Check for updates

SOFTWARE TOOL ARTICLE

REVISED FastQ Screen: A tool for multi-genome mapping and quality control [version 2; referees: 4 approved]

Steven W. Wingett, Simon Andrews

Bioinformatics, Babraham Institute, Cambridge, CB22 3AT, UK

v2 **First published:** 24 Aug 2018, **7**:1338 (doi: 10.12688/f1000research.15931.1)
Latest published: 17 Sep 2018, **7**:1338 (doi: 10.12688/f1000research.15931.2)

Abstract

DNA sequencing analysis typically involves mapping reads to just one reference genome. Mapping against multiple genomes is necessary, however, when the genome of origin requires confirmation. Mapping against multiple genomes is also advisable for detecting contamination or for identifying sample swaps which, if left undetected, may lead to incorrect experimental conclusions. Consequently, we present FastQ Screen, a tool to validate the origin of DNA samples by quantifying the proportion of reads that map to a panel of reference genomes. FastQ Screen is intended to be used routinely as a quality control measure and for analysing samples in which the origin of the DNA is uncertain or has multiple sources.

Keywords

Bioinformatics Contamination FastQC Illumina Metagenomics NGS QC Sequencing

Open Peer Review

Referee Status: ✓ ✓ ✓ ✓

	Invited Referees			
	1	**2**	**3**	**4**
REVISED **version 2** published 17 Sep 2018			✓ report	
version 1 published 24 Aug 2018	✓ report	✓ report	? report	✓ report

1 **Russell S. Hamilton**, University of Cambridge, UK

2 **Ian J. Donaldson**, University of Manchester, UK

3 **Stéphane Le Crom**, Univ Antilles, Univ Nice Sophia Antipolis, France
UMS Omique, Plateforme Post-génomique de la Pitié-Salpêtrière, France
Ecole normale supérieure, France
Laurent Jourdren, Ecole normale supérieure, France

4 **Matthew D. Teasdale**, University of York, UK

Discuss this article

Comments (0)

Corresponding author: Steven W. Wingett (steven.wingett@babraham.ac.uk)

Author roles: Wingett SW: Conceptualization, Software, Writing – Original Draft Preparation, Writing – Review & Editing; **Andrews S**: Conceptualization, Funding Acquisition, Software, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the the Medical Research Council (G0801156) and the Biotechnology and Biological Sciences Research Council of the UK (BBS/E/B/000C05).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2018 Wingett SW and Andrews S. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Wingett SW and Andrews S. **FastQ Screen: A tool for multi-genome mapping and quality control [version 2; referees: 4 approved]** *F1000Research* 2018, **7**:1338 (doi: 10.12688/f1000research.15931.2)

First published: 24 Aug 2018, **7**:1338 (doi: 10.12688/f1000research.15931.1)

REVISED **Amendments from Version 1**

Corrected typographical errors in the Figure 1 caption text.

See referee reports

Introduction

In general, reaching sound conclusions from sequencing experiments requires the origin of a sample to be identified correctly prior to mapping. To reduce the risk of contaminants leading to incorrect inferences, it is advisable to map sequencing results against not only the expected reference genome but also against reasonable sources of contamination. Common reasons for contamination include amplifying the wrong target molecule, unwanted DNA being present in reagents used in library generation, carry-over from samples previously loaded onto a sequencing machine or sample swaps.

The tool utilises either Bowtie[1], Bowtie 2[2] or BWA[3], as preferred by the user, to map reads against pre-specified genomes. FastQ Screen presents the mapping results in both text and graphical formats, thereby allowing the user to confirm the genomic origin of a sample or identify sources of DNA contamination. The tool summarises the proportion of reads that map to a single genome or to multiple genomes. In addition, it reports whether those alignments are to a unique position, or to more than one location, within the genome of interest (Figure 1).

FastQ Screen functionality is generally independent of the laboratory protocol followed and so can be used to analyse genomic DNA, RNA-Seq[4], ChIP-Seq or Hi-C experiments. In addition, FastQ Screen is compatible with Bismark[5], and so can also be used to process bisulfite sequence data.

