ProfileGrids solve the large alignment visualization problem: influenza hemagglutinin example [v1; ref status: approved with reservations 3, <http://f1000r.es/10XU1h2>]

Alberto I Roca¹, Aaron C Abajian¹ and David J Vigerust²

¹ ProfileGrid.org, P.O. Box 6414, Irvine, CA 92612, USA

² Department of Veteran Affairs, and Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN 37212, USA

Correspondence to Alberto Roca: aroca@profilegrid.org

v1 First published: 4 Jan 2013, 2:2 (doi: 10.3410/f1000research.2-2.v1)

© Usage Licensed by Creative Commons CC-BY 3.0

How to cite this article: Roca A Abajian AC, Vigerust DJ (2013) ProfileGrids solve the large alignment visualization problem: influenza hemagglutinin example [v1; ref status: approved with reservations 3, <http://f1000r.es/10XU1h2>] *F1000Research* 2:2 (doi: 10.3410/f1000research.2-2.v1)

Abstract Large multiple sequence alignments are a challenge for current visualization programs. ProfileGrids are a solution that reduces alignments to a matrix, color-shaded according to the residue frequency at each column position. ProfileGrids are not limited by the number of sequences and so solves this visualization problem. We demonstrate the new metadata searching and grep filtering features of the JProfileGrid version 2.0 software on an alignment of 11,900 hemagglutinin protein sequences. JProfileGrid is free and available from <http://www.ProfileGrid.org>.

Introduction

The explosion in biological sequence information has led to the generation of large multiple sequence alignments (MSA). For example, the biggest protein family alignment currently in the Pfam database (Wellcome Trust-Sanger Institute) has over 288,000 sequences¹. A new generation of alignment programs, such as Clustal Omega², are available that allow the routine calculation of such large alignments. However, a *Nature Methods* review³ noted the lack of software tools for visualizing the results of large alignment calculations. Specifically, there was a call for overcoming the conceptual and technical limitations of large data sets to allow one to navigate visually both an overview and the details of an alignment, while having mechanisms to query annotated data. We point out that this conceptual limitation was solved in late 2008 by the introduction of ProfileGrids as a new paradigm for visualizing large multiple sequence alignments⁴. Here, we report that the remaining technical limitations have been overcome with version 2.0 of the JProfileGrid software, and that therefore, the large alignment visualization problem has now been solved. We use the influenza hemagglutinin protein family as a case study to demonstrate the new features of the software.

Previous MSA visualization paradigms^{3,5} do not provide both alignment overviews together with the details of each character's frequency distribution at each homologous position in the entire alignment. A particular technical limitation of stacked sequence representations is that as the alignment size grows, a printed or digital visualization can run out of convenient space. This compounds the conceptual limitation where the user cannot grasp the overall conservation trends and the observed variation details in a large data set. The ProfileGrid paradigm was introduced as a solution to this conceptual problem by converting a multiple sequence alignment to a color-coded matrix of the residue frequency occurring at every homologous position across the entire length of a MSA. Therefore, all MSA information is represented both at variable regions and of infrequent residues that may yield clues about biological function.

Improvements to JProfileGrid software

Comparisons of the stacked sequence and ProfileGrid visualization paradigms highlight the challenges of large MSAs (Fig. 1). We downloaded 11,981 full-length, non-redundant hemagglutinin protein sequences from the NCBI influenza virus resource⁶,

Figure 1 Stacked sequence versus ProfileGrid visualization paradigms. The alignment of hemagglutinin sequences is compared side-by-side using the new JProfileGrid 2.0 “overview” feature for both stacked sequence (a) and ProfileGrid (c) visualization methods. An example detailed view of a stacked sequence alignment shows only 12 sequences (b) with the one-letter amino acid codes colored according to the Taylor scheme. While space limitations restrict the stacked sequence overview to only 600 sequences, the ProfileGrid overview represents the entire residue content of all 11,981 sequences. See text for details.

