Sequence analysis

**DASher: a stand-alone protein sequence client for DAS, the Distributed Annotation System**

David N. Messina and Erik L. L. Sonnhammer*

Stockholm Bioinformatics Centre, Stockholm University, 10691 Stockholm, Sweden

Received on November 10, 2008; revised on February 26, 2009; accepted on March 14, 2009

Advance Access publication March 17, 2009

Associate Editor: Dmitrij Frishman

1 INTRODUCTION

Biological data are accumulated and provided by a large number of laboratories across the world. As a result, today we have hundreds of different web sites with different interfaces and little integration among them. Researchers wanting to compare annotations from disparate sources, even those that relate to the same sequence, must aggregate those annotations themselves, and this is typically a manual, tedious and time-consuming process.

The Distributed Annotation System (DAS) was developed to overcome these problems by creating a standard protocol by which source databases could serve and client programs could access that data. DAS establishes two huge advantages over the previous system: (i) DAS creates a standard way to access data, so all sources which comply with that standard can be viewed with the same software, and that can run as a self-contained application outside of a web browser. The speed, flexibility and extensibility that come with a stand-alone application motivated us to create DASher, an open-source Java DAS client. Given a UniProt sequence identifier, DASher automatically queries DAS-supporting servers worldwide for any information on that sequence and then displays the annotations in an interactive viewer for easy comparison. DASher is a fast, Java-based DAS client optimized for viewing protein sequence annotation and compliant with the latest DAS protocol specification 1.53E.

**Availability**: DASher is available for direct use and download at http://dasher.sbc.su.se including examples and source code under the GPLv3 licence. Java version 6 or higher is required.

**Contact**: erik.sonnhammer@sbc.su.se

**Supplementary information**: Supplementary data and all figures in color are available at Bioinformatics online.

2 OVERVIEW

2.1 Implementation

DASher is written in the object-oriented, platform-independent programming language Java (http://java.sun.com). It is based on the Sfixem platform (Chalk et al., 2004), and uses the Dasobert library to handle DAS stream input and output (http://dasmi.bioinf.mpi-inf.mpg.de/), Jalview (Clamp et al., 2004), PeppeR (http://bioscomp.cnb.uam.es/das/PeppeR/), IGB (http://genoviz.sourceforge.net/), Pfam (Finn et al., 2008), MaDas (http://madas.bioinfo.cnio.es), Dasyt2 (http://www.ebi.ac.uk/dasty) and Ensemble protview (Birney et al., 2006). However, only Dasyt2, Ensemble and Pfam are focused on protein sequence annotation, yet run in a web browser rather than as a separate application, which can limit their flexibility.

Here we introduce DASher, a lightweight, stand-alone Java DAS client optimized for viewing protein sequence features. As a stand-alone application, DASher offers advantages in responsiveness and interactivity. For example, users can zoom in on a region, and zoom-dependent sequence rendering displays residues automatically at a sufficiently high zoom level. Also, users have full control to customize the appearance of the data being displayed.

Supplementary information:

*To whom correspondence should be addressed.

© The Author 2009. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org
D.N.Messina and E.L.L.Sonnhammer

Fig. 1. The DASher main window. When a UniProt identifier is entered in the top left box, DASher fetches annotations from relevant DAS servers which are each displayed on a separate row, color-coded by feature type. In this example, we analyze the transmembrane topology of a largely unclassified human protein. Yellow denotes cytoplasmic regions; white, non-cytoplasmic regions; and brown, transmembrane regions. Tracks from top to bottom are the query sequence, UniProt, three transmembrane topology predictors, Kyte-Doolittle hydrophobicity (blue) and predicted solvent accessibility (green).

offer protein feature annotations, asks each of those servers for any information it may have on the requested sequence, and then displays those results in an integrated view in the main DASher window. For example, Figure 1 shows how one can examine in detail the transmembrane topology of a protein. In this case, three transmembrane domain prediction programs, a solvent accessibility predictor and a hydrophathy plot all argue against the second transmembrane domain (residues 268–288) annotated in the UniProt record (track 2).

Each server’s annotation appears as a separate track labeled with the name of the server. Depending on the nature of the data, a feature will be shown either as a box or as a line plot. Segments, such as protein domains, use the former representation, while annotations composed of continuous values such as hydrophobicity use the latter. All features of a given type are color-coded identically so that it is easy to identify when multiple servers have made the same annotation for a given region of a protein. There is an ongoing effort to standardize DAS servers around a sequence feature ontology (Jenkinson et al., 2008), and DASher uses this information to match annotations where possible.

We have developed DASher to provide a lightweight, stand-alone DAS client application specialized for comparing protein sequence annotations quickly and easily. DASher is compliant with the latest DAS specification 1.53E (Jenkinson et al., 2008) and is freely available as open source.

ACKNOWLEDGEMENTS
The authors would like to thank Andreas Pricic and the Sonnhammer lab for valuable assistance and insightful discussion.

Conflict of Interest: none declared.

REFERENCES
Pricic,A. et al. (2005) Adding some SPICE to DAS. Bioinformatics, 21 (Suppl. 2), ii51-ii54.