High quality draft genome of *Lactobacillus kunkeei* EFB6, isolated from a German European foulbrood outbreak of honeybees

Marvin Djukic1, Anja Poehlein1, Juliane Strauß1, Fabian Jannik Tann1, Andreas Leimbach1, Michael Hoppert2 and Rolf Daniel1*

**Abstract**

The lactic acid bacterium *Lactobacillus kunkeei* has been described as an inhabitant of fructose-rich niches. Here we report on the genome sequence of *L. kunkeei* EFB6, which has been isolated from a honeybee larva infected with European foulbrood. The draft genome comprises 1,566,851 bp and 1,417 predicted protein-encoding genes.

**Keywords:** *Lactobacillus kunkeei*, Lactic acid bacteria, European foulbrood, Honeybee, Cellular surface protein, Biofilm formation

**Introduction**

Honeybees are the most economically valuable pollinators of agricultural crops [1]. A disappearance of honeybees would result in an approximately 90% decrease in production of some fruits [2]. European foulbrood (EFB) and American foulbrood (AFB) are the two most important honeybee diseases affecting the brood [3]. While the AFB is caused by the spore-forming, Gram positive bacterium *Paenibacillus larvae* [4], EFB is caused by the capsule-producing *Melissococcus plutonius* [5]. It has been shown that members of the lactic acid bacteria (LABs) inhibit the growth of *M. plutonius* [6] and *P. larvae* [7]. LABs are found in a variety of habitats, including human and animal microbiomes, and are used as food additives.

The honeybee crop microbiome consists of 13 bacterial species belonging to the genera *Lactobacillus* and *Bifidobacterium* [8]. These bacteria play a key role in the production of honey and bee bread. The latter serves as long-term food storage for adult honeybees and larvae. *L. kunkeei* is a common symbiont for *Apis* and the dominating LAB member in bees [6]. The organism is a specialist for colonization of the honeybee crop and interacts with the epithelial layer of the crop. *L. kunkeei* has been described as a fructophilic LAB [9]. Initially, it was isolated from wine [10], but it has also been found on flowers and in honey.

*L. kunkeei* EFB6 is the first LAB isolated from a German EFB-diseased larva. Here, we describe genomic features of this organism, focusing on factors that improve competition with bacteria such as *M. plutonius* and *P. larvae*. In addition, potential cell surface proteins that might play a role in cellular adhesion and biofilm formation are analyzed.

**Organism information**

In October 2012, an EFB outbreak in Bavaria (Germany) was confirmed. EFB-diseased larvae from this outbreak were collected, immediately frozen in liquid nitrogen and stored at −80°C for further investigation. Several EFB-infected larvae were dissected under sterile conditions. To obtain LAB the guts of the larvae, which formed a yellow, glue-like slime, were suspended in MRS medium (Carl Roth GmbH & Co KG, Karlsruhe, Germany) and subsequently streaked on solidified MRS to isolate single colonies. Strain *L. kunkeei* EFB6 (Table 1, Additional file 1: Table S1) was isolated from these agar plates after aerobic incubation at 35°C.

*L. kunkeei* EFB6 is a non-sporulating, low G+C Gram positive member of the *Lactobacteriaceae* and taxonomically related to the genus *Pediococcus*. The strain exhibited...
a 100% 16S rRNA gene nucleotide sequence identity to the type strain \textit{L. kunkeei} YH-15 (Table 1, Figure 1). Cells harvested in exponential growth phase exhibited a length ranging from 0.7 to 1.3 \(\mu\)m and a diameter ranging from 0.3 to 0.5 \(\mu\)m as determined by transmission electron microscopy (TEM) of either negatively stained or ultrathin-sectioned samples (Figure 2). Preparations for ultrathin sectioning and negative staining of cells were performed as described by [23]. The \textit{L. kunkeei} EFB6 cell wall is approximately 12 nm thick. This value is rather thin compared to cell walls of other Gram positives [24]. Three distinct wall layers of \textit{L. kunkeei} EFB6 (two darker stained outer and inner layers and a brighter layer in between) could be distinguished by TEM. Surface layers and cellular appendages (pili, fimbriae) were not detected.

