



Peeping into genomic architecture by re-sequencing of *Ochrobactrum intermedium* M86 strain during laboratory adapted conditions



Kushal N. Gohil^{a,b}, Priya S. Neurgaonkar^a, Aditi Paranjpe^a, Syed G. Dastager^{a,b}, Mahesh S. Dharne^{a,b,*}

^a National Collection of Industrial Microorganisms (NCIM), CSIR-National Chemical Laboratory (NCL), Pune, India

^b Academy of Scientific and Innovative Research (AcSIR), New Delhi, India

ARTICLE INFO

Article history:

Received 31 December 2015

Received in revised form 21 March 2016

Accepted 11 April 2016

Available online 19 April 2016

Keywords:

Genome

Re-sequencing

Ochrobactrum intermedium

Laboratory adapted strain

ABSTRACT

Advances in *de novo* sequencing technologies allow us to track deeper insights into microbial genomes for restructuring events during the course of their evolution inside and outside the host. Bacterial species belonging to *Ochrobactrum* genus are being reported as emerging, and opportunistic pathogens in this technology driven era probably due to insertion and deletion of genes. The *Ochrobactrum intermedium* M86 was isolated in 2005 from a case of non-ulcer dyspeptic human stomach followed by its first draft genome sequence in 2009. Here we report re-sequencing of *O. intermedium* M86 laboratory adapted strain in terms of gain and loss of genes. We also attempted for finer scale genome sequence with 10 times more genome coverage than earlier one followed by comparative evaluation on Ion PGM and Illumina MiSeq. Despite their similarities at genomic level, lab-adapted strain mainly lacked genes encoding for transposase protein, insertion elements family, phage tail-proteins that were not detected in original strain on both chromosomes. Interestingly, a 5 kb indel was detected in chromosome 2 that was absent in original strain mapped with phage integrase gene of *Rhizobium* spp. and may be acquired and integrated through horizontal gene transfer indicating the gene loss and gene gain phenomenon in this genus. Majority of indel fragments did not match with known genes indicating more bioinformatic dissection of this fragment. Additionally we report genes related to antibiotic resistance, heavy metal tolerance in earlier and re-sequenced strain. Though SNPs detected, there did not span urease and flagellar genes. We also conclude that third generation sequencing technologies might be useful for understanding genomic architecture and re-arrangement of genes in the genome due to their ability of larger coverage that can be used to trace evolutionary aspects in microbial system.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Human gastrointestinal (GI) tract is compartmentalized into different sections, which allows digestion and absorption of nutrients in the body. The distinct part of the gut has different physiological conditions that shape the gut microbiome (or microbial diversity within the gut). The human intestinal microbiota is composed of 10^{13} to 10^{14} microorganisms whose collective genome ("microbiome") contains at least 100 times as many genes as our own genome [1]. As in most mammals, the gut microbiome is dominated by four bacterial phyla namely *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, and *Proteobacteria* [2], representing more than 1000 different molecular species or phylotypes. Stomach of the GI tract has very low pH ~ 2.5, urea stress and microaerophilic environment making it extreme niche. To adopt up with this condition organism have to have some additional machinery to survive and perform its functions. There has been an estimate

that human stomach comprises 32 phylotypes from *Proteobacteria*, including 3 uncharacterized phylotypes [3]. *Helicobacter pylori* (*H. pylori*), known causative agent of peptic ulcers and gastric cancer, colonize the stomach of approximately 50% of the world's population [4]. However now there are few recent reports indicating presence of bacteria other than *H. pylori* in the human stomach. Our group had isolated a bacterium from Non ulcer dyspeptic individual from north India in 2005 and was identified as *Ochrobactrum intermedium* by 16S rRNA sequencing and its first draft genome sequencing was reported by using Ion torrent personal genome machine (PGM) [4,5]. To gain insights into genomic re-architecturing in terms of gain and loss during continuous passaging for several generations and also to evaluate sequencing platform for getting more refined data (in terms of coverage), we re-sequenced the *Ochrobactrum intermedium* M86 strain with Illumina MiSeq and compared with 10 year earlier draft genome of the same strain.

