

Erin R. Butterfield¹, James C. Abbott² & Mark C. Field^{3,4}

1: Wellcome Centre for Anti-infectives Research, School of Life Sciences, University of Dundee, UK, 2: Data Analysis Group, Division of Computational Biology, School of Life Sciences, University of Dundee, UK, 2: Data Analysis Group, Division of Computational Biology, School of Life Sciences, University of Dundee, UK, 2: Data Analysis Group, Division of Computational Biology, School of Life Sciences, University of Dundee, UK, 2: Data Analysis Group, Division of Computational Biology, School of Life Sciences, University of Dundee, UK, 2: Data Analysis Group, Division of Computational Biology, School of Life Sciences, University of Dundee, UK, 2: Data Analysis Group, Division of Computational Biology, School of Life Sciences, University of Dundee, UK, 2: Data Analysis Group, Division of Computational Biology, School of Life Sciences, University of Dundee, UK, 2: Data Analysis Group, Division of Computational Biology, School of Life Sciences, University of Dundee, UK, 2: Data Analysis Group, Division of Computational Biology, School of Life Sciences, University of Dundee, UK, 2: Data Analysis Group, Division of Computational Biology, School of Life Sciences, University of Dundee, UK, 2: Data Analysis Group, Division of Computational Biology, School of Life Sciences, UNIVERSITY of Dundee, UK, 2: Data Analysis Group, Division of Computational Biology, School of Life Sciences, UNIVERSITY of Dundee, UK, 2: Data Analysis Group, Division of Computational Biology, School of Life Sciences, UNIVERSITY of Dundee, UK, 2: Data Analysis Group, Division of Computational Biology, School of Life Sciences, UNIVERSITY of Dundee, UK, 2: Data Analysis Group, Division of Computational Biology, School of Life Sciences, UNIVERSITY of Dundee, UK, 2: Data Analysis Group, Division of Computational Biology, School of Life Sciences, UK, 2: Data Analysis Group, Division of Computational Biology, School of Life Sciences, UNIVERSITY of Dundee, Division of Computational Biology, School of Life Sciences, UK, 2: Data Analysis Group, Division of Co Dundee, UK, 3: School of Life Sciences, University of Dundee, Dundee, UK, 4: Biology Centre, Institute of Parasitology, Faculty of Sciences, University of South Bohemia, České Budějovice, Czech Republic

Comparative genomics

Identifying orthologs - genes sharing an ancestor in different organisms¹ - is vital to WCAIR research from assigning functions to unknown proteins, and identifying best candidates for crystallisation, to finding opportunities for pathogen hopping and identifying potential compound off-target effects. Currently there are many tools available for ortholog identification but all have various limitations including organism availability, scalability, command-line knowledge and issues with divergence leading researchers to



use manual best-reciprocal BLAST, a highly time consuming and unscalable method. To combat this, we have created the WOH, a website which automates manual best reciprocal BLAST drastically reducing user time from months to minutes.

Figure 1: Orthologs have the same ancestor before speciation



Figure 2: batch_brb workflow

We previously developed batch_brb², a command-line tool for rapid ortholog identification and validation using automated best reciprocal BLAST³ and FastTree⁴. Automation of the data collection process drastically reduces user time, while increasing the scalability and reproducibility of the analysis. The WOH provides a graphical interface for batch_brb.

WOH website



Data management



data, tasks and projects and download source



Figure 3: CompGen landing page

Help section



Figure 5: Help section of the WOH

A: The WOH contains extensive help documentation including help guides covering all pages of the website as well as a step-by-step tutorial detailing how to obtain the data needed by the site and performing a full comparative genomics analysis using the site and finally, a link to the batch_brb manuscript, the software the site runs also detailing how to analyse a multiple sequence alignment and phylogenetic tree - the outputs of the WOH. B: Example help guide.

Users can upload and create their own custom organism databases. Identify orthologs using automated best reciprocal BLAST and validate their predictions through FastTree phylogenetic analysis.

Site availability

- Site address: <u>https://uod.ac.uk/woh</u>
- Available to members of WCAIR Accessible on the University network or the VPN

References: 1; Stamboulian et al. 2020, Bioinformatics, 36(Supplement_1):i219-i226, 2; Butterfield et al. 2021, In: de Pablos & Sotillo eds, Parasite Genomics. Methods in Molecular Biology, NY, Humana, 41-63, 3; Altschul et al. 1997, Nucleic Acids Res, 25:3389-3402, 4; Price et al. 2010, PLoS One, 5(3):e9490. Acknowledgements: We thank all involved in the project including the Data Protection Team, current and previous Barton Group members, Tim Butterfield, Michele Tinti, Frederik Droest and all our beta testers.