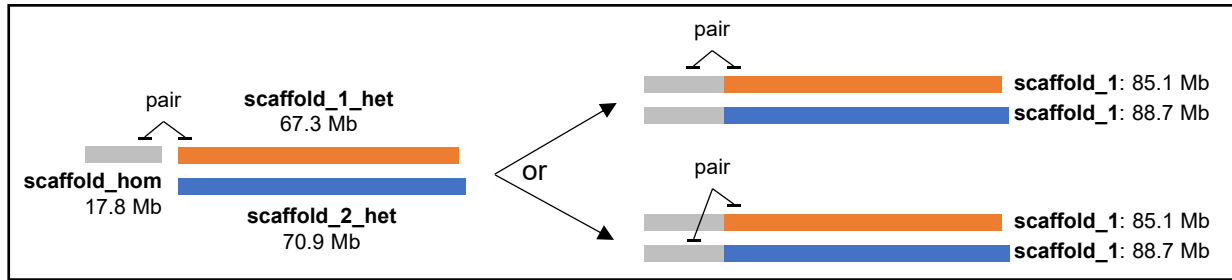## Transforming the matrices

What I really need is for all the reads mapping to scaffold_hom to be assigned to either scaffold_1 or scaffold_2 containing the homozygous region. To do this, I used a Python script (`Dpa-canu-het00c-YaHS-v202304-hethomosep_PoreC_correction.py`) that changes the .pairs file that records all the PoreC mapped pairs. The .pairs file has this structure:

read	chr1	pos1	chr2	pos2	Other attributes…

The script goes through the file line-by-line. At `chr1` or `chr2`, if it encounters "`scaffold_1_het`" or "`scaffold_2_het`", it simplifies the name to "`scaffold_1`" or "`scaffold_2`", and it changes `pos1/2` from the original value to that value + 17839851 (length of the homozygous region). On the other hand, if at `chr1` or `chr2`, if it encounters "`scaffold_hom`", the `pos1/2` value is kept the same, but the `chr1/2` value is assigned either to "`scaffold_1`" or "`scaffold_2`" at random (using python's `random.choice(["scaffold_1", "scaffold_2"])`), with the rationale being that **it is truly unknown which scaffold it came from.**

With these values changed, the `.pairs` file is sorted using `pairtools sort` and turned into a hic matrix using `juicer_tools pre` command.