2. Results and discussion

Genome sequencing features and subsystem distribution among these organisms were also compared. The 16S rRNA based phylogenetic

* Corresponding author at: National Collection of Industrial Microorganisms (NCIM), CSIR-National Chemical Laboratory (NCL), Pune, Maharashtra, India.
E-mail address: ms.dharne@ncl.res.in (M.S. Dharne).

analysis showed *Ochrobactrum* and *Brucella* are closely related than *Helicobacter*. Urease gene (*UreC*) based tree clearly indicated the presence of a distinct class of urease in *Ochrobactrum* when compared to *Helicobacter* and *Brucella* (data not shown). Whole-genome re-sequencing of *O. intermedium* M86 was conducted with the Illumina MiSeq. We obtained 5,384,581 paired-end reads of 300 bp using Illumina MiSeq which was 4.1 fold higher than the reads obtained from Ion torrent PGM machine [5]. Sequence yield per run is higher in Illumina MiSeq than Ion torrent PGM machine and reports also showed the error rate is higher in PGM 1.71% than Illumina MiSeq, 0.4%. The number of error-free reads, without a single mismatch or indel, was 76.45%, 15.92% for, MiSeq, Ion Torrent respectively [6]. This comparison revealed efficiency in terms of quantity and quality MiSeq is somewhat better than PGM in terms of output data. The genome of passaged/re-sequenced/laboratory adapted M86 comprised of two circular chromosomes with a total size of 4,683,452 bases with a 57% G + C content and no plasmids. Of the 4415 (CDSs) genes predicted, 38 tRNA and 5 rRNA genes were found (Table 1). The majority of genes were assigned a putative function. The genes were distributed into 470 subsystems using RAST.

To check the quality of assemblies obtained from velvet with different k-mer values; Genome coverage, N50 values and Maximum/median/average contig sizes were compared. We implemented velvet to optimize k-mer value to achieve maximum coverage. It was observed that at k-mer = 25, estimated coverage was 53%, 2,642,663 bp nucleotides were assembled and N50 value was decreased as shown in Fig. 2(A), (D), and (E). In case of k-mer = 57, N50 value was maximum but the coverage attained was 79%, moreover number of assembled nucleotides was less as compared to k-mer = 31 as mentioned in Fig. 2(C), (D) and (E). At k-mer = 31, all the parameters were optimized and maximum coverage was obtained (Fig. 2(B), (D), and (E)). This k-mer value was used for further analysis. The *De novo* assembly using the Velvet and SPAdes method generated 121 contigs, which were further assembled into 34 scaffolds. Sequencing statistics of earlier genome (did on Ion PGM) and resequenced strain (on Illumina) was also compared. Resequenced genome on Illumina was better on Ion PGM on all technical aspects like number of reads (5 times), genome coverage (10 times), N50 value (1.25 × higher), number of scaffolds (equal) except run time (4 times more). This study was also aimed to enhance the quality of reads by re-sequencing for futuristic view of functional genomics. In our study, Illumina MiSeq and PGM were performed to sequence genome of *O. intermedium* M86 and we could conclude that the Illumina probably had higher coverage as compared to PGM with respect to total reads obtained and coverage, which could also be useful for larger genomes.

For chromosome 1 and 2, the IS finder BLAST results showed unique insertion sequences/tandem repeats which were not reported in earlier draft genome but appeared in current analysis. Overall 107 and 103 such repeats were recorded in earlier and resequenced genomes. Unique insertion sequences that we detected only in re sequenced strain were IS2, IS51, IS151, IS407, ISAS1, IS30, and IS1595. The RAST annotation system showed absence of prophage encoding genes

(confirmed by PHAST) in earlier strain, but were present in re sequenced strain, although they were not complete indicative of a possibility of gene exchange via transduction mechanism. Reference based mapping of *O. intermedium* M86 draft genome and resequenced genome with *O. intermedium* LMG3301^T revealed that resequenced genome appeared to have undergone insertion as well as deletion of some genes or islands (Fig. 1). Further investigation of the deleted portion belonged to phage proteins that were excised probably during passaging. No CRISPR loci were detected in the genome indicative of absence of cas system. Anti-SMASH analysis revealed interesting findings with respect to absence of Trps-T3pks system in the resequenced strain, with replacement of Terpene and Arylpolyene. Such differences in biosynthetic gene cluster could assist the strains in developing and combating strategies. As was previously reported the strains of *O. intermedium* were resistant to all β-lactam antibiotics except imipenem [5]. This was confirmed in resequenced genome by RAST annotation server and it showed the presence of β-lactamase. In addition, Metal-dependent hydrolases of the beta-lactamase superfamily I and Beta-lactamase class C, other penicillin binding proteins, fluoroquinolones and Streptothricin resistant genes were also found in resequenced as well as earlier genome but not reported so far. These results indicate the essentiality of these genes during exposure to various antimicrobial agents.