aligned them using MUSCLE⁷, and visualized the MSA⁸ with the new JProfileGrid 2.0 software. Both paradigms depict the entire width of the alignment, *i.e.*, 650 columns from the approximately 570 protein residues and the 80 inserted gap characters. However, due to space limitations, the stacked sequence overview (Fig. 1a) is only a sampling of 600 sequences from the whole alignment. In the stacked representation, each row is an individual protein sequence with the amino acids represented as pixels colored according to the Taylor scheme⁹. Below the stacked sequence overview is a similarly colored magnified view (Fig. 1b) from a much smaller alignment with the amino acid one-letter codes shown. By contrast, the ProfileGrid overview representation (Fig. 1c) divides the alignment width into multiple “tiers” of which six are 100 characters wide. Within each tier are 21 rows for the 20 amino acids and a gap character. Each cell in the matrix is a color shaded count of the character occurrence at the corresponding MSA column position from low (white) to high (dark blue) frequency. Thus, all of the sequences from the entire alignment are represented in the ProfileGrid by the

frequency color shading of each cell in the matrix. As a reference, the left-most column of each ProfileGrid tier is colored according to the Taylor scheme for each amino acid row. ProfileGrids solve the visualization problem of large alignments since there is no limit to the number of sequences that can be represented. Note that a stacked sequence representation of all 11,981 hemagglutinin sequences in this example would be twenty times larger than the number of rows shown in the overview (Fig. 1a).

The first version of the Java JProfileGrid software was designed for alignments with hundreds of sequences such as the ubiquitous RecA protein involved in bacterial recombination⁴. Upon analyzing virus protein families such as influenza hemagglutinins with thousands of sequences, it became apparent that the larger data sets were taxing certain software technical limits. We completely overhauled the program to improve the code with respect to object oriented design, memory management, calculation efficiency, and speed. We strategically reduced memory usage and introduced parallelized code for computing amino

acid frequency counts. For example, we were able to reduce the memory requirements by addressing a technical limitation of the Java programming language. Java uses 2 bytes (16-bits) of memory for every Unicode character (UTF-16) reflecting the need to support thousands of characters from hundreds of languages. However, typical MSAs consist of only ASCII characters. Thus, we were able to reduce the memory use 2-fold by introducing a byte (8-bit) map between the integers 0 to 20 and the twenty common amino acid codes and a gap character. Parallelization was possible due to the nature of the ProfileGrid format. Amino acid counts within each column are independent of each other and so were partitioned into separately running threads thereby taking advantage of multi-core processor environments for enhanced ProfileGrid calculation speed.

New features in JProfileGrid v2.0 accommodate viewing and analyzing very large MSAs. The software now allows sorting and searching the menu list of sequence names for finding a specific homolog to serve as the reference sequence for the ProfileGrid. The new “overview” modes enable the user to visualize the entire MSA within one window. The detailed ProfileGrid window with the character counts has a new second pane that facilitates simultaneous viewing of different parts of the MSA. To easily

focus on rare residues, the “highlight” functionality now identifies residues that occur greater or less than a user-defined threshold of residue frequency. We introduced data sampling to accelerate similarity plot calculations rather than needlessly including every single sequence from large MSAs. Regular expression functionality has also been implemented to find sequences with particular names or amino acid patterns. Finally, JProfileGrid can import simple sequence annotations from flat file spreadsheet databases¹⁰ for metadata filtering to reduce large MSAs to subsets of interest.

