).

### 2.2. in silico identification of SNPs and Indels

Three previously sequenced draft genomes, including the EV76-CN (ADSA00000000), and two Madagascar isolates, MG05 (NZ_AAYS00000000) and IP275 (NZ_AAOS00000000) (Morelli et al., 2010), were used for comparative genomics analysis. The first completed genome of *Y. pestis*, strain CO92 (NC_003143) was used as the outgroup for phylogenetic analysis (Table 1).

SNPs were identified by comparing the scaffolds of EV76-CN assemblies to the genomes of IP275 and MG05, with the MUMmer package (Kurtz et al., 2004). MUMmer results were filtered according to the criteria described elsewhere (Cui et al., 2013), which excludes unreliable SNPs locating in repeated regions, with low quality scores (<20) or supported by few reads (<10 paired-end reads).

We then aligned the EV76-CN scaffolds to the two reference genomes of Madagascar wild isolates (IP275 and MG05) to identify Indels shorter than 10 bp by using LASTZ software (http://www.bx.psu.edu/~rsharris/lastz/). The Indels were validated by mapping the reads to the EV76-CN scaffolds, using the BWA v0.5.9 software (Li and Durbin, 2010). All identified alignment errors were amended manually. Finally, we extracted the sequences of the Indels and their breakpoints on EV76-CN assembly and the two reference genomes. To minimize possible assembly errors, we excluded the Indels locating at the end of contigs, and the ones with mismatched reads in the flanking 20 bp range.

### 2.3. Luminex-based SNP typing

The SNPs identified in the EV76-CN genome were tested by a multiplex Luminex assay as previously described (Song et al., 2010). Firstly, genomic fragments containing the SNP sites were amplified by multiplex PCR. Each 10 μl multiplex PCR system consisted of 5 μl of 2× Multiplex PCR Master Mix (Qiagen, Germany), 1 μl of 0.5 μmol/L PCR primers mixture for the six SNPs (Table S1), and 4 μl of template DNA (5 ng/μl). PCR cycling was performed in DNA Engine Tetrad 2 Thermocycler (Bio-Rad Inc., USA), with an initial denaturation at 95 °C for 10 min, followed by 30 cycles at 94 °C for 30 s, 58 °C for 30 s and 72 °C for 30 s, and a final extension step at 72 °C for 3 min. One μl of Exo-SAP containing 5 U of exonuclease I (Exo) and 0.5 U of shrimp alkaline phosphatase (SAP) (USB Corp., Germany) was added to the multiplex PCR product and incubated at 37 °C for 30 min, followed by 10 min at 80 °C to remove the unincorporated PCR primers and dNTPs.

For multiple Allele Specific Primer extension (ASPE), 10 μl APSE mixture including 5 μl of Exo-SAP-treated multiplex PCR product, 0.3 U of *Tsp* DNA polymerase (Invitrogen), ASPE primer mixture for all 6 SNPs with final concentration as 25 nmol/L (Table S1), and 5 dATP/dTTP/dGTP/biotin-dCTP mixture as 5 μmol/L (Invitrogen, USA). Thermo-cycling was performed at 95 °C for 5 min, followed by 25 cycles of 94 °C for 30 s, 55 °C for 30 s and 72 °C for 1 min, and a final extension at 72 °C for 3 min.

