Trends of the Major Porin Gene (ompF) Evolution: Insight from the Genus Yersinia

Anna M. Stenkova1*, Marina P. Isaeva1, Felix N. Shubin2, Valeri A. Rasskazov1, Alexander V. Rakin3

1 Pacific Institute of Bioorganic Chemistry, Far Eastern Branch of Russian Academy of Sciences, Vladivostok, Russian Federation, 2 Scientific Research Institute of Epidemiology and Microbiology, Siberian Branch of Russian Academy of Medical Sciences, Vladivostok, Russian Federation, 3 Max von Pettenkofer Institute for Hygiene and Clinical Microbiology of Ludwig Maximilians-University, Munich, Germany

Abstract

OmpF is one of the major general porins of Enterobacteriaceae that belongs to the first line of bacterial defense and interactions with the biotic as well as abiotic environments. Porins are surface exposed and their structures strongly reflect the history of multiple interactions with the environmental challenges. Unfortunately, little is known on diversity of porin genes of Enterobacteriaceae and the genus Yersinia especially. We analyzed the sequences of the ompF gene from 73 Yersinia strains covering 14 known species. The phylogenetic analysis placed most of the Yersinia strains in the same line assigned by 16S rDNA-gyrB tree. Very high congruence in the tree topologies was observed for Y. enterocolitica, Y. kristensenii, Y. ruckeri, indicating that intragenic recombination in these species had no effect on the ompF gene. A significant level of intra- and interspecies recombination was found for Y. alevsicae, Y. intermedia and Y. mollaretii. Our analysis shows that the ompF gene of Yersinia has evolved with nonrandom mutational rate under purifying selection. However, several surface loops in the OmpF porin contain positively selected sites, which very likely reflect adaptive diversification Yersinia to their ecological niches. To our knowledge, this is a first investigation of diversity of the porin gene covering the whole genus of the family Enterobacteriaceae. This study demonstrates that recombination and positive selection both contribute to evolution of ompF, but the relative contribution of these evolutionary forces are different among Yersinia species.

Introduction

The genus Yersinia, a member of the Enterobacteriaceae family, is currently composed of 14 known species: Y. pestis, Y. pseudotuberculosis, Y. enterocolitica Y. aldovae, Y. alevsicae, Y. berovieri, Y. frederiksenii, Y. intermedia, Y. kristensenii, Y. massiliensis, Y. mollaretii, Y. rohden, Y. ruckeri, and Y. similis [1–3]. Three of them are well documented human pathogens. Y. pestis is the etiologic agent of plague while Y. pseudotuberculosis and Y. enterocolitica are known to cause a variety of gastrointestinal symptoms [4]. The characterization of the remaining 11 species is more limited. However, these species accepted as human nonpathogenic possess novel virulence mechanisms, and some of them have been associated with human cases [5,6]. Yersinia are disseminated all over the world in terrestrial and aquatic environments, and associated with many different hosts (plants, animals, insects, fish and so on). Despite recent advances in our understanding of the pathobiology of Yersinia, the molecular-genetic mechanisms by which Yersinia colonizes and adapts to various host or environmental conditions are still poorly understood. In this context, membrane surface molecules are considered the major targets of the membrane-environment interaction.

General bacterial porins (GBPs) are one of the most abundant proteins (up to 10^5 copies per cell) in the outer membrane of the gram-negative bacteria [7,8]. Structurally, a typical GBP subunit consists of 16 antiparallel β-strands forming a β-barrel, with short turns facing the periplasmic space and long loops facing the external surface of bacterial membrane [9–11]. Three porin subunits are assembled into stable homotrimers. The best-studied GBPs, which include OmpF, OmpC and PhoE of E. coli, differ in their solute selectivity, porin activity and gene expression in response to many environmental factors, such as osmotic pressure, temperature and pH [12–14]. Porins are one of the first molecules responding to environmental changes and at least for some bacteria have been found to reflect their ecological niche by the sequence type [15,16]. As the major components of the outer membrane, some pore-forming proteins play a role in bacterial pathogenesis, such as adherence, invasion, and serum resistance [17–20].

Little is known about evolution and diversity of GBPs of the Enterobacteriaceae at all and the Yersinia especially. Scattered reports showed that Yersinia’s major porin is the β-structured protein resistant to high temperature, proteases, and detergents [21–23]. Primary structure and topology of the OmpF porin of pathogenic Yersinia was determined and demonstrated 55% homology with E. coli and 70% homology with Serratia marcescens OmpFs, respectively [24,25]. Here we conducted an in-depth study of the ompF gene diversity in all currently known Yersinia paying special attention to
evolution inference and phylogenetic relationships of these bacteria.

**Results and Discussion**

16S rDNA and gyrB sequence variations and genetic relationships among Yersinia species

To justify evolutionary relationships and taxonomic position, 16S rDNA and gyrB genes sequences were analysed in all Yersinia strains used in this study (Table 1). The 16S rDNA gene sequencing has definitely allowed Yersinia identification [26] and recognizing novel species and subspecies within the genus [5, 27, 28]. However, the 16S rDNA sequence analysis cannot resolve the phylogenetic relationships between closely related Yersinia species [2, 29]. Recently, gyrB has been successfully applied to characterization of Y. frederiksenii genospecies [30] and was included as one of the MLST gene targets for studying genetic relationships among Yersinia species [29].

In this study, the total number of Yersinia strains was 65, covering all Yersinia species, and originating from different sources and geographic locations (Table 1). 16S rDNA and gyrB sequences from all the above strains were PCR amplified and sequenced. Eight additional sequences of each gene were obtained from publicly available Yersinia genomes (http://www.ncbi.nlm.nih.gov). In total, 73 16S rDNA and gyrB sequences were analysed. The sequences were aligned and adjusted to 750 bp for 16S rDNA and to 838 bp for gyrB. Each unique sequence, differing in one or more nucleotide or amino acids sites, was assigned as a distinctive allele, resulting in 31 alleles for 16S rDNA and 48 alleles for gyrB (Table 1). The number of the detected alleles for 16S rDNA ranged from 1 of 16S rDNA per species (or 21 alleles for GyrB (Table 1). The number of the detected alleles for gyrB has been successfully applied to characterization of Y. frederiksenii genospecies [30].

It reaffirmed that Y. frederiksenii sp. nov., since, Y. frederiksenii was recently separated from Y. kristensenii [1] and Y. massiliensis is more closely related to Y. frederiksenii [2]. Therefore, these strains were designated as Y. aleksiciae-like and Y. massiliensis-like, respectively. Based on the 16S rDNA-gyrB tree, most Y. intermedia clustered together into one of two branches; four Russian strains (6044, 5954, 6270 and 601) were located on the line leading to the rest Y. intermedia, shared the intraspecies distances up to 0.007.

Taken together, species identification of Yersinia strains based on the 16S rDNA-gyrB concatenated tree was in relative agreement with the MLST tree reported previously [29]. Three Y. kristensenii strains (991, Y332 and 6266) were designated as Y. aleksiciae-like and one Y. frederiksenii strain (2043) was as Y. massiliensis-like. Six Yersinia species (Y. pestis, Y. pseudotuberculosis, Y. hovdei, Y. ruckeri, Y. rohdei and Y. alcocia) and to 13 alleles of gyrB (Table 1). The number of allele variants slightly varied from those published previously [29] possibly because of inclusion of more distant strains and/or increasing the lengths of the analyzed fragments.

In order to correctly identify each strain examined, a neighbour-joining tree was constructed from the 16S rDNA-gyrB concatenated sequences (Fig. 1). Ten Yersinia species (Y. aldovae, Y. bercovieri, Y. enterocolitica, Y. intermedia, Y. mollaretii, Y. pestis, Y. pseudotuberculosis, Y. similis, Y. rohdei and Y. ruckeri) were clearly grouped into relatively distinct clusters. The intraspecies genetic distance means of these species were up to 0.012. Y. pestis strains clustered tightly with the Y. pseudotuberculosis strains and the distance mean for this group was 0.029, and the group means of 0.097–0.166. Phylogenetic tree. We found 62 unique nucleotide alleles of the ompF gene (table 1), which clustered into 18 groups on the tree (Fig. 2). Though different algorithms and clustering methods produced similar topologies of the ompF tree, phylogenetic clustering of the strains performed by neighbor-joining method with Kimura 2-parameter algorithm gave the highest bootstrap values. With the exception of five species, Y. similis, Y. intermedia, Y. mollaretii, Y. frederiksenii and Y. aleksiciae, all strains that belong to the same Yersinia species were clustered in one group. Five strains of Y. ruckeri clustered together in a distinct group V and showed the intragroup distance mean by 0.002 and the largest intergroup genetic distance means from 0.166 to 0.197. It reaffirmed that Y. ruckeri has been fairly clonal and genetically the most distant species within the genus [26, 29]. For Y. enterocolitica (group VII) the intraspecies genetic distance mean was 0.029, and the groups means of 0.097–0.166. Phylogenetic grouping of Y. enterocolitica ompF genes exactly replicated that of 16S rDNA-gyrB sequences with division in two subspecies, Y. enterocolitica subsp. palea (Y11, 1234, 2974/81, 6579, 1245, 2720/87, and 1215) and Y. enterocolitica subsp. enterocolitica (WA220 and ATCC 8081), supported by a high bootstrap value (100%). Interestingly, in both phylogenetic trees, Y. enterocolitica subsp. palea clearly splits into two lines (bootstrap value 100%), one of them was only formed by Y. enterocolitica strains (1215, 1234, and 1245) isolated in Russian Far-East. Strains of Y. kristensenii formed group X with intragroup distance mean 0.020, and intergroup distance means 0.072–0.183. The strains of Y. pseudotuberculosis (group VI), Y. ruckeri (group IV), Y. rohdei (group XII) and Y. aldovae (group IV) were represented by only two strains and the within and between group distance means were up to 0.009 and 0.068–0.188, respectively. The strains of Y. pestis, Y. pseudotuberculosis and Y. similis grouped together (group VIII) with intragroup distance mean of 0.037, and between group distance means being 0.138–0.196. The VIII group splits into two subgroups with bootstrap value of 100%. One of these subgroups included two Y. pseudotuberculosis strains IP32953, IP31758 and Y. similis Y239, while the other—all Y. pestis strains and Y. pseudotuberculosis YPIII. This ompF tree topology did not correlate
| Species                  | Strain | Serotype | Source       | Country     | Allele type (NT/AT) | 16S RNA | gyrB    | ompF    |
|-------------------------|--------|----------|--------------|-------------|---------------------|---------|---------|---------|
| Y. aldovae              | Y112   |          |              |             |                     | 1       | 16/11   | 10/14   |
| ATCC 35236              |        |          | Water        | Czechoslovakia |                     | 1       | 46/20   | 58/14   |
| Y. aleksicæ             | Y159   |          |              | Germany     |                     | 2       | 15/11   | 11/15   |
| Y. bercovieri           | ATCC 43970 | | Human feces | France      |                     | 17      | 1/2     | 13/16   |
| H632-36/85              |        |          |              |             |                     | 14      | 1/2     | 12/17   |
| Y. enterocolitica subsp. | Y11    | O:3      |              |             |                     | 3       | 2/3     | 1/1     |
| palearctica             |        |          |              |             |                     |         |         |         |
| subsp. entenocolitica   |        | O:8      |              |             |                     |         |         |         |
| ATCC 8081               |        | O:8      | Human        | USA         |                     | 4       | 4/4     | 2/3     |
| Y. frederiksenii        | H56-36/81 | O:60   |              | Germany     |                     | 5       | 18/5    | 20/20   |
| 4648                    |        |          | Human feces  | Russia       |                     | 5       | 5/6     | 22/21   |
| 4849                    |        |          | Russia       |             |                     | 5       | 20/13   | 24/22   |
| ATCC 33641              |        |          | Sewage       | Denmark     |                     | 18      | 19/12   | 21/23   |
| 176–36                  |        |          |              |             |                     | 19      | 5/6     | 23/24   |
| Y. massiliensis³         | 2043   |          | Russia       |             |                     | 20      | 21/14   | 25/25   |
| Y. intermedia           | 5631   |          | Lemming      | Russia      |                     | 1       | 6/7     | 37/26   |
| 5934                    |        |          | Citellus     | Russia      |                     | 1       | 14/7    | 30/27   |
| 6325                    |        |          | Lemming      | Russia      |                     | 1       | 27/7    | 6/7     |
| ATCC 29909              |        |          | Human urine  |             |                     | 6       | 7/7     | 28/28   |
| 5373                    |        |          | Water        | Russia      |                     | 6       | 6/7     | ¼       |
| 6390                    |        |          | Lemming      | Russia      |                     | 6       | 9/7     | 5/6     |
| 5593                    |        |          | Lemming      | Russia      |                     | 6       | 24/16   | 5/6     |
| 5986                    |        |          | Field mouse  | Russia      |                     | 6       | 7/7     | 34/29   |
| H537/85                 | O:3    |          |              |             |                     | 6       | 10/7    | 35/30   |
| N27/84                  |        | 52,53:2q | Water        | Germany     |                     | 6       | 9/7     | 5/6     |
| H9-36/83                |        | O:17     |              | Germany     |                     | 6       | 7/7     | 26/31   |
| N13/84                  |        | 37:q     | Human        | Germany     |                     | 7       | 8/7     | 27/32   |
| 1948                    |        |          | Water        | Russia      |                     | 7       | 25/7    | 4/5     |
| 5828                    |        |          | Field mouse  | Russia      |                     | 7       | 8/7     | 36/33   |
| 6043                    |        |          | Russia       |             |                     | 7       | 26/7    | 38/34   |
| 5375                    |        |          | Water        | Russia      |                     | 7       | 8/7     | 4/5     |
| 5638                    |        |          | Lemming      | Russia      |                     | 7       | 6/7     | ¼       |
| 6270                    |        |          | Lemming      | Russia      |                     | 13      | 14/7    | 29/35   |
| 6044                    |        |          | Field mouse  | Russia      |                     | 13      | 23/15   | 33/36   |
| N9/83                   |        | 17:q     | Human        | Germany     |                     | 21      | 10/7    | 31/37   |
| 601                     |        |          | Russia       |             |                     | 22      | 22/7    | 32/38   |
| Y. kristensenii         | 6276   |          | Lemming      | Russia      |                     | 23      | 28/17   | 6/7     |
| 5306                    |        |          | Sorex araneus| Russia      |                     | 15      | 35/8    | 42/9    |
| 5862                    |        |          | Field mouse  | Russia      |                     | 15      | 34/8    | 7/9     |
| 5932                    |        |          | Field mouse  | Russia      |                     | 15      | 36/8    | 43/40   |
| 6032                    |        |          | Sorex araneus| Russia      |                     | 16      | 33/8    | 7/9     |
| 5868                    |        |          | Anas acuta   | Russia      |                     | 16      | 32/8    | 7/9     |
| 6572                    |        |          | Carrot       | Russia      |                     | 24      | 31/8    | 41/41   |
with the 16S rDNA-\textit{gyrB} tree branching, possibly indicating interspecies recombination between \textit{Y. pseudotuberculosis} and \textit{Y. similis}, or/and diversification of the \textit{ompF} gene of \textit{Y. pseudotuberculosis} before emergence of \textit{Y. pestis} by adaptive evolution.

The \textit{ompF} sequences of the remaining species, \textit{Y. intermedia}, \textit{Y. mollaretii}, \textit{Y. frederiksenii} and \textit{Y. aleksiciae}, exhibited different phylogenetic relationships and produced incongruent molecular phylogenies with the 16S rDNA-\textit{gyrB} tree. The \textit{Y. frederiksenii} strains, that were genetically distinct and not closely related to each other according to the 16S rDNA-\textit{gyrB} tree, split into three groups; two groups (XI and XII) were presented by single strains, and IX group was by strains with intragroup of 0.043 and intergroup from 0.112 to 0.178 distance means. From previously characterized \textit{Y. frederiksenii} genomic groups [31], \textit{Y. frederiksenii} IX and XI groups of \textit{ompF} could corresponded to genomic groups 1b and 1a, respectively, and XII group (\textit{Y. massiliensis})-to genomospecies 2. A mixed branching pattern was found in \textit{Y. mollaretii} and \textit{Y. aleksiciae} strains. Two \textit{Y. aleksiciae} strains (Y159 and Y332) grouped together with three \textit{Y. mollaretii} strains (H279-36/85, N850/89 and ATCC43969), whereas two others \textit{Y. aleksiciae} strains (991 and 6266) grouped together with three other \textit{Y. mollaretii} (87-36/87, H87/82 and N846/89). Therefore, \textit{Y. aleksiciae} (groups XVI and XVII) and \textit{Y. mollaretii} (groups XV and XVIII) strains split into two relatively distinct groups with intragroup distance means up to 0.007, and intergroup distance means of 0.048–0.197. Interestingly, \textit{Y. aleksiciae} recently isolated from \textit{Y. kristensenii} was more closely related to \textit{Y. bercovieri} and \textit{Y. mollaretii} than to \textit{Y. kristensenii}, and that was confirmed by the 16S rDNA-\textit{gyrB} tree. Previously, based on the concatenated \textit{tufA-tufB} tree, \textit{Y. aleksiciae} type strain LMG 22254 was found to be distinct from the \textit{Y. kristensenii} cluster and clearly grouped with \textit{Y. bercovieri} and \textit{Y. mollaretii} [32]. The most genetically heterogeneous was \textit{Y. intermedia} that formed four different groups (I-III, XIII) with intragroup distance means up to 0.039, and with between groups means being 0.076–0.195. Moreover, XIII group, formed by five \textit{Y. intermedia} strains (601, Nr12/84, 1948, 5375 and 5631) was separated from the rest \textit{Y. intermedia} groups by a number of genetic clades.

As mentioned above, some species produced incongruent 16S rDNA-\textit{gyrB} and \textit{ompF} phylogenies. A mix branching pattern can be a sign of recombination, whereas in the case of mutation the gene trees look the same [33]. To verify this assumption, we used four tests (RDP, MaxChi, Chimera, and Geneconv) in the RDP3.34 package for investigation of the \textit{ompF} gene of all \textit{Yersinia} groups. We detected four recombination events with break-points involving three species, \textit{Y. intermedia} (groups I, II, XIII), \textit{Y. aleksiciae} (groups XVI and XVII) and \textit{Y. mollaretii} (groups XVIII and XV).

| Species               | Strain | Serotype | Source       | Country | Allele type (NT/AT)A | 16S RNA | gyrB | ompF |
|-----------------------|--------|----------|--------------|---------|----------------------|---------|------|------|
| \textit{Y. pestis}     | 91001  |          | Human        | USA     | 11                   | 12/9    | 52/43|      |
| \textit{Y. pseu dotuberculosis} | IP 32953 | 1         | Human        | France  | 11                   | 44/9    | 55/44|      |
| \textit{Y. rohdei}     | H274-36/78 | O:76    | Dog feces    | USA     | 28                   | 45/19   | 57/47|      |
| \textit{Y. ruckeri}    | Nr 34/85 |          | Fish         | Germany | 12                   | 13/10   | 9/13 |      |
| \textit{Y. similis}    | Y239   |          | Germany      |         | 31                   | 48/21   | 62/50|      |
| **Total allele number** |        |          |              |         | 31                   | 48/21   | 62/50|      |

\textit{A–NT-nucleotide sequence type, AT-amino acid sequence type. B–Species identity corrected by 16S-\textit{gyrB} genotype.}

doi:10.1371/journal.pone.0020546.t001

Table 1. Cont.
(Fig. 3). From the ompF tree, one can suppose that a recombination event between ompFs of Y. alexsiae and Y. mollaretii occurred twice. In the first case, ompF of Y. mollaretii group XV served as a donor and ompF of Y. alexsiae group XVII was a recipient, producing a recombinant ompF allele of Y. alexsiae group XVI. And vice versa, ompF of Y. alexsiae group XVII served as a donor and ompF of Y. mollaretii group XV was a recipient, giving a recombinant ompF allele of Y. mollaretii group XVIII. This explanation comes from comparison of the branch length and sequence diversity of the group members. To our data, interspecies intragenic recombination was detected for the first time in the genus Yersinia. We observed a complex pattern of recombination in Y. intermedia ompF (groups I, II, XIII). Group I mainly played a parental role in different recombination events, giving ompF variants of Y. intermedia groups II and XIII; other players of the events were not identified in this analysis. It should be noted, that group I strains are most numerous and widely geographically distributed. So it can be supposed, that this ompF variant is more spread and successful in coexistence with mammals including evolutionary newcomers, humans. Acquisition of regions of a successful allele by recombination can be preferred for minor variants (groups II and XIII) when bacterium get into a new niche such as mammals. Interestingly, an extraordinary position of the XIII group on the phylogenetic tree indicates a new origin of the ompF gene not represented by any known Yersinia species. The fact, that this group includes a human isolate (Nr13/84) may be an evidence of occurrence in new niche, human. Noteworthy, it was extremely difficult to reconstruct a scenario of recombination events for all Y.
intermedia ompF. This might be a subject of further research, as well as investigation of associations within a specific niche.

It is very interesting to note that one of the brake-points of all recombinant ompF is located in the same region (431–501 bp), corresponding to 6-th β-strand of OmpF. The reconstruction of the ompF tree for the region 1–501 bp produced very similar branch pattern with that of the 16S rDNA-gyrB tree (data not shown). We suspect that significant nucleotide similarity in this region (with the exception of the external loops) reflects a strong selective pressure (purifying selection) due to an important functional role of this region as a zone of monomer’s contacts in a porin trimer. This might be an evident example of protein structural constraints.

