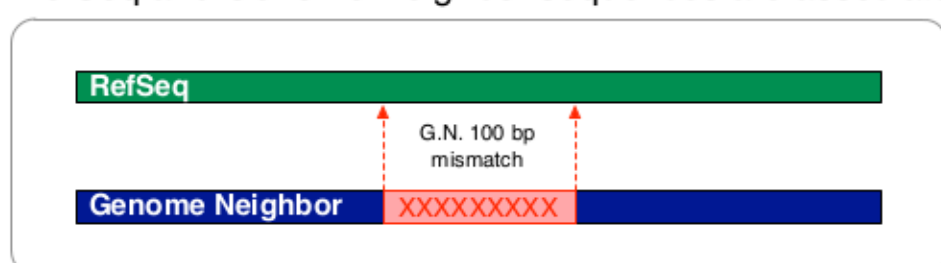
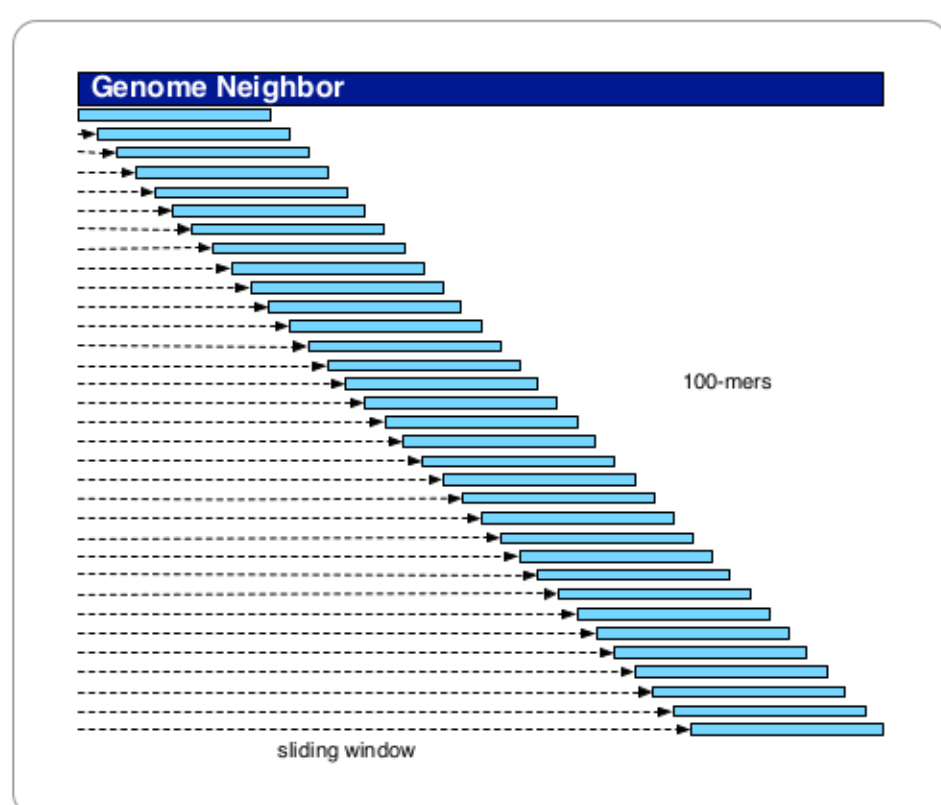