**Table 1**
*Y. pestis* strains and genomes used in this study.

| Strain ID | Genotype | Country | Source and characteristics | Purpose |
|---|---|---|---|---|
| *Wild type isolates* | | | | |
| IP1513 | EVT00 | Madagascar | Institut Pasteur Madagascar | Genotyping |
| IP241 | EVT00 | Madagascar | Institut Pasteur Madagascar | Genotyping |
| IP304 | EVT00 | Madagascar | Institut Pasteur Madagascar | Genotyping |
| IP528 | EVT00 | Madagascar | Institut Pasteur Madagascar | Genotyping |
| IP529 | EVT00 | Madagascar | Institut Pasteur Madagascar | Genotyping |
| IP643 | EVT00 | Madagascar | Institut Pasteur Madagascar | Genotyping |
| IP666 | EVT00 | Madagascar | Institut Pasteur Madagascar | Genotyping |
| *EV derivatives* | | | | |
| 1319 | EVT01 | China | Beijing Institute of Microbiology and Epidemiology | Genotyping |
| IP28 | EVT02 | France | Institut Pasteur Paris, EV 76 pYV+ | Genotyping |
| IP805 | EVT02 | France | Institut Pasteur Paris, EV76 | Genotyping |
| IP808 | EVT02 | France | Institut Pasteur Paris, EV33 | Genotyping |
| IP811 | EVT02 | France | Institut Pasteur Paris, EV40 | Genotyping |
| IP812 | EVT02 | France | Institut Pasteur Paris, EV40 | Genotyping |
| IP813 | EVT02 | France | Institut Pasteur Paris, EV40 | Genotyping |
| IP814 | EVT02 | France | Institut Pasteur Paris, EV76 | Genotyping |
| 1324 | EVT02 | China | National Institute for the Control of Pharmaceutical and Biological products | Genotyping |
| EV76 $Nal^R$ | EVT02 | Finland | Haartman Institute, University of Helsinki | Genotyping |
| EV76-B-SHU | EVT03 | China | Qinghai Institute for Endemic Diseases Prevention and Control | Genotyping |
| EV NIIEG-ISS | EVT04 | Former USSR | Tarasevich State Institute for Standardisation and Control of Biomedical Preparations, industry standard sample strain | Genotyping |
| EV NIIEG-NIIS | EVT04 | Former USSR | Microbiology Research Institute of the Ministry of Defense of the Russian Federation, vaccine from EV line NIIEG-ISS | Genotyping |
| EV NIIEG-Stavropol | EVT04 | Former USSR | Stavropol Anti-Plague Research Institute, vaccine from EV line NIIEG-ISS | Genotyping |
| KM215 | EVT04 | Former USSR | Russian Anti-Plague Research Institute "Microbe", laboratory line of EV | Genotyping |
| EV-SHU | EVT04 | China | Qinghai Institute for Endemic Diseases Prevention and Control | Genotyping |
| EV-YUAN | EVT04 | China | Qinghai Institute for Endemic Diseases Prevention and Control | Genotyping |
| 2307 | EVT04 | China | Xinjiang Center for Disease Control and Prevention | Genotyping |
| 2308 | EVT04 | China | Xinjiang Center for Disease Control and Prevention | Genotyping |
| YN401 | EVT04 | China | Yunnan Institute for Epidemic Diseases Control and Research | Genotyping |
| YN404 | EVT04 | China | Yunnan Institute for Epidemic Diseases Control and Research | Genotyping |
| CMCC120020 | EVT04 | China | Hebei Anti-Plague Research Institute | Genotyping |
| 74014 | EVT04 | China | Gansu Center for Disease Control and Prevention | Genotyping |
| 830572 | EVT04 | China | Qinghai Institute for Endemic Diseases Prevention and Control | Genotyping |
| 830567 | EVT04 | China | Qinghai Institute for Endemic Diseases Prevention and Control | Genotyping |
| EV76-CN[a] | EVT05 | China | Lanzhou Institute of Biological Products, accession number ADSA00000000 47685 | Genotyping |
| *Suspected EV strains* | | | | |
| IP678 | EVT02 | Former Belgian Congo | Institut Pasteur Paris, documented as isolated from a horse in Congo | Genotyping |
| IP283 | EVT04 | India | Institut Pasteur Paris, original name 9/95 from Surat | Genotyping |
| IP1893 | EVT04 | India | Institut Pasteur Paris, original name FN 95-419 from Surat | Genotyping |
| *Genomes* | | | | |
| CO92 | – | USA | Isolated from a fatal case of human plague, accession number NC_003143 | Genome comparison |
| MG05 | – | Madagascar | Accession number NZ_AAYS00000000 | Genome comparison |
| IP275 | – | Madagascar | Accession number NZ_AAOS00000000 | Genome comparison |

[a] EV76-CN genome (Cui et al., 2013) was used for genome comparison along with the other three genomes.

The ASPE products were then mixed and incubated with 12 types of xTAG beads targeting the ASPE primers (Luminex, USA). Finally the mixture were treated with 2 mg/L streptavidin-R-phycoerythrin (Invitrogen, USA), and subjected to Luminex 200 station (Luminex co, USA). SNP calling was achieved by comparing the median fluorescence intensity (MFI) of both beads for one SNP.

### 2.4. Indel profiling

Primers spanning each Indel were designed to amplify the targeted fragments (Table S1).

The PCR amplifications for individual Indel were performed in a 30 μl volume containing 10 ng of DNA template, 0.5 μmol/L of each primer, 1 unit of *Taq* DNA polymerase, 200 μmol/L of dNTPs, 50 mmol/L KCl, 10 mmol/L Tris HCl (pH 8.3) and 2.5 μmol/L MgCl2. The cycling condition were 95 °C for 5 min, followed by 30 cycles of 94 °C for 30 s,58 °C for 30 s and 72 °C for 60 s, and a final extension step at 72 °C for 3 min. All amplicons were sequenced by using both primers by BGI Inc. (China). The sequences were aligned against genomes of CO92, MG05, IP275 and EV76-CN. The alleles found in the wild type strains were defined as ancestor status, and those found in EV76-CN were scored as derived status.

### 2.5. Data storing and phylogenic analysis

The identified SNPs and Indels were both treated as bi-allelic polymorphic sites and entered into Bionumerics 6.6 software (Applied Maths, Belgium) as binary datasets (ancestor status was defined as "0" and derived status as "1"). The minimum spanning tree was constructed by combing SNP and Indel characteristics using the clustering module of Bionumerics 6.6 software, with CO92 genome as the outgroup (all the SNPs and Indels were defined as "0") as described elsewhere (Cui et al., 2013; Morelli et al., 2010).

## 3. Results and discussion

### 3.1. EV76-CN-specific SNPs and Indels

To identify the EV76 lineage-specific polymorphisms, we aligned the genome sequences of EV76-CN with two

Madagascar-origin wild strains, MG05 and IP275, by using CO92 as the outgroup. The phylogenetic relationship of the three strains was constructed based on both SNPs and Indels (Fig. 1).

There are six SNPs and six Indels on the EV76-CN genome branch, which is much less than the genomes of MG05 and IP275 (16 SNPs, 41 Indels and 16 SNPs, 49 Indels respectively, Fig. 1). There are more genetic polymorphisms in Madagascar wild isolates than in EV76-CN since they diverged from their common ancestor. EV76-CN is a seed strain used for vaccine manufacturing in China, implying very limited laboratory passages since its import to China. One possible explanation for the higher genomic diversity in the wild isolates might be that they experienced more generations (genome replications) than lab-stored EV76 strains, although we do not have direct evidences to support this speculation.

As expected, Indels were more frequent found than SNPs in wild Madagascar isolates (Fig. 1). Indels including variable number tandem repeats (VNTRs) are known to happen in higher rates than other genetic polymorphism makers (IS-mediated deletions and SNPs) in *Y. pestis* and other pathogenic bacteria (Li et al., 2013; Pearson et al., 2009). For example, copy number variation event in single-locus VNTR occurred at estimated rates ranging from $1.0 \times 10^{-5}$ to $3.5 \times 10^{-4}$/generation for the 14 VNTR loci in *Y. pestis* strain A1122, while deletion mediated by IS*100* occurred at a much lower rate ($6.5 \times 10^{-7}$/generation) in the same strain (Vogler et al., 2007). Our previous study estimated the substitution rate for SNPs ranging from $2.09 \times 10^{-9}$ to $2.30 \times 10^{-8}$/year in Madagascar *Y. pestis* isolates, and it is much lower than that of VNTR given the fact that *Y. pestis* will go through many generations per year in the plague natural foci (Morelli et al., 2010).

Among the six EV76-CN specific SNPs, five were in coding regions and caused amino acid substitutions (non-synonymous, Table 2). The functions of these genes have not yet been studied, and some of them are annotated as encoding regulators, transporters, and membrane proteins. Of the six Indels identified in the EV76-CN genome, four caused either a frame shift in the coding sequence or amino acid changes in the gene product. Therefore, most EV76-CN specific polymorphisms (9/12) potentially caused the alteration of gene functions. Five of the six Indels identified in EV76-CN occurred within VNTRs, and the only Indel outside of the VNTR region was Indel5 (Table 2).

To validate the specificity of the 6 SNPs and 6 Indels identified in EV76-CN, we *in silico* screened these polymorphic sites in 129 other published *Y. pestis* genomes (Cui et al., 2013). None of them had these polymorphisms, implying that these non-homoplasic markers can be used for genotyping of EV76 lineage strain.

### 3.2. Genotyping of EV76 derivatives

The 12 polymorphic sites (SNPs and Indels) differentiating EV76-CN from the two sequenced wild Madagascar isolates were also absent in seven other wild Malagasy *Y. pestis* isolates, confirming the specificity of these polymorphisms for EV76-CN. These seven strains were grouped as EVT00 (Table 3).

Since different EV76 lineages were stored and used in different countries in the past 80 years (Feodorova and Corbel, 2009), we then wanted to determine whether these mutation events accumulated early in the EV76 ancestor strain or occurred subsequently in different lineages. For this purpose, we screened the status of the EV76-CN specific polymorphisms in 25 other EV-lineage strains stored in China, Russia, France and Finland (Table 1). Furthermore, three strains documented as wild isolates, which were suspected as belonging to the EV lineage in a previous analysis (Morelli et al., 2010), were also tested.

None of the 29 tested EV-lineage strains belongs to the EVT00 genotype, and 5 polymorphisms (SNPs 1, 2 and 5; and Indels 2 and 4) were found in all EV derivatives. This finding and the fact that these five polymorphisms were already present in EV33 and EV40 isolates (EVT02), which are the subcultures anterior to EV76 (Girard, 1963), demonstrate that they all occurred early during the evolution of the EV strains.

The seven other polymorphisms allowed grouping the EV strains into five genotypes (EVT01 to 05). Most isolates were grouped into two genotypes, EVT02 (10 strains) and EVT04 (16 strains), while the three remaining genotypes had only one strain each (Table 3).

### 3.3. Mutation hotspots during laboratory passages

As aforementioned, Indel5 is the only Indel outside of VNTR regions. This 6 bp in-frame deletion TGCCGA (changing coding amino acid sequence from NAQ to K) was located at the 3′ terminus of *pyrE*, which encodes an orotate phosphoribosyltransferase (OPRTase) involved in pyrimidine biosynthesis (Fig. 2).

To determine the Indel5 profiles, we sequenced partial *pyrE* gene of the tested strains and revealed that EVT04 and EVT05 genotypes (17 strains in total) have this in-frame deletion (Indel5 status as 1). Of the twelve other strains, ten have the same intact ancestral *pyrE* (no deletion, Indel5 status as 0) as CO92 and Madagascar wild isolates. While the remaining two strains, 1319 (EVT01) and IP678 (EVT02), although there was no TGCGCA in-frame deletion (Indel5 status as 0), an IS*1661* element was found inserted right next to the TGCGCA block, causing a truncated *pyrE* (Fig. 2). These two strains with *IS1661* interrupted *pyrE* belong to different genotypes (1319 as EVT01 and IP678 as EVT02), and they should get IS*1661* insertion independently (after diverging into different genotypes), or they shared one common ancestor with IS*1661* insertion (and then diverged into different genotypes). Indel5 might have only happened once and resulted in a modified *pyrE* in EVT04 and EVT05. That is to say, at least two independent mutation events (Indel5 and IS*1661* insertion) happened in the *pyrE* gene in our EV76 collections.

Taking the limited polymorphisms found in this study into account, these two independent events in *pyrE* are unlikely to happen by chance, and *pyrE* seems to be a mutation hotspot of *Y.*

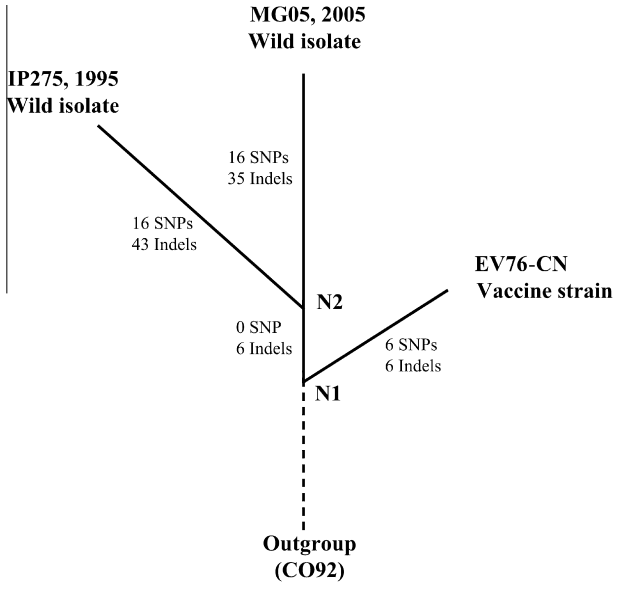


**Fig. 1.** Phylogenetic relationship between EV76-CN and two wild *Y. pestis* strains from Madagascar. The genome of CO92 was used as the outgroup of the minimum spanning tree. The number of polymorphisms (SNPS and Indels) between the hypothetical nodes (N1 and N2), and between the nodes-isolates are indicated besides each branch.