Examples of the homologous recombination in porin genes have been recognized for some bacteria, mainly for naturally transformable species as Pseudomonas (OprD), Neisseria (PorB, OmpA), Chlamydia (OmpA), and Leptospira (OmpL1) [16,34–38]. For these genes different mosaic patterns have been identified. The intragenic recombination has been frequently observed within species due to the transfer of a portion or an entire gene. As a rule, the exchanges occur only in the loop regions and do not affect the transmembrane domains. Moreover, rare cases of interspecies recombination of porin genes have been described in literature. It was suggested that porB2, an allele of porB, arose in meningococci by interspecies recombination between ancestral pathogenic and commensal Neisseria species [39]. Also, an interspecies recombination in ompA between a mouse strain of C. trachomatis and a horse strain of C. pneumoniae was documented [35]. Multiple interspecies recombination patterns were observed among ompL1 genes, belonging to four different Leptospira species [38].

The phylogenetic analysis of ompF sequences placed most of the Yersinia strains in the same line assigned by 16S rDNA-gyrB tree with the exception of six species, Y. pseudotuberculosis, Y. similis, Y. frederiksenii, Y. intermedia, Y. mollaretii, and Y. aleksicue. The incongruence of ompF and 16S rDNA-gyrB trees indicated the inter- and intraspecies recombination. Despite extensive recomb-
nation events in the *Yersinia ompF* genes, this seems to happen not so often to remove all phylogenetic signals.

**Adaptive evolution of the ompF gene in Yersinia**

As it was shown above, the *ompF* gene of *Yersinia* is more divergent than the 16S rDNA and *gyrB* genes. The nucleotide diversity for all *ompF* genes (0.131 ± 0.005) is twofold higher than for housekeeping genes (0.051 ± 0.004). The common alignment of 73 *ompF* sequences contains 40% (479/1200 bp) of polymorphic nucleotide sites, which distributed strikingly nonrandom and formed hypervariable and conserved regions (Fig. 4). We have divided *yersinia’s* *ompF* gene into 18 regions, according to domain organization of *Escherichia coli* OmpF protein [9]. Loops L2, L4–L7 were characterized by nucleotide deletions and/or insertions. Comparative analysis of surface-exposed loops exhibited significant heterogeneity of L4 and L5 (46 ± 4.5%). The highest homology was conserved in L3 (8.2 ± 1.6%). The same nonrandom heterogeneity with characteristic conserved regions forming the β-barrel structure of the proteins, and variable regions, making up the putative surface-exposed loops, has been shown in some other porins [39,40].

To estimate deviation in codon usage, the codon adaptation index (CAI) was calculated for the *ompF* gene. CAI is a measure of the relative adaptiveness of the codon usage of a gene towards the codon usage of highly expressed genes of that organism: the higher the index value, the greater the codon usage bias [41]. As a reference for highly expressed genes, we used the 27 concatenated ribosomal genes for ten *Yersinia* species. The genes of the ribosomal proteins had a CAI value from 0.52 to 0.56 for all species, but CAI values for the *ompF* gene were higher (from 0.64 to 0.75). Therefore, there is a strong codon usage bias in the *ompF* gene in all *Yersinia* species, as expected for highly expressed genes. This is another reason to assume that the high level of *ompF* transcription may be also responsible for nonrandom heterogeneity in the gene.

To determine how the level of selective constraint varies along the *ompF* gene, we estimated the numbers of synonymous substitutions per synonymous site (dS) and nonsynonymous substitutions per nonsynonymous site (dN) and calculated the dS/dN ratio for the *ompF* gene. If purifying selection has occurred, a gene has a dS/dN > 1. Absence of selection should generate dS/dN = 1. A ratio dS/dN < 1 indicates diversifying selection or accelerated evolution [42,43]. We excluded *Yersinia* groups with recombination events from analysis and dealt only with six *ompF* groups of *Yersinia* (VII, VIII, IX, I, X, XIII). The dS/dN ratio was calculated as an average over all of the codon sites in each *ompF* group using the Nei-Gojobori method by MEGA 4 of Jukes-Cantor model. Statistical significance was tested by Codon-based Z-test. For all groups we detected approximately identical dS/dN.
means from 4.224 to 5.748 with p<0.05 of purifying selection. Thus, ompF gene is under strong purifying selection in all six Yersinia groups. Nielsen and Yang method [44], compiled in Site wise likelihood ratio estimation programme [45], was used to identify the sites with the evidence of positive selection in selected ompF groups. The porin protein structures for these groups have been simulated and sites with weak or strong positive selection have been located on the models (Fig. 5).

When these selected sites were mapped onto three-dimensional structural models, it becomes clear that the majority fell within regions predicted to encode surface-exposed loop regions. It is important to note that these sites were located in different surface loops of analyzed Yersinia groups. For example, three residues in putative loop L1 were shown to be under strong selection in the group VIII, whereas there is no evidence of positive selected sites in putative loop L1 for groups VII, XIII, I and X. Smith N.H. observed unlike distribution of positive selected regions in porB genes in N. meningitides and N. gonorrhoeae [46]. Authors explained this by differences in the immune response to these two organisms. The impact of diversifying selection on ompC, ompF, lamB and fhuA omp’s genes of Escherichia and Shigella [47,48], ompC, ompS1 and ompS2 genes of Salmonella [49] has been demonstrated. Authors proposed that positive selection in omp genes may be an important mechanism that facilitates adaptation of bacterial pathogens allowing them to escape recognition by the host immune system, phages and penetration of antibiotics.

Our analyses demonstrated that the Yersinia ompF gene has evolved with nonrandom mutational rate under purifying selection in overall. However, the surface loops of the OmpF porin contain sites subjected to positive selection. Interestingly, such sites are located in different surface loops in different Yersinia species. We suppose that the ompF genes of different Yersinia species have evolved under individual constraints associated with unlike environmental challenges. Existence of both positive selection and recombination in porin genes has previously been reported for Neisseria porB and porA genes [37,50] as well as for ompA from Chlamidia [35] and Wolbachia [51]. In case of Yersinia ompF gene we consider that horizontally acquired fragments of some surface loops may be fixed by positive selection in process of species adaptation to new ecological niches. Such recombinant genes might supply their new hosts with benefits allowing to escape a deadly response of the immune system as well as lethal attacks of phages and antimicrobials. This might be more easily achieved by gene recombination rather than by random mutations. Moreover, these mechanisms seem to operate in evolution of porins genes of all taxonomic groups.

