

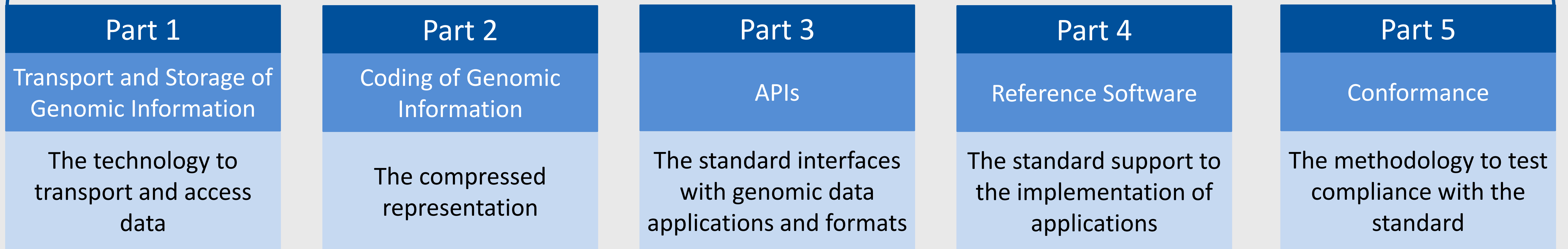
# MPEG-G: The Emerging Standard for Genomic Data

