# Use of duplicate homozygous alleles in STRmix™ evidence text format input files

27 November 2018

Some post-electrophoresis analysis software programs (for example GeneMapper® and GeneMapper® *ID-X*) have the option to duplicate a peak if it is the only peak detected at a locus. Recently, a member of our user community identified that use of the 'Duplicate Homozygous Alleles' feature affects drop-in modelling and the Likelihood Ratio (*LR*) calculation within STRmix™ under a very specific set of circumstances.

If the 'Duplicate Homozygous Alleles' feature of GeneMapper® is turned on, any locus where a single peak has been detected will have the resulting peak data duplicated in the exported text file (i.e. the STRmix™ input file). If the height of a duplicated peak is below the STRmix™ drop-in cap, then the peak may be considered as drop-in during deconvolution. When this occurs STRmix™ must consider both instances of the peak in the input file as drop-in. This results in an additional drop-in penalty being applied and will impact the genotype weights assigned by STRmix™. Furthermore, if genotype sets including drop-in are accepted during deconvolution, additional term/s for the allele frequency of the drop-in peak will be included in the *LR* calculation.

The double counting of drop-in only occurs in STRmix™ under the following conditions:

1. A single peak at a locus within an evidence input file is duplicated. This will occur if the 'Duplicate Homozygous Alleles' option is enabled in GeneMapper® (or equivalent feature in alternative analysis software programs)
2. Drop-in is enabled within the STRmix™ interpretation
3. The single peak detected at a locus is below the STRmix™ drop-in cap.

This has been shown to affect STRmix™ versions 2.0 through 2.6 inclusive.

We expect that the number of interpretations affected will be small given the conditions described above. In these few instances where an *LR* has been calculated to a POI and drop-in is not required under $H_p$, the impact on the size of the *LR* will be negligible. Where an *LR* has been calculated to a POI and drop-in is required under $H_p$, then the magnitude of the effect is determined by the allele frequency of the drop-in peak.

## Effect on deconvolution and *LR*

Under the conditions listed above, STRmix™ will assign an additional drop-in penalty when considering that the detected peak is drop-in. This will mean that genotype combinations that include drop-in will be assigned less weight than they otherwise would be (or may not be accepted at all). In extreme cases, this could result in an exclusion of a POI if drop-in is required under $H_p$, however we would expect that careful review of the STRmix™ output would alert the user to this possibility. Furthermore, if genotype combinations that include drop-in have been accepted (i.e. have been given weight in the STRmix™ output), additional term/s for the allele frequency of the drop-in peak will be included in the *LR* calculation. This may lead to a wide HPD interval being generated if the proposed drop-in peak has a rare frequency or has not been observed in the relevant population.

Because both STRmix™ interpretations and *LR* calculations are affected, we advise that if a laboratory wishes to reanalyse a sample they think may have been affected, they carry out both a new deconvolution and *LR* with a revised text input file (where single peaks have not been duplicated).

The current STRmix™ v2.4, v2.5 and v2.6 Operation Manuals will be re-issued advising that single peaks should not be duplicated within STRmix™ evidence input files. Given that the double counting of drop-in peaks is caused by an issue with the input file format rather than with the modelling used within STRmix™, revised versions of the software will not be issued.

## Examples of the magnitude of the effect

We demonstrate an example of the difference in *LR* for a single-source profile at one locus where a single 14 peak (at 34 rfu) was duplicated versus not duplicated in the input file. In order to replicate this problem, the drop-in rate was artificially inflated to 0.1 (with a uniform 0,0 distribution). For comparison purposes, the same genotype weights were used for both calculations:

14,14 weight 0.08

14,Q weight 0.89

Q,Q (14 drop-in) weight 0.03

The locus *LR* for POI = 14,14 (i.e. not requiring drop-in) is provided in the following table. The *LR*s produced with and without peak duplication are within one order of magnitude, with peak duplication producing a slightly larger *LR*.

| **Scenario** | **Pr(E\|$H_p$)** | **Pr(E\|$H_d$)** | ***LR*** |
|---|---|---|---|
| Peak not duplicated in input file | 0.08 | 0.42161 | 0.18975 |
| Peak duplicated in input file | 0.08 | 0.41882 | 0.19101 |

The locus *LR* for POI = 15,15 (i.e. requiring drop-in of the 14 allele) is provided in the following table. The *LR*s produced with and without peak duplication are within one order of magnitude, with peak duplication producing a smaller *LR*.

| **Scenario** | **Pr(E\|$H_p$)** | **Pr(E\|$H_d$)** | ***LR*** |
|---|---|---|---|
| Peak not duplicated in input file | 0.01045 | 0.41038 | 0.02546 |
| Peak duplicated in input file | 0.00364 | 0.40741 | 0.00893 |

The example was repeated modifying the single peak detected to a rare allele. The locus *LR* for POI = 14,14 (requiring drop-in of the rare allele) is provided in the following table.
The *LR* produced when the peak was duplicated is several orders of magnitude smaller than the figure produced when the peak was not duplicated. This demonstrates that the magnitude of the effect is increased when the proposed drop-in peak is rare.

| **Scenario** | **Pr(E\|$H_p$)** | **Pr(E\|$H_d$)** | ***LR*** |
|---|---|---|---|
| Peak not duplicated in input file | 5.6980E-6 | 3.3392E-4 | 1.7064E-2 |
| Peak duplicated in input file | 1.0822E-9 | 3.2823E-4 | 3.2972E-6 |