**Conclusion.** Genetic diversity of outer membrane proteins might result from bacterial adaptation to different ecological

---

**Figure 5. Location of positively selected sites in OmpF porins of Yersinia.** Group VII- Y. enterocolitica WA220; Group XIII-Y. intermedia 1948; Group IX-Y. frederiksenii 4648; Group I-Y. intermedia ATCC 29909; Group X-Y. kristensenii 5868; Group VIII-Y. pseudotuberculosis IP 31758. Sites that show positive selection (P<0.05) are depicted as yellow spheres and (P<0.01)-as red spheres. doi:10.1371/journal.pone.0020546.g005
Niche. Porins are surface-exposed and their structure strongly reflects the history of multiple interactions with the environmental changes in their ecological niches. The evolution of the ompF gene of *Yersinia* clearly demonstrates a combination of diversifying selection (recombination and positive selection) and function-structure constraint (translational selection and purifying selection). The data can be important for clarification the role of porin’s surface exposed loops on bacterial adaptation and development of broad-spectrum *Yersinia* vaccine antigens and serological methods of diagnostics.

**Materials and Methods**

**Bacterial strains, growth conditions, and DNA isolation**

A total of 65 *Yersinia* strains from the collections of Max von Pettenkofer Institute (Munchen, Germany) and Research institute of epidemiology and microbiology, Siberian branch of Russian academy of medical sciences, (Vladivostok, Russia) were used in this study. Strain selection was intended to include strains of all known *Yersinia* species with a high degree of diversity. All strains were grown overnight at 30°C or 37°C under aerobic conditions on LB medium. Bacterial DNA was isolated from overnight cultures of the selected strains using Genomic DNA Purification Kit (Fermentas, EU). The DNA concentration was determined by agarose gel electrophoresis. The gels were scanned and the signals were analyzed with the VersaDoc 4000 MP system (Bio-Rad Laboratories AG, Switzerland). Additionally, eleven *Yersinia* strains for which the genome sequences are available on the GeneBank of NCBI website were analyzed.

**PCR amplification and DNA sequencing**

PCR amplification of 16S rDNA gene from all strains of *Yersinia* was performed using the primers, BF-20 (5’-ATCACCGCT-TAAAAATCCT-3’) and BR2-22 (5’-CCGGAATATCATGGTG-TGGT-3’). The expected amplicon size was 1500 bp. The part of gyrB gene was amplified using primers YglyF (5’-CCGAGTTTA-TACCT-3’) and YglyR (5’-CCGAGCTGTTATACCT-3’). The expected amplicon size was 980 bp. The ompF gene was amplified using primers Fcds-F (5’-CCGAGTTTATACCT-3’) and Fcds-R (5’-CCGAGCTGTTATACC-3’). These were designed by aligning sequences of ompF genes of *Y. enterocolitica* 0881 (AM286405), *Y. intermedia* ATCC 29909 (AALF02000005), *Y. mollaretii* ATCC 33645 (NZ_AALD02000003) and *Y. frederiksenii* ATCC 33641 (NZ_AALE02000015). The expected amplicon size was 1100 bp. PCR conditions for all genes were as follows: initial denaturation at 95°C for 5 min followed by 30 cycles each at 94°C for 30 s, 55°C for 30 s, 72°C for 2 min and a final extension step at 72°C for 5 min. PCR products were evaluated on a 1% agarose gel stained with ethidium bromide. Unincorporated primers and dNTPs were removed from PCR products with NucleoSpin Extract II kit (Macherey-Nagel). Purified DNA was sequenced using the dideoxynucleotide chain-termination method with fluorescent dNTPs from Applied Biosystems on an ABI 310 Prism automated DNA sequencer, in accordance with the manufacturer’s instructions. Sequence data for the appropriate loci from *Y. hovorevii* ATCC 43970 (NZ_AALL00000000), *Y. enterocolitica* 0881 (NC_008800), *Y. frederiksenii* ATCC 33641 (NZ_AALE00000000) and *Y. intermedia* ATCC 29909 (NZ_AALF00000000), *Y. mollaretii* ATCC 43969 (NZ_AALL00000000), *Y. pestis* 91001 (NC_005810), *Y. pseudotuberculosis* IP32953, *Y. enterocolitica* ATCC 8081, *Y. intermedia* ATCC 29909, *Y. rohdei* ATCC 43380, *Y. kristensenii* ATCC 33638, *Y. frederiksenii* ATCC 33641, *Y. mollaretii* ATCC 43969, *Y. ruckeri* ATCC 29473, *Y. hovorevii* ATCC 43970, *Y. aldovae* ATCC 35296 (NZ_ACCB00000000) were obtained from GenBank (http://ncbi.nlm.nih.gov) and analyzed together with other *Yersinia* isolates (Table 1).

**Comparative sequence analysis and phylogeny inference**

Nucleotide sequence data from forward- and reverse-strand chromatograms were assembled into single contiguous sequences using the Vector NTI Advance 9.1.0 software. Sequences were aligned by ClustalW 2.0.10 [52]. MEGA version 4.1 [53] was used to calculate genetic distances between sequences and to produce phylogenetic trees. To construct the tree from nucleotide sequences, all three coding positions were examined and the Neighbour-Joining model with Kimura 2-parameter method [54] was applied. The reliability of the inferred trees was assessed using the bootstrap test (1000 replications) [55]. Alignment gaps were excluded using function “Pairwise Deletion” from all analyses.

