

# Calcordance: A Genotyping Concordance Tool

Brad Tibbils, Sean M.L. Griffith, Michelle Z. Mawhinney, Ivy A. McMullen, Mary Jewell, Alysen B. Robinson, Chrissie M. Ongaco, Marie E. Hurley, Jane M. Romm, Lee Watkins, Jr. and Kimberly F. Doheny  
Center for Inherited Disease Research (CIDR), Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD

## Key Terms

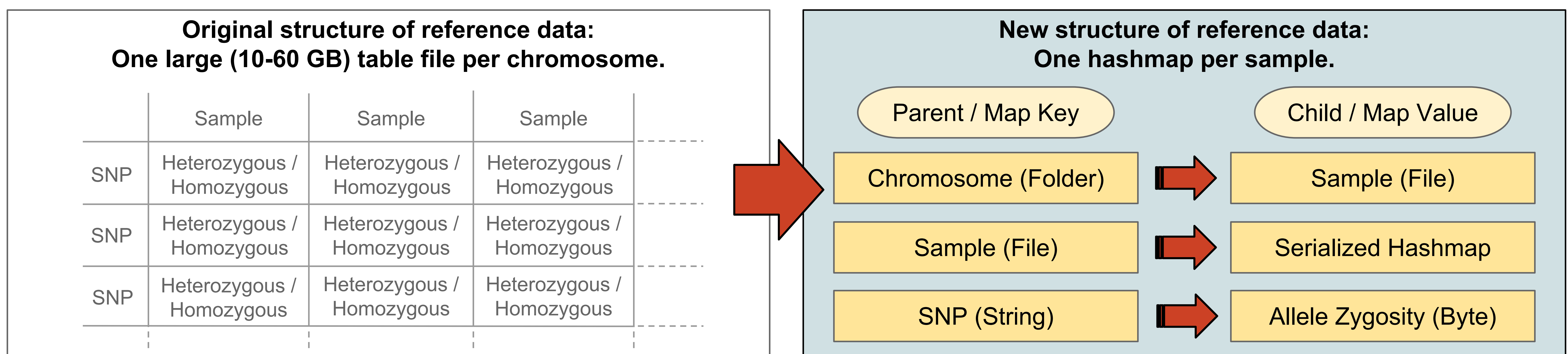
**Concordance** (in genotyping) involves determining the percentage of identical genotype calls between two samples from the same individual. In this particular use case, one sample is always from a constant set of reference samples and the other is from a set of samples submitted by the investigator. Calcordance (a portmanteau of calculation and concordance) is a software application created to complete this task as quickly as possible with a very large reference data set (the 1000 Genomes Project Phase 3 call set).

**concordance** *n.* (in genotyping) The process of determining the percentage of identical genotype calls between two independently generated data sets for the same individual.

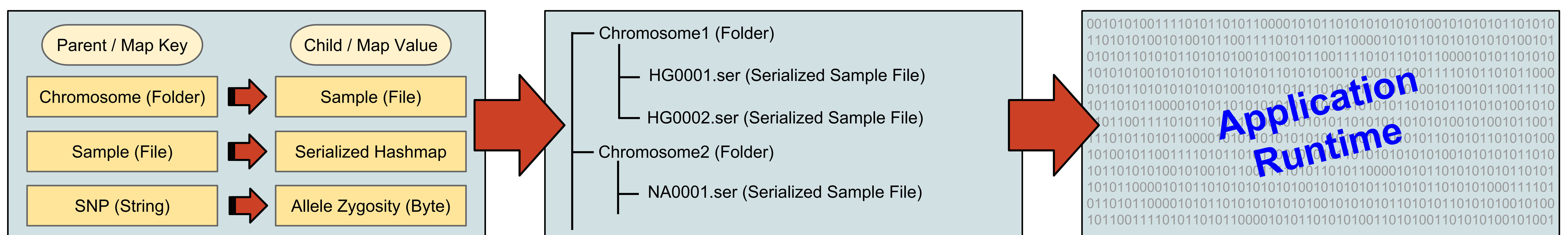
**serialize** *v.* To translate a data structure or object state into a format that can be stored.

**hashmap** *n.* A data structure that allows very quick lookups between two pieces of information.

## Step 1: Create relationships between the reference data using hashmap structures and file organization.

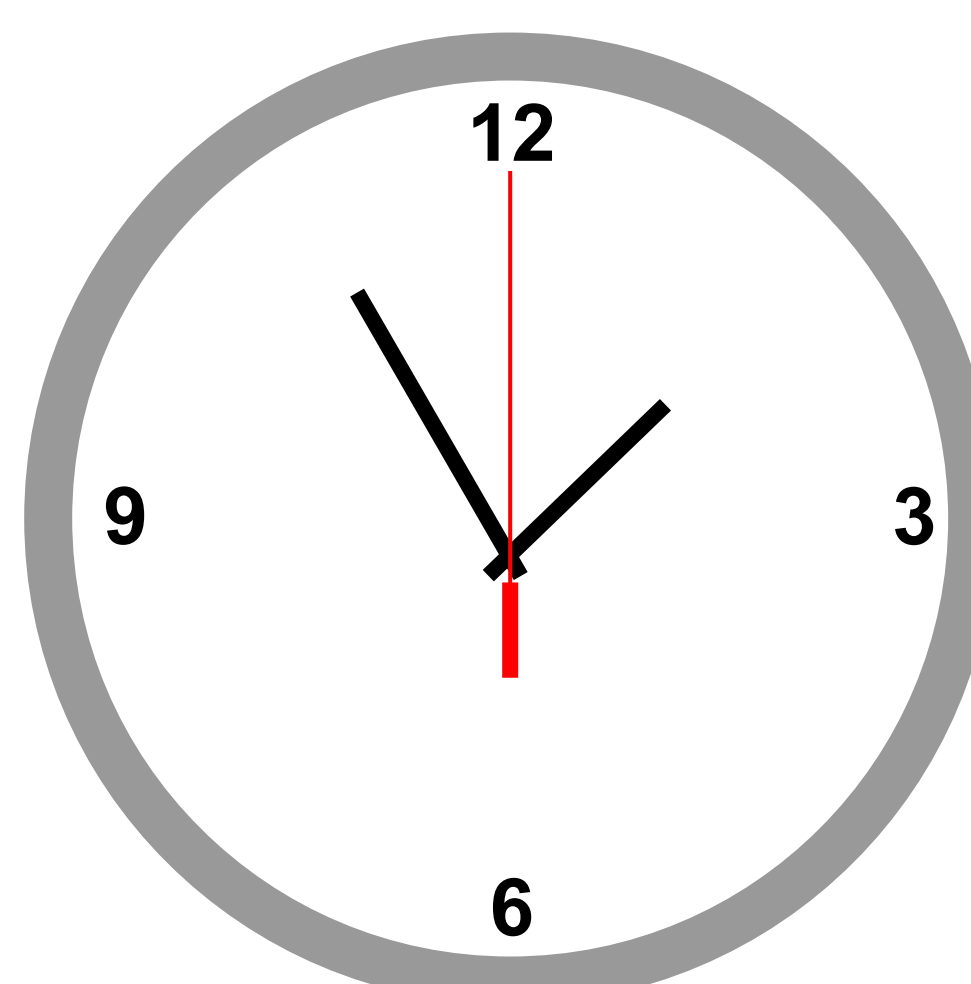


## Step 2: Serialize the new data structure and pull it into the application during runtime.



## Results (Application Runtime)

Original Application Runtime:  
2 Days



New Application Runtime:  
48 Minutes