**Table 2**
EV76 lineage-specific polymorphisms (SNPs and Indels).

| ID | CO92 position | Gene Designation | Gene Production | Type | | | | |
|---|---|---|---|---|---|---|---|---|
| SNP | | | | | Anc Nt | Der Nt | Anc AA | Der AA |
| SNP1 | 456222 | YPO0435 | Putative Na+ dependent nucleoside transporter-family protein | nsyn | G | T | M | I |
| SNP2 | 921145 | YPO0841 | Putative regulatory protein | nsyn | C | A | A | E |
| SNP3 | 1680198 | YPO1482 | Putative membrane protein | nsyn | G | T | G | W |
| SNP4 | 2241486 | YPO1973 | Putative GntR-family transcriptional regulatory protein | nsyn | C | T | G | S |
| SNP5 | 2268596 | YPO1996 | Hypothetical protein | nsyn | C | A | M | I |
| SNP6 | 4448411 | Intergenic | NA | intergenic | T | G | NA | NA |
| Indels | | | | | Indel sequence | VNTR* | | |
| Indel1 | 1839450-1839457 | Intergenic | NA | Deletion | ATGACCAT | Y | | |
| Indel2 | 630388-630389 | Intergenic | NA | Insertion | A | Y | | |
| Indel3 | 2013319-2013319 | YPO1768 | Putative 4-hydroxyphenylacetate permease | Deletion | G | Y | | |
| Indel4 | 965401-965402 | YPO0880 | Putative primase | Insertion | TGCT | Y | | |
| Indel5 | 58824-58829 | YPO0045 | Orotate phosphoribosyltransferase | Deletion | TGCGCA | N | | |
| Indel6 | 358208 | YPO0347 | Anaerobic C4-dicarboxylate transporter | Deletion | C | Y | | |

Anc: ancestral; Der: derived; Nt: Nucleotide; AA: amino acid; nsyn: non synonymous SNP; NA: not applicable.
* Y: yes; N, no.

**Table 3**
Genotypes of various wild *Y. pestis* strains and EV76 derivatives based on the 16 EV76-CN SNPs and Indels

| Genotype | Country | Number | SNPs* | | | | | | Indels* | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
| EVT00 | M | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EVT01 | C | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| EVT02 | Fr, C, Fi, BC | 10 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |
| EVT03 | C | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| EVT04 | R, C, I | 16 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| EVT05 | C | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

M, Madagascar; C, China, Fr, France; Fi, Finland; BC, former Belgian Congo; R, former USSR; I, Indian.
* 0, ancestral; 1, derived.

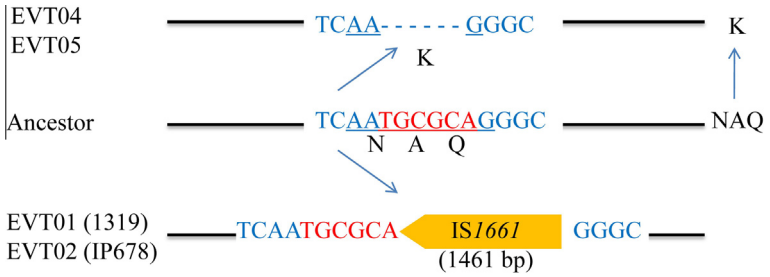


**Fig. 2.** The mutation hotspot in *pyrE*. The central line indicates the ancestral status of *pyrE*, with the position of indel5 highlighted in red fonts. For EVT04 and 05, the six nucleotides were deleted and the amino acid sequences altered from NAQ to K in the corresponding region. For strain 1319 and IP678, IS*1661* interrupted *pyrE* on the position nearby Indel5, although Indel5 loci itself remain ancestral.

*pestis* during its laboratory passages. Interestingly, an 82 bp deletion was also observed in the *pyrE* gene of 7/11 endpoint strains of *Escherichia coli* that underwent 60-day laboratory adaptive evolution (Conrad et al., 2009). This deletion conferred *ca.* 15% increase in the growth rate, and gave an increase of *ca.* 30% when it was introduced into a strain harboring two other adaptive mutations (Conrad et al., 2009). Our results thus suggest a similar selection for *pyrE* inactivation in the EV76 strain that has been subjected to numerous and prolonged laboratory passages. Additional studies will be helpful to clarify the underlying biological mechanisms.

### 3.4. Evolutionary scenarios of EV strains

Due to the mutation discovery bias (Pearson et al., 2009), the sequenced EV76-CN genome will be located at the tip of the tree reconstructed by the identified polymorphisms and other strains will collapse successively into nodes on the way from the root (Madagascar isolates, EVT00) to the tip (EV76-CN, EVT05). Assuming all the 12 mutation events happened only once amongst the genotyped strains, the fewer derived alleles a strain has, the more ancient it would be during the evolution process. Since EVT01 had the least derived alleles (5 sites), it is likely to be the most ancient EV lineage genotype amongst those studied. The most parsimonious scenario for EV76 lineage evolution is that five polymorphisms (SNP1, 2, 5 and Indel 2, 4) accumulated on the way from genotype EVT00 to EVT01, followed by Indel6 giving rise to EVT02, and then additional polymorphisms sequentially led to genotypes EVT03, 04 and 05 respectively, as described on Fig. 3A.

However, Fig. 3A was purely based the polymorphisms profiling results, and it did not take the strain history and origins into

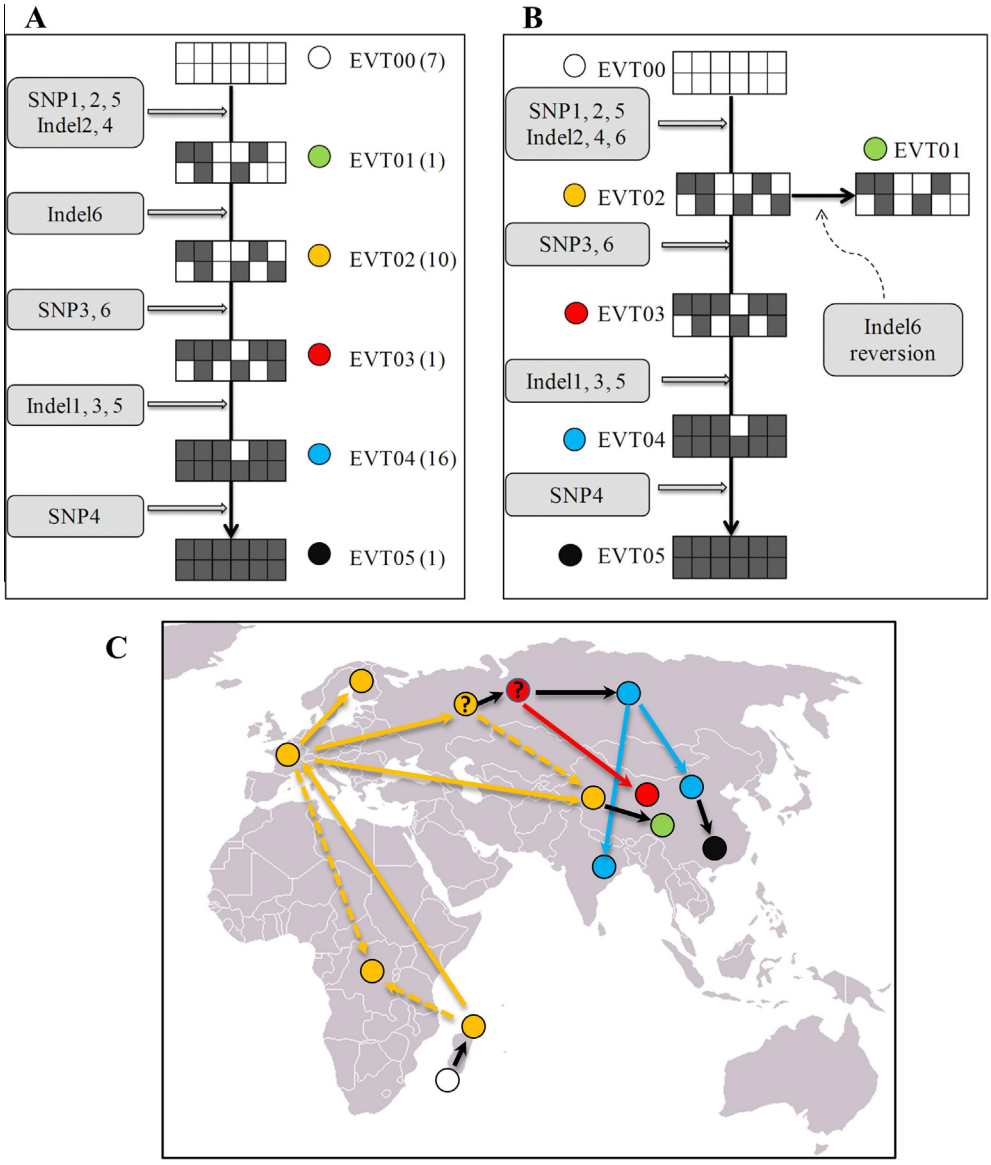


**Fig. 3.** Evolutionary scenario of EV76 lineage strains inferred from twelve polymorphic sites (A). Six genotypes defined by twelve polymorphisms. Each 12-block rectangle represents one genotype with upper and lower blocks indicating status of SNPs and Indels respectively (White and black blocks standing for ancestral and derived status respectively, 0 and 1 in Table 3). Gray blocks along the branches illustrate proposed mutation events between the genotypes. (B). Similar to panel A, except for a reversion mutation event in EVT01. (C). The proposed dissemination scenario of EV76 derivatives. The colored circles are genotypes shown in panels A and B. The circles with question mark in former USSR indicate hypothetical missing genotypes. Directions of strain export were indicated by arrows, with black ones coupled with mutation events, colored ones showing direct links of the same genotypes, and dashed one as possible links. As Institut Pasteur (Madagascar) and Institut Pasteur (Paris) both get involved in the promoting of EV76 strains and it is difficult to differentiate these two sources exactly according to historical records, Madagascar is set as the source of parent EV strain and Paris as sources for developed EV76 deviates.

account. If Fig. 3A was true, the only EVT01 strain (strain 1319 from China) would be more ancient than any other EV76 lineage strains tested in this study. However, four EVT02 strains from Institut Pasteur (IP808, 811, 812 and 813), as indicated by their alternate IDs (EV33 and EV40, namely 33 and 40 passages from the original EV strain), are intermediate strains from the parent strain EV to the developed EV76 vaccine strain (Girard, 1963). These four strains should be more ancient, and the Chinese genotype EVT01 (strain 1319) cannot be the ancestor of them.