**Evolution analyses**

Nucleotide divergence (Pi) along ompF sequences was determined by DNA SP v5 [56] using Sliding window with length of 20 and step size of 7. Adaptive evolution of *ompF* gene was calculated as proportion of synonymous (silent; ds) and non-synonymous (amino acid-changing; dn) substitution rates in MEGA 4 using the Nei-Gojobori method with Jukes-Cantor correction and SLR [45] software. Recombinant *ompF* sequences were detected with the RDP v3.34 software [57] using four automated recombination detection methods including RDP [58], Genconv [59], Chimaera [60], Maximum Chi Square [60,61]. For the RDP method, internal reference sequences were used, the window size was set to 20, and 0–100 sequence identity was used. For both the MaxChi and the Chimaera methods, the number of variable sites was set to 40. For the GENCONV method, we used standard settings. A maximum P value of 0.01 and a Bonferroni correction were used. Results were then checked by visual inspection. CAI index was calculated by CodonW 1.3 (http://tmolbiol.ox.ac.uk/codonW.tar.Z) software for 11 *Yersinia* species (*Y. pestis* CO92, *Y. pseudotuberculosis* IP32953, *Y. enterocolitica* ATCC 8081, *Y. intermedia* ATCC 29909, *Y. rohdei* ATCC 43380, *Y. kristensenii* ATCC 33638, *Y. frederiksenii* ATCC 33641, *Y. mollaretii* ATCC 43969, *Y. ruckeri* ATCC 29473, *Y. hovorevii* ATCC 43970, *Y. aldovae* ATCC 35296). As a reference for highly expressed genes, we used the 26 concatenated ribosomal genes for each organism.

Nucleotide sequence accession numbers

The novel sequences determined in this study have been deposited in GenBank under accession no. GQ241361-GQ4 21424; FJ641877-FJ641894; 146 HM142614-HM142721.

**Acknowledgments**

We would like to acknowledge Prof. Nellya F. Timchenko (Research institute of epidemiology and microbiology SD RAMS, Russia) who kindly supplied us with *Yersinia* strains (1215, 4648, 5373, 5375, 1948 and 6572) and also other researchers who provided strains to the Max von Pettenkofer Institute. We would like to thank Konstantin Gusev, Evgeniya Bystritskaya and Vladimir Antonenko for the excellent technical assistance. Finally, we would like to thank our reviewers and the editor for the fruitful comments and suggestions on the manuscript.

**Author Contributions**

Conceived and designed the experiments: AMS. MPI. Performed the experiments: AMS. Analyzed the data: AMS MPI. Contributed reagents/materials/analysis tools: AVR VAR FNS. Wrote the paper: AMS MPI AVR.
References

1. Sprague LD, Neubauer H (2005) Yersinia aereoviridis sp. nov. Int J Syst Evol Microbiol 55: 831–835.

2. Merhej V, Adekbaci T, Pageau I, Raoult D, Drancourt M (2008) Yersinia pseudotuberculosis sp. nov., isolated from fresh water. Int J Syst Evol Microbiol 58: 779–784.

3. Sprague LD, Scholz HC, Amann S, Busse HJ, Neubauer H (2008) Yersinia simia sp. nov. Int J Syst Evol Microbiol 58: 932–938.

4. Bottez EJ (1999) Yersinia enterocolitica: overview and epidemiologic correlates. Microbes Infect 1: 323–333.

5. Sahilwade A (2000) Yersinia other than Y. enterocolitica, Y. pseudotuberculosis, and Y. pestis: the ignored species. Microbes Infect 2: 497–513.

6. Loftus CG, Harewood GC, Cockerill FR 3rd., Murray JA (2002) Clinical analyses. FEMS Microbiol Lett 238: 423–428.

7. van Putten JP, Duensing TD, Carlson J (1998) Gonococcal invasion of epithelial cells: Review. J Clin Microbiol 36: 673–681.

8. Nikaido H (1996) Outer membrane. In: Neidhardt FC, et al., ed. Escherichia and Salmonella: Cellular and molecular biology. Washington: ASM Press DC. pp 47–57.

9. Cowan SW, Schirmer T, Rummel G, Steiert M, Ghosh R, et al. (1992) Crystal structure of osmoporin OmpC from E. coli at 2.0 A. J Mol Biol 227: 493–509.

10. Weiss MS, Schulz GE (1992) Structure of porin refined at 1.8 515 A resolution. J Mol Biol 227: 493–509.

11. Bottone EJ (1989) Physiological and genetic responses of bacteria to osmotic stress. Microbiol Rev 53: 121–147.

12. Bodilis J, Hedde M, Orange N, Barray S (2006) OprF polymorphism as a functional activity of genotype. FEMS Microbiol Lett 257–260.

13. Bodilis J, Hedde M, Orange N, Barray S (2006) OprF polymorphism as a functional activity of genus Yersinia. FEMS Microbiol Lett 146: 73–78.

14. Csonka LN (1989) Functional and genetic responses of bacteria to osmotic stress. Microbiol Rev 53: 121–147.

15. van Putten JP, Duensing TD, Carlson J (1998) Gonococcal invasion of epithelial cells: Review. J Clin Microbiol 36: 673–681.

16. Vostrikova OP, Novikova OD, Kim NYu, Likhatskaya GN, Solovjeva TF (2003) Yersinia frederiksenii strain by means of 16S rDNA and Y. enterocolitica species. J Clin Microbiol 41: 2674–2684.

