A peer-reviewed version of this preprint was published in PeerJ on 27 March 2018.

Bochkareva OO, Dranenko NO, Ocheredko ES, Kanevsky GM, Lozinsky YN, Khalaycheva VA, Artamonova II, Gelfand MS. 2018. Genome rearrangements and phylogeny reconstruction in *Yersinia pestis*. PeerJ 6:e4545 https://doi.org/10.7717/peerj.4545

Genome rearrangements and phylogeny reconstruction in *Yersinia pestis*

Olga O Bochkareva Corresp., [1] , **Natalia O Dranenko** [2] , **Elena S Ocheredko** [3] , **German M Kanevsky** [4] , **Yaroslav N Lozinsky** [3] , **Vera A Khalaycheva** [5] , **Irena I Artamonova** [1, 6] , **Mikhail S Gelfand** [1, 3, 7, 8]

[1] Kharkevich Institute for Information Transmission Problems, Moscow, Russia

[2] Department of Molecular and Chemical Physics, Moscow Institute of Physics and Technology, Moscow, Russia

[3] Department of Bioengineering and Bioinformatics, Moscow State University, Moscow, Russia

[4] Higher Chemical College at the Russian Academy of Sciences, Moscow, Russia

[5] Stavropol State Agrarian University, Stavropol, Russia

[6] Vavilov Institute of General Genetics, Moscow, Russia

[7] Faculty of Computer Science, Higher School of Economics, Moscow, Russia

[8] Skolkovo Institute of Science and Technology, Moscow, Russia

Corresponding Author: Olga O Bochkareva
Email address: olga.bochkaryova@iitp.ru

Genome rearrangements have played an important role in the evolution of *Yersinia pestis* from its progenitor *Yersinia pseudotuberculosis*. Traditional phylogenetic trees for *Y. pestis* based on sequence comparison have short internal branches and low bootstrap supports as only a small number of nucleotide substitutions have occurred. On the other hand, even a small number of genome rearrangements may resolve topological ambiguities in a phylogenetic tree.

We reconstructed the evolutionary history of genome rearrangements in *Y. pestis*. We also reconciled phylogenetic trees for each of the three CRISPR-loci to obtain an integrated scenario of the CRISPR-cassette evolution. We detected numerous parallel inversions and gain/loss events by the analysis of contradictions between the obtained evolutionary trees. We also tested the hypotheses that large within-replichore inversions tend to be balanced by subsequent reversal events and that the core genes less frequently switch the chain by inversions. Both predictions were not confirmed.

Our data indicate that an integrated analysis of sequence-based and inversion-based trees enhances the resolution of phylogenetic reconstruction. In contrast, reconstructions of strain relationships based on solely CRISPR loci may not be reliable, as the history is obscured by large deletions, obliterating the order of spacer gains. Similarly, numerous parallel gene losses preclude reconstruction of phylogeny based on gene content.

Genome rearrangements and phylogeny reconstruction in *Yersinia pestis*

Olga O. Bochkareva[1], Natalia O. Dranenko[2], Elena S. Ocheredko[3], German M. Kanevsky[4], Yaroslav N. Lozinsky[3], Vera A. Khalaycheva[5], Irena I. Artamonova[6,1], and Mikhail S. Gelfand[1,3,7,8]

[1]**Kharkevich Institute for Information Transmission Problems, Moscow, Russia**
[2]**Moscow Institute of Physics and Technology, Moscow, Russia**
[3]**Lomonosov Moscow State University, Moscow, Russia**
[4]**Higher Chemical College at the Russian Academy of Sciences, Moscow, Russia**
[5]**Stavropol State Agrarian University, Stavropol, Russia**
[6]**Vavilov Institute of General Genetics, Moscow, Russia**
[7]**Higher School of Economics, Moscow, Russia**
[8]**Skolkovo Institute of Science and Technology, Moscow, Russia**

Corresponding author:
Olga O. Bochkareva[1]

Email address: olga.bochkaryova@iitp.ru

ABSTRACT

Genome rearrangements have played an important role in the evolution of *Yersinia pestis* from its progenitor *Yersinia pseudotuberculosis*. Traditional phylogenetic trees for *Y. pestis* based on sequence comparison have short internal branches and low bootstrap supports as only a small number of nucleotide substitutions have occurred. On the other hand, even a small number of genome rearrangements may resolve topological ambiguities in a phylogenetic tree.

We reconstructed the evolutionary history of genome rearrangements in *Y. pestis*. We also reconciled phylogenetic trees for each of the three CRISPR-loci to obtain an integrated scenario of the CRISPR-cassette evolution. We detected numerous parallel inversions and gain/loss events by the analysis of contradictions between the obtained evolutionary trees. We also tested the hypotheses that large within-replichore inversions tend to be balanced by subsequent reversal events and that the core genes less frequently switch the chain by inversions. Both predictions were not confirmed.

Our data indicate that an integrated analysis of sequence-based and inversion-based trees enhances the resolution of phylogenetic reconstruction. In contrast, reconstructions of strain relationships based on solely CRISPR loci may not be reliable, as the history is obscured by large deletions, obliterating the order of spacer gains. Similarly, numerous parallel gene losses preclude reconstruction of phylogeny based on gene content.

INTRODUCTION

Yersinia pestis, causing fulminant plague, has evolved clonally from an enteric pathogen, *Yersinia pseudotuberculosis*, that, in contrast, causes a relatively benign enteric illness. Horizontal gene acquisition, massive gene loss, and genome rearrangement events all have played important roles in the evolution of *Y. pestis* from its progenitor (Achtman et al., 1999).*Y. pseudotuberculosis* and *Y. pestis* differ radically in their pathogenesis despite sharing >97% identity in 75% of their chromosomal genes (Martínez-Chavarría and Vadyvaloo, 2015). As only a small number of nucleotide substitutions have occurred, traditional phylogenetic trees of *Y. pestis* strains based on sequence comparison have short internal branches and low bootstraps. They are also significantly affected by extensive horizontal gene flow between strains due to homologous recombination. Genome rearrangements are less sensitive to homologous recombination and hence allow for an alternative approach to construction of phylogenetic trees, as even a small number of genome rearrangements may resolve topological ambiguities in a phylogenetic tree. The*Y. pestis* chromosome contains a large variety and number of insert sequences (ISs) that may have caused frequent

chromosome rearrangements (Liang et al., 2014).

Comparison of the KIM genome sequence with *Y. pestis* strain CO92 allowed to divide both genomes into 27 conserved segments and the most parsimonious series of inversions for three multiple-inversion regions were described (Deng et al., 2002). Further, large-scale genome rearrangements were described in strains Antiqua, Nepal and Angola (Chain et al., 2006; Eppinger et al., 2010). Multiple genome alignment of nine *Y. pestis* and *Y. pseudotuberculosis* genomes featured universal Locally Collinear Blocks (LCBs) and yielded seven parsimonious scenarios of the inversion history (Darling et al., 2008). Later, the LCB model has been used to infer the phylogenetic relationships among eight complete *Y. pestis* genomes from the breakpoint distance matrix, yielding the conclusion that the pattern of *Y. pestis* chromosome rearrangements reflects the genetic features of specific geographical areas and could be applied to distinguish *Y. pestis* isolates (Liang et al., 2010). A set of gene families from 13 *Yersinia* species was used to reconstruct a complete genome sequence for the ancestor, integrating information from the sequences, the species tree, and the gene order (Duchemin et al., 2015).

Being a traditional object for the spoligotyping, a special type of genotyping based on the spacer nucleotide analysis, CRISPR systems of *Y. pestis* strains often served as a model for CRISPR-based evolutionary studies. All three separate genomic CRISPR loci were described in detail (Pourcel et al., 2005), including numerous strains without complete genomes (Vergnaud et al., 2007; Cui et al., 2008; Riehm et al., 2012; Barros et al., 2014; Riehm et al., 2015). Relationships between strains were studied using the distance based on shared and differential spacers content only (Barros et al., 2014) or taking into account the principles of evolutionary cassette dynamics. In particular, the evolutionary history of *Y. pestis* based on CRISPR polymorphism was reconstructed in the form of an acyclic oriented graph (Cui et al., 2008). Later, a general mathematical model of CRISPR evolution was applied to reconstruct the relationships of strains for each of the three CRISPR loci (Kupczok and Bollback, 2013).

Here, we integrate the history at different levels of genome evolution, including gene flux, sequence divergence, chromosome segmental inversions, and spacer acquisitions and deletions in CRISPR cassettes, for genomes of twelve completely sequenced *Y. pestis* strains and four *Y. pseudotuberculosis* strains.

MATERIALS AND METHODS

Genomes

Complete genome sequences of four *Yersinia pseudotuberculosis* and twelve *Yersinia pestis*, all available as of August 1st, 2013, were taken from the NCBI Genome database (Benson et al., 2013) and are listed in Supplementary Table 1.

Construction of orthologs

Bidirectional best hits (BBHs) were constructed for each pair of strains using BLASTP (Zhang and Madden, 1997). BLASTP hits with identity $<50\%$ or coverage of the shorter sequence $<67\%$ were ignored. At the next step, if paralogs were more similar to each other than to either BBH partner, both paralogs were added to the ortholog group. Then, maximal connected components were constructed. This was done using ad hoc software based on the Relational Database Management System (RDBMS) Oracle Database Express Edition.

Phylogenetic trees

Single-copy universal genes were used to construct a phylogenetic tree. Multiple alignments were obtained using the Muscle software (Edgar, 2004). For concatenation of alignments, long insertions/deletions at gene boundaries were ignored. A phylogenetic tree was constructed using the NJ algorithm with the Mega7 software (Tamura et al., 2013).

Synteny blocks and rearrangements history

Synteny blocks were constructed using the Sibelia algorithm (Minkin et al., 2013) with default parameter and the length of blocks more than 5000 bp. Blocks that were found in any genome more than once were filtered out. The history of rearrangements was reconstructed using the MGRA algorithm (Avdeyev et al., 2016).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3223v1 | CC BY 4.0 Open Access | rec: 5 Sep 2017, publ: 5 Sep 2017

The origins and terminus of replication

The origins and terminators of replication for six *Y. pestis* and two *Y. pseudotuberculosis* strain were previously identified (Darling et al., 2008). We used these data to identify synteny blocks that contains origins and terminators.

