Centromere **(nucleotide sequence)**

Monomer **(consensus alpha satellite)**

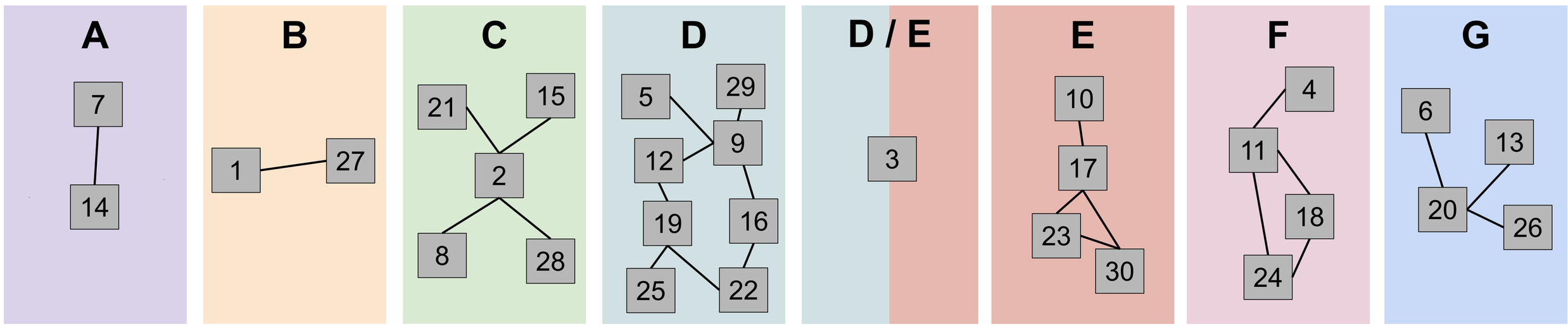
StringDecomposer(*Centromere, Monomer*)

Monomer-blocks

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

CentromereArchitect

Generating monomers (connected components of the block graph) and identifying hybrid monomers (D/E)

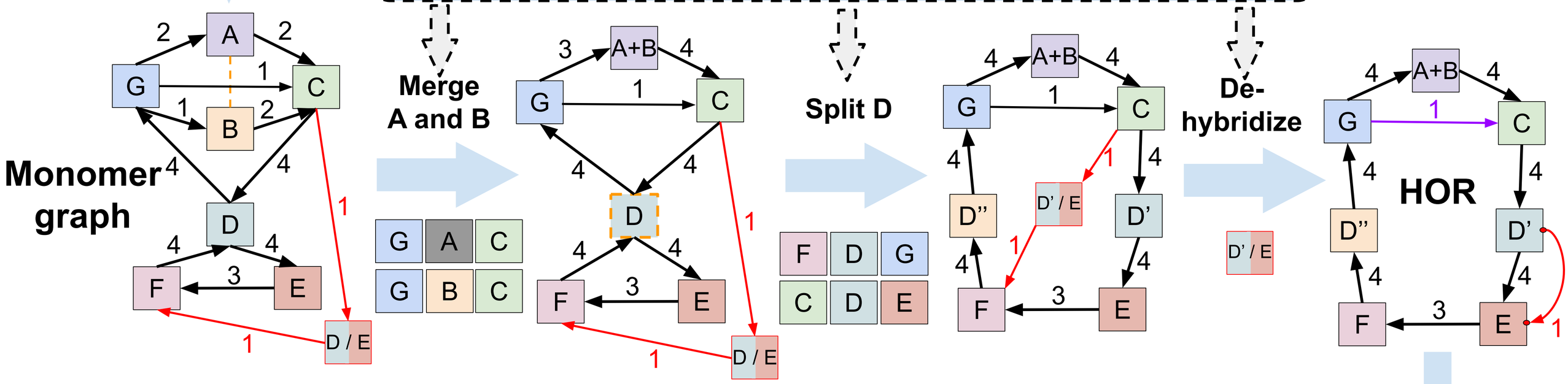


Monocentromere *Centromere in alphabet of 8 monomers:** A B C D E F G D / E

B C D / E F D G A C D E F D G A C D E F D G C D E F D G B C D E

Centromere Evolution Postulate

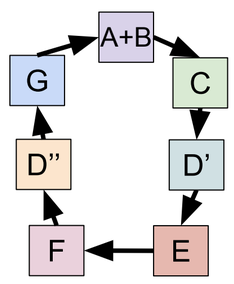
Monomer graph



Monocentromere *Centromere* in alphabet of 7 monomers:** A+B C D' E F G D''

A+B C D' / E F D'' G A+B C D' E F D'' G A+B C D' E F D'' G C D' E F D'' G A+B C D' E

HOR decomposition of *Centromere* with HOR**



$p_{(A+B)C}$ D' / E c_F c_F p_{FG} c_C p_{CE}

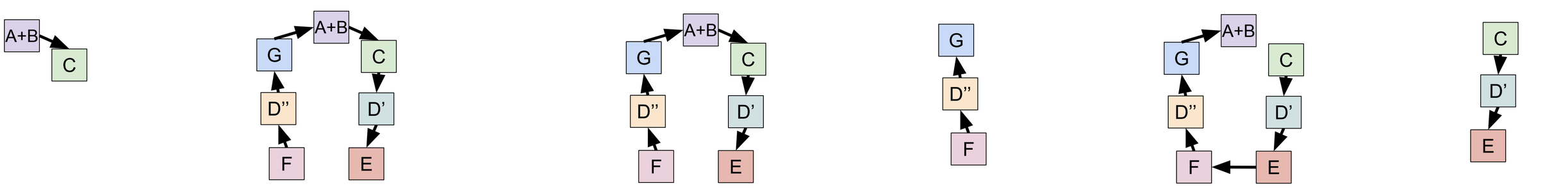


Figure 2. HORmon pipeline. Given the nucleotide sequence *Centromere* and a consensus alpha satellite sequence *Monomer*, HORmon iteratively launