A Little Book of Bioinformatics

Leighton Pritchard

2023-01-01

Table of contents

Preface to *A Little Book of Bioinformatics*

Welcome to *A Little Book of Bioinformatics.* This is an online book, under continual development, which I am building as and when topics come to mind or prominence.

My goal is that this online book will come to be a fairly transparent and honest reference for students in bioinformatics, and maybe for some researchers, too.

I would be very grateful for feedback by email or through the GitHub repository Issues page

1 Introduction

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

```
1 + 1
```

[1] 2

2 Summary

In summary, this book has no content whatsoever.

```
1 + 1
```

```
[1] 2
```

Part I

multiple_sequence_alignment.qmd

MAFFT

MAFFT is a very fast consistency-based iterative multiple sequence alignment method, suitable for aligning large sequence sets.

MAFFT stands for Multiple Alignment using Fast Fourier Transform, and was first described in 2002 (Katoh et al. 2002). Since then it has received multiple enhancements and extensions to improve scalability to larger sequence sets, and integration of other data such as protein structures [Katoh2008-hj; Katoh and Standley (2013); Rozewicki et al. (2019); Katoh, Rozewicki, and Yamada (2019)].

How does MAFFT work?

References

Katoh, Kazutaka, Kazuharu Misawa, Kei-Ichi Kuma, and Takashi Miyata. 2002. "MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform." *Nucleic Acids Res.* 30 (14): 3059–66. https://doi.org/10.1093/nar/gkf436.

Katoh, Kazutaka, John Rozewicki, and Kazunori D Yamada. 2019. "MAFFT Online Service: Multiple Sequence Alignment, Interactive Sequence Choice and Visualization." *Brief. Bioinform.* 20 (4): 1160–66. https://doi.org/10.1093/bib/bbx108.

Katoh, Kazutaka, and Daron M Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Mol. Biol. Evol.* 30 (4): 772–80. https://doi.org/10.1093/molbev/mst010.

Knuth, Donald E. 1984. "Literate Programming." *Comput. J.* 27 (2): 97–111. https://doi.org/10.1093/comjnl/27.2.97.

Rozewicki, John, Songling Li, Karlou Mar Amada, Daron M Standley, and Kazutaka Katoh. 2019. "MAFFT-DASH: Integrated Protein Sequence and Structural Alignment." *Nucleic Acids Res.* 47 (W1): W5–10. https://doi.org/10.1093/nar/gkz342.