Permutation testing for inversions

For each *Y. pestis* strain, we constructed a null distribution for the percentage of chromosome length that switched its location between the leading and lagging chain compared to the common ancestor. Given a set of inversion lengths for each strain, we selected inversion start positions at random in the range from 1 to $4.5 \cdot 10^6$ (corresponding to the average genome length). Then, for each strain we obtained 10^4 of random inversion sets to calculate the p-value.

CRISPR analysis

CRISPR-cassettes were downloaded from CRISPRdb (Grissa et al., 2007). Phylogenetic trees were reconstructed manually based on the CRISPR-cassettes evolution rules. At that, two types of events were allowed, addition of a new spacer at the leader end, and deletion of one or several adjacent spacers from any part of a cassette. We further assumed (1) no independent addition of the same spacer to two different cassettes; (2) rare, but possible independent deletions of the same cassette segments; and (3) more probable single deletion of a segment including several adjacent spacers compared to several subsequent deletions of the segment parts.

RESULTS AND DISCUSSION

Phylogenetic trees and evolutionary events

The phylogenetic tree for the analyzed *Y. pseudotuberculosis* and *Y. pestis* was constructed based on 2408 single-copy universal genes using a concatenation of individual nucleotide alignments (Fig. 1A). We used the *Y. enterocolitica* genome to root the tree. We observed that *Y. pestis* strains form a clade within the *Y. pseudotuberculosis* subtree, in agreement with previous genome analyses (Chain et al., 2006; Rasmussen et al., 2015). As the phylogeny of *Y. pestis* was not completely resolved, we added 76 genes universal for these strains and reconstructed the *Y. pestis* branch separately (Fig. 1B). There seemed to be several key noise factors. A small number of nucleotide substitutions resulted in low bootstrap values in several vertices, e.g. for Z176003, D182038, and D106004. On the other hand, branches of the Angola, Microtus, and Antiqua strains could have been placed incorrectly due to long branch attraction.

To analyze the history of rearrangements we constructed whole-genome alignment using the Sibelia software. We identified 166 synteny blocks, with 130 blocks found in all strains, and the remaining blocks reflecting gain/loss events. The coverage of genomes by blocks exceed 87%.

Fixing the tree to the one based on the concatenated sequence alignment of universal genes, we reconstructed 161 inversions and 62 insertions/deletions. (Fig. 2A) Two inversions occurred twice, in D106004 and Z176003, and in A1122 and D182038. Parallel events could be explained by homologous recombination (horizontal transfer between strains) involving a segment containing the inverted fragments. Indeed, in the tree constructed using only the genes involved in the first inversion, A1122 and D182038 formed a separate branch (Fig. 3A), indicating a close affinity limited to this fragment. Another possibility is incomplete resolution of the sequence-based tree, e.g. in the case of the D106004, Z176003, D182038 group with internal branches having relatively low bootstrap support (Fig. 1B). In the inversion analysis, D106004 and Z176003 are sister strains with D182038 being an outgroup (Fig. 2B), whereas the sequence-based tree for the genes involved in the inversion is poorly resolved and hence provides no information about possible horizontal transfer.

To calculate the inversion rate, we performed regression analysis (Fig. 4). On average one inversion occurs per 20 substitutions per genome, $R^2 = 0,87$.

About a dozen of parallel insertions/deletions were found. We applied optimization methods (Avdeyev et al., 2016) to obtain an alternative topology with a lower number of parallel events. The optimal topology based on location of the synteny blocks results in 160 inversions and 58 insertions/deletions for *Y. pestis* (Fig. 2B). Here, Antiqua moves to the Microtus node, in agreement with the fact that, according to the ability to ferment glycerol and to reduce nitrate, strains Antiqua, Pestoides, Microtus and Angola belong to the Antiqua biovar (Chain et al., 2006). This analysis demonstrates that parallel gain and loss events

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3223v1 | CC BY 4.0 Open Access | rec: 5 Sep 2017, publ: 5 Sep 2017