Jan Voges • Jörn Ostermann  
 {voges,office}@tnt.uni-hannover.de

## MPEG-G at a Glance

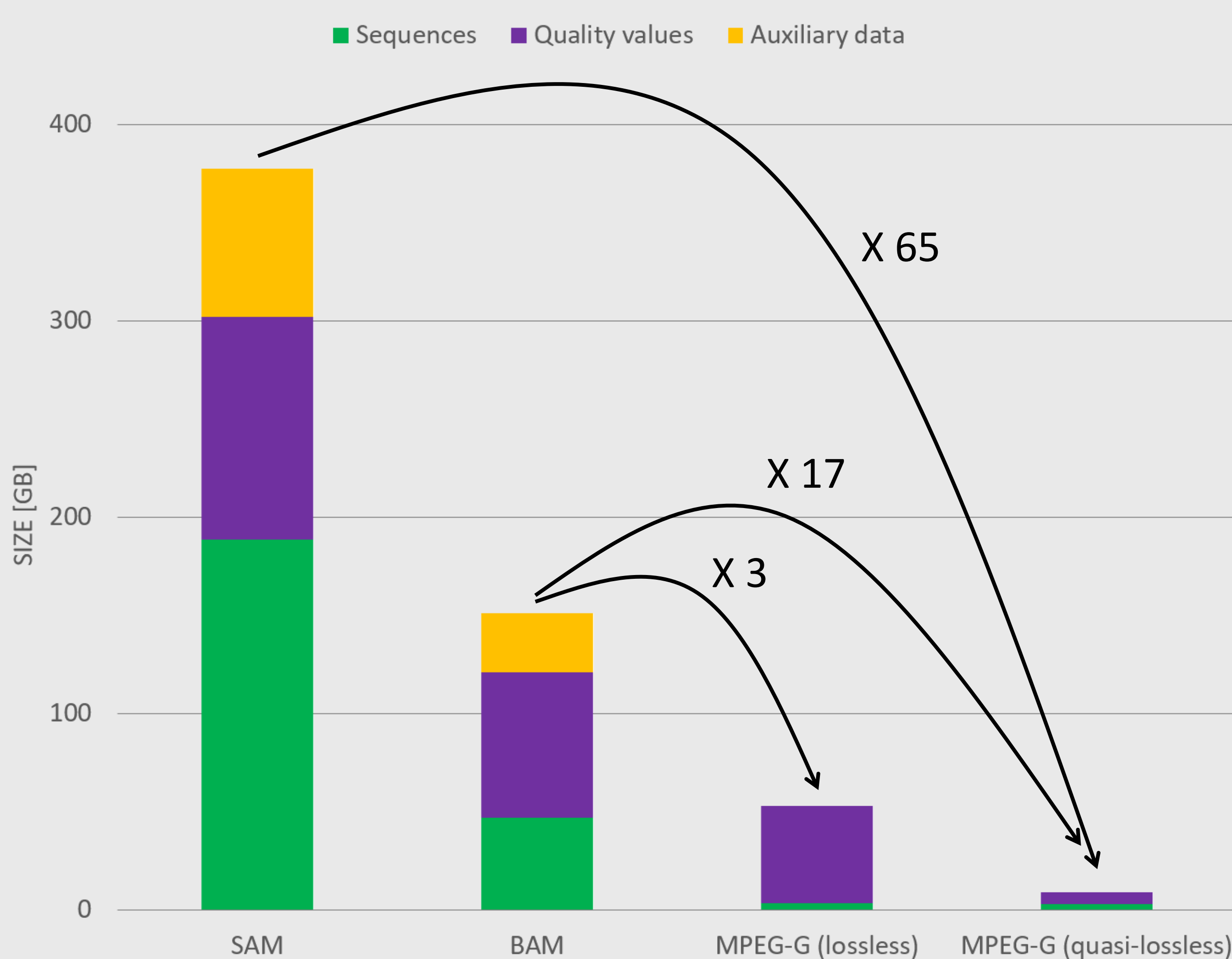
ISO/IEC 23092

Developed by ISO/IEC JTC 1/SC 29/WG 11  
 a.k.a. Moving Picture Experts Group (MPEG)