Diverse *Ochrobactrum* sps. have been reported from contaminated habitat which showed its role in bioremediation and tolerance to different heavy metals like Arsenic, Nicotine [7,8] and were found in earlier as well as resequenced strain. In addition, presence of genes coding for Arsenic (*ArsH*, *Acr3*, *Arsenate reductase*) Copper (*CutE*, *CcmH*, *CopC*, *CopD*, *Copper chaperone*, *CopG*, *CcmM*, *Copper-translocating P-type ATPase*, *Blue copper oxidase CueO precursor*), Cobalt-zinc-cadmium resistance (*Co/Zn/Cd efflux system membrane fusion protein*) and Mercuric reductase were also detected in earlier and resequenced strain indicating the necessity of these genes for growth under in-vitro as well as in-vivo environment. Strain differences under selective or non-selective pressure may lead to formation of SNPs leading to inter and intra-strain diversity. Around 8878 SNPs were detected among previous genome and resequenced *O. intermedium* M86 (both chromosomes), wherein no SNPs detected in *rpoB*, *trpE*, *recA*, *dnaK*, *aroC* and *gap*. Genes responsible for survival in acidic condition of stomach (*ureC*, *ureD*, *ureE*, *ureF* and *ureG* of urease operon) and flagellar assembly (*FliC*) were devoid of SNPs indicating their stability under in-vitro conditions outside simulated niche.

In Chromosome I, we detected 2 gaps from which one fragment is absent in both the genomes, previous genome and resequenced genome. This fragment mainly includes genes of transposase protein and IS66 and IS911 family protein. These proteins either stimulate or inhibit the jumping of insertion sequences. And the other gap which is specific to resequenced genome includes phage related protein like phage tail protein, assembly protein. These mainly affect the assembly of phage. Interestingly in chromosome 2, a 5 kb fragment which was absent in previous genome and also in *O. intermedium* LMG 3301^T strain was detected in a re-sequenced strain. Further, tblastx analysis of this fragment showed the presence of integrase gene from *Rhizobium* spp. We speculate that the insertion might be result of a horizontal gene transfer event in *O. intermedium* M86. Integrase gene product mainly helps in the site specific recombination through *attB* and *attP* site [9]. Construction of vector containing attachment site and integrase gene can be used as engineering tool for insertion of desired fragment of gene through site specific recombination. We found presence of tRNA gene upstream to integrase gene. It is just because crossover of nucleic acid fragment and host integration site can occur at any 3 sublocation within a tRNA can occur, two with flanking symmetry (anticodon-loop and T-loop tDNA) and the third at the asymmetric 3' end of the gene [10]. Apart from 5 kb indel, genes absent only in resequenced genome were phage encoded protein like phage tail coding protein, DNA packaging protein, HK97 family phage genes which mainly play an important

Table 1
General features of *O. intermedium* M86 re-sequenced draft genome.

Attributes	Values
Genome size (bp)	4,683,452
Total number of scaffolds	34
ORFs	4415
G + C content	57
tRNA	38
rRNA	5
Data accessibility	This Whole Genome Shotgun project has been deposited at GenBank under the accession LPQX00000000 and Bioproject number PRJNA302728.