Therefore we came to another interpretation (Fig. 3B). Genotype EVT00 evolved into EVT02 with the accumulation of 6 mutations (SNP1, 2, 5 and Indel 2, 4, 6), followed by a reversion of Indel6 leading to genotype EVT01. Another two SNPs led to EVT03, and then three Indels to EVT04. The only EV76-CN specific SNP (SNP4) defines EVT05. Indel6 happened in a CCCCCC region of the gene encoding putative anaerobic C4-dicarboxylate transporter, and the mutated strains have five C instead of six. Given the fact that copy numbers of single base stretches and short repeats are prone to vary due to slipped-strand mispairing during DNA replication (Levinson and Gutman, 1987), we deem that the scenario in Fig. 3B is more reasonable than Fig. 3A.

### 3.5. Disseminations of EV76 lineage strains

Based on Fig. 3B, we proposed dissemination routes for various EV76 lineages found in France, Finland, China, the former USSR, India and the former Belgian Congo (Fig. 3C). All the seven EV strains from Institut Pasteur clustered into EVT02 (Table 1), which is at the base of the tree (Fig. 3B) and contains EV strains ancestral to EV76. This indicates that the original Madagascar EV strain was

most likely EVT02 genotype, or its ancestral genotype which is not identified yet.

By subculture transferring or vaccination projects, the EV76 derivatives were exported to other countries, such as Finland, the former Belgian Congo and China. Strains from Finland and the former Belgian Congo had the genotype EVT02 as French ones (Table 1), suggesting that these countries received the strain either directly from the Institut Pasteur in Madagascar or through the Institut Pasteur in France (Fig. 3C). A direct transfer of EV76 from France to China is historically established, and this subculture exported to China might experience a reverse mutation in Indel6 leading to genotype EVT01 (strain 1319) as discussed before.

All EV strains from various institutes in the former USSR were grouped into a single genotype EVT04. In the year of 1936, a subculture of the EV76 vaccine strain was established at the NIIEG in the former USSR. Scientists in the former USSR then managed to get many derivatives of the received EV76 subcultures, including the widely-used NIIEG line (Feodorova and Corbel, 2009). Compared with EVT02, EVT04 incorporated at least two SNPs (SNP3 and 6) and three Indels (Indel1, 3 and 5).

EVT04 genotype strains were also found in China and India (Table 1). China basically copied the plague surveillance and control system from the former USSR since 1949, including importing vaccine strains. It is not surprising to found out that seventeen out of nineteen EV76 lineage strains in China (EVT04 and EVT05) are the descendants of USSR lineages, by one or multiple imports.

Although no USSR EVT002 strains were identified in this study, the former USSR scientist did receive EV76 strains directly from the Institut Pasteur (Girard, 1963). The USSR EVT04 strains would thus have been derived from EVT02. However there is still a missing link between these two genotypes, i.e. EVT03 strains, which is only found in China in this study. As China received EV76 strains from USSR and the transfer of EV76 strains from China back to the USSR is not documented, it is thus most likely that the EVT03 strain evolved in USSR and then transferred to China. We failed to identify EVT03 strains in the former USSR, and it might result from the sampling bias due to the limited number of former USSR strains included here, or the extinction of this genotype in the former USSR.

The scenario describe above suggests at least three imports of the EV76 strains to China: one from France (EVT02), and two from the former USSR (EVT03 and EVT04). Finally, the two unique EVT01 and EVT05 genotypes found in China may result from their local divergence from EVT02 and EVT04 strains respectively, with a single nucleotide insertion (Indel6) or SNP (SNP4). Therefore, the genetic polymorphisms of Chinese EV76 strains are shaped by multiple imports from abroad, and mutation events occurring during laboratory passages within China as well.

### 3.6. Suspected EV76 strains

Three strains documented as wild isolates (IP678, IP283 and IP1893) were suspected to belong to the EV lineage in our previous analysis (Morelli et al., 2010).

Strain IP678 was isolated from the nasal mucus of a sick horse in Elisabethville Belgian Congo in 1950. It exhibited an attenuated virulence and possessed vaccine properties, which led scientists to correlate it to EV76 (Girard and Robic, 1942). Our results concur with previous data and confirm that IP678 is actually an EV76 strain (genotype EVT02, Table 2), clustered with the ancestral EV strains from Institut Pasteur Paris. Large scale vaccination with EV76 took place in the former Belgian Congo from 1939 to 1946 (Girard, 1963). It is possible that the vaccine strain was used to immunize horses and produce protective antibodies for serotherapy, as initially done by Alexandre Yersin (Hawgood, 2008), and got isolated from the sick horse.

The other two strains were from India. IP1893 was sent to Institut Pasteur in 1995, and it was reported to have been isolated from the lungs of a patient who died of plague during the 1994 pneumonic plague outbreak in the region of Surat, India. IP283 was also documented as an isolate from the Surat outbreak. Intriguingly, some *Y. pestis* strains isolated from patients in Surat appeared to be $pgm^-$ (Welkos et al., 2002), which is an important genetic feature of EV76 strain (Sun and Curtiss, 2014). Although *pgm- Y. pestis* strains are lethal in hereditary hemochromatosis individuals (Quenee et al., 2012), it seems unlikely that they had caused the large plague outbreak in the normal population of Surat. IP283 and IP1893 were genotyped as EVT04 together with EV76 lineage strains from China and former USSR (Table 2). One possible explanation would be that these two strains were resulted from laboratory contamination by EV76. However additional data, especially epidemiological data and historical records, are needed to draw a solid conclusion on the origin of these two strains.