### Table 1 Classification and general features of \textit{Lactobacillus kunkeei} EFB6

| MIGS ID | Property             | Term                           | Evidence code |
|---------|----------------------|--------------------------------|---------------|
|         | Classification       | Domain \textit{Bacteria}       | TAS [11]      |
|         |                      | Phylum \textit{Firmicutes}     | TAS [12-15]   |
|         |                      | Class \textit{Bacilli}         | TAS [16]      |
|         |                      | Order \textit{Lactobacillales} | TAS [17]      |
|         |                      | Family \textit{Lactobacillaceae}| TAS [18]      |
|         |                      | Genus \textit{Lactobacillus}   | TAS [18-21]   |
|         |                      | Species \textit{Lactobacillus kunkeei} strain: EFB6 | TAS (this study) |
|         | Gram stain           | Positive                       | TAS [10]      |
|         | Cell shape           | Rod-shaped                     | IDA           |
|         | Motility             | Non-motile                     | IDA           |
|         | Sporulation          | Non-sporulating                | NAS           |
|         | Temperature range    | Mesophile                      | TAS [10]      |
|         | Optimum temperature  | 30°C                           | NAS           |
|         | pH range; Optimum    | 4.5-6.2; 6                     | NAS           |
|         | Carbon source        | Varied                         | NAS           |
| MIGS-6  | Habitat              | Honeybee larva                 | IDA           |
| MIGS-6.3| Salinity             | 5% NaCl (w/v)                  | TAS [10]      |
| MIGS-22 | Oxygen requirement   | Facultative                    | IDA           |
| MIGS-15 | Biotic relationship  | Host-associated                | TAS [6]       |
| MIGS-14 | Pathogenicity        | Non-pathogen                   | NAS           |
|         | Biosafety level      | 1                              | TAS [22]      |
| MIGS-23 | Isolation            | EFB-diseased honeybee larva    | IDA           |
| MIGS-4  | Geographic location  | Bavaria, Germany               | IDA           |
| MIGS-5  | Sample collection    | October 1, 2012                | IDA           |
| MIGS-4.1| Latitude             | 49°14′ N                       | IDA           |
| MIGS-4.2| Longitude            | 11°05′ E                       | IDA           |
| MIGS-4.4| Altitude             | 400 m a.s.l                    | IDA           |

### Genome sequencing and annotation

#### Genome project history

The organism was selected for sequencing on the basis of its use as potential inhibitor for the primary agents of AFB and EFB [6,7]. The aim was to investigate potential factors to increase bacterial competition fitness and cell surface proteins, which might be important for cellular adhesion and biofilm formation.

A summary of the project information is shown in Table 2.

#### Growth conditions and DNA isolation

To isolate genomic DNA \textit{L. kunkeei} EFB6 was grown aerobically in 50 ml MRS medium at 35°C with shaking at 150 rpm (Lab-Therm Lab-Shaker, Adolf Kühner AG, Birsfelden, Switzerland). Cells were harvested in
exponential growth phase using a Beckman Coulter Allegra™ X-12R centrifuge (Beckman Coulter GmbH, Krefeld, Germany) for 25 minutes at 2,750 g and 4°C. Genomic DNA was isolated using the Epicentre® MasterPure™ DNA Purification kit (Epicentre®, Madison, WI, USA).

**Genome sequencing and assembly**

Whole-genome sequencing of *L. kunkeei* EFB6 was performed by employing the Genome Analyzer II (Illumina, San Diego, CA). The shotgun library was prepared according to the manufacturer’s protocols. For *de novo*
assembly, we used 2,000,000 paired-end Illumina reads (112 bp) and the SPAdes 2.5 software [27]. The final assembly contained 55 contigs larger than 500 bp and revealed an average coverage of 142.96.

**Genome annotation**

For automatic gene prediction the software tools YACOP [28] and Glimmer [29] were used. Identification of rRNA and tRNA genes was performed by employing RNAmmer [30] and tRNAscan [31], respectively. The annotation provided by the IMG-ER system [32] was corrected manually. For this purpose, data obtained from different databases (Swiss-Prot [33], TrEMBL [34] and InterPro [35]) were used to improve the quality of the annotation.