Example: influenza hemagglutinin

As a demonstration of the new metadata (Fig. 2a) and regular expression (Fig. 2b) software features, the 11,981 sequence alignment (Fig. 2c) was filtered to just 60 hemagglutinin homologs (Fig. 2d) by using a metadata search for “human” hosts and a regular expression search for a “Mexico” country location. The ProfileGrid views are positioned at alignment column 333 where there is a potential glycosylation site (asn-thr-thr-cys underlined in red) that is found among this sequence subset. Glycosylation is a key post-translational modification that is vital to the proper folding and trafficking of viral proteins. In addition to the role in protein folding and processing, glycosylation in influenza is also important to virus immune evasion

Figure 2 Screenshot of the JProfileGrid 2.0 interactive viewer and search features. The metadata (a) and regular expression filters (b) allow the entire sequence alignment (c) to be filtered to a select subset of sequences (d). Note that in panel (d), the amino acids are colored and sorted according to the Taylor classification scheme. A glycosylation site (red underline; MSA column position 333) is located within bioinformatic element variable_5a. See text for details.

strategy¹¹. The majority of antibody recognition is dedicated to the surface antigen on the globular head although stem-directed antibodies have recently been described¹². Glycosylation in the stem region could prevent recognition of these antibodies and allow for virus to escape the immune response. The region visualized by ProfileGrid analysis lies in the stem region of the hemagglutinin molecule. ProfileGrids, therefore, permit observing rare natural sequence variation (Fig. 2d) within the context of an entire multiple sequence alignment (Fig. 2c). ProfileGrids have this unique advantage over other compressed alignment visualization methods such as sequence logos¹³.

In conclusion, ProfileGrids have solved the problem of visualizing very large alignments. As sequence data sets grow, both for the end user and in central database repositories, we anticipate that ProfileGrids will simplify the dissection and analysis of MSAs. Parenthetically, we note that some bioinformatic studies about influenza proteins have lacked figures depicting the alignment upon which the analyses are based¹⁴. This omission was probably due to technical limitations that now no longer exist. Thus, we propose that ProfileGrid representations of alignment datasets be included as part of publication to assist science communication. This may initiate establishing a future standard for MIMSA: Minimum Information about a Multiple Sequence Alignment¹⁵. Likewise, database curators can use JProfileGrid and its new PNG image output to include ProfileGrid visualizations with protein family descriptions and search results. JProfileGrid v2.0 is available under a GNU General Public License and can be downloaded from <http://www.ProfileGrid.org>.

Multiple sequence alignment of influenza hemagglutinin protein sequences

1 Data File

<http://dx.doi.org/10.6084/m9.figshare.94031>

Meta data for multiple sequence alignment of influenza hemagglutinin protein sequences

1 Data File

<http://dx.doi.org/10.6084/m9.figshare.94032>

Author contributions

AIR and DJV conceived the study. AIR designed the software, collected hemagglutinin sequences, performed the bioinformatic analysis and biocuration, and prepared the first draft of the manuscript. ACA wrote Java code and contributed to writing the manuscript. DJV interpreted the bioinformatic observations and contributed to writing the manuscript. AIR and ACA contributed equally to the work. All authors were involved in the revision of the draft manuscript and have agreed to the final content.

Competing interests

We have developed the software but have no competing interests.

Grant information

AIR was supported by the Erasmo Foundation (grant TSC13702); and DJV was supported by a Veterans Administration Career Development Grant. This material is based upon work supported in part by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development with resources and the use of facilities at the VA Tennessee Valley Healthcare System.

Acknowledgements

We thank David Talalayevsky and Andy Pitt for programming support; Tamara Munzner and Adam Steinberg for visualization design suggestions; and the ISMB KillerApp Award, VIZ-BI, and SACNAS Postdoc committees for helpful discussions.

