



(RESEARCH ARTICLE)



New tool for phylogenetic analysis of *Helicobacter pylori*

Sorokin Vladimir Mihajlovich ^{1,*}, Pisanov Ruslan Vjacheslavovich ¹, Vodop'janov Aleksej Sergeevich ¹ and Golubkina Elena Vadimovna ²

¹ Rostov-on-Don Research Institute for Plague Control, Rostov-on-Don 344002, Russia.

² Astrakhan State Medical University, Astrakhan 414000, Russia.

Publication history: Received on 26 April 2020; revised on 03 May 2020; accepted on 05 May 2020

Article DOI: <https://doi.org/10.30574/wjarr.2020.6.2.0128>

Abstract

The aim of this study was to detect canonical insertion/deletion (INDEL) markers in the genome of *Helicobacter pylori* and offer INDEL-typing method for differentiation of *H. pylori* strains. For comparative analysis of the genomes of *H. pylori* presented in the GenBank database, a local database of nucleotide sequences of 69 *H. pylori* strains was created. For detecting all INDEL markers with a preset size of 6-20 bp a pairwise comparison of more than 1500 open reading frames (ORF) in the genomes of local database strains was performed. Ten loci containing INDEL markers were founded. The five most variable loci were tested in silico with 21 strains with known geographical origin from the most common populations of hpEurope, hspWAfrica, and hspEAsia. Fifteen individual genotypes with a high diversity index (DI=0.95) were identified. For cluster analysis, the minimal spanning tree (MST) method was used, which demonstrated a clear distribution of clusters according to the geographical origin of the strains tested. INDEL-typing of 21 regional strains from the Astrakhan region was performed in vitro. It was shown that an extensive majority of them belong to the population hpEurope. The findings in this study indicate that the proposed INDEL-typing method almost perfectly reflects the geographical distribution of *H. pylori* strains determined by the multilocus sequence typing (MLST) method, despite the fact that the primary object of research is completely different genes. Further research is needed to determine the geographical origin of *H. pylori* strains in Russia.

Keywords: *Helicobacter pylori*; PCR; INDEL; MST dendrogram; Geographical origin

1. Introduction

The most important factor of *H. pylori* providing the high genetic variability of strains is a significant frequency of endogenous mutations and respective low occurrence of identical alleles [1]. In addition, *H. pylori* cells are naturally competent and can integrate foreign DNA into their own genome through recombination and exchange genetic information between different strains in the same patient [2]. Therefore, *H. pylori* is a species without a significant clonal structure [2, 3]. The real genetic diversity of *H. pylori* was confirmed by various methodological approaches [3-8].

Previously, Achtman et al. [3] established the global structure of the population of *H. pylori* according to their geographical origin using the method of multi-locus sequencing-typing (MLST) of seven life-support genes (*ureI*, *mutY*, *efp*, *ppa*, *yphC*, *atpA*, *trpC*). Initially, they described two groups of strains with a weak clonal structure: Asian Clone (Asian strains) and Clone2 (strains from other parts of the world). Later, Falush et al. [9] identified four groups of strains with different geographical origins: hpEurope, hpAfrica 1 (later divided into hspWAfrica and hspSAfrica), hpAfrica 2, and hpEastAsia (consisting of hspAmerind, hspEAsia, and hspMaori) [9]. Currently, the MLST-typing method is the most common method for determining the geographical origin of *H. pylori* strains in the world, but it has some disadvantages. The method requires determining the DNA sequence size of 3850 nucleotides with further determination of SNP substitutions in the DNA sequence size of more than 1400 nucleotides for each studied strain, which necessitates the use of high-tech equipment and complex software. Perhaps, this determines the fact that the structure of the population

* Corresponding author: Sorokin Vladimir Mihajlovich Email: soroka53@mail.ru

of *H. pylori* strains circulating in Russia is practically not studied today. At the same time, the INDEL-typing method [10-13], based on the available PCR method, is widely used to study the phylogenetic relationships between strains of some other microorganisms, in addition to the MLST typing method. To date, the method of INDEL-typing of *H. pylori* strains was not been described in the world literature.