Supplemental Figure S4

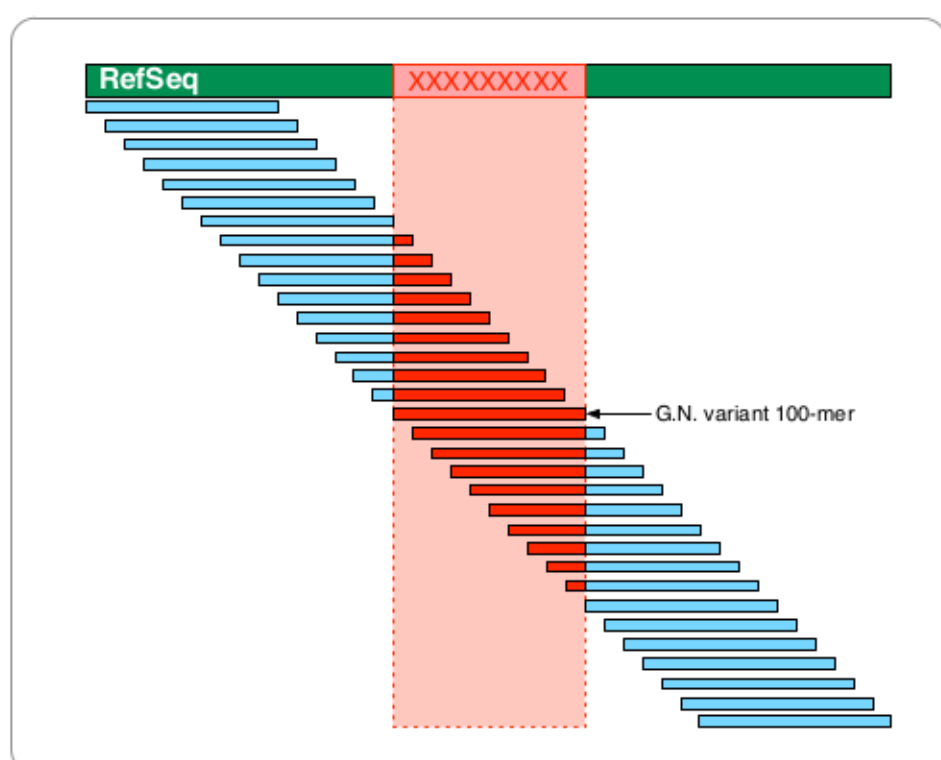
1. RefSeq and Genome Neighbor sequences are associated.



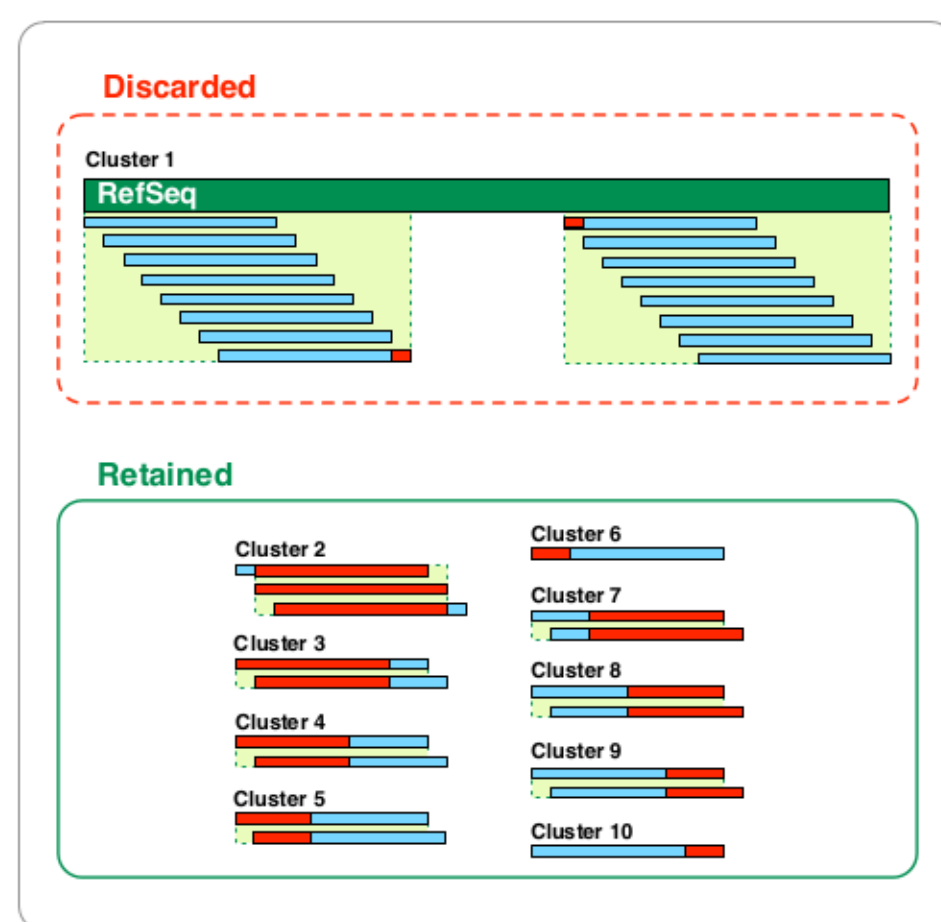
2. Genome Neighbor 100-mer sequences are generated.



3. Compared to RefSeq, some 100-mers contain variation.



4. Sequence clustering identifies homologous RefSeq 100-mers.



5. Variant 100-mers are merged into contiguous target sequences.

