Dr. Moh.Gheras
Email: mohghattas@gmail.com

Assembly Conversion Approaches

Abstract

The genome assembly is updated and changed regularly. Usually, researchers always need to convert the results obtained from the old assembly to the new version, and vice versa. There are some conversion tools that can convert genomes to facilitate meta-analysis, data integration and visualization. In this document, some methods are explored and analyzed.

Genome Conversion Methods

Many studies have been contacted to analyze different versions of genomic transformation. With the use of more genomic references, there is an urgent need to integrate new and old genomics results. Therefore, there are two general methodologies for genomic conversion between different assemblies.

The first method involves re-aligning all data to the same reference genome. Compared with other methods, the main advantage of this method is that it can provide the best results. However, this is time-consuming, and in some cases, if the original data is not available or does not contain a sequence, the re-alignment process cannot be performed.

The second method is to use mapping files to convert coordinates between assemblies. This method is faster, but may lose a small amount of information. The good approach balances between performance and accuracy. Recently, the three most popular tools for comparing genome assemblies by coordinates: LiftOver (Gao, 2018) from the University of California, Cross Map (assemblies, 2014) from Zhao, Remap from NCBI (Coordinators, 2016), and Ensembl assembly converter (pudich, Springer).

Dr. Moh.Gheras
Email: mohghattas@gmail.com

LiftOver tools are available through web services and command line utilities. It facilitates the conversion between assemblies for different organisms. The result generated by CrossMap is the same as liftOver, but the accuracy of converting genomic coordinates is low. Finally, the remapping tool provides cross-species mapping for list of major assemblies, but it can be used for a limited number of organisms.

All of the aforementioned methods can convert coordinates between different assemblies and provide almost the same results. When the genome segments are no longer continuous in the reference assembly, the main challenge arises, as shown in Figure 1. Both CrossMap and liftOver split the segment into smaller sub-segments and map it to different locations on the target assembly. The remapping tool maintains the integrity of the segment and maps the span to the target assembly. This tool can only convert continuous segments, and is not suitable for large-scale or pipeline applications, because only one file can be submitted through its web service.

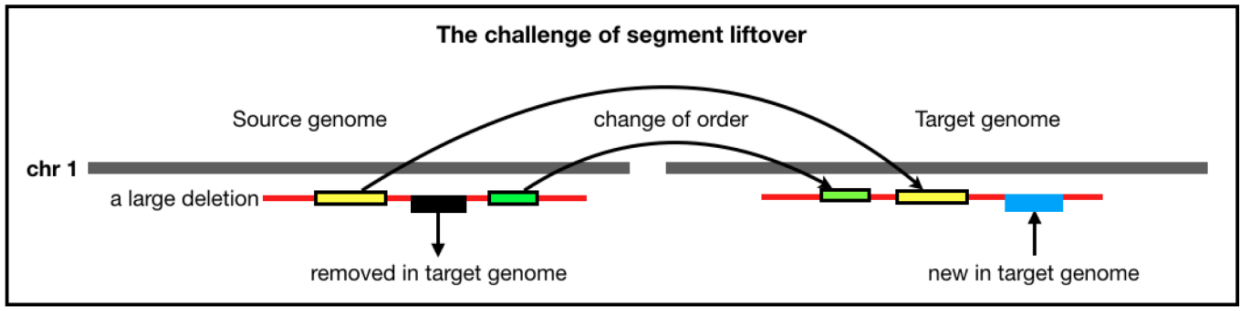

Figure 1 The challenge of segment leftover *(Gao, 2018)*

Summary

In bioinformatics, coordinate conversion between different genomic versions is a tedious task. Existing methods aim to balance performance and accuracy. One of the challenges of genome conversion is that the segment is no longer continuous in the target genome. To overcome this challenge, some methods divide the segment into small blocks and map it to different locations. With the help of automatic batch processing, the liftOver method can greatly reduce the data processing time.

Dr. Moh.Gheras
Email: mohghattas@gmail.com

Bibliography

assemblies, C. a. (2014). Zhao, Hao and Sun, Zhifu and Wang, Jing and Huang, Haojie and Kocher, Jean-Pierre and Wang, Liguo. *Bioinformatics*, 1006--1007.

Coordinators, N. R. (2016). *Database resources of the National Center for.*

Gao, B. a. (2018). segment liftover: a Python tool to convert segments between genome assemblies. *F1000Research*.

pudich, G. M.-S. (Springer). Touring Ensembl: a practical guide to genome browsing. *BMC genomics*, 295.