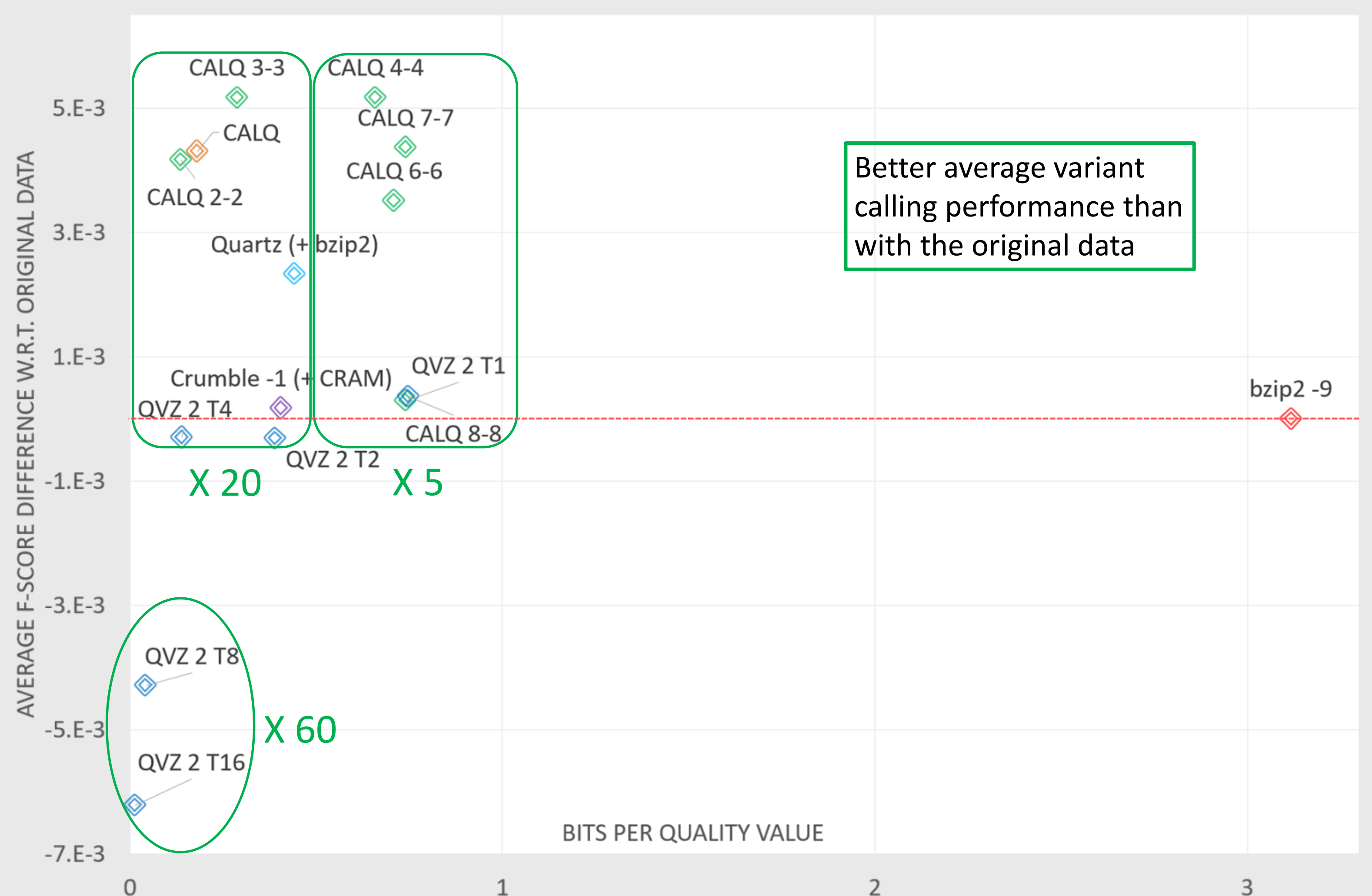
## Performance

### Overall MPEG-G performance on human high-coverage WGS data



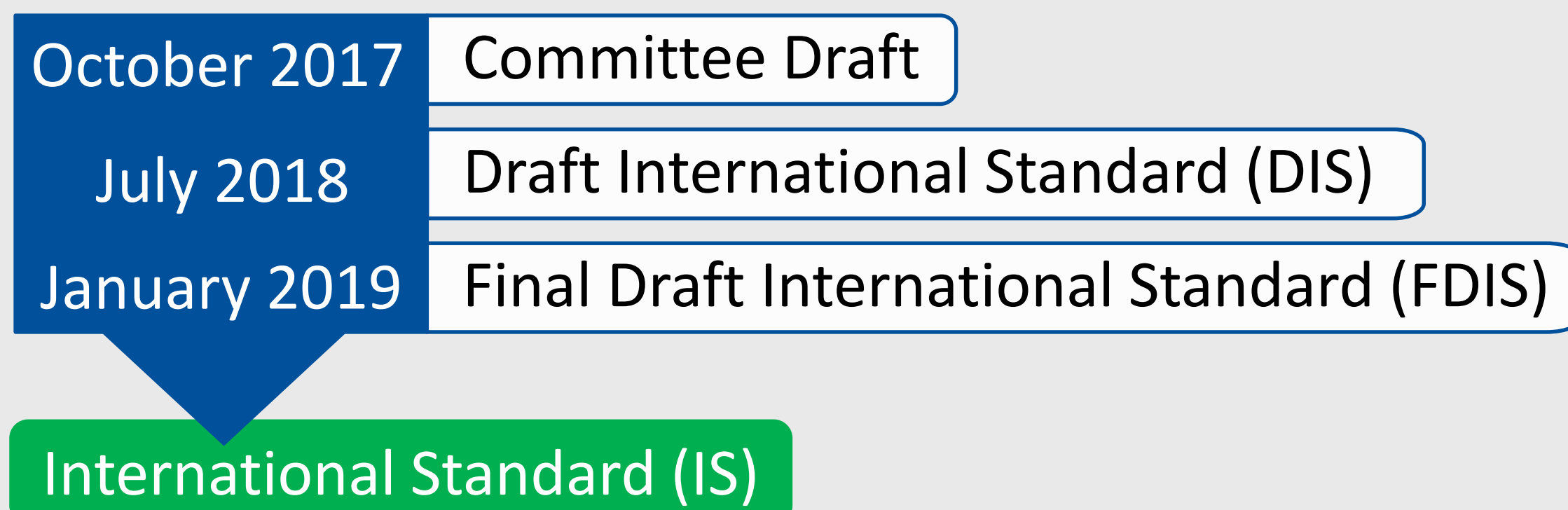
- Data:**
- whole genome from individual NA12878
  - re-sequenced with Illumina HiSeq @ 53x

### Impact of MPEG-G quasi-lossless quality value compression on variant calling



- Experiment setup:**
- customized variant calling pipeline following the “GATK Best Practices” [2]
  - comparison of the called variants to the NIST consensus set
- Data:**
- chromosomes 11 and 20 from individual NA12878
  - re-sequenced with Illumina HiSeq 2000 @ 14x

## Standardization Workplan



## References

- [1] Alberti, C. et al. (2016). An Evaluation Framework for Lossy Compression of Genome Sequencing Quality Values. In *Proceedings 2016 Data Compression Conference (DCC 2016)*, pages 221–230.
- [2] McKenna, A., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), pages 1297–1303.
- [3] Hernaez, M., Ochoa, I., and Weissman, T. (2016). A Cluster-Based Approach to Compression of Quality Scores. In *Proceedings 2016 Data Compression Conference (DCC 2016)*, pages 261–270.
- [4] Yu, Y. W., Yorukoglu, D., Peng, J., and Berger, B. (2015). Quality score compression improves genotyping accuracy. *Nature Biotechnology*, 33(3), pages 240–243.
- [5] Voges, J., Ostermann, J., Hernaez, M. (2017). CALQ: compression of quality values of aligned sequencing data. *F1000Research*, 6, page 1382 (poster).
- [6] Bonfield, J. K. (2014). The Scramble conversion tool. *Bioinformatics*, 30(19), pages 2818–2819.
- [7] Li, H., Handsaker, B., Wysoker, A. et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), pages 2078–2079.

## Advantages of an International Standard

