**References**

1. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477. <http://dx.doi.org/10.1007/978-3-642-37195-0_13>
2. Boc A, Diallo AB, Makarenkov V (2012) T-REX: a Web server for inferring validating and visualizing phylogenetic trees and networks. *Nucleic Acids Res.* **40,**  W573–W579. <http://dx.doi.org/10.1093/nar/gks485>
3. Chen L, Zheng D, Liu B, Yang J and Jin Q (2016) VFDB 2016: hierarchical and refined dataset for big data analysis - 10 years on. *Nucleic Acids Res* **44,** D694–D697. <http://dx.doi.org/10.1093/nar/gkv1239>
4. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T *et al.* (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118 ; iso-2; iso-3. *Fly* (Austin) **6**, 80–92. <http://dx.doi.org/10.4161/fly.19695>
5. Darling AE, Mau B and Perna NT (2010) progressiveMauve: Multiple Genome Alignment with Gene Gain Loss and Rearrangement. *PLoS ONE* **5**, e11147. <http://dx.doi.org/10.1371/journal.pone.0011147>
6. Gibson MK, Forsberg KJ, Dantas G (2015) Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J* **9,** 207–216. <https://dx.doi.org/10.1038/ismej.2014.106>
7. Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075. <http://dx.doi.org/10.1093/bioinformatics/btt086>
8. Li H (2011) A statistical framework for SNP calling mutation discovery association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993. <http://dx.doi.org/10.1093/bioinformatics/btr509>
9. Li H and Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595. <http://dx.doi.org/10.1093/bioinformatics/btp698>
10. Luo R, Liu B, Xie Y, Li Z, Huang W *et al.*  (2012) SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* **1**, 18. <http://dx.doi.org/10.1186/2047-217x-1-18>
11. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10-12. <http://dx.doi.org/10.14806/ej.17.1.200>
12. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S *et al.* (2015) Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693. <http://dx.doi.org/10.1093/bioinformatics/btv421>
13. Price MN, Dehal PS and Arkin AP (2010) FastTree 2 - Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **5**, e9490. <http://dx.doi.org/10.1371/journal.pone.0009490>
14. Seemann T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinforma. Oxf. Engl.* **30***,* 2068–2069. <http://dx.doi.org/10.1093/bioinformatics/btu153>
15. Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. **14**, 178–192. <http://dx.doi.org/10.1093/bib/bbs017>