**Genome properties**

The genome statistics are provided in Table 3. The high quality draft genome sequence consists of 55 contigs that account for a total of 1,566,851 bp and a G + C content of 37 mol%. Of the 1,455 predicted genes, 1,417 were putatively protein-encoding, 35 represented putative tRNA genes and three putative rRNA genes. For the majority of the protein-encoding genes (75%) a function could be assigned. The distribution of these genes into COG functional categories [36] is shown in Table 4.

**Insights into the genome**

Five different *Lactobacillus* species were used for genome comparisons with *L. kunkeei* EFB6 based on blastp [37]. Results are shown in Figure 3. All five species are of interest as probiotics, part of the gastrointestinal tract of animals or humans, or used in the production of fermented food.

The identification of orthologous proteins was performed with the program Proteinortho 5.04 [39] by using

### Table 2 Genome sequencing project information

| MIGS ID | Property                  | Term                                    |
|---------|---------------------------|-----------------------------------------|
| MIGS-31 | Finishing quality         | Improved high-quality draft             |
| MIGS-28 | Libraries used            | One Illumina paired-end library with 1 kb insert size |
| MIGS-29 | Sequencing platforms      | Illumina GAI                            |
| MIGS-31.2| Fold coverage            | 142.96 x Illumina                       |
| MIGS-30 | Assemblers                | SPAdes 2.5                              |
| MIGS-32 | Gene calling method       | YACOP, Glimmer                          |
|         | Locus Tag                 | LAKU                                    |
|         | Genbank ID                | A2B100000000                            |
|         | GenBank Date of Release   | May, 2014                               |
|         | GOLD ID                   | GI0053745                               |
|         | NCBI project ID           | 227106                                  |
|         | BIOPROJECT                | PRJNA227106                              |
|         | Project relevance         | Host-associated                         |

### Table 3 Genome statistics

| Attribute                | Value   |
|--------------------------|---------|
| Genome size (bp)         | 1,566,851 |
| DNA coding (bp)          | 1,413,077 |
| DNA G + C (bp)           | 578,359  |
| DNA scaffolds            | 55      |
| Total genes              | 1,455   |
| Protein coding genes     | 1,417   |
| RNA genes                | 38      |
| Pseudo Genes             | 0       |
| Genes in internal clusters| 20     |
| Genes with function prediction | 1,012   |
| Genes assigned to COGs   | 1,195   |
| Genes assigned Pfam domains| 1,221   |
| Genes with signal peptides| 62      |
| Genes with transmembrane helices | 419    |
| CRISPR repeats           | 0       |

### Table 4 Number of genes associated with the general COG functional categories

| Code | Value | % age | Description                                                  |
|------|-------|-------|--------------------------------------------------------------|
| J    | 137   | 10.57 | Translation, ribosomal structure and biogenesis             |
| A    | 0     | 0.00  | RNA processing and modification                              |
| K    | 95    | 7.33  | Transcription                                                |
| L    | 94    | 7.25  | Replication, recombination and repair                        |
| B    | 0     | 0.00  | Chromatin structure and dynamics                              |
| D    | 24    | 1.85  | Cell cycle control, cell division, chromosome partitioning   |
| V    | 18    | 1.39  | Defense mechanisms                                           |
| T    | 32    | 2.47  | Signal transduction mechanisms                               |
| M    | 88    | 6.79  | Cell wall/membrane biogenesis                                 |
| N    | 10    | 0.77  | Cell motility                                                |
| U    | 25    | 1.93  | Intracellular trafficking and secretion                      |
| O    | 45    | 3.47  | Post translational modification, protein turnover, chaperones |
| C    | 49    | 3.78  | Energy production and conversion                             |
| G    | 67    | 5.17  | Carbohydrate transport and metabolism                        |
| E    | 112   | 8.64  | Amino acid transport and metabolism                          |
| F    | 68    | 5.25  | Nucleotide transport and metabolism                          |
| H    | 34    | 2.62  | Coenzyme transport and metabolism                            |
| I    | 35    | 2.70  | Lipid transport and metabolism                               |
| P    | 61    | 4.71  | Inorganic ion transport and metabolism                       |
| Q    | 13    | 1.00  | Secondary metabolites biosynthesis, transport and catabolism  |
| R    | 155   | 11.96 | General function prediction only                             |
| S    | 134   | 10.34 | Function unknown                                             |
| -    | 260   | 17.87 | Not in COGs                                                  |
the protein content deduced from 232 lactobacilli genomes as references (GenBank database as of 28.02.2014). For this purpose ncbi_ftp_download v0.2, cat_seq v0.1 and cds_extractor v0.6 were used [40]. With an identity cutoff of 50%, we identified 425 proteins in _L. kunkeei_ EF B6 without orthologs in any other _Lactobacillus_ species. Among these unique _L. kunkeei_ EF B6 proteins, we selected 7 proteins for detailed analyses.