Other tools exist with similar functionality to FastQ Screen, most notably Multi Genome Alignment (MGA)[6]. FastQ Screen has a number of advantages over these tools, including directly reporting the proportion of multi-mapping reads, thereby helping identify DNA populations rich in low-complexity sequences. Another benefit of our program is the capability to create filtered FASTQ files. FastQ Screen is also the only quality control (QC) tool that aligns reads to multiple bisulfite reference genomes.

Methods

Implementation

The program utilises a short read sequence aligner to map FASTQ reads against pre-defined reference genomes. The tool records against which genome or genomes each read maps and summarises the results in graphical and text formats.

Operation

We coded FastQ Screen in Perl and made use of the CPAN module GD::Graph for the generation of summary bar plots. The software requires a functional version of Bowtie, Bowtie 2 or BWA, and should be run on a Linux-based operating system.

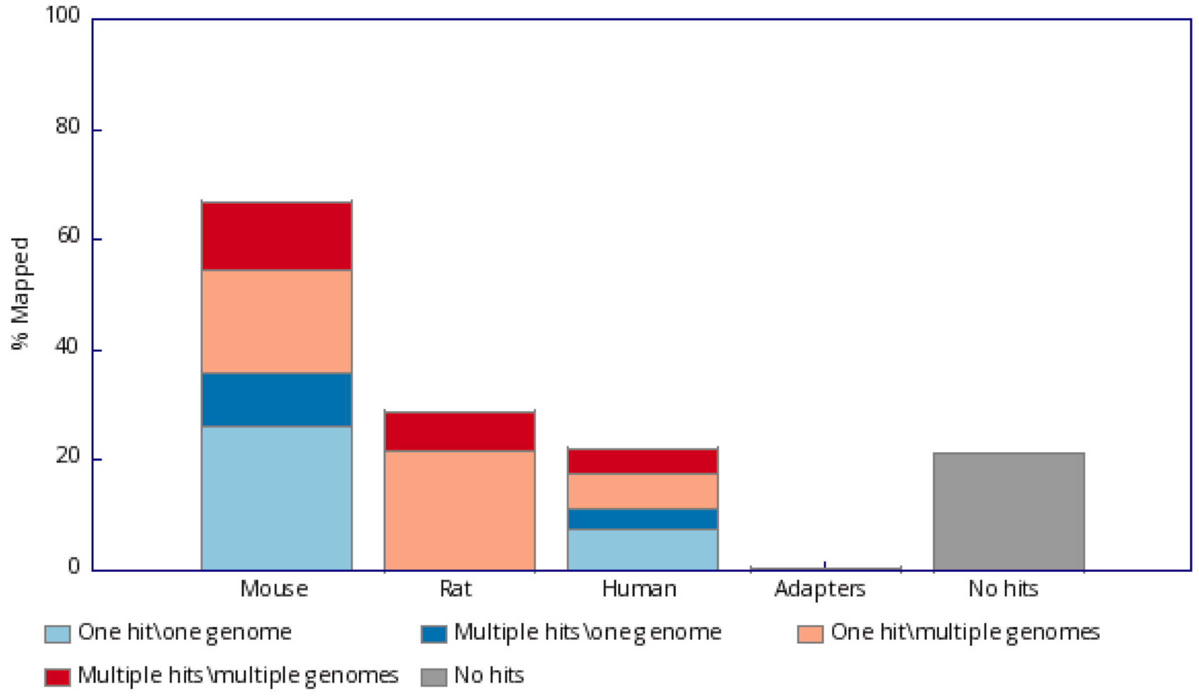

Figure 1. Graphical output from FastQ Screen after mapping a publicly available RNA-Seq sample (SRR5100711) against several reference genomes. Reads either i) mapped uniquely to one genome only (light blue), ii) multi-mapped to one genome only (dark blue), iii) mapped uniquely to a given genome and mapped to at least one other genome (light red), or iv) multi-mapped to a given genome and mapped to at least one other genome (dark red). The reads represented by blue shading are significant since these are sequences that align only to one genome, and consequently, if are observed in an unexpected genome they suggest contamination.

FASTQ Screen uses Plotly to enable visualisation of results in a web browser. The tool takes as input a text configuration file and FASTQ files, which are sub-sampled by default to 100,000 reads to reduce running times, and then mapped to a panel of pre-specified genomes.

