OXFORD

Sequence analysis

Kablammo: an interactive, web-based BLAST results visualizer

Jeff A. Wintersinger* and James D. Wasmuth

Department of Ecosystem and Public Health, Faculty of Veterinary Medicine, University of Calgary, Calgary, Alberta, Canada

*To whom correspondence should be addressed.
Associate Editor: John Hancock

Abstract

Motivation: Kablammo is a web-based application that produces interactive, vector-based visualizations of sequence alignments generated by BLAST. These visualizations can illustrate many features, including shared protein domains, chromosome structural modifications and genome misassembly.
Availability and implementation: Kablammo can be used at http://kablammo.wasmuthlab.org. For a local installation, the source code and instructions are available under the MIT license at http://github.com/jwintersinger/kablammo.
Contact: jeff@wintersinger.org

1 Introduction

The Basic Local Alignment Search Tool (BLAST) rapidly finds similar subsequences shared between a query and database sequences (Altschul *et al*., 1997). Its uses include inferring homology from shared sequence similarity, identifying the species associated with an uncharacterized sequence, and locating domains shared between proteins.

BLAST's ease-of-use and diverse applications have made it a ubiquitous tool in molecular biology, especially for researchers without extensive bioinformatics training. However, its output is predominantly text-based. In our experience running training workshops, non-bioinformaticians often have difficulty finding features of interest, restricting BLAST to serving as a mere tool for identifying genes or species. Various applications attempt to resolve this by providing graphical interpretations of the BLAST results (Neumann*et al*., 2013), but these suffer important shortcomings.

The NCBI BLAST service (http://blast.ncbi.nlm.nih.gov) provides a simple visual summary of alignment results, but works only for BLAST searches performed through the web service. The BLAST Output Visualization tool (BOV) creates a static rather than interactive rendering of the BLAST alignments, renders only one query at a time, and works only with BLAST version 2.2.24 or earlier, which is almost four years out-of-date. We have tried other methods reviewed by Neumann *et al*. (2013), finding that no single tool satisfied our needs. To remedy these issues, we present Kablammo, an application that leverages modern web technologies to produce an expedient visual overview of BLAST results.

2 Implementation

2.1 Rapid BLAST result visualization

Kablammo visualizes BLAST results, permitting the identification of domains shared between proteins, structural modifications to chromosomes such as duplications and translocations, and technical issues such as genome misassembly. The results of all queries used in a BLAST run are rendered simultaneously and ordered by alignment score, allowing the user to quickly obtain a visual overview. A given query sequence and the matched database sequence are represented by horizontal bars (Fig. 1). The trapezoids drawn between these bars indicate the portions of the sequences that align to each other. The quality of each alignment is indicated by the colour of its trapezoid, with stronger alignments shaded darker. The user may also select individual alignments to view BLAST parameters, including sequence coordinates, reading frames, bit scores, and E values, as well as the raw aligned sequences. Alignments can be filtered by E value, bit score, length, and the names of subject or query sequences. This allows the user to display only alignments exceeding a quality threshold or those belonging to a particular gene family.
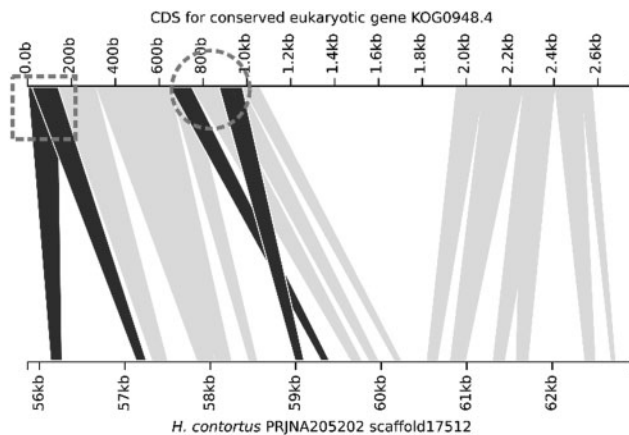
Fig. 1. Using Kablammo to investigate gene structure in an assembly of the *Haemonchus contortus* genome. The top axis represents the reference CDS for KOG0948.4, a highly conserved eukaryotic gene taken from CEGMA 2.4 (Parra *et al.*, 2007). The bottom axis represents a portion of one scaffold in the *H.contortus* genome PRJNA205202 (Schwarz *et al.*, 2013). Trapezoids drawn between these axes illustrate individual BLAST alignments between the two sequences. The darker alignments indicated by the square and circle denote structural features of interest described in Section 3