Analysis of the 89-kb region shown in Figure 3 revealed five ORFs (LAKU_4c00030-LAKU_4c00070) without orthologs in any genomes derived from lactobacilli deposited in GenBank (as of 28.02.2014). Furthermore, no homologs could be identified in any other sequenced microbial genome (NCBI nr-database as of 05.03.2014) by using blastp (e-value cutoff of 1e-20). Except for LAKU_4c00060 (7,521 amino acids), we could identify an N-terminal signal peptide and a non-cytoplasmic domain (Figure 4A) using Phobius’ domain prediction software [41]: LAKU_4c00040 (4,579 amino acids) and LAKU_4c00070 (3,129 amino acids) contain
coiled coil structures. Except of LAKU_4c00050 (8,342 amino acids), all ORFs show weak similarity to large surface proteins or extracellular matrix-binding proteins found in bacteria such as Staphylococcus, Streptococcus, Burkholderia, Weissella, Mannheimia, and Marinomonas, but also in Lactobacillus and Pediococcus. Since, L.

Figure 4 Domain prediction (A) of the 89-kb region found in L. kunkeei EFB6 and its presence in other lactobacilli (B). A combined transmembrane topology and signal peptide predictor [41] was used to determine putative domains. The yellow blocks represent signal peptides, the white color of the arrows show the non-cytoplasmic part, red blocks represent transmembrane regions and blue blocks predicted coiled-coil structures. To test whether this region exists in other L. kunkeei strains, we designed specific primer-pairs for each ORF (Table 5, Figure 4A). Predicted PCR product sizes are depicted in white boxes. The presence of the genes were tested for L. kunkeei EFB6, L. kunkeei HI3 (isolated from honey), L. kunkeei DSM 12361 (isolated from wine), and L. johnsonii DSM 10533 (isolated from human blood) (Figure 4B). The obtained PCR product sizes correlated with the predicted sizes (Table 5, Figure 4A). For L. johnsonii DSM 10533, no PCR product could be obtained.
**Table 5 Primer used in this study**

| Primer            | DNA sequence (5′-3′)             | Open reading frame | Product size |
|-------------------|----------------------------------|--------------------|--------------|
| LKU_ORSF1A_for   | AACCAAGAGTAACGATGCCC             | LAKU_4c00030       | 536 bp       |
| LKU_ORSF1A_rev   | CTTTGTTATCGGCTTGGGC              |                    |              |
| LKU_ORSF1B_for   | CGATGCGACAACACTGCTTACG           | LAKU_4c00030       | 355 bp       |
| LKU_ORSF1B_rev   | CATCCTTTTGGCCTGCTTGTG            |                    |              |
| LKU_ORSF2_for    | AGCTTCTTTAGCTGCCGCTTGG           | LAKU_4c00040       | 323 bp       |
| LKU_ORSF2_rev    | TATGCGCTTTGGTCTTTGTGCC           |                    |              |
| LKU_ORSF3_for    | GCGACTTTTCTGTGTTGGG              | LAKU_4c00050       | 358 bp       |
| LKU_ORSF3_rev    | ATAGCCCCCCAGCATATCCAGC           |                    |              |
| LKU_ORSF4_for    | CTAGCCTGAGGTTCGCTG               | LAKU_4c00060       | 566 bp       |
| LKU_ORSF4_rev    | GTTGAGTACCTTGGCCACC              |                    |              |
| LKU_ORSF5_for    | TCCAGTATGAAACAGTAACACC           | LAKU_4c00070       | 358 bp       |
| LKU_ORSF5_rev    | AAGCGGTGATTTCCATTACC             |                    |              |