Use cases

Preliminary sequencing QC: FastQ Screen provides preliminary evidence on whether a sequencing run has been successful, as demonstrated in Figure 1, which shows results using a publicly available RNA-Seq sample (SRR5100711) labelled as mouse. The software processed the deposited FASTQ file to generate summary results in text, HTML and PNG format. As expected, the dataset contained a substantial proportion of reads that mapped only to the mouse genome, and although a sizeable proportion of reads mapped to both the mouse and rat genomes, that may have also been expected considering the close evolutionary relationship between those two species. Of concern, however, was the discovery that 11.4% of the reads mapped solely to the human genome, suggesting the sample was contaminated. This may prove problematic if human-derived reads that also align to the mouse reference genome are not removed, since differences between mouse samples may then actually reflect the variation in the degree of contamination between the samples rather than genuine biological differences. Very few reads aligned to adapter sequences which was an encouraging observation.

Identifying sample origin from a range of alternatives: FastQ Screen was recently used by researchers to identify the origin of the clothes of the Tyrolean Iceman (popularly named Ötzi), a famous 5,300 year old natural mummy discovered in 1991 in the Italian Ötztal Alps. By screening sequences against probable sources of preserved leathers, the research team showed that the iceman's hat came from Brown Bear, his quiver from Roe deer and his loincloth came from sheep[7]. In a similar fashion, FastQ Screen has been used to determine the animal origin of vellum found in 13th century Bibles[8].

Filtering results: FastQ Screen can also be used to filter reads mapping (or not mapping) to specified genomes. This has numerous applications, most typically to remove DNA contaminants, as exemplified by a recent clinical microbial metagenomics study in which nucleic acids were extracted from porcine faeces[9]. FastQ Screen was then used to filter-out host sequences, and the remaining reads were then mapped, leading to the identification of over 1,600 bacterial and Archaea species and strains of virus.

In contrast, in some experiments the source of contamination may be completely unpredictable and so we have incorporated a setting in which all unsuccessfully mapped reads are written to a FASTQ format output file. This may then be used by other resources, such as BLAST, to determine the origin of those sequences.

Summary

Since its release, FastQ Screen has been used to analyse a myriad of sequencing datasets. We initially envisioned the software as a QC tool to complement our related program FastQC, but we subsequently used the software to confirm the origin of samples and added functionality for filtering FASTQ reads. The program may be used in conjunction with several common aligners, including Bismark for processing bisulfite libraries. FastQ Screen has been incorporated by other groups into bioinformatics workflows, was reimplemented in the recently released QC tool Aozan[10], and is compatible with MultiQC[11], a tool to aid comparison of samples with respect to a large number of QC metrics.

Software availability

FastQ Screen is available from: https://www.bioinformatics.babraham.ac.uk/projects/fastq_screen

Source code available from: https://github.com/StevenWingett/FastQ-Screen

Archived source code as at time of publication: https://doi.org/10.5281/zenodo.1346672[12]

License: GNU GPL 3.0

Grant information

This work was supported by the the Medical Research Council (G0801156) and the Biotechnology and Biological Sciences Research Council of the UK (BBS/E/B/000C05).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

The authors would like to thank Felix Krueger who helped with making FastQ Screen compatible with Bismark and Philip Ewels who gave advice on generating the HTML format summary results. Mikhail Spivakov and Jonathan Cairns both assisted with the manuscript preparation.

References

1. Langmead B, Trapnell C, Pop M, *et al.*: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol.* 2009; **10**(3): R25.
PubMed Abstract | Publisher Full Text | Free Full Text

2. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2.** *Nat Methods.* 2012; **9**(4): 357–359.
PubMed Abstract | Publisher Full Text | Free Full Text

3. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics.* 2009; **25**(14): 1754–1760.
PubMed Abstract | Publisher Full Text | Free Full Text

4. Woodham EF, Paul NR, Tyrrell B, *et al.*: **Coordination by Cdc42 of Actin, Contractility, and Adhesion for Melanoblast Movement in Mouse Skin.** *Curr Biol.* 2017; **27**(5): 624–637.
PubMed Abstract | Publisher Full Text | Free Full Text