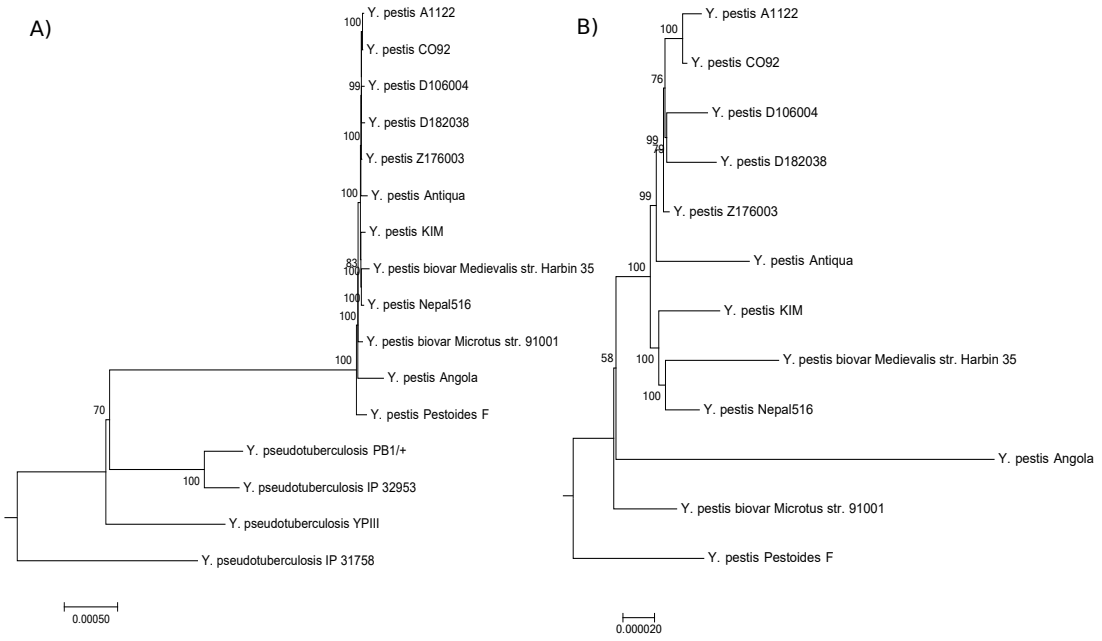

Figure 1. Phylogenetic trees (A) *Yersinia spp.*, based on nucleotide alignments of 2408 single-copy universal genes. (B) *Yersinia pestis*, based on nucleotide alignments of 2484 single-copy universal genes including 76 *Y. pestis*-specific universal genes.

are not rare in the evolution of *Yersinia* spp., and these events should be considered only in the context of other criteria.

Selection acting on inversions

We further analyzed the selection pressure on within-replichore inversions. Inversions of common synteny blocks were separated into 41 inter-replichore and 49 within-replichore events. We constructed a null distribution for the fraction of within-replichore inversions in the *Y. pestis* history (see Methods). In the considered genomes, within-replichore inversions were over-represented with p-value $< 10^{-4}$.

We tested the hypothesis that large within-replichore inversions are usually balanced by subsequent (partial) reversal events. Indeed, inversions within a replichore change the leading/lagging strand A/T and G/C biases, relative gene density, and gene expression levels. Hence, they may introduce many slightly deleterious traits and be detrimental (Darling et al., 2008).

For strains whose evolution involved more than three within-replichore inversions, we calculated the fraction of the chromosome length that switched its chain compared to the common ancestor and calculated the p-values of the null distributions. No significant tendency for reversal could be observed (data not shown).

Only 560 of 2300 universal OGs have never switched the chain compared to the common ancestor. We tested whether the core genes less frequently switch the chain by inversions using the list of 139 bacterial core genes (Rinke et al., 2013). No bias could be observed, as the fractions of core genes in stable and inverted synteny blocks were roughly equal (data not shown).

CRISPR analysis

CRISPR cassettes of the considered *Y. pestis* strains are shown in Fig. 5. Initially, we constructed separate phylogenetic trees for each of the three CRISPR-loci using the parsimony approach (Fig. 6, see Methods). As the number of events in each locus was small, the history of each locus could be reconstructed unambiguously.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3223v1 | CC BY 4.0 Open Access | rec: 5 Sep 2017, publ: 5 Sep 2017

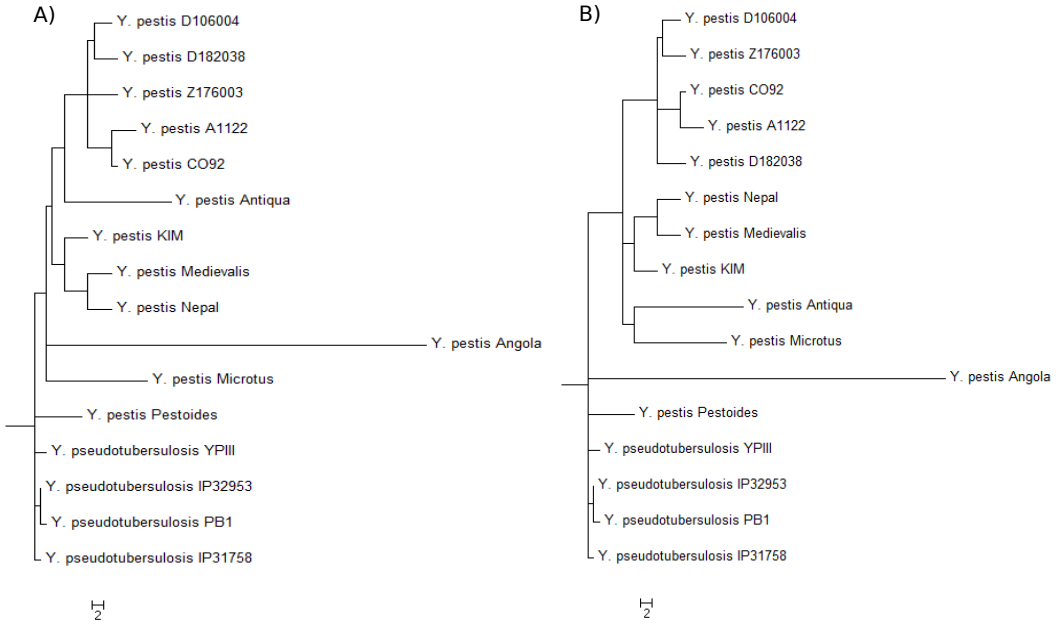

Figure 2. Phylogenetic trees *Yersinia spp.* reflecting the history of rearrangements. Branch lengths reflect the number of inversions, red top (blue bottom) numbers indicate the number of insertions (resp. deletions). (A) Optimal set of rearrangements for the sequence-based topology. (B) Optimal topology based on rearrangements.

