# biomodal software release notes duet pipeline 1.7.0 and CLI 2.2.0: November 2026

## Contents


duet pipeline 1.7.0 and CLI 2.2.0 ..... 2
Metadata, parameters, and pipeline usage ..... 2
Resolution algorithm ..... 2
Deduplication ..... 3
Methylation quantification ..... 3
Variant calling ..... 3
Fragmentomics ..... 3
Publishing rules ..... 4
Support for alternative pipeline modes and use cases ..... 4
Support for alternative reference genomes ..... 4
Resource utilisation and runtime efficiency ..... 4





biomodal Limited is registered in England and Wales, registered number: 08005377, registered address: The Trinity Building, Chesterford Research Park, Cambridge, CB10 1XL; VAT no: GB 0141 4564 31.

## duet pipeline 1.7.0 and CLI 2.2.0

We have issued the following feature improvements and bug fixes in the current duet pipeline 1.7.0 and biomodal CLI 2.2.0 release (covering changes from duet pipeline 1.6.0 and biomodal CLI 2.1.1).

Parameter overrides described in these release notes are applied when running the pipeline with the biomodal CLI, using the `--additional-params` option, e.g. `biomodal run duet … --additional-params variant_caller=deepvariant`. To apply several overrides, combine them in a single comma-separated list with no spaces, e.g:
`--additional-params variant_caller=both,fragment_length_dedup=false`.

### Metadata, parameters, and pipeline usage

#### Standardised metasheets

- Metasheets now use a standardised format, consistent with modality XPLR, and allow more flexible naming of input files.
- From pipeline 1.7.0, use one header row of field names followed by one row per sample. This is a transpose of the format required in earlier pipeline versions.
- A `sample_id` column is required. Any other metadata fields are passed through to the Excel and CSV summary reports.
- You can also specify input files in the metasheet using `R1`, `R2` and `lane` columns (for example, `L001`). Column names are case-sensitive. If all three are present, the pipeline uses them to find the input files. Otherwise, it matches input filenames against the expected patterns and takes the sample ID and lane from the filename, as before.

### Resolution algorithm

#### Use the maximum, not minimum, observed read length during Q-model/Q-table selection

- The Q-table or Q-model is now selected using the maximum read length observed in the first 10,000 reads, rather than the minimum.
- Previously, extensively trimmed reads could cause the wrong Q-table to be chosen (for example, a PE111 Q-table for PE151 data), resulting in out-of-bounds errors.

#### Q-models as an alternative to Q-tables

- Q-models are now available as an optional alternative to Q-tables for assigning base quality to resolved bases. Unlike Q-tables, Q-models also consider factors such as read cycle, proximity to the fragment end, and proximity to resolution errors. This gives more accurate scores for low-quality bases, which can then be used to filter out possible errors.
- By default (`prelude.phred=auto`), the pipeline uses Q-tables, except for NovaSeqX data in `mosaic_6bp` mode, which uses Q-models.
- To override this, set `--additional-params prelude.phred=qmodel` (Q-models), `=qtable` (Q-tables) or `=min` (the lowest instrument-reported quality of the bases being resolved).
- Q-models may be useful in conjunction with the hmC quality filtering option described below.



biomodal Limited is registered in England and Wales, registered number: 08005377, registered address: The Trinity Building, Chesterford Research Park, Cambridge, CB10 1XL; VAT no: GB 0141 4564 31.

## Deduplication

### Fragment-length-aware deduplication enabled by default

- Fragment-length-aware deduplication has been available to customers as an experimental feature since release 1.5.0 and following positive customer feedback, it has now been enabled by default.
- Fragment-length aware deduplication uses fragment length information inferred from the identification and removal of hairpin sequences to reduce the duplication rate among fragments that are shorter than the number of sequencing cycles. This increases the proportion of usable bases on a run, and increases coverage.
- Fragment-length-aware deduplication can be disabled by setting `--additional-params fragment_length_dedup=false`.

## Methylation quantification

### hmC quality filtering

- You can now set a minimum hmC quality score for hmC calls to be included in methylation quantification. Calls with a lower score are filtered out.
- Set the threshold with the optional parameter `--additional-params epiquant.min_hmc_quality` (for example, `=20`). Setting it to `=0` disables hmC quality filtering.
- By default, hmC quality filtering is applied only to NovaSeqX data processed in `mosaic_6bp` mode.
- This option is intended for use with the Q-models described above.

## Variant calling

### DeepVariant as an alternative germline variant caller

- DeepVariant, trained on duet data, is now available as an alternative to GATK for germline variant calling. GATK remains the default.
- To use DeepVariant instead of GATK, set `--additional-params variant_caller=deepvariant`.
- To run both callers in parallel and compare their outputs, set `--additional-params variant_caller=both`.

## Fragmentomics

### Multi-sample fragmentomics Zarr stores

- Multi-sample fragmentomics Zarr stores are now generated by the pipeline. These are compatible with biomodal's modality XPLR tertiary analysis suite.
- A multi-sample zarr store features statistics about 5-base or 6-base end motifs and fragment lengths across each sample.

### Plain-text fragmentomics exporters

- Fragmentomics workflows can export plain-text fragment-length and end-motif data for each sample. These are single-sample plain-text alternatives to the multi-sample fragmentomics Zarr store.
- Activate by setting: `--additional-params run_by_default.modality_fragmentomics_export=true`.



biomodal Limited is registered in England and Wales, registered number: 08005377, registered address: The Trinity Building, Chesterford Research Park, Cambridge, CB10 1XL; VAT no: GB 0141 4564 31.

## Publishing rules

### Disable default publishing of resolved FASTQs

- Resolved FASTQ files are no longer published by default. This reduces the publish directory size for runs that do not require these intermediate or diagnostic outputs.
- Resolved FASTQ files can still be published by setting `--additional-params publish_by_default.prelude=true`.

## Support for alternative pipeline modes and use cases

### Remove the `dedup` suffix requirement from aligned BAM entry-point filenames

- Aligned BAM entry-point filenames no longer require `*dedup.bam` and `*dedup.bam.bai` as their filename suffixes: `*bam` and `*bam.bai` is sufficient.

## Support for alternative reference genomes

### Deprecate GRCm38 and remove obsolete reference genomes

- The legacy `GRCm38` reference is deprecated in favour of `GRCm38p6`.
- The unused `BDGP6-32`, `R64-1-1`, `TAIR10`, `Spirodela_polyrhiza_9509`, `stickleback_v5`, `Wm82_a4` and `Ptrichocarpa_v4-1` references have been removed from the supported configuration and reference-packaging workflow.
- You can still use the reference file pipeline to generate reference files for any genome or species, including those no longer provided as standard.

## Resource utilisation and runtime efficiency

### Increase memory for lane merging in the `super_seq` profile

- The memory allocation for `SAMTOOLS_MERGE_LANES` has been increased in the `super_seq` profile to support larger or more demanding lane-merge jobs.



biomodal Limited is registered in England and Wales, registered number: 08005377, registered address: The Trinity Building, Chesterford Research Park, Cambridge, CB10 1XL; VAT no: GB 0141 4564 31.