5. Krueger F, Andrews SR: **Bismark: a flexible aligner and methylation caller for**

Bisulfite-Seq applications. *Bioinformatics.* 2011; **27**(11): 1571–1572.
PubMed Abstract | Publisher Full Text | Free Full Text

6. Hadfield J, Eldridge MD: **Multi-genome alignment for quality control and contamination screening of next-generation sequencing data.** *Front Genet.* 2014; **5**: 31.
PubMed Abstract | Publisher Full Text | Free Full Text

7. O'Sullivan NJ, Teasdale MD, Mattiangeli V, *et al.*: **A whole mitochondria analysis of the Tyrolean Iceman's leather provides insights into the animal sources of Copper Age clothing.** *Sci Rep.* 2016; **6**: 31279.
PubMed Abstract | Publisher Full Text | Free Full Text

8. Fiddyment S, Holsinger B, Ruzzier C, *et al.*: **Animal origin of 13th-century uterine vellum revealed using noninvasive peptide fingerprinting.** *Proc Natl Acad Sci U S A.* 2015; **112**(49): 15066–15071.
PubMed Abstract | Publisher Full Text | Free Full Text

9. Rose G, Wooldridge DJ, Anscombe C, *et al.*: **Challenges of the Unknown: Clinical Application of Microbial Metagenomics.** *Int J Genomics.* 2015; **2015**: 292950.
PubMed Abstract | Publisher Full Text | Free Full Text

10. Perrin S, Firmo C, Lemoine S, *et al.*: **Aozan: an automated post-sequencing data-processing pipeline.** *Bioinformatics.* 2017; **33**(14): 2212–2213.
PubMed Abstract | Publisher Full Text

11. Ewels P, Magnusson M, Lundin S, *et al.*: **MultiQC: summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics.* 2016; **32**(19): 3047–3048.
PubMed Abstract | Publisher Full Text | Free Full Text

12. Wingett S: **StevenWingett/FastQ-Screen: Released v0.12.2 (Version 0.12.2).** *Zenodo.* 2018.
http://www.doi.org/10.5281/zenodo.1346672

Open Peer Review

Current Referee Status: ✓ ✓ ✓ ✓

Version 2

Referee Report 18 September 2018

doi:10.5256/f1000research.17815.r38403

✓ **Stéphane Le Crom** [1,2,3], **Laurent Jourdren** [3]

[1] Sorbonne Université, Univ Antilles, Univ Nice Sophia Antipolis, Paris, France

[2] Sorbonne Université, UMS Omique, Plateforme Post-génomique de la Pitié-Salpêtrière, Paris, France

[3] Institut de biologie de l'Ecole normale supérieure (IBENS), Ecole normale supérieure, Paris, France

With this second version the authors answer all the concerns we raised about their FastQ Screen article. We will approve it for indexing.

Competing Interests: No competing interests were disclosed.

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Referee Report 17 September 2018

doi:10.5256/f1000research.17398.r37623

✓ **Matthew D. Teasdale**

BioArCh, University of York, York, UK

In this paper Wingett and Andrews describe FastQ Screen a program for quality control and source species identification. The paper is well written with clear example use cases and I am very happy to recommend FastQ Screen for indexing.

I have personally used FastQ Screen for over 6 years and now consider it to be an essential part of my analysis pipelines. The documentation for the program is excellent and it is under active development.

Is the rationale for developing the new software tool clearly explained?

Yes

Is the description of the software tool technically sound?

Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

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.

> Author Response 17 Sep 2018
> **Steven Wingett**, Babraham Institute, UK
>
> We are delighted that FastQ Screen has proven useful in your research and we hope it will remain part of your analysis pipeline as we add new features to the software.
>
> ***Competing Interests:*** No competing interests were disclosed.

Referee Report 06 September 2018

doi:10.5256/f1000research.17398.r37622

?

