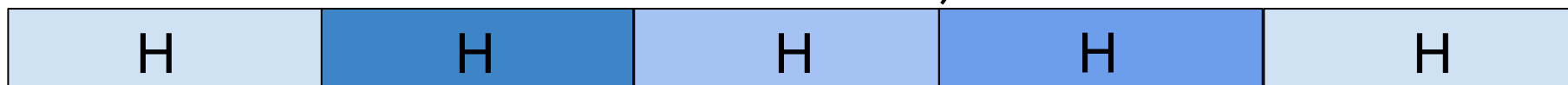


Stacked tandem repeat



Centromere (chrX)

Figure 1. The architecture of centromere on chromosome X. The centromere of chromosome X (cenX) consists of ~18100 monomers of length ≈ 171 bp each based on the cenX assembly in Bzikadze and Pevzner, 2020 (the latest T2T assembly (Nurk et al., 2021) represents a minor change to this assembly). These monomers are organized into ~1500 units. Five units in the Figure are colored by five shades of blue illustrating unit variations. Each unit is a stacked tandem repeat formed by various monomers. The vast majority of units in cenX correspond to the canonical HOR which is formed by twelve monomers (shown by twelve different colors). The figure on top represents the dot plot of the nucleotide sequence for the canonical HOR that reveals twelve monomers. While the canonical units are 95–100% similar, monomers are only 65–88% similar. In addition to the canonical 12-monomer units, cenX has a small number of partial and auxiliary HORs with varying numbers of monomers.