Therefore, the objective of this study is to detect INDEL markers in the genome of *Helicobacter pylori* and develop INDEL-typing method for differentiation of *H. pylori* strains.

2. Material and methods

2.1. Methods of bioinformation analysis

Original software was used to create a local database of *H. pylori* genomes. Genomes represented as reads were assembled using the Spades program [14]. Comparative analysis of open reading frames (ORF) was performed using the original program Gene Expert, written in the Java script. The variability of detected insertions/deletions (INDEL) was estimated using the Simpson method (diversity index, DI), calculated using the formula [15]. Primers were designed and PCR was performed *in silico* using the author's programs Primer M and Virtual PCR. Cluster analysis and dendrogram construction were performed using the UPGMA method. The MEGA 5 program was used to construct the dendrogram [16]. Optimization of the set of INDEL loci was performed using the program AuSeTTS (Automated Selection of Typing Target Subsets) [17]. The phylogenetic tree was constructed using the MST (minimal spanning tree) method (Bionumerix 7.6 software package).

2.2. Bacterial strains and DNA isolation

Twenty- one *H. pylori* isolates were obtained from collection of Astrachan state medical university. Genomic DNA was extracted from bacterial isolates using a Probe NA Kit (DNA-Technology, Russia), according to the manufacturer's instructions.

2.3. PCR amplification

Amplification was carried out as described earlier [7], with one exception: annealing at 50 °C for locus hp3660 and 55 °C for other loci. Each PCR product was resolved by 8% polyacrylamide gel electrophoresis and allelic sizes were estimated using a pBlueScript DNA / MspI (MBI Fermentas, Vilnius, Lithuania) as a size marker. Gels were visualised using UV transillumination and the images captured using the ChemiDoc XRS System (BioRad).

3. Results and discussion

3.1. *In silico* INDEL-typing

For comparative analysis of the genomes of *H. pylori* presented in the GenBank database, a local database of nucleotide sequences of 69 *H. pylori* strains was created using the author's software. Using the Gene Expert program, a pairwise comparison of more than 1,500 open reading frames in the genomes of local database strains was performed to detect all INDEL markers with a preset size of 6-20 bp and 10 loci containing INDEL markers were detected.

The variability of each individual locus was determined for all 10 loci containing INDEL markers. Six most variable loci were selected for further research (hp3330, hp5605, hp6405, hp340, hp1390, hp3660). To optimize a set of 6 INDEL loci the program AuSeTTS (Automated Selection of Typing Target Subsets) [17] was applied to obtain the maximum number of individual genotypes.

It was shown that the exclusion of the hp3330 locus does not affect the resolution of the method and allows detecting fifteen individual genotypes with a high diversity index (DI=0.95) in 21 *H. pylori* strains. For PCR *in silico* primers were constructed using the original program Primer M (Table 1).

Table 1 Oligonucleotides used for PCR amplifications included in the INDEL typing scheme for *H. pylori* and expected product sizes for each reaction.

INDEL locus	forward	reverse	product sizes (bp)
hp3330	AATGTTTGCATGGGCAAAG	GAATAGCGTTTAGAGGGCGTTA	60/69
hp5605	AGGGAAGTGAGCGAAGAAGA	ACAGCCCCACCTTTCTTT	93/102
hp6405	ATTATTTCCGTTGGCTCGTG	GCATAAATGGGCATGGTAGC	100/106
hp340	CCCCAAAAC TCCAAAGACA	AACACCGCATCATTGACTTG	79/85
hp1390	CGCTCCTATCTCATGCCCTA	GATCTCGAGCTTAAGCCCTCT	64/76
hp3660	TGTCAATAAGATTAAGGGCGAAG	TGGATGTCAAAGCCAATCAA	96/108

Recently, the most accurate method of constructing a minimal spanning tree (MST) has been increasingly used to clarify the phylogenetic relationships between microbial strains of different species [18-20]. The five INDEL loci (hp5605, hp6405, hp340, hp1390, hp3660) were used to type *in silico* 21 strains of *H. pylori* from the GenBank database. The results of the MST analysis are shown in Figure 1.