References

1. Punta, M., Coggill, P.C., Eberhardt, R.Y., Misty, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J. *et al.* (2012) *The Pfam protein families database*. *Nucleic Acids Research*, **40**, D290–301.
2. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, **7**, 539.
3. Procter, J.B., Thompson, J., Letunic, I., Creevey, C., Jossinet, F. and Barton, G.J. (2010) Visualization of multiple alignments, phylogenies and gene family evolution. *Nature Methods*, **7**, S16–25.
4. Roca, A.I., Almada, A.E. and Abajian, A.C. (2008) ProfileGrids as a new visual representation of large multiple sequence alignments: a case study of the RecA protein family. *BMC Bioinformatics*, **9**, 554.
5. Puntervoll, P. and Aasland, R. (2005) In Cesareni, G., Gimona, M., Sudol, M. and Yaffe, M. (eds.), *Modular Protein Domains*. Wiley-VCH, Weinheim, Germany, pp. 477–486.
6. Bao, Y., Bolotov, P., Dernovoy, D., Kiryutin, B., Zaslavsky, L., Tatusova, T., Ostell, J. and Lipman, D.J. (2008) The influenza virus resource at the National Center for Biotechnology Information. *Journal of Virology*, **82**, 596–601.
7. Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.
8. Roca, A.I. (2012) Multiple sequence alignment of influenza hemagglutinin protein sequences. *figshare*, <http://dx.doi.org/10.6084/m9.figshare.94031>.
9. Taylor, W.R. (1997) Residual colours: a proposal for aminochromography. *Protein Engineering*, **10**, 743–746.

10. Roca, A.I. (2012) Meta data for multiple sequence alignment of influenza hemagglutinin protein sequences. *figshare*, <http://dx.doi.org/10.6084/m9.figshare.94032>.
11. Vigerust, D.J. and Shepherd, V.L. (2007) Virus glycosylation: role in virulence and immune interactions. *TRENDS in Microbiology*, **15**, 211–218.
12. Fleishman, S.J., Whitehead, T.A., Ekiert, D.C., Dreyfus, C., Corn, J.E., Strauch, E.M., Wilson, I.A. and Baker, D. (2011) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science*, **332**, 816–821.
13. Schneider, T.D. and Stephens, R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Research*, **18**, 6097–6100.
14. Amaro, R.E., Swift, R.V., Votapka, L., Li, W.W., Walker, R.C. and Bush, R.M. (2011) Mechanism of 150-cavity formation in influenza neuraminidase. *Nature Communications*, **2**, 388.
15. Taylor, C.F., Field, D., Sansone, S.A., Aerts, J., Apweiler, R., Ashburner, M., Ball, C.A., Binz, P.A., Bogue, M., Booth, T. *et al.* (2008) Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. *Nature Biotechnology*, **26**, 889–896.

Current Referee Status:

Referee Responses for Version 1

Christos Ouzounis, Computational Genomics Unit and Institute of Agrobiotechnology, National Center for Research & Technology, Thessalonica, Greece

Approved with reservations: 29 January 2013

Ref Report: 29 January 2013

Computing and visualizing multiple sequence alignments might have been considered a solved issue for bioinformatics a few years ago, until the flood of sequence data we have been experiencing recently. In that respect, efforts to address the issue on another scale altogether are indeed necessary. This work reports on a newer version of ProfileGrid that attempts to solve the visualization of very large alignments and a demonstration of its capabilities with the influenza hemagglutinin as an example.

JProfileGrid 2.0 appears to perform well, captures some elements of large alignments and provides a useful interface for the exploration of large homologous sequence data-sets. Offering a number of options to users, e.g. color schemes and frequency diagrams might indeed be helpful, but the claim of 'solving' the visualization issues with these data is a bit of an over-statement. For funding agencies and beginners in the field, that might have implications such as not appreciating the importance of this line of research.

Beyond this extraordinary claim, the work is fine and provides sufficient detail for interested readers.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Jaap Heringa, Netherlands Bioinformatics Centre (NBIC), Amsterdam, Netherlands

Approved with reservations: 29 January 2013

Ref Report: 29 January 2013

Visualising multiple sequence alignments (MSA) is of great importance, and has only become more essential in the wake of next generation sequencing (NGS) projects. The authors report on a new improved version of their software called JProfileGrid 2.0, originally published in [BMC Bioinformatics in 2008](#). In their new implementation, the authors add the possibility to view very large alignments by means of a profile abstracted from the alignment, such that the number of sequences (typically running into the hundreds or even thousands in the NGS era) is compressed to just 21 rows in a protein MSA, corresponding to the 20 amino acid types and a gap character.

