

WHAT TO TEST/QUESTIONNAIRE

BIOHDF V0.4 ALPHA RELEASE

Thanks for trying out BioHDF and providing feedback! We would appreciate it if you could fill out this questionnaire (or just email derobins@hdfgroup.org) so we can get a better idea of where your needs lie. We don't need you to answer every question – just fill in anything that you had trouble with or that you feel strongly about.

WHAT WE NEED TESTED

Build issues – Did the configure script and makefiles (POSIX) or Visual Studio projects (Windows) work for you or did you have to make changes? Did 'make check' (POSIX) run correctly?

FASTQ and SAM import/export – Do the tools fail or generate incorrect data for your typical data files?

BioHDF API – If you are a programmer (especially in C) please let us know if the API seems useful and intuitive. If you wrap the API calls via SWIG, XS, etc. please let us know of your experiences. We'll be providing official wrappers in the future, but early feedback would be appreciated.

Performance – What sort of compression ratio are you getting? Is the I/O speed adequate? (Keep in mind that if you are using variable-length data storage, your I/O speed and file size will suffer.)

BUILDING THE SOFTWARE

What operating system are you using? If you are using linux, which distribution? 32 or 64 bit?

Did the software compile correctly? If not, please send us a copy of the output and description of your build system. BioHDF should compile quietly with no warnings.

Did you have a difficult time obtaining and installing the HDF5 libraries for your system?

THE BIOHDF TOOLS

Can you use `bioh5g_import_alignments` to import a SAM file? Can you use `bioh5g_export_alignments` to export a SAM file? Does a diff of the original and exported SAM files show any differences?¹

What sort of compression ratios are you getting?

Does a range query on an alignment hits collection return correct data? MAPQ filters? FLAGS bitmasks?

Is there anything confusing about the command line tool options?

Are there any tools or functionality that you would like to see?

Some examples I can think of: A tool that lists all collections in a file (`biohdf_ls`, perhaps), storing pileup.

THE API

If you are a C programmer or have experience in language interoperability (Perl, Python, etc.), we'd appreciate it if you could have a look at our API.

Does the API seem like it would be difficult or easy to use?

Can you follow the BioHDF API calls in one of the export tools? Is anything unclear?

¹ Note that there may be some small differences between the input and output SAM files (case, etc.) that may make diff difficult to use.

Can you spot anything in the API that seems like it would be a pain point for interoperability with higher-level languages?