However, the genome of *Y. pestis* evolves as a whole and the individual histories of the loci should be reconciled. In this case the reconstruction is ambiguous, as there are two equivalent reconstructions of the common ancestors and five equal positions of the Nepal strain on the maximum parsimony tree. Two maximum parsimony trees most compatible with the sequence tree are shown in Fig. 7. The trees constructed based on nucleotide sequences or rearrangements satisfy the rules of CRISPR-cassettes evolution (see Methods) but each of them implies two additional losses of cassette segments in comparison with the maximum parsimony tree. In particular, the sequence-based tree implies two independent parallel losses of the same segments of the main locus on the Angola and Antiqua strains branches.

CONCLUSIONS

Detailed reconstruction of evolution of bacterial strains provides a framework for epidemiological studies and analysis of acquired pathogenesis loci and drug resistance determinants.

Using *Y. pestis* as an example, we demonstrate that integrated analysis of sequence-based and inversion-based trees enhances the resolution of phylogenetic reconstruction. At that, inversions may resolve branches with low bootstrap support; on the other hand, sequence analysis may distinguish between parallel inversions and single inversion propagated by homologous recombination of a larger block.

In contrast, reconstructions of strain relationships based on solely CRISPR loci may not be reliable, as the history is greatly obscured by large deletions, obliterating the order of spacer gains. Even less reliable seem to be reconstructions based on shared spacer content. Similarly, numerous parallel gene losses preclude reconstruction of phylogeny based on gene content.

The hypothesis that large within-replichore inversions are usually balanced by subsequent events was not confirmed. However, it might be caused by a lack of data as inter-replichore inversions occurred rarely. The hypothesis that core genes tend not to change their chain during evolution was discarded.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3223v1 | CC BY 4.0 Open Access | rec: 5 Sep 2017, publ: 5 Sep 2017

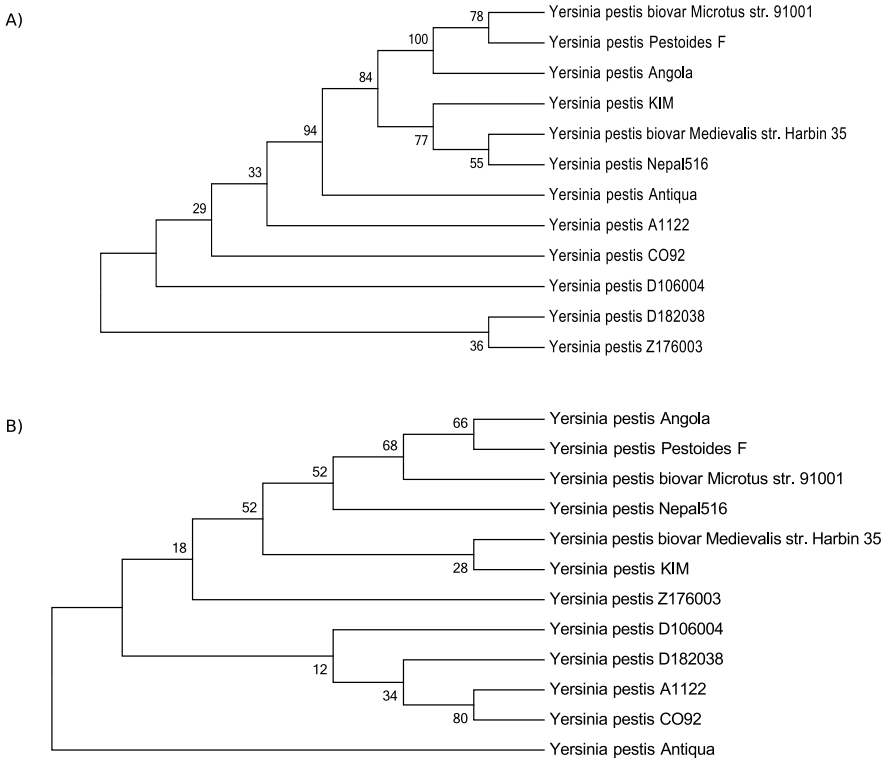

Figure 3. Phylogenetic tree for *Yersinia pestis* genes involved in the parallel inversions (A) in A1122 and D182038 and (B) in Z176003 and D106004.

ACKNOWLEDGMENTS

This study was supported by the Russian Science Foundation under grant 14-50-00150. It was initiated at the Summer School of Molecular and Theoretical Biology supported by the Dynasty foundation.

REFERENCES

Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiyoule, A., and Carnie, E. (1999). *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci U S A*, 96(24). 14043–14048.

Avdeyev, P., Jiang, S., Aganezov, S., Hu, F., and Alekseyev, M. A. (2016). Reconstruction of ancestral genomes in presence of gene gain and loss. *Journal of Computational Biology*, 23(3):150–164.

Barros, M., França, C., Lins, R., Santos, M., Silva, E., Oliveira, M., Silveira-Filho, V., Rezende, A., Balbino, V., and Leal-Balbino, T. (2014). Dynamics of CRISPR loci in microevolutionary process of *Yersinia pestis* strains. *PLoS One*, 9(9):e108353.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. (2013). GenBank. *Nucleic Acids Res.*, 45(Database issue):D30–5.

Chain, P. S., Hu, P., Malfatti, S. A., Radnedge, L., Larimer, F., Vergez, L. M., Worsham, P., Chu, M. C., and Andersen, G. L. (2006). Complete genome sequence of *Yersinia pestis* strains Antiqua and Nepal516: Evidence of gene reduction in an emerging pathogen. *J Bacteriol.*, 188(12).

Cui, Y., Li, Y., Gorgé, O., Platonov, M., Yan, Y., Guo, Z., Pourcel, C., Dentovskaya, S., Balakhonov, S., Wang, X., Song, Y., Anisimov, A., Vergnaud, G., and Yang, R. (2008). Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced short palindromic repeats. *PLoS One*, 3(7):e2652.

Darling, A. E., Miklós, I., and Ragan, M. A. (2008). Dynamics of genome rearrangement in bacterial populations. *PLoS Genet.*, 4(7):e1000128.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3223v1 | CC BY 4.0 Open Access | rec: 5 Sep 2017, publ: 5 Sep 2017

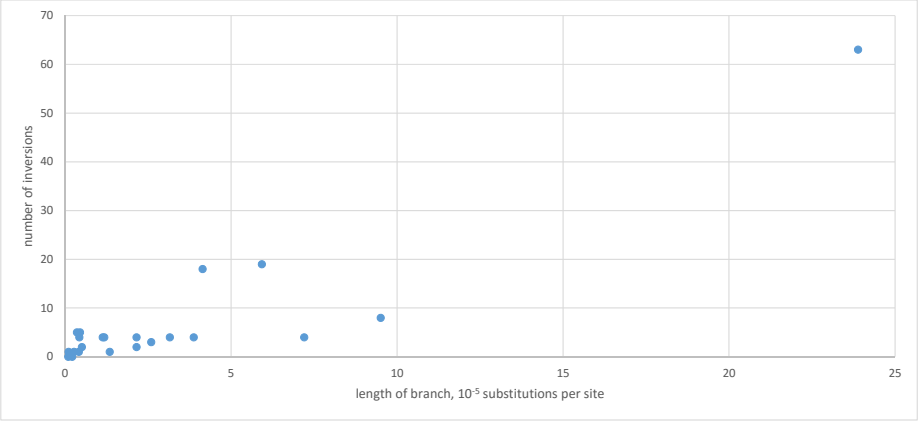

Figure 4. Correlation between inversion rates and mutation rates. Each dot corresponds to a branch in the *Y. pestis* phylogenetic tree. Horizontal axis branch length in substitution per site. Vertical axis shows the number of inversions.

Deng, W., Burland, V., Plunkett, G. I., Boutin, A., Mayhew, G., Liss, P., Perna, N., Rose, D., Mau, B., Zhou, S., Schwartz, D., Fetherston, J., Lindler, L., Brubaker, R., Plano, G., Straley, S., McDonough, K., Nilles, M., Matson, J., Blattner, F., and Perry, R. (2002). Genome sequence of *Yersinia pestis* KIM. *J Bacteriol*, 184(16):4601–4611.

Duchemin, W., Daubin, V., and Tannier, E. (2015). Reconstruction of an ancestral *Yersinia pestis* genome and comparison with an ancient sequence. *BMC Genomics*, 16.

Edgar, R. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, 32(5):1792–1797.

Eppinger, M., Worsham, P. L., Nikolich, M. P., Riley, D. R., Sebastian, Y., Mou, S., Achtman, M., Lindler, L. E., and Ravel, J. (2010). Genome sequence of the deep-rooted *Yersinia pestis* strain Angola reveals new insights into the evolution and pangenome of the plague bacterium. *J Bacteriol.*, 192(6):1685–1699.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics*, 8:172.

Kupczok, A. and Bollback, J. (2013). Probabilistic models for CRISPR spacer content evolution. *BMC Evol Biol.*, 13(54).

Liang, Y., Hou, X., Wang, Y., Cui, Z., Zhang, Z., Zhu, X., Xia, L., Shen, X., Cai, H., Wang, J., Xu, D., Zhang, E., Zhang, H., Wei, J., He, J., Song, Z., Yu, X., Yu, D., and Hai, R. (2010). Genome rearrangements of completely sequenced strains of *Yersinia pestis*. *J Clin Microbiol*, 48(5):1619–1623.

Liang, Y., Xie, F., Tang, X., Wang, M., Zhang, E., Zhang, Z., Cai, H., Wang, Y., Shen, X., Zhao, H., Yu, D., Xia, L., and Hai, R. (2014). Chromosomal rearrangement features of *Yersinia pestis* strains from natural plague foci in China. *Am J Trop Med Hyg.*, 91(4):722–728.

Martínez-Chavarría, L. C. and Vadyvaloo, V. (2015). *Yersinia pestis* and *Yersinia pseudotuberculosis* infection: a regulatory RNA perspective. *Front Microbiol.*, 6:956.

Minkin, I., Patel, A., Kolmogorov, M., Vyahhi, N., and Pham, S. (2013). Sibelia: A scalable and

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3223v1 | CC BY 4.0 Open Access | rec: 5 Sep 2017, publ: 5 Sep 2017

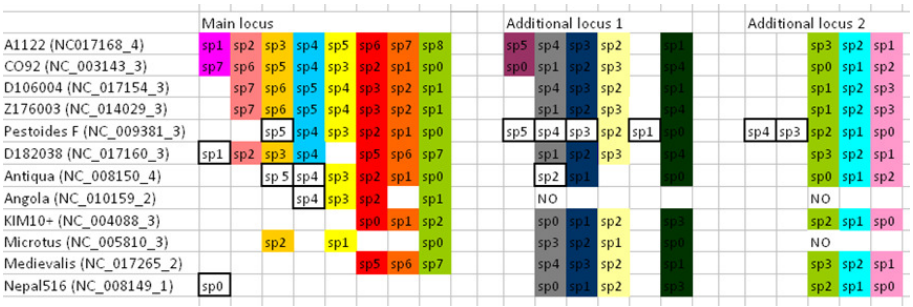

Figure 5. CRISPR-cassettes of completely sequenced *Y. pestis* strains. Cassette IDs and spacer numbers are given according to CRISPRdb (Grissa et al., 2007). Identical spacers are shown by the same color; unique spacers are set in frames.

comprehensive synteny block generation tool for closely related microbial genomes. *13th Workshop on Algorithms in Bioinformatics (WABI2013)*.

Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, 151(Pt 3):653–63.

Rasmussen, S., Allentoft, M. E., Nielsen, K., Orlando, L., Sikora, M., Sjögren, K.-G., Pedersen, A. G., Schubert, M., Dam, A. V., Kapel, C. M. O., Nielsen, H. B., Brunak, S., Avetisyan, P., Epimakhov, A., Khalyapin, M. V., Gnuni, A., Kriiska, A., Lasak, I., Metspalu, M., Moiseyev, V., Gromov, A., Pokutta, D., Saag, L., Varul, L., Yepiskoposyan, L., Sicheritz-Pontén, T., Foley, R. A., Lahr, M. M., Nielsen, R., Kristiansen, K., and Willerslev, E. (2015). Early divergent strains of *Yersinia pestis* in Eurasia 5,000 years ago. *Cell*, 163(3):571–582.

Riehm, J., Projahn, M., Vogler, A., Rajerison, M., Andersen, G., Hall, C., Zimmermann, T., Soanandrasana, R., Andrianaivoarimanana, V., Straubinger, R., Nottingham, R., Keim, P., Wagner, D., and Scholz, H. (2015). Diverse genotypes of *Yersinia pestis* caused plague in Madagascar in 2007. *PLoS Negl Trop Dis*, 9(6):e0003844.

Riehm, J., Vergnaud, G., Kiefer, D., Damdindorj, T., Dashdavaa, O., Khurelsukh, T., Zöller, L., Wölfel, R., Flèche, P. L., and Scholz, H. (2012). *Yersinia pestis* lineages in Mongolia. *PLoS One*, 7(2):e30624.

Rinke, C., Darling, A., Malfatti, S., Tsiamis, G., Stepanauskas, R., Schwientek, P., Sievert, S. M., Rubin, E. M., Sczyrba, A., Ivanova, N. N., Swan, B. K., Liu, W.-T., Eisen, J. A., Hugenholtz, P., Woyke, T., Gies, E. A., Dodsworth, J. A., and Hallam, S. J. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, 499(7459):431–7.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, 30(12):2725–2729.

Vergnaud, G., Li, Y., Gorgé, O., Cui, Y., Song, Y., Zhou, D., Grissa, I., Dentovskaya, S., Platonov, M., Rakin, A., Balakhonov, S., Neubauer, H., Pourcel, C., Anisimov, A., and Yang, R. (2007). Analysis of

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3223v1 | CC BY 4.0 Open Access | rec: 5 Sep 2017, publ: 5 Sep 2017

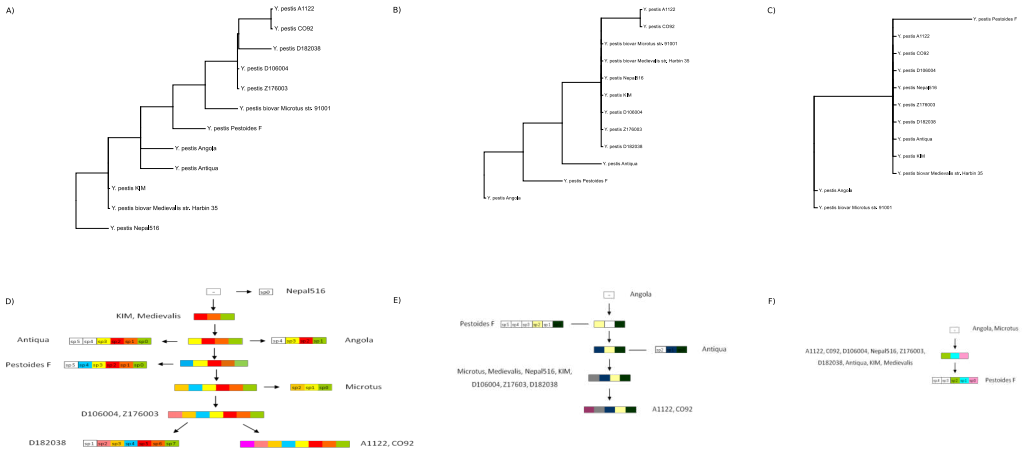

Figure 6. Cladograms (A, B, C) and schemas of evolution (D, E, F) of three CRISPR loci of *Y. pestis*. (A, D) The main, most variable, locus; (B, E) additional locus 1; (C, D) additional locus 2

the three *Yersinia pestis* CRISPR loci provides new tools for phylogenetic studies and possibly for the investigation of ancient DNA. *Adv Exp Med Biol*, 603:327–38.

Zhang, J. and Madden, T. (1997). PowerBLAST: A new network BLAST application for interactive or automated sequence analysis and annotation. *Genome Res.*, 7(6):649–656.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3223v1 | CC BY 4.0 Open Access | rec: 5 Sep 2017, publ: 5 Sep 2017

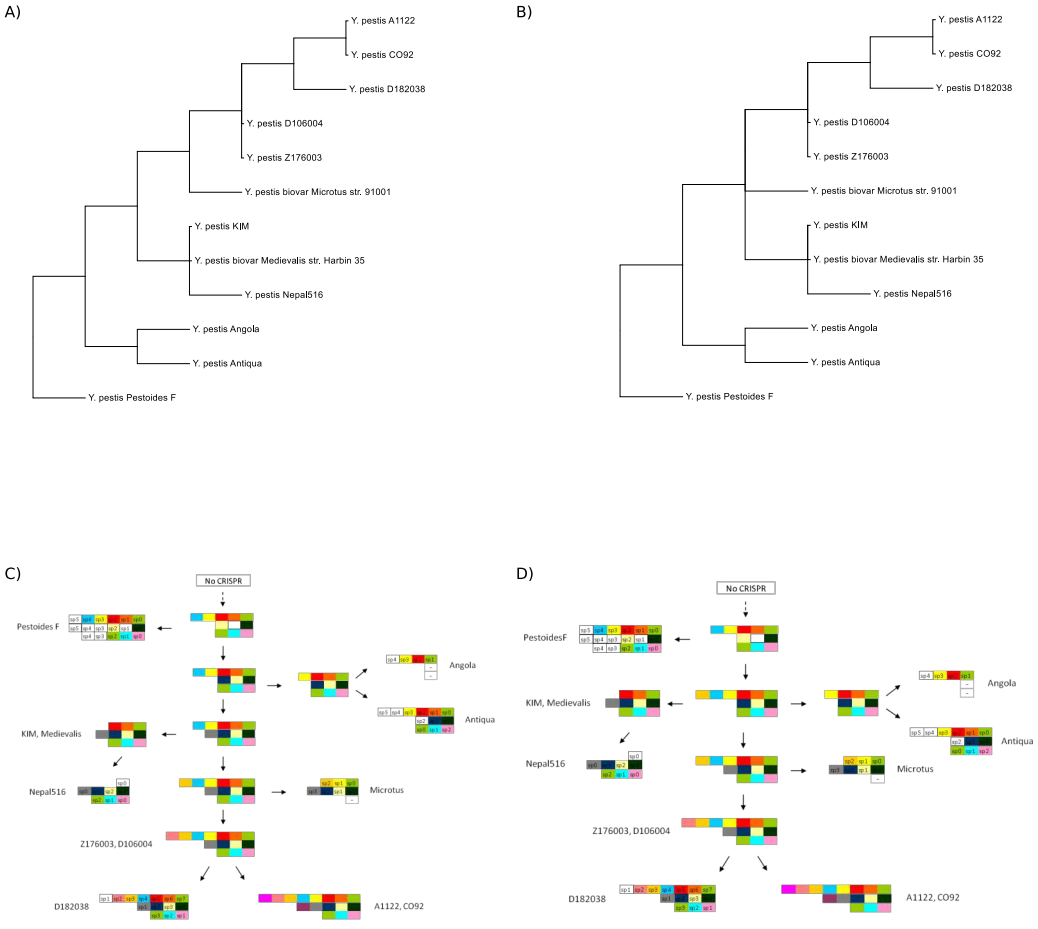

Figure 7. Cladograms (A, B) and schemas of evolution (C, D) of two integrated CRISPR-based maximum parsimony phylogenetic trees most compatible with the sequence tree

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3223v1 | CC BY 4.0 Open Access | rec: 5 Sep 2017, publ: 5 Sep 2017