17. Derrick JP, Urwin R, Suker J, Feavers IM, Maiden MCJ (1999) Structural and evolutionary inference from molecular variation in Neisseria. Microbiol Rev 63: 255–65.

18. Hall RM, Doolittle WF, Brown WJ (1995) Sequence evolution of the porA gene of Neisseria gonorrhoeae and Neisseria meningitidis: evidence of positive Darwinian selection. Mol Biol Evol 12: 363–370.

19. Chen SL, Hung CS, Xu J, Reigstad CS, Magrini V, et al. (2006) Identification of genes subject to positive selection in uropathogenic strains of Escherichia coli: a comparative genomics approach. Proc Natl Acad Sci 103: 5977–5982.

20. Petersen L, Bollback JP, Dinmich M, Hubisz M, Nielsen R (2007) Genes under positive selection in Escherichia coli. Genome Res 17: 1336–1343.

21. Sniehotta AM, Orr HV, Rodriguez-Rivera DL, Sun Q, Wiedmann M (2009) Genome wide evolutionary analyses reveal serotype specific patterns of positive selection in selected Salmonella serotypes. BMC Evol Biol 9: 264.

22. Posada D, Crandall KA, Nguyen M, Demuth JC, Visuti RP (2000) Population Genetics of the porA Gene of Neisseria gonorrhoeae. Different Dynamics in Different Homologs. Mol Biol Evol 17: 423–436.

23. Maynard SJ (1992) Analyzing the mosaic structure of genes. J Mol Evol 34: 609–618.

24. Martin DP, Rybicki E (2000) RDP: detection of recombination amongst aligned DNA polymorphism data. Bioinformatics 25: 1451–1452.

25. Martin DP, Williamson C, Posada D (2005) RDP2: recombination detection and phylogenetic analyses. Bioinformatics 21: 332–334.

26. Baldo L, Desjardins CA, Russell JA, Stahlhut JK, Werren JH (2010) Accelerated microevolution in an outer membrane protein (OMP) of the intracellular bacteria Wolbachia. BMC Evol Biol 10: 48.

27. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) ClustalW and ClustalX version 2. Bioinformatics 23: 2947–2948.

28. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596–1599.

29. Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16: 111–120.

30. Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783–791.

31. Martin DP, Williamson C, Posada D (2005) RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21: 260–262.

32. Martin DP, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. Bioinformatics 16: 562–563.

33. Padman M, Sawyer S, Fauquet CM (1999) Possible emergence of new gemiviruses by frequent recombination. Virolology 265: 218–223.

34. Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. Proc Natl Acad Sci 98: 13757–13762.

35. Maynard SJ (1992) Analyzing the mosaic structure of genes. J Mol Evol 34: 126–129.

36. Sprague LD, Neubauer H (2005) Yersinia aereoviridis sp. nov. Int J Syst Evol Microbiol 55: 831–835.

37. Merhej V, Adekbaci T, Pageau I, Raoult D, Drancourt M (2008) Yersinia pseudotuberculosis sp. nov., isolated from fresh water. Int J Syst Evol Microbiol 58: 779–784.

38. Sprague LD, Scholz HC, Amann S, Busse HJ, Neubauer H (2008) Yersinia simia sp. nov. Int J Syst Evol Microbiol 58: 932–938.

39. Bottez EJ (1999) Yersinia enterocolitica: overview and epidemiologic correlates. Microbes Infect 1: 323–333.

40. Sahilwade A (2000) Yersinia other than Y. enterocolitica, Y. pseudotuberculosis, and Y. pestis: the ignored species. Microbes Infect 2: 497–513.

41. Loftus CG, Harewood GC, Cockerill FR 3rd., Murray JA (2002) Clinical analyses. FEMS Microbiol Lett 238: 423–428.

42. van Putten JP, Duensing TD, Carlson J (1998) Gonococcal invasion of epithelial cells: Review. J Clin Microbiol 36: 673–681.

43. Nikaido H (1996) Outer membrane. In: Neidhardt FC, et al., ed. Escherichia and Salmonella: Cellular and molecular biology. Washington: ASM Press DC. pp 47–57.

44. Cowan SW, Schirmer T, Rummel G, Steiert M, Ghosh R, et al. (1992) Crystal structure of osmoporin OmpC from E. coli at 2.0 A. J Mol Biol 227: 493–509.

45. Weiss MS, Schulz GE (1992) Structure of porin refined at 1.8 515 A resolution. J Mol Biol 227: 493–509.

46. Bodilis J, Hedde M, Orange N, Barray S (2006) OprF polymorphism as a functional activity of genus Yersinia. FEMS Microbiol Lett 146: 73–78.

47. van Putten JP, Duensing TD, Carlson J (1998) Gonococcal invasion of epithelial cells: Review. J Clin Microbiol 36: 673–681.

48. Bodilis J, Hedde M, Orange N, Barray S (2006) OprF polymorphism as a functional activity of genus Yersinia. FEMS Microbiol Lett 146: 73–78.

49. Neubauer H, Stojanov A, Andreas H, Ernst-Jugen F, Hermann M (2000) Yersinia enterocolitica OmpA DNA gene types belong to the same genospecies but form three homology groups. Int J Med Microbiol 290: 61–64.

50. Kotetsvili M, Kegret A, Wauters G, Morris JG Jr, Saralidze A, et al. (2005) Multilocus sequence typing for studying genetic relationships among Yersinia species. J Clin Microbiol 43: 2674–2684.

51. Demonstrar A, Respinas SDE, Dolina M, Peduzzi R (2004) Molecular typing of Yersinia frederiksenii strains by means of 16S rDNA and gvd genes sequence analyses. FEMS Microbiol Lett 238: 423–428.