

# CS 175 Problem set 1

Rianna Patricia Cruz (2013-59075)

1. By inspection using alignment visualization application AliView<sup>1</sup>, it can be determined that the number of base pairs in each of the given sequences was 10,179, except for sequence *seq\_PK*, which had 10,176 base pairs.
2. Using a Python program, it has been determined that there are 6226 fully conserved positions in the resulting alignment of the given sequences.
3. From automated inspection using a Python program, it has been found that the longest fully conserved substring in the resulting alignment is 21 nucleotides long, and there are two such instances of these.

These substrings are

- *tggatgacaacagaagacatg*; and
  - *ctttcaatatgctgaaacgcg*.
4. For this item, the description “*pair of sequences have the most number of matched regions*” has been taken as those which quantitatively had the highest percent query covering given a query of the given sequences, or those which had the highest ‘ident’ or ‘identity’ parameter in a BLAST query of the given sequences among themselves.

BLAST<sup>2</sup> defines identity as ‘[t]he extent to which two (nucleotide or amino acid) sequences have the same residues at the same positions in an alignment, often expressed as a percentage,’ or simply as the extent to which two given sequences are identical.

	Sequence	100% covered	99% covered only	Highest ‘identity’
1	seq_CN	seq_CN seq_KH seq_VN seq_TH seq_JP seq_TW	seq_ID seq_PH seq_INseq_SG	seq_KH seq_VN seq_TH seq_JP seq_TW (98% all)

---

<sup>1</sup> A. Larsson, “AliView: a fast and lightweight alignment viewer and editor for large datasets,” *Bioinformatics*, vol. 30, no. 22, pp. 3276–3278, May 2014.

<sup>2</sup> Fassler J, Cooper P. BLAST Glossary. 2011 Jul 14. In: BLAST® Help [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2008-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK62051/>

		seq_MM		
2	seq_ID	seq_PH seq_IN seq_SG	seq_MM seq_KH seq_VN seq_CN seq_TH seq_JP seq_TW	seq_PH (95%)
3	seq_IN	seq_SG seq_ID seq_PH	seq_CN seq_MM seq_KH seq_VN seq_TH seq_TW seq_JP	seq_SG (95%)
4	seq_JP	seq_TW seq_CN seq_KH seq_VN seq_TH seq_MM	seq_ID seq_PH seq_IN seq_SG	seq_TW seq_CN seq_KH (98% all)
5	seq_KH	seq_TH seq_VN seq_CN seq_JP seq_TW seq_MM	seq_ID seq_PH seq_IN seq_SG	seq_TH seq_VN (99% all)
6	seq_MM	seq_CN seq_KH seq_VN seq_TH seq_JP seq_TW	seq_ID seq_PH seq_IN seq_SG	seq_CN seq_KH seq_VN (97% all)
7	seq_PH	seq_ID seq_IN seq_SG	seq_MM seq_KH seq_VN seq_CN seq_TH seq_TW seq_JP	seq_ID (95%)
8	seq_PK	NONE	NONE	seq_PH (73%)
9	seq_SG	seq_IN seq_ID seq_PH	seq_VN seq_KH seq_CN seq_MM seq_TH seq_TW seq_JP	seq_IN (95%)
10	seq_TH	seq_KH seq_VN seq_CN seq_JP seq_TW seq_MM	seq_ID seq_PH seq_IN seq_SG	seq_KH seq_VN (99% all)
11	seq_TW	seq_JP seq_CN seq_KH seq_TH seq_VN seq_MM	seq_ID seq_PH seq_IN seq_SG	seq_JP seq_CN seq_KH (98% all)
12	seq_VN	seq_VN seq_KH	seq_ID seq_PH	seq_KH seq_TH

		seq_TH seq_CN seq_JP seq_TW seq_MM	seq_IN seq_SG	(99% all)
--	--	--	---------------	-----------

5. Based on a BLAST sequence search of the given sequences, most aligned sequences are of the mosquito-borne dengue virus. This virus is from the family Flaviviridae; genus Flavivirus<sup>3</sup>.
6. From inspection of a BLAST sequence search using the given dataset, it can be determined that the dataset describes selected sequences of a genotype of the dengue virus.

The particular strain of the virus appears to have been found in certain samples from both East Asia (Taiwan, certain provinces of China such as the Hanzhu province) and Southeast Asia (the Philippines, Thailand, Singapore), as well as South Asia (Sri Lanka), as well as Polynesian territories such as Hawaii and French Polynesia<sup>4</sup>, which circulated around the first few decades of the 2000s.

While most sequences from the dataset describe the genotype of dengue type 1 virus, some (eg. seq\_PK) describe dengue type 2, which may explain why such genotype has little similarity with the others from the same set, though this genotype also appears to have been sampled in some of the above mentioned regions (specifically in the Guangzhou province of China<sup>5</sup>), albeit somewhat more recently than other sequences in this dataset.

---

<sup>3</sup> Dengue virus, 2018 Aug 24. [Online]. Available: [https://en.wikipedia.org/wiki/Dengue\\_virus](https://en.wikipedia.org/wiki/Dengue_virus). [Accessed: 01-Oct-2018].

<sup>4</sup> From inspection; validated by A. Imrie, C. Roche, Z. Zhao, S. Bennett, M. Laille, P. Effler, and V.-M. Cao-Lormeau, "Homology of complete genome sequences for dengue virus type-1, from dengue-fever- and dengue-haemorrhagic-fever-associated epidemics in Hawaii and French Polynesia," *Annals of Tropical Medicine & Parasitology*, vol. 104, no. 3, pp. 225–235, 2010.

<sup>5</sup> T. Jiang, X.-D. Yu, W.-X. Hong, W.-Z. Zhou, M. Yu, Y.-Q. Deng, S.-Y. Zhu, E.-D. Qin, J. Wang, C.-F. Qin, and F.-C. Zhang, "Co-circulation of two genotypes of dengue virus serotype 3 in Guangzhou, China, 2009," *Virology Journal*, vol. 9, no. 1, p. 125, 2012.

## References

- A. Imrie, C. Roche, Z. Zhao, S. Bennett, M. Laille, P. Effler, and V.-M. Cao-Lormeau, "Homology of complete genome sequences for dengue virus type-1, from dengue-fever- and dengue-haemorrhagic-fever-associated epidemics in Hawaii and French Polynesia," *Annals of Tropical Medicine & Parasitology*, vol. 104, no. 3, pp. 225–235, 2010.
- A. Larsson, "AliView: a fast and lightweight alignment viewer and editor for large datasets," *Bioinformatics*, vol. 30, no. 22, pp. 3276–3278, May 2014.
- Dengue virus, 2018 Aug 24. [Online]. Available: [https://en.wikipedia.org/wiki/Dengue\\_virus](https://en.wikipedia.org/wiki/Dengue_virus). [Accessed: 01-Oct-2018].
- Fassler J, Cooper P. BLAST Glossary. 2011 Jul 14. In: BLAST® Help [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2008-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK62051/>
- T. Jiang, X.-D. Yu, W.-X. Hong, W.-Z. Zhou, M. Yu, Y.-Q. Deng, S.-Y. Zhu, E.-D. Qin, J. Wang, C.-F. Qin, and F.-C. Zhang, "Co-circulation of two genotypes of dengue virus serotype 3 in Guangzhou, China, 2009," *Virology Journal*, vol. 9, no. 1, p. 125, 2012.