2.2 Interactive, vector-based output

The user can pan and zoom the alignment graphs to focus on the most interesting alignments. This operation is performed interactively and in real time, enabling refinement of rendered alignments even when the user does not know the precise coordinates of the regions of interest.

Kablammo produces vector-based output in the scalable vector graphics (SVG) file format. Thus, its graphs scale to any resolution and are suitable for use in print. Furthermore, these graphs can be easily imported into vector-image editors such as Inkscape or Adobe Illustrator, whereupon the user can change axis titles, alter colours, remove irrelevant alignments, or otherwise customize the graphs. Alternatively, the user can export high-resolution, print-ready rasterized PNG images directly from Kablammo.

Individual alignments can be selected to view their nucleotide or amino acid sequences. These alignments can also be exported to FASTA format for processing in other applications.

2.3 Web-based and entirely client-side

As Kablammo runs entirely within a web browser, it does not need to be installed by the user. Its processing is done entirely by the client, and so requires only a web server that can serve static files. This means that Kablammo can be used with almost any web server, such as Apache, Nginx or even the minimal static server included in every Python installation.

Kablammo is useful both for individual users and as a communal resource. In the individual mode, the user loads BLAST XML files directly into Kablammo. These BLAST results may originate from a local installation of BLAST or be downloaded from the NCBI BLAST web service. For communal use, BLAST results can be viewed without the need to load XML files directly into Kablammo. To achieve this, a static file that lists the available BLAST results is used. This file can be created and updated via a simple script

provided with Kablammo, which may be integrated into a lab's existing BLAST pipeline.

Kablammo's minimal dependencies on other software and permissive MIT license mean it can be easily incorporated into other projects. More expansive web-based bioinformatics tools may thus leverage Kablammo to interactively display sequence alignments.

3 Example usage

Figure 1 demonstrates one use of Kablammo, in which we have investigated gene structure in a genome assembly for the parasitic nematode *Haemonchus contortus* (Schwarz *et al.*, 2013). Kablammo makes clear three notable structural features of the predicted *Haemonchus* gene when compared to the reference coding sequence (CDS). First, the overlapping trapezoids in the region indicated by the square show that the same 120-bp portion of the CDS appears at two different positions in the *Haemonchus* scaffold. Second, the crisscrossed trapezoids originating in the circled region suggest that two portions of the CDS have been assembled out-of-order on this scaffold. Finally, the absence of trapezoids in the CDS region between 1.0 and 2.0 kb illustrate that there were no alignments for this part of the gene.

Though the subtleties of these inferences may be difficult to appreciate from the static nature of Figure 1, they are abundantly clear in the interactive graph rendered by Kablammo. The user may highlight individual alignments while making others transparent, as well as view the exact coordinates, bit scores and *E* values of each alignment. By illustrating these structural features, Kablammo lets the user rapidly determine that this portion of the genome is likely misassembled. Kablammo can be of equal utility in other analyses requiring a rapid visual overview of BLAST results.

Acknowledgements

Funding

References

Altschul,S.F. *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Neumann,R.S. *et al.* (2014). BLAST output visualization in the new sequencing era. *Briefings Bioinf.*, **15**, 484–503.

Parra,G.*et al.* (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics (Oxford, England)*, **23**, 1061–1067.

Schwarz,E.M. *et al.* (2013). The genome and developmental transcriptome of the strongylid nematode *Haemonchus contortus*. *Genome Biol.*, **14**, R89.