## 4. Conclusion

Here we compared the genome sequence of EV76-CN, a live attenuated vaccine seed strain of *Y. pestis* used in China, with two genomes from Madagascar isolates. In this study, we mainly focus on SNPs and Indels. Of the 12 polymorphisms (6 SNPs and 6 Indels) identified in EV76 lineages, all 5 SNPs within genes are nonsynonymous, and the 4 Indels inside genes changed the coding proteins (Table 2). However, it is difficult to present solid conclusions whether there is diversifying selection or relaxed purifying selection going on in EV76 lineage, based on this small number of polymorphic sites. We also found out that *pyrE* gene is a mutation hotspot in our EV strain collection, although the biological effects are still unclear.

The EV76 vaccine strains accumulated a number of genetic polymorphisms during laboratory passage. Based on the polymorphisms profiles, we are able to identify different EV76 groups and to reconstruct a scenario for their dissemination. For instance, we identified at least three independent imports of EV strains into China. We also confirmed three suspected EV76 strains, which provided an example of how state of the art technologies like fine genotyping can be used to resolve old doubts about strain origins.

Twenty-nine EV-76 lineage strains from various sources are clustered into 5 genotypes by the SNPs and Indels identified here. Despite the limitation by its sample size and phylogenetic discovery bias (only mutation events occurring in the sequenced EV76-CN strain were investigated) (Pearson et al., 2009), this genotyping scheme can potentially be helpful in identifying the source of EV76 strains closely related to the ones in this study. Analyzing more strains in EV76 lineage will help to confirm and refine the dissemination routes of EV76 strains proposed in this study, and possibly developing novel protocols for quality controlling of plague vaccine manufacturing.

## Acknowledgements

We are grateful to Dr Daniel Falush for his help in preparing this manuscript. This work was supported by National Natural Science Foundation of China (NSFC 31000015 and 31071111), National Key Program for Infectious Diseases of China (2012ZX10004215, 2013ZX10004221-002), Industry Research Special Foundation of China, Ministry of Health (201202021), and Science Foundation of Ireland (05/FE1/B882). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.meegid.2014.05.023.

## References


Achtman, M., 2008. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. Annu. Rev. Microbiol. 62, 53–70.

Alvarez, M.L., Cardineau, G.A., 2010. Prevention of bubonic and pneumonic plague using plant-derived vaccines. Biotechnol. Adv. 28, 184–196.

Anisimov, A.P., Lindler, L.E., Pier, G.B., 2004. Intraspecific diversity of *Yersinia pestis*. Clin. Microbiol. Rev. 17, 434–464.

Conrad, T.M., Joyce, A.R., Applebee, M.K., Barrett, C.L., Xie, B., Gao, Y., Palsson, B.O., 2009. Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. Genome Biol. 10, R118.

Cui, Y., Yu, C., Yan, Y., Li, D., Li, Y., Jombart, T., Weinert, L.A., Wang, Z., Guo, Z., Xu, L., Zhang, Y., Zheng, H., Qin, N., Xiao, X., Wu, M., Wang, X., Zhou, D., Qi, Z., Du, Z., Wu, H., Yang, X., Cao, H., Wang, H., Wang, J., Yao, S., Rakin, A., Falush, D., Balloux, F., Achtman, M., Song, Y., Wang, J., Yang, R., 2013. Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. Proc. Natl. Acad. Sci. U.S.A. 110, 577–582.

Derbise, A., Cerda Marin, A., Ave, P., Blisnick, T., Huerre, M., Carniel, E., Demeure, C.E., 2012. An encapsulated *Yersinia pseudotuberculosis* is a highly efficient vaccine against pneumonic plague. PLoS Negl. Trop. Dis. 6, e1528.

Feodorova, V.A., Corbel, M.J., 2009. Prospects for new plague vaccines. Expert Rev. Vaccines 8, 1721–1738.

Gallut, J., Girard, G., 1961. Action of chlorpromazine on experimental poisoning and infection of the mouse by *Pasteurella pestis* (EV vaccinal strain). Ann. Inst. Pasteur. (Paris) 100, 672–676.

Girard, G., 1963. Immunity in plague. Acquisitions supplied by 30 years of work on the "*Pasteurella pestis* EV" (Girard and Robic) strain. Biol. Med. (Paris) 52, 631–731.

Girard, G., Robic, J., 1942. Current status of the plague in Madagascar and vaccinal prophylaxis with the aid of the EV virus-vaccine. Bull. Soc. Path. Exot. 35, 43–49.

Greenfield, R.A., Drevets, D.A., Machado, L.J., Voskuhl, G.W., Cornea, P., Bronze, M.S., 2002. Bacterial pathogens as biological weapons and agents of bioterrorism. Am. J. Med. Sci. 323, 299–315.

Guiyoule, A., Grimont, F., Iteman, I., Grimont, P.A., Lefevre, M., Carniel, E., 1994. Plague pandemics investigated by ribotyping of *Yersinia pestis* strains. J. Clin. Microbiol. 32, 634–641.

Hawgood, B.J., 2008. Alexandre Yersin (1863–1943): discoverer of the plague bacillus, explorer and agronomist. J. Med. Biogr. 16, 167–172.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., Salzberg, S.L., 2004. Versatile and open software for comparing large genomes. Genome Biol. 5, R12.

Kutyrev, V.V., Filippov, A.A., Shavina, N., Protsenko, O.A., 1989. Genetic analysis and simulation of the virulence of *Yersinia pestis*. Mol. Gen. Mikrobiol. Virusol., 42–47.

Levinson, G., Gutman, G.A., 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. 4, 203–221.

Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–595.