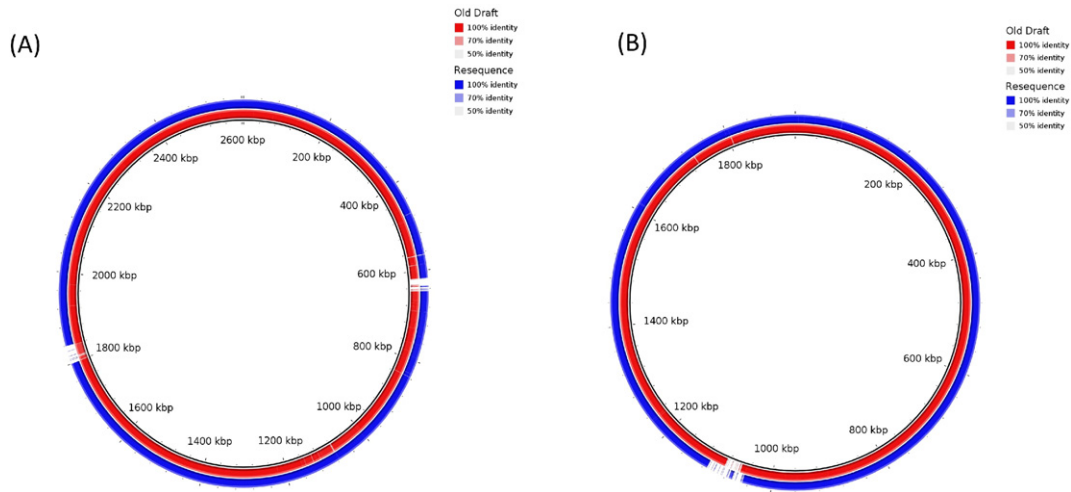


Fig. 1. Reference based mapping indicating gaps in passaged *O. intermedium* M86 strain chromosome I (A) and chromosome II (B) using unpassaged M86 strain and *O. intermedium* LMG 3301^T.

role in connecting the DNA filled head to tail via connector [11]. Another gene Holotricin-3 precursor which is a defense response gene [12] and Chloramphenicol acetyltransferase which provides resistant to chloramphenicol antibiotic [13] were also absent. The stability and effect of indel events on the overall physiology, regulatory and colonization mechanisms in the passaged strain needs to be investigated for acid resistance, colonization, and persistence in hostile environment of acid and urea present in stomach.

Considering the opportunities and challenges of niche specialized *O. intermedium* strain (among genus *Ochrobactrum*) that has a tendency of genome reduction in the technology driven era. Our study is significant for this isolate of gastric origin in terms of genome architecture and genome rearrangement in terms of deletion as well as insertional events during routine maintenance under laboratory conditions. Given the understanding of having such property of recombination and specialized process of adaptation in diverse environment, the more