**Figure 5** Tblastx comparison of *L. kunkeei* ORF LAKU_24c00010 to matrix binding proteins of *L. rhamnosus* ATCC 53103 and ”*L. zae*” KCTC 3804. The graphical presentation was done with Easyfig software (minimum blast hit length of 200 bp and a maximum e-value of 1e−105) [44]. LAKU_24c00010 shows similarities to WP_010490864, WP_010490864 and WP_010490862 of ”*L. zae*” KCTC 3804, but also to YP_005866289 (*L. rhamnosus* ATCC 53103). The ORFs used for comparison are labeled with NCBI accession numbers. The blast identity is shown in a colored scale ranging from 31 % (yellow) to 100 % (red).
to possess lysozyme activity. Two of the deduced proteins (LAKU_13c00160 and LAKU_32c00010) contain a signal peptide, indicating secretion of the proteins. LAKU_19c00290 harbors transmembrane helices and is probably anchored in the cell wall. LAKU_6c00080 did not contain a putative signal peptide or transmembrane helices.

Rapid test PCR
Specific primer pairs have been designed to test other strains by PCR for the presence of an 89 kb region, which harbors five open reading frames (ORFs). Genomic DNA of the L. kunkeei strains EFB6, HI3 and DSM 12361, and Lactobacillus johnsonii DSM 10533 was used as template for PCR amplifications employing the thermal cycler peqSTAR 2X (PEQLAB Biotechnologie GmbH, Erlangen, Germany). PCR amplification was performed with the BIO-X-ACT™ Short DNA Polymerase (Bioline, Luckenwalde, Germany) and an initial denaturation step at 98°C for 2 min, followed by 30 cycles of denaturation for 20 s at 96°C, annealing for 20 s at 60°C and elongation for 30 s at 68°C. Subsequently, a final elongation step of 10 min at 68°C was performed. PCR products were purified employing the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany).

Conclusion
In this study, we characterized the genome of L. kunkeei strain EFB6 isolated from an EFB-diseased larva. In a recent study was shown that L. kunkeei has the potential for biofilm formation and adhesion to the honey crop [6]. Our genome analysis supports these results. Using large surface proteins or extracellular matrix-binding proteins, L. kunkeei might be able to attach to eukaryotic epithelial cells. Furthermore, due to the presence of polysaccharide biosynthesis proteins and several enzymes with lysozyme activity, it is possible that L. kunkeei is actively protecting its niche against bacterial competitors. As LABs have been shown to have an inhibitory growth effect on M. plutonius, the use of LABs as probiotic additive against the EFB-causing agent is conceivable.

Additional file
Additional file 1: Associated MIGS Record.

Abbreviations
AFB: American foulbrood of honeybees; EFB: European foulbrood of honeybees; LABs: Lactic acid bacteria; TEM: Transmission electron microscopy.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
MD, AP and RD designed research, MD, JS and FJT isolated and characterized strain EFB6, MD, AP and AL carried out genome analyses, MH performed electron microscopy, MD and RD wrote the manuscript with help of AP. All authors read and approved the final manuscript.

Acknowledgements
We thank the "Bundesministerium für Bildung und Forschung" and the "Niedersächsisches Ministerium für Wissenschaft und Kultur" for support. Additionally, we acknowledge support by the German Research Foundation and the Open Access Publication Funds of the Göttingen University.

Author details
1Department of Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University of Göttingen, Göttingen, Germany. 2Department of General Microbiology, Institute of Microbiology and Genetics, Georg-August University of Göttingen, Göttingen, Germany.

Received: 20 May 2014 Accepted: 4 December 2014 Published: 19 February 2015

References
1. Klein A-M, Vaissière BE, Cane JH, Steffan-Dewenter I, Cunningham SA, Kremen C, et al. Importance of pollinators in changing landscapes for world crops. Proc R Soc B. 2007;274:303–13.
2. Southwick EE, Southwick Jr L. Estimating the economic value of honey bees (Hymenoptera: Apidae) as agricultural pollinators in the United States. J Econ Entomol. 1992;85:621–33.
3. Forsgren E. European foulbrood in honey bees. J Invertebr Pathol. 2010;103 (Suppl):S5–9.
4. Djúkic M, Brauskiewicz E, Fürnhaus A, Voss J, Gollnow K, Poppagina L, et al. How to Kill the Honey Bee Larva: Genomic Potential and Virulence Mechanisms of *P. gambiense* var. *P. apeiropalaeimba*. DNA Res. 2014;21:90–103.