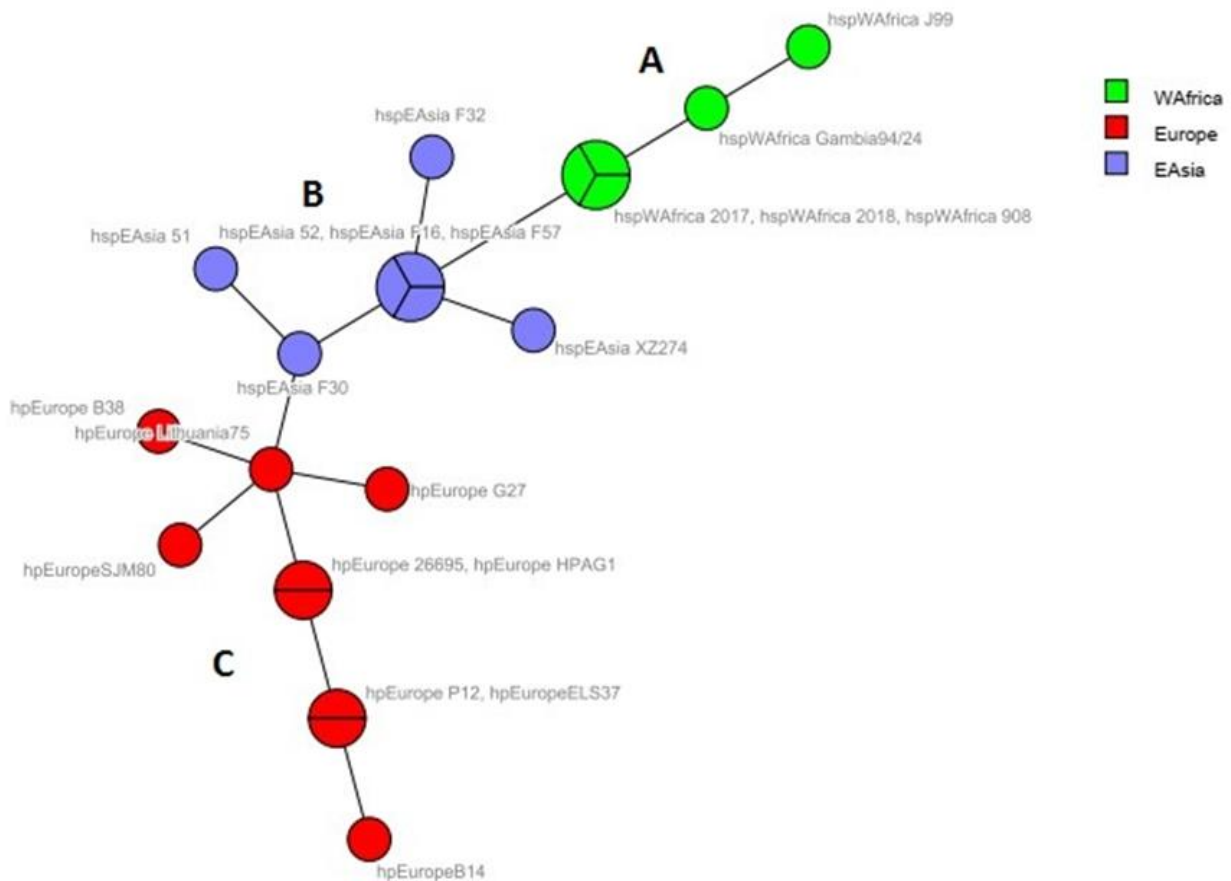


Figure 1 MST phylogenetic tree of 21 *H. pylori* strains from the GenBank database

Cluster A is represented exclusively by strains belonging to the hspWAfrica population. Cluster B is represented only by strains belonging to the hspEAsia population, and cluster C contains European strains (hpEurope). MST analysis *in silico*

demonstrated an accurate division of strains into clusters according to their geographical origin. It should be noted that there is a significant variability of individual genotypes within each of the populations: 3 individual genotypes in hspWAfrica, 5 genotypes in hspEAsia, and 7 genotypes in hpEurope. The high degree of genetic diversity in hpEurope is due to the mixing of two ancestral populations, AE1 and AE2, which may have reached Europe at different times and from different sources: AE1 was indeed present throughout Central Asia in addition to Europe, and AE2 came from North Africa and Central Asia [21]. In a recent study by Vale et al. [22] could to differentiate two European populations represented mainly in Northern Europe (with genetic affinity to AE1) and southern Europe (with genetic affinity to AE2). The study was based on typing the nucleotide sequence of two *H. pylori* prophages. The distribution of INDEL genotypes in the hpEurope strain cluster also indicates the existence of two distinct groups of strains (Fig.1).