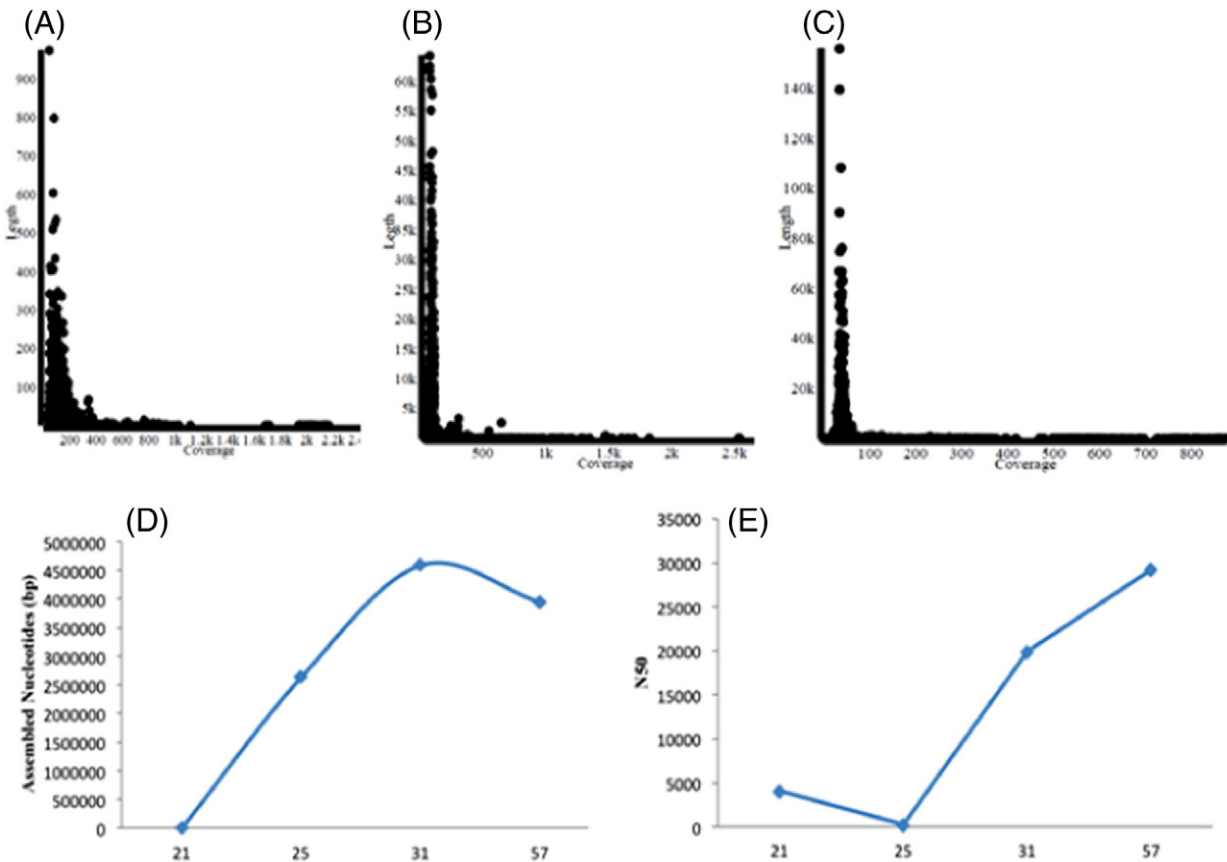


Fig. 2. Selection of k-mer based on coverage (A, B, and C) assembled nucleotides (D) and N50 (E).

understanding of *Ochrobactrum* genus at genomic level could be important to understand if specialized set of genes are assigned for environmental and clinical adaptations.

3. Materials and methods

3.1. Culture conditions and phylogenetic analysis

Strain *O. intermedium* M86 was weekly passaged for about 100 generations for this study and used as laboratory adapted strain in this study. Then grown in MacConkey broth (HiMedia, India) and after overnight incubation at 37 °C purity of the culture was checked and pellet was collected by centrifugation. Chromosomal DNA isolation was done by using Pure-link DNA isolation kit (Invitrogen, Carlsbad, USA) and quality of the DNA was checked by recording absorbance at A260/280 nm and agarose gel electrophoresis. Phylogenetic analysis of 16S rRNA genes for *O. intermedium* M86, *Brucella melitensis biovar Suis*, *Helicobacter pylori* ATCC 43504 and *O. intermedium* LMG 3301^T was done using MEGA v 6.0 with UPGMA method (Bootstrap = 1000) as described [4]. Similarly Urease alpha subunit (*Ure C*) of genera *Ochrobactrum*, *Helicobacter* and *Brucella* was used to build UPGMA tree using MEGA.

3.2. Genome sequencing and assembly

Genome re-sequencing of M86 strain was done using Illumina MiSeq M02845 sequencer was carried out using 2 × 300 bp chemistry. To check the quality of paired end reads obtained from Illumina MiSeq platform FastQC tool was used. Poor quality reads (Phred quality below Q30) and adapter sequences were removed using FASTQ Trimmer [14] and Flexbar [15] pre-processing tool respectively. To preserve mate integrity and to filter sequences by composition Paired-end compositional filtering tool was used [16]. *De novo* assembly of preprocessed FASTQ reads was performed using Velvet 1.2.10 (k-mer = 25, 31, 57) [17] and SPAdes 3.0.0 (k-mer 21, 33, 55) [18]. Contigs generated from Velvet and SPAdes were integrated with CISA contig integrator toolkit [16,19]. Further the quality of integrated contigs/scaffolds was evaluated using Check bacterial contigs tool [16]. Whole pipeline of assembly was performed in NGS Galaxy Oriome CRS4 [16]. The contigs produced by Galaxy Oriome CRS4 were merged into scaffolds using SSPACE tool [20]. Contigs of *O. intermedium* M86 draft genome and resequenced genome were mapped against chromosome I and II of *O. intermedium* LMG 3301^T and visualized using BRIG (Fig. 1A and B) [21].