Stéphane Le Crom [1,2,3], **Laurent Jourdren** [3]

[1] Sorbonne Université, Univ Antilles, Univ Nice Sophia Antipolis, Paris, France

[2] Sorbonne Université, UMS Omique, Plateforme Post-génomique de la Pitié-Salpêtrière, Paris, France

[3] Institut de biologie de l'Ecole normale supérieure (IBENS), Ecole normale supérieure, Paris, France

When dealing with multiple high throughput sequencing experiments, especially for core facilities, you need to pay great attention to quality controls. Contaminations from different species you are working with are one of the potential problems you can encounter. The FastQ Screen software as been designed by Steven Wingett and Simon Andrews in order to solve this drawback. Using different mapping softwares, FastQ Screen allows to identify from the reads present in your samples the different species they came from. The graphical and text outputs provided, detailed information on the potential level of contamination obtained.
The software objectives are clearly explained just as the way to use it and how to interpret its outputs. FastQ Screen source code is available through GitHub and a documentation is provided on the authors' website.

This software is publicly available since several years and is used today by many genomics laboratories. The tool is very stable, the command line help is easy to understand and we not found any issue when we launch it in all our tests. Finally, the output report is informative and very clear.

FastQ Screen is a must have tool for everyone working with multiple species samples or who want to prevent unpredicted contamination of its samples.

Remarks
1. In order to more clearly explain the way FastQ Screen is working, more information should be provided on pre-defined reference genomes. How can it be chosen before running FastQ Screen. Is there a limitation among the number of genomes selected? Is there a significant impact on software running time according to the number or size of reference genomes selected? Is there a list of already available pre-defined reference genomes? Does it works using subset sequence database or with the whole genome?

2. It could also be interesting to get some information about the running time of several analyses. What are the specifications required for the computer needed to run FastQ Screen? I didn't findd this information through the documentation.

3. It seems that FastQ Screen when processing large dataset (FASTQ files more than 100 million reads) use a large amount of memory as it store the identifier of each read. It may be useful to advise users in documentation about this issue.

Minor remarks
1. For future evolution of FastQ Screen software it could be interesting to provide a clue for the "No Hits" reads obtained. In the paper the authors suggest to run BLAST analyses. Perhaps there is a way to use a subset of "NoHits" reads in order to guess where the possible contamination is coming from?

2. The link to the "Babraham Bioinformatics download page" in the documentation "https://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/_build/html/index.html#download" is not working

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

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

Author Response 17 Sep 2018

Steven Wingett, Babraham Institute, UK

Thank you for your detailed feedback. Both you and the reviewer Dr Hamilton pointed out that FastQ Screen would be better served if we made pre-made genome indices available. As described in our response to Dr Hamilton, FastQ Screen now has a --get_genomes option to obtain Bowtie2 genome indices. We have also provided more information in the documentation (https://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/_build/html/index.html), explaining how a user may create the desired aligner index files.

We now also mention in the documentation that FastQ Screen has an upper limit to the number of genomes that may be used. This value, which is 32, is the result of how the script records read/genome mapping data as a 32-bit variable. This limit far exceeds every task we have performed with FastQ Screen. Should more genomes be required at a future date however, we would be able to modify the code to increase the maximum allowed value.

The time taken to process a dataset varies substantially depending on the input data and specified parameters. For example, larger files take longer to process than those with fewer reads. Similarly, screening against more genomes will increase processing times, as will the complexity and size of the genomes to which the aligner maps FASTQ reads. Using FastQ Screen in the default quality control (QC) mode, in which only a subset of the data is processed, is significantly quicker than using the tool to filter a dataset. Screening bisulfite libraries is also a more computationally intensive task and therefore takes significantly longer to complete. In contrast, times may be reduced by muti-threading submitted jobs. The hardware of a system will of course also substantially impact running times, as will the competition for system resources from the jobs being run concurrently with FastQ Screen.

There are similar points to consider when evaluating memory overheads. Most notably, running the program in QC mode will require substantially less memory than filtering a dataset. The software needs to hold in memory whether a read maps to any of the reference genomes. In QC mode, this will only be necessary for approximately 100,000 reads, but when filtering this will be required for every read, simultaneously, in the FASTQ file – and FASTQ files may comprise hundreds of millions of reads.

To help the user make sense of these considerations, we have now included in the documentation a report of the memory requirements and time taken to process different data files, using different parameters. Obviously, it is impossible to cover every scenario, but as a general rule using the tool to QC a dataset should take minutes whereas filtering a large dataset may take several hours.