3.2. INDEL-typing of clinical isolates

At the stage of constructing primers, we faced the problem of extreme polymorphism of the *H. pylori* genome and, consequently, the very high variability of flanking regions of INDEL markers. However, the problem was solved, and the vast majority of primers were specific also *in vitro* (Fig.2). For each locus, different alleles are clearly separated using electrophoresis in 8% polyacrylamide gel and can be easily identified (Fig. 2).

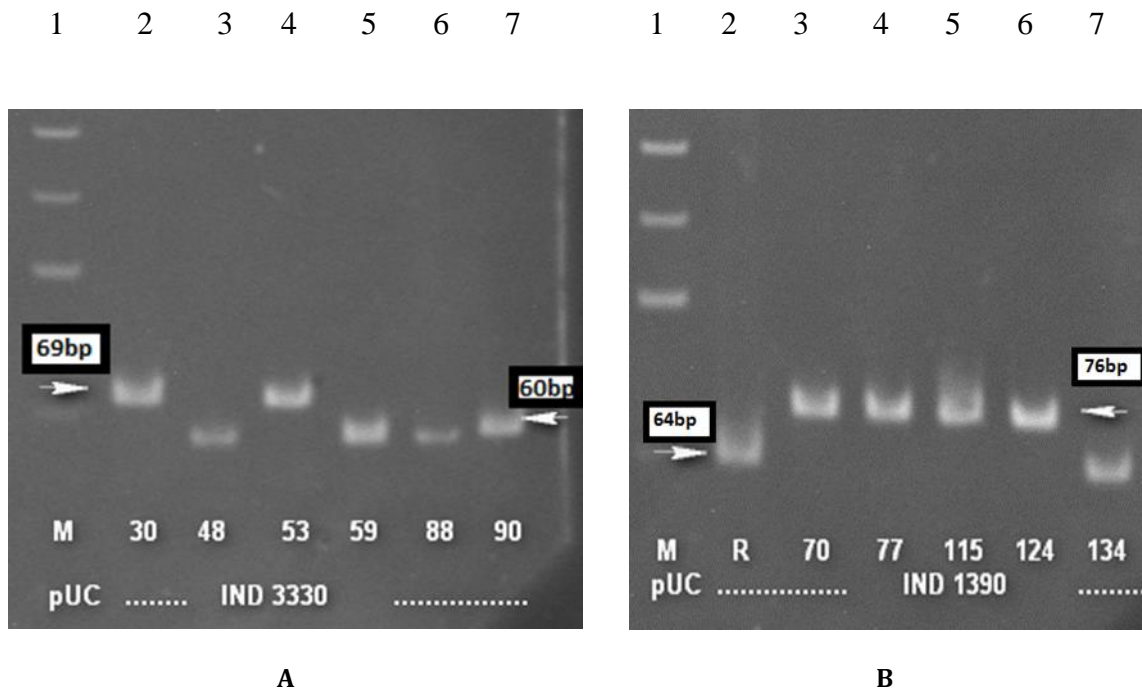


Figure 2 Electrophoregram of amplification products by locus hp3330 (A) and locus hp1390 (B): A - track 1-marker of molecular masses; tracks 2-7 (DNA of clinical isolates of *H. pylori*); B - track 1-marker of molecular masses; tracks 2-7 (DNA of clinical isolates of *H. pylori*). Arrows indicate different alleles of each locus.

INDEL-typing of DNA of 21 *H. pylori* strains isolated in the Astrakhan region was performed (Table 2). According to the AuSeTTS program [17], 21 strains are represented by 16 individual INDEL genotypes, which indicates a significant heterogeneity of the strains studied.