However, visualising a profile is an entirely different thing than visualising an alignment. The authors' claim, therefore, that they have 'solved' the alignment visualisation problem is unwarranted. They should adapt their title and main conclusion to reflect this fact.

Profiles are an abstraction of an MSA, but losing quite a bit of information: for example, correlations between alignment positions, or specific subgroups discernible in an MSA are not visible in a profile. The authors could have opted for alternative ways to compress MSAs containing many sequences, for instance, by clustering the sequences based on sequence similarity, and selecting a representative sequence for each cluster group. The number of clusters could then be selected by the user, or even set automatically based upon optimality criteria concerning the clustering.

Many techniques exist already for visualising profiles, such as the widely used SequenceLogos and related representations. As such, just representing the raw amino acid frequencies using a color coding scheme can hardly be considered novel. The authors could make their profile visualization more useful by using other frequency-derived schemes, such as the aforementioned SequenceLogo (entropy-based), or the log likelihood used in PSSMs calculated by BLAST. On a more general note, the authors should cite related work and software by other researchers in the field.

Summarising, the authors have compiled a useful software package to visualise alignment profiles, but have not solved the MSA visualisation problem for large sequence sets. Further, they could enhance the usability of the software by implementing more profile scoring schemes as indicated above.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Alex Bateman, EMBL European Bioinformatics Institute, Hinxton, Cambridge, UK

Approved with reservations: 04 January 2013

Ref Report: 04 January 2013

The authors have selected an important problem to address which is how to visualise very large multiple sequence alignments. The manuscript describes a new improved version of their software called JProfileGrid originally published in [BMC Bioinformatics in 2008](#). The software ran on a Stockholm formatted alignment I selected from Pfam and so appears to work largely as advertised.

My main critique of the work is the claim that ProfileGrids solve the large alignment visualisation problem as stated in the title. In my opinion the visualisation given does not solve the problem. The problem is recast as a visualisation of a profile. Profiles give the frequency of amino acids at each position calculated from a multiple sequence alignment and so their size is independent of the number of sequences in the alignment. So visualising the profile is one way of gaining an overview of a multiple sequence alignment. But others have already done this such as [HMM-logos](#) and [SUPERFAMILY profile visualisation](#).

Both of these visualisations are more intuitive than the ProfileGrid because it is easier to map these visualisations to a real sequence. However, none of these are really alignment visualisations. They all lose a lot of information that is implicit in the alignment. For example, the correlations between amino acid positions in subfamilies. The JProfileGrid software does provide a compact stacked alignment representation, but that is also possible in other alignment viewers by choosing a sufficiently small font.

Overall I think that this is a well engineered update of an existing software package that has some utility. The authors need to remove the claims that this software solves the large alignment visualisation problem. This is in my opinion is not justified. In particular the current title is not acceptable. The text of the paper also needs to include mention that other profile visualisation software exists.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Open Peer Review

Current Peer Review Status: ? ? ?

Version 1

Reviewer Report 29 January 2013

<https://doi.org/10.5256/f1000research.306.r731>

© 2013 Ouzounis C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Christos Ouzounis

Computational Genomics Unit and Institute of Agrobiotechnology, National Center for Research & Technology, Thessalonica, Greece

Computing and visualizing multiple sequence alignments might have been considered a solved issue for bioinformatics a few years ago, until the flood of sequence data we have been experiencing recently. In that respect, efforts to address the issue on another scale altogether are indeed necessary. This work reports on a newer version of ProfileGrid that attempts to solve the visualization of very large alignments and a demonstration of its capabilities with the influenza hemagglutinin as an example.

JProfileGrid 2.0 appears to perform well, captures some elements of large alignments and provides a useful interface for the exploration of large homologous sequence data-sets. Offering a number of options to users, e.g. color schemes and frequency diagrams might indeed be helpful, but the claim of 'solving' the visualization issues with these data is a bit of an over-statement. For funding agencies and beginners in the field, that might have implications such as not appreciating the importance of this line of research.

