

## Approaching full production for CRAM

Since our proof-of-principle publication of reference-based sequence read data compression in 2011 (<http://genome.cshlp.org/content/21/5/734>), the EBI has been moving towards a production release of CRAM, a framework technology comprising file format and toolkit in which we combine highly efficient and tunable compression with a data format that is directly available for computational use ([http://www.ebi.ac.uk/ena/about/cram\\_toolkit](http://www.ebi.ac.uk/ena/about/cram_toolkit)).

Having been openly available since its early versions, CRAM has attracted a substantial community interest. Robust review and testing by groups external to EBI has led to substantial usability improvements and has been informative for the development of the most appropriate data reduction models for optimal compression with minimal loss of signal integrity.

Currently in version 0.9 of CRAM, we expect to be in a position to circulate candidate specifications for our full production release, version 1.0, in late August 2012. Depending on feedback, we will move quickly from this point to release. Amongst the features we plan to offer for CRAM version 1.0 will be

- Full support for lossless and lossy compression
- Indexing and streamability for direct computational access
- Support for arbitrary tags and read names
- Integration with Picard and suitability for future GATK integration
- Readiness for deployment in production services\* at EBI, including support for data submissions and documentation on the use of CRAM as an archive format
- Use of CRAM at two major global sequencing centres

CRAM will be released under the fully open Apache 2.0 License and we will document a release schedule that is conservative for format stability and backwards compatibility.

The release will be announced through the `cram-dev` mailing list (see <http://listserv.ebi.ac.uk/mailman/listinfo/cram-dev> to subscribe) and on the ENA news page (<http://www.ebi.ac.uk/ena/about/news>).

\*See [http://www.ebi.ac.uk/ena/about/ena\\_compression\\_policy](http://www.ebi.ac.uk/ena/about/ena_compression_policy) for details of ENA policy on compression of data.

ENA Team, EMBL-EBI 3.8.12