We added extra information in the documentation pertaining to the system requirements necessary to run FastQ Screen. To summarise these, FastQ Screen should be run on an up-to-date Linux operating system that has Perl installed and has a working version of either Bowtie, Bowtie2 or BWA installed.

So far as we can tell, the following web link is functional: https://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/_build/html/index.html#download . In case there is any confusion, this is an anchor to the “Download” section in the documentation and not a link to the software download page.

Thank you for bringing to our attention the need for follow-on support for reads that map to no genomes (extracted when using the --no_hits parameter). We intend to address this in future FastQ Screen releases by adding new functionality to the software.

Competing Interests: No competing interests were disclosed.

Referee Report 04 September 2018

doi:10.5256/f1000research.17398.r37624

✓ **Ian J. Donaldson**
Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK

FastQ Screen by Wingett and Andrews is a tool to map a sample of sequenced reads against a panel of reference genomes.

The tool is comprehensively documented and is available from the authors' web site and via Github.

I have personally used this tool since 2011 and it is incorporated in the quality control pipeline for our core facility on all sequencing runs. The tool has been used to detect contamination from a panel of commonly used genomes, and to estimate the contribution of sequence from mixed genome samples. It is also used to detect rRNA contamination in RNA-seq protocols. Recently the ability of FastQ Screen to filter individual reads has been incorporated into our single cell analysis pipeline.

The article is clear, well explained, and gives interesting use cases.

<u>Minor correction</u>:
In the legend of Figure1 -
'Reads either i) mappped' should be 'Reads are either i) mapped'
'ii) mapped uniquely' should be 'iii) mapped uniquely'
'(light red) or multi-mapped' should be '(light red), or iv) multi-mapped'

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

Referee Expertise: Genomics

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.

Author Response 17 Sep 2018

Steven Wingett, Babraham Institute, UK

Thank you for your comments and we are pleased that you find FastQ Screen useful in your research. We have updated the manuscript to correct the typographical errors.

Competing Interests: No competing interests were disclosed.

Referee Report 29 August 2018

doi:10.5256/f1000research.17398.r37620

✓ **Russell S. Hamilton**
University of Cambridge, Cambridge, UK

Wingett and Andrews present FastQ Screen for mapping sequencing reads to multiple genomes with the goal of identifying the genome of origin. FastQ Screen is well documented, open source and freely available via GitHub and their own website.

I have been using the software routinely as part of my NGS pipelines for several years and find it to be an invaluable QC tool. Sample mix ups, if from different species, are easily detected. I have also found FastQ Screen useful as a proxy indicator of rRNA contamination in total RNA-Seq, due to rRNAs being similar between related species, they show up as multi-genome aligned reads, thus identifying samples for further processing.

I therefore have no reservations in recommending FastQ Screen for indexing.

Suggestions:

1. Each lab or facility will have their own unique requirements for genomes to screen against, but having a suggested “starter set” of genomes may ease the burden of installation / configuration for first time users. The bowtie website has some pre-made common, bowtie indexed, genomes for convenient download (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml). Or more

comprehensively, Illumina's iGenome project contains a wide range of bowtie/bowtie2/bwa indexed genomes (http://support.illumina.com/sequencing/sequencing_software/igenome.ilmn).

2. One of the most useful features of FastQ Screen is its compatibility with MultiQC, where multiple samples are plotted together for assessment of entire sequencing runs or batches. Mentioning this in the documentation would alert users to this very useful feature.

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

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.

Author Response 17 Sep 2018

Steven Wingett, Babraham Institute, UK

We agree with the excellent suggestion to create pre-built genomes for users. Indeed, the latest version of FastQ Screen (v0.13.0) now has a new option (--get_genomes) which instructs the script to download commonly used pre-built Bowtie2 reference genomes deposited on the Babraham Bioinformatics website. Along with the reference genomes, FastQ Screen also downloads a configuration file, which it subsequently edits to list the full path of the downloaded genomes as stored on the user's machine. This setup should be ready-to-use and the selection of genomes should suit most requirements.

We have updated the documentation to alert users that our tool is compatible with MultiQC.

Competing Interests: No competing interests were disclosed.

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