Beyond this extraordinary claim, the work is fine and provides sufficient detail for interested readers.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 29 January 2013

<https://doi.org/10.5256/f1000research.306.r740>

© 2013 Heringa J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Jaap Heringa

Netherlands Bioinformatics Centre (NBIC), Amsterdam, The Netherlands

Visualising multiple sequence alignments (MSA) is of great importance, and has only become more essential in the wake of next generation sequencing (NGS) projects. The authors report on a new improved version of their software called JProfileGrid 2.0, originally published in [BMC Bioinformatics in 2008](#). In their new implementation, the authors add the possibility to view very large alignments by means of a profile abstracted from the alignment, such that the number of sequences (typically running into the hundreds or even thousands in the NGS era) is compressed to just 21 rows in a protein MSA, corresponding to the 20 amino acid types and a gap character.

However, visualising a profile is an entirely different thing than visualising an alignment. The authors' claim, therefore, that they have 'solved' the alignment visualisation problem is unwarranted. They should adapt their title and main conclusion to reflect this fact.

Profiles are an abstraction of an MSA, but losing quite a bit of information: for example, correlations between alignment positions, or specific subgroups discernible in an MSA are not visible in a profile. The authors could have opted for alternative ways to compress MSAs containing many sequences, for instance, by clustering the sequences based on sequence similarity, and selecting a representative sequence for each cluster group. The number of clusters could then be selected by the user, or even set automatically based upon optimality criteria concerning the clustering.

Many techniques exist already for visualising profiles, such as the widely used SequenceLogos and related representations. As such, just representing the raw amino acid frequencies using a color coding scheme can hardly be considered novel. The authors could make their profile visualization more useful by using other frequency-derived schemes, such as the aforementioned SequenceLogo (entropy-based), or the log likelihood used in PSSMs calculated by BLAST. On a more general note, the authors should cite related work and software by other researchers in the field.

Summarising, the authors have compiled a useful software package to visualise alignment profiles, but have not solved the MSA visualisation problem for large sequence sets. Further, they could enhance the usability of the software by implementing more profile scoring schemes as indicated above.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 04 January 2013

<https://doi.org/10.5256/f1000research.306.r465>

© 2013 Bateman A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Alex Bateman

EMBL European Bioinformatics Institute, Hinxton, Cambridge, UK

The authors have selected an important problem to address which is how to visualise very large multiple sequence alignments. The manuscript describes a new improved version of their software called JProfileGrid originally published in [BMC Bioinformatics in 2008](#). The software ran on a Stockholm formatted alignment I selected from Pfam and so appears to work largely as advertised.

My main critique of the work is the claim that ProfileGrids solve the large alignment visualisation problem as stated in the title. In my opinion the visualisation given does not solve the problem. The problem is recast as a visualisation of a profile. Profiles give the frequency of amino acids at each position calculated from a multiple sequence alignment and so their size is independent of the number of sequences in the alignment. So visualising the profile is one way of gaining an overview of a multiple sequence alignment. But others have already done this such as [HMM-logos](#) and [SUPERFAMILY profile visualisation](#).

Both of these visualisations are more intuitive than the ProfileGrid because it is easier to map these visualisations to a real sequence. However, none of these are really alignment visualisations. They all lose a lot of information that is implicit in the alignment. For example, the correlations between amino acid positions in subfamilies. The JProfileGrid software does provide a compact stacked alignment representation, but that is also possible in other alignment viewers by choosing a sufficiently small font.

Overall I think that this is a well engineered update of an existing software package that has some utility. The authors need to remove the claims that this software solves the large alignment visualisation problem. This is in my opinion is not justified. In particular the current title is not acceptable. The text of the paper also needs to include mention that other profile visualisation software exists.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research