5. Bailey L, Collins MD. Reclassification of *Streptococcus plautii* (White) in a new genus *Melissococcus*, as *Melissococcus plautii* nom. rev. comb. nov. J Appl Microbiol. 1982;52:215–26.

6. Vásquez A, Forsgren E, Fries I, Paxton RJ, Flaberg E, Szekely L, et al. Symbiosis as major modulators of insect health: lactic acid bacteria and honeybees. PLoS One. 2012;7:e33188.

7. Forsgren E, Olofsson TC, Vásquez A, Fries I. Novel lactic acid bacteria inhibiting *P. gambiense* var. *P. apeiropalaeimba* in honey bee larva. Apidologie. 2012;43:99–108.

8. Olofsson TC, Vásquez A. Detection and identification of a novel lactic acid bacterial flora within the honey stomach of the honeybee *A. mellifera*. Curr Microbiol. 2008;57:35–63.

9. Endo A, Iriyama T, Futagawa-Endo Y, Takanu K, du Toit M, Okada S, et al. Characterization and emended description of *Lactobacillus kunkeei* as a fructophilic lactic acid bacterium. Int J Syst Evol Microbiol. 2012;62:500–4.

10. Edwards CG, Haag KM, Collins MD, Hutson RA, Huang YC. *Kunkeelia*. Int J Syst Bacteriol. 1980;30:225–30.

11. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal of the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci. 1990;87:4576–9.

12. Gibbons NE, Murray RG. Proposals Concerning the Higher Taxa of Bacteria. Int J Syst Bacteriol. 1978;28:1–6.

13. Garrity G, Holt J. The Road Map to the Manual. In: Bergey’s Manual of Systematic Bacteriology, vol. 1. 2nd ed. New York: Springer; 2001. p. 119–69.

14. Murray R. The Higher Taxa, or a Place for Everything? In: Bergey’s Manual of Systematic Bacteriology, vol. 1. 1st ed. Baltimore: The Williams and Wilkins Co; 1984. p. 31–4.

15. List Editor. List of new names and new combinations previously effectively, but not validly, published. List no. 122. Int J Syst Evol Microbiol. 2011;61:469–72. doi:10.1099/ijs.0.022855-0.

16. Ludwig W, Schleifer K, Whitman W. Class I. Bacilli class nov. In: Bergey’s Manual of Systematic Bacteriology, Fourth ed. Baltimore: The Williams and Wilkins Co; 1994. p. 300–5.

17. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. Technical rules for single-cell sequencing. J Comput Biol. 2012;19:455–77.

18. Donovan C, Martin MJ, Gattiker A, Gasteiger E, Bairoch A, Apweiler R. High-quality protein knowledge resource: SWISS-PROT and TrEMBL. Brief Bioinform. 2003;4:41.

19. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics. 2011;12:142.

20. Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of co-orthologs in large-scale analysis. BMC Bioinformatics. 2011;12:4.

21. Leimbach A. bac-genomics-scripts. 2014. Available at: https://github.com/aleimb/bac-genomics-scripts.

22. Olofsson M, Rönnberg L, Forsgren E, Fries I. *Lactobacillus kunkeei* sp. nov. – a novel lactic acid bacteria from the honey stomach of the honeybee *A. mellifera*. Curr Microbiol. 2008;57:35–63.

23. Hoppert M. Microscopic Techniques in Biotechnology. Weinheim, FRG; Wiley-VCH Verlag GmbH & Co. KGaA. 2003. Available at: http://doi.wiley.com/10.1002/3527602615.

24. Mayer F. Cytology and Morphogenesis of Bacteria (Handbook of plant Anatomy IV, Z). Stuttgart: Borntraeger; 1988.

25. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22:4673–80.

26. Kumar S, Nei M, Dudley J, Tamura K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform. 2008;9:290–300.

27. Bankievich A, Nuk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77.