3.3. Genome annotation

To predict gene coding regions (CDS) Prokka [22] and Rapid Annotation Subsystem Technology (RAST) [23] were used. Metabolic pathway prediction was performed on KAAS to assign KEGG Orthology (KO) numbers to each predicted CDS [24]. Comparative analysis was carried out between previous draft genome and resequenced genome for different features such as presence of insertion sequences, Tandem repeats, phage genes, clustered regularly interspaced short palindromic repeats (CRISPR), Plasmid and secondary metabolite production genes using IS Finder [25], Tandem Repeats Finder [26], PHAST [27], CRISPRFinder [28], BLAST and Anti-SMASH 2.0 [29] respectively. *O. intermedium* M86 draft genome and resequenced genome were compared with *O. intermedium* LMG 3301^T for single nucleotide polymorphism (SNPs) detection using parsnp pipeline [30].

Competing interest

The authors declare that they have no competing interests.

Acknowledgements

KG is thankful to University Grant Commission (UGC), and Academy of Scientific and Innovative Research (AcSIR), New Delhi, India for providing fellowship and support. This work was funded by Science and Engineering Research Board (SERB), New Delhi, India (SB/YS/LS-347/2013) and NCL start up grant (MLP027326). Authors acknowledge Genotypic Technologies, Bengaluru, India for generation of sequencing data..