Li, Y., Cui, Y., Cui, B., Yan, Y., Yang, X., Wang, H., Qi, Z., Zhang, Q., Xiao, X., Guo, Z., Ma, C., Wang, J., Song, Y., Yang, R., 2013. Features of variable number of tandem repeats in *Yersinia pestis* and the development of a hierarchical genotyping scheme. PLoS One 8, e66567.

Li, Y., Dai, E., Cui, Y., Li, M., Zhang, Y., Wu, M., Zhou, D., Guo, Z., Dai, X., Cui, B., Qi, Z., Wang, Z., Wang, H., Dong, X., Song, Z., Zhai, J., Song, Y., Yang, R., 2008. Different region analysis for genotyping *Yersinia pestis* isolates from China. PLoS One 3, e2166.

Meyer, K.F., 1970. Effectiveness of live or killed plague vaccines in man. Bull. World Health Organ. 42, 653–666.

Meyer, K.F., Cavanaugh, D.C., Bartelloni, P.J., Marshall Jr., J.D., 1974. Plague immunization. I. Past and present trends. J. Infect. Dis. 129 (Suppl.), S13–18.

Morelli, G., Song, Y., Mazzoni, C.J., Eppinger, M., Roumagnac, P., Wagner, D.M., Feldkamp, M., Kusecek, B., Vogler, A.J., Li, Y., Cui, Y., Thomson, N.R., Jombart, T., Leblois, R., Lichtner, P., Rahalison, L., Petersen, J.M., Balloux, F., Keim, P., Wirth, T., Ravel, J., Yang, R., Carniel, E., Achtman, M., 2010. *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. Nat. Genet. 42, 1140–1143.

Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T., Prentice, M.B., Sebaihia, M., James, K.D., Churcher, C., Mungall, K.L., Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdeno-Tarraga, A.M., Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Karlyshev, A.V., Leather, S., Moule, S., Oyston, P.C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S., Barrell, B.G., 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. Nature 413, 523–527.

Pearson, T., Okinaka, R.T., Foster, J.T., Keim, P., 2009. Phylogenetic understanding of clonal populations in an era of whole genome sequencing. Infect. Genet. Evol. 9, 1010–1019.

Podladchikova, O.N., Rykova, V.A., Ivanova, V.S., Eremenko, N.S., Lebedeva, S.A., 2002. Study of *pgm* mutation mechanism in *Yersinia pestis* (plague pathogen) vaccine strain EV76. Mol. Gen. Mikrobiol. Virusol., 14–19.

Quenee, L.E., Hermanas, T.M., Ciletti, N., Louvel, H., Miller, N.C., Elli, D., Blaylock, B., Mitchell, A., Schroeder, J., Krausz, T., Kanabrocki, J., Schneewind, O., 2012. Hereditary hemochromatosis restores the virulence of plague vaccine strains. J. Infect. Dis. 206, 1050–1058.

Saltykova, R.A., Faibich, M.M., 1975. Experience from a 30-year study of the stability of the properties of the plague vaccine strain EV in the USSR. Zh. Mikrobiol. Epidemiol. Immunobiol., 3–8.

Smiley, S.T., 2008. Current challenges in the development of vaccines for pneumonic plague. Expert Rev. Vaccines 7, 209–221.

Song, Y., Roumagnac, P., Weill, F.X., Wain, J., Dolecek, C., Mazzoni, C.J., Holt, K.E., Achtman, M., 2010. A multiplex single nucleotide polymorphism typing assay for detecting mutations that result in decreased fluoroquinolone susceptibility in *Salmonella enterica* serovars Typhi and Paratyphi A. J. Antimicrob. Chemother. 65, 1631–1641.

Stenseth, N.C., Atshabar, B.B., Begon, M., Belmain, S.R., Bertherat, E., Carniel, E., Gage, K.L., Leirs, H., Rahalison, L., 2008. Plague: past, present, and future. PLoS Med. 5, e3.

Sun, W., Curtiss, R., 2014. Rational considerations about development of live attenuated *Yersinia pestis* vaccines. Curr. Pharm. Biotechnol. 14, 878–886.

Sun, W., Roland, K.L., Curtiss, R.I., 2011. Developing live vaccines against *Yersinia pestis*. J. Infect. Dev. Ctries. 5, 614–627.

Titball, R.W., Williamson, E.D., 2004. *Yersinia pestis* (plague) vaccines. Expert. Opin. Biol. Ther. 4, 965–973.

Vogler, A.J., Keys, C.E., Allender, C., Bailey, I., Girard, J., Pearson, T., Smith, K.L., Wagner, D.M., Keim, P., 2007. Mutations, mutation rates, and evolution at the hypervariable VNTR loci of *Yersinia pestis*. Mutat Res. 616, 145–158.

Wang, X., Zhang, X., Zhou, D., Yang, R., 2013. Live-attenuated *Yersinia pestis* vaccines. Expert Rev. Vaccines 12, 677–686.

Welkos, S., Pitt, M.L., Martinez, M., Friedlander, A., Vogel, P., Tammariello, R., 2002. Determination of the virulence of the pigmentation-deficient and pigmentation-/plasminogen activator-deficient strains of *Yersinia pestis* in non-human primate and mouse models of pneumonic plague. Vaccine 20, 2206–2214.

WHO, 2009. Human plague: review of regional morbidity and mortality, 2004–2009. Wkly. Epidemiol. Rec. 85, 40–45.

Zhou, D., Han, Y., Dai, E., Song, Y., Pei, D., Zhai, J., Du, Z., Wang, J., Guo, Z., Yang, R., 2004. Defining the genome content of live plague vaccines by use of whole-genome DNA microarray. Vaccine 22, 3367–3374.