Table 2 INDEL-genotypes of Astrakhan region strains

strain	hp5605	hp6405	hp340	hp1390	hp3660
AO 1.03.13	0	100	85	76	0
AO 4.03.13	102	100	79	76	0
AO 6.03.13	0	100	79	76	96
AO 11.03.13	102	100	79	76	0
AO 12.03.13	0	100	79	76	108
AO 16.03.13	0	100	85	76	108
AO 6.10.13	102	106	85	0	0
AO 8.10.13	102	100	79	76	96
AO 1.04.15	93	100	79	76	0
AO 2.04.15	102	100	85	76	0
AO 3.04.15	102	100	79	76	108
AO 2.01.17	102	100	79	0	96
AO 3.01.17	102	100	85	76	0
AO 4.01.17	102	106	79	76	96
AO 1.06.17	0	106	79	76	96
AO 8.06.17	0	100	79	76	0
AO 3.01.18	102	100	79	64	0
A04.01.18	102	100	79	64	0
A05.01.18	0	100	79	76	96
A06.01.18	102	106	79	76	96
A08.01.18	102	100	79	64	96

3.3. Building and analyzing a phylogenetic tree

To clarify the structure of phylogenetic relationships between clinical isolates and strains from the GenBank database with a known geographical origin, the MST procedure was applied, the result of which is shown in Figure 3. It should be noted, that there are some changes in the structure of phylogenetic relationships compared to the original one (Figure 1). Cluster A, represented by hspWafrika strains was supplemented by one Astrakhan strain. Cluster B, represented by strains of the hspEAsia population, has undergone some changes, namely, one strain was "displaced" to cluster C. The remaining 20 strains of the Astrakhan population belong to cluster C, which contains hpEurope strains. The same cluster contains hpEurope population strains from the GenBank database. It should be noted, that the strains of the Astrakhan population are represented by both INDEL genotypes shared with GenBank strains (4) and unique individual ones (11), which emphasizes the high genetic heterogeneity of hpEurope population strains. These additional genotypes are represented by three phylogenetic lineages that extend the genetic diversity of European strains. In total, the studied 42 strains are represented by 27 individual genotypes with a high diversity index ($DI=0.95$).

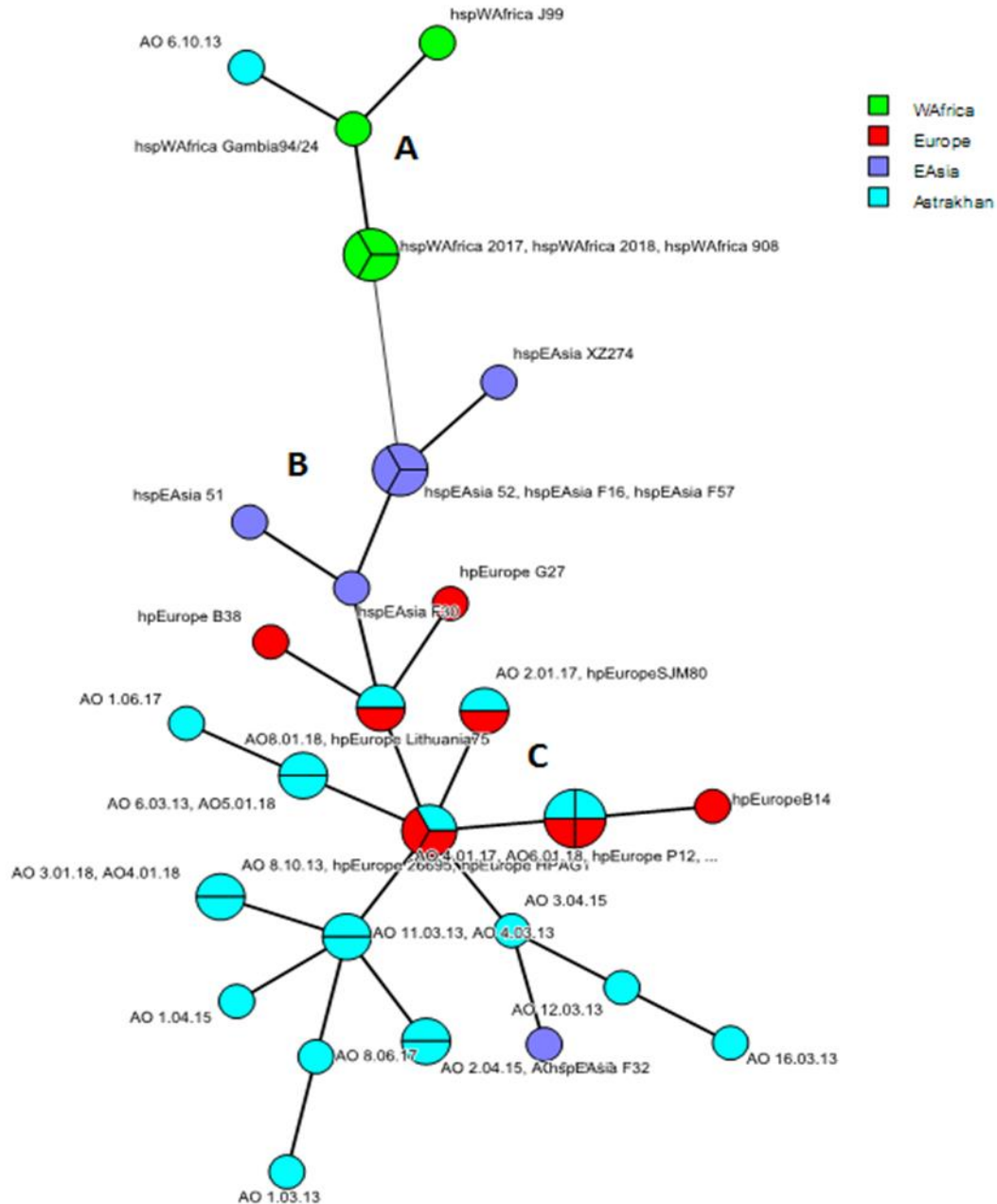


Figure 3 Common MST phylogenetic tree of 21 *H. pylori* strains from the GenBank database and 21 Astrakhan strains

Thus, the proposed INDEL-typing method almost perfectly reflects the geographical distribution of *H. pylori* strains determined by the MLST method, despite the fact that the primary object of research is completely different genes.

4. Conclusion

The INDEL-typing method for differentiation of *H. pylori* strains based on the available PCR method is proposed for the first time. This method includes the determination of alleles of five INDEL loci by PCR, followed by the study of their phylogenetic relationship by MST. Twenty-one *H. pylori* strains from a database GenBank with a known geographical origin were typed. The results obtained shows very good similarity with MLST typing data. The method was successfully tested *in vitro* in the study of regional Russian strains.

Compliance with ethical standards

Acknowledgments

Work performed in the framework of the industry program Rospotrebnadzor.

Disclosure of conflict of interest

No disclosure of any conflict of interest.

References

- [1] Wang G, Humayun MZ and Taylor DE. (1999). Mutation as an origin of genetic variability in *Helicobacter pylori*. Trends Microbiol, 7, 488-93.
- [2] Suerbaum S, Smith JM, Bapumia K, Morelli G, Smith NH, Kunstmann E, Dyrek I and Achtman M. (1998). Free recombination within *Helicobacter pylori*. Proc Natl Acad Sci USA, 95, 12619-24.
- [3] Achtman M, Azuma T, Berg DE, Ito Y, Morelli G, Pan ZJ, Suerbaum S, Thompson SA, van der Ende A and van Doorn LJ. (1999). Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. Mol Microbiol, 32, 459-70.
- [4] Guo C, Liao Y, Li Y, Duan J, Guo Y, Wu Y and Cui Y. (2011). Genotyping analysis of *Helicobacter pylori* using multiplelocus variable-number tandem-repeats analysis in five regions of China and Japan. BMC Microbiol, 11, 197.
- [5] Kang J and Blaser MJ. (2006). Bacterial populations as perfect gases: genomic integrity and diversification tensions in *Helicobacter pylori*. Nat Rev Microbiol, 4, 826-36.
- [6] Han SR, Zschausch HC, Meyer HG, Schneider T, Loos M, Bhakdi S and Maeurer MJ. (2000). *Helicobacter pylori* : clonal population structure and restricted transmission within families revealed by molecular typing. J Clin Microbiol, 38, 3646-51.
- [7] Sorokin V, Pisanov R, Golubkina E, Berezhnyak E and Prozorova L. (2017). Comparative Multiple-Locus Variable-Number Tandem Repeat Analysis of *Helicobacter pylori* Isolates from South of Russia. International Journal of Microbiology and Biotechnology. 2(3), 135-138.
- [8] Sorokin VM, Pisanov RV and Vodop'janov AS. Improvement of Multiple-Locus VNTR Analysis Typing Scheme for *Helicobacter pylori*. (2018). Asian Journal of Biochemistry, Genetics and Molecular Biology, 1(4), 1-7.
- [9] Falush D, Wirth T, Linz B, Pritchard JK and Stephens M. (2003). Traces of human migrations in *Helicobacter pylori* populations. Science, 299, 1582-5.
- [10] Liu F, Hu Y and Wang Qi. (2014). Comparative genomic analysis of Mycobacterium tuberculosis clinical isolates. BMC Genomics, 15, 469.
- [11] Larsson P, Svensson K and Karlsson L. (2007). Canonical Insertion-Deletion markers for rapid DNA typing of Francisella tularensis. Emerging Infectious Diseases, 13(11), 1725-1732.
- [12] Li K, Gu W and Liang J. (2014). Gene polymorphism analysis of Yersinia enterocolitica outer membrane protein A and putative outer membrane protein A family protein. BMC Genomics, 15(201), 1471-2164.
- [13] Vodop'janov AS, Vodop'janov SO and Olejnikov IP. (2017). INDEL-typing of Vibrio cholerae strains. Epidemiologiya i infekcionnye bolezni, 22(4), 95-100.
- [14] Bankevich A, Nurk S, Antipov D, Gurevich A and Dvorkin M. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology, 19(5), 455-477.
- [15] Simpson EH. (1949). Measurement of diversity. Nature (London), 163(4148), 668-688.
- [16] K Tamura D, Peterson N, Peterson G, Stecher and M Nei. (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution, 28, 2731-2739.
- [17] O'Sullivan MV, Sintchenko V and Gilbert GL. (2013). Software for selecting the most informative sets of genomic loci for multi-target microbial typing. BMC bioinformatics, 14(1), 148-158.

- [18] Wilkinson DA, O'Donnell AJ, Akhter, RN Fayaz and Mack HJ. (2018). Updating the genomic taxonomy and epidemiology of *Campylobacter hyointestinalis*. *Scientific reports*, 8(1), 1-12.
- [19] Blouin Y, Hauck Y, Soler C, Fabre M and Vong R. (2012). Significance of the Identification in the Horn of Africa of an Exceptionally Deep Branching *Mycobacterium tuberculosis* Clade. *PLoS one*, 7(12).
- [20] Boujemaa S, Allaya AB, Mlik B, Mardassi H and Mardassi BB. (2018). Phylogenetics of *Mycoplasma hominis* clinical strains associated with gynecological infections or infertility as disclosed by an expanded multilocus sequence typing scheme. *Scientific reports*, 8(1), 1-10.
- [21] Megraud F, Lehours P and Vale FF. (2016). The history of *Helicobacter pylori*: from phylogeography to paleomicrobiology. *Clinical Microbiology and Infection*, 22(11), 922-927.
- [22] Vale FF, Vadivelu J, Oleastro M, Breurec S and Engstrand L. (2015). Dormant phages of *Helicobacter pylori* reveal distinct populations in Europe. *Sci Reports*, 5, 14333.

How to cite this article

Sorokin VM, Pisanov RV, Vodop'janov AS and Golubkina EV. (2020). New tool for phylogenetic analysis of *Helicobacter pylori*. *World Journal of Advanced Research and Reviews*, 6(2), 60-67.