References

- [1] Steven R. Gill, Mihai Pop, Robert T. DeBoy, Paul B. Eckburg, Peter J. Turnbaugh, Buck S. Samuel, Jeffrey I. Gordon, David A. Relman, Claire M. Fraser-Liggett, Karen E. Nelson, Metagenomic analysis of the human distal gut microbiome. *Science* 312 (5778) (2006) 1355–1359.
- [2] R.E. Ley, M. Hamady, C. Lozupone, P.J. Turnbaugh, R.R. Ramey, J.S. Bircher, M.L. Schlegel, T.A. Tucker, M.D. Schrenzel, R. Knight, et al., Evolution of mammals and their gut microbes. *Science* 320 (2008) 1647–1651.
- [3] E.M. Bik, P.B. Eckburg, S.R. Gill, K.E. Nelson, E.A. Purdom, F. Francois, et al., Molecular analysis of the bacterial microbiota in the human stomach. *Proc. Natl. Acad. Sci. U. S. A.* 17 (2006) 732–737.
- [4] Mahesh S. Dharne, Sri Prakash Misra, Vatsala Misra, Manisha Dwivedi, Milind S. Patole, Yogesh S. Shouche, Isolation of urease-positive *Ochrobactrum intermedium* in the stomach of a non-ulcer dyspeptic patient from north India. *J. Microbiol. Immunol. Infect.* (2008) 183–186.
- [5] Girish Kulkarni, Dhairaj Dhotre, Mahesh Dharne, Sudarshan Shetty, Somak Chowdhury, Vatsala Misra, Sriprakash Misra, Milind Patole, Yogesh Shouche, Draft genome of *Ochrobactrum intermedium* strain M86 isolated from non-ulcer dyspeptic individual from India. *Gut Pathog.* 5 (2013) 7.
- [6] Michael A. Quail, Miriam Smith, Paul Coupland, Thomas D. Otto, Simon R. Harris, Thomas R. Connor, Anna Bertoni, Harold P. Swerdlow, Gu. Yong, A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13 (1) (2012) 341.
- [7] Yiren Yang, Xuefei Yu, Ren Zhang, Draft genome sequence of *Ochrobactrum pseudogrignone* strain CDB2, a highly efficient arsenate-resistant soil bacterium from arsenic-contaminated cattle dip sites. *Genome Announc.* 1 (2) (2013) e00173-13.
- [8] Y.J. Yuan, Z.X. Lu, N. Wu, L.J. Huang, F.X. Lü, X.M. Bie, Isolation and preliminary characterization of a novel nicotine-degrading bacterium, *Ochrobactrum intermedium* DN2. *Int. Biodeterior. Biodegrad.* 56 (1) (2005) 45–50.
- [9] Szabolcs Semsy, István Papp, Zsuzsanna Buzas, Andras Patthy, Laszlo Orosz, Peter P. Papp, Identification of site-specific recombination Genesint and xis of the rhizobium temperate phage 16-3. *J. Bacteriol.* 181 (14) (1999) 4185–4192.
- [10] Kelly P. Williams, Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. *Nucleic Acids Res.* 30 (4) (2002) 866–875.
- [11] Lia Cardarelli, Robert Lam, Ashleigh Tuite, Lindsay A.A. Baker, Paul D. Sadowski, Devon R. Radford, John L. Rubinstein, et al., The crystal structure of bacteriophage HK97 gp6: defining a large family of head–tail connector proteins. *J. Mol. Biol.* 395 (4) (2010) 754–768.
- [12] Anthony J. De Lucca, Thomas J. Walsh, Antifungal peptides: novel therapeutic compounds against emerging pathogens. *Antimicrob. Agents Chemother.* 43 (1) (1999) 1–11.
- [13] L.C. Sands, W.V. Shaw, Mechanism of chloramphenicol resistance in staphylococci: characterization and hybridization of variants of chloramphenicol acetyltransferase. *Antimicrob. Agents Chemother.* 3 (2) (1973) 299–305.
- [14] D. Blankenberg, A. Gordon, G. Von Kuster, N. Coraor, J. Taylor, Nekrutenko A; Galaxy team. Manipulation of FASTQ data with Galaxy. *Bioinformatics* 26 (14) (2010) 1783–1785.
- [15] Matthias Dodt, Johannes T. Roehr, Rina Ahmed, Christoph Dieterich, Flexbar — flexible barcode and adapter processing for next-generation sequencing platforms. *Biology* 1 (3) (2012) 895–905.
- [16] G. Cuccuru, M. Orsini, A. Pinna, A. Sbardellati, N. Soranzo, A. Travaglione, P. Uva, G. Zanetti, G. Fotia, Oriome, a web-based framework for NGS analysis in microbiology. *Bioinformatics* 30 (13) (2014) 1928–1929 (2013).
- [17] Daniel R. Zerbino, Ewan Birney, Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18 (5) (2008) 821–829.
- [18] Anton Bankevich, Sergey Nurk, Dmitry Antipov, Alexey A.A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, et al., SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19 (5) (2012) 455–477.
- [19] Shin-Hung Lin, Yu-Chieh Liao, CISA: contig integrator for sequence assembly of bacterial genomes. *PLoS One* 8 (3) (2013), e60843.
- [20] Marten Boetzer, Christiaan V. Henkel, Hans J. Jansen, Derek Butler, Walter Pirovano, Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27 (4) (2011) 578–579.
- [21] Nabil-Fareed Ali Khan, Nicola K. Petty, Nouri L. Ben Zakour, Scott A.A. Beatson, BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12 (1) (2011) 402.
- [22] Torsten Seemann, Prokka: rapid prokaryotic genome annotation. *Bioinformatics* (2014) btu153.

- [23] Ramy K. Aziz, Daniela Bartels, Aaron A. Best, Matthew DeJongh, Terrence Disz, Robert A. Edwards, Kevin Formsma, et al., The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9 (1) (2008) 75.
- [24] Minoru Kanehisa, Susumu Goto, KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28 (1) (2000) 27–30.
- [25] Patricia Siguier, Jocelyne Pérochon, L. Lestrade, Jacques Mahillon, Michael Chandler, ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 34 (Suppl. 1) (2006) D32–D36.
- [26] Gary Benson, Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27 (2) (1999) 573.
- [27] You Zhou, Yongjie Liang, Karlene H. Lynch, Jonathan J. Dennis, David S. Wishart, PHAST: a fast phage search tool. *Nucleic Acids Res.* (2011) gkr485.
- [28] Ibtissem Grissa, Gilles Vergnaud, Christine Pourcel, CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35 (Suppl. 2) (2007) W52–W57.
- [29] Kai Blin, Marnix H. Medema, Daniyal Kazempour, Michael A. Fischbach, Rainer Breitling, Eriko Takano, Tilmann Weber, antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.* (2013) gkt449.
- [30] Todd J. Treangen, Brian D. Ondov, Sergey Koren, Adam M. Phillippy, The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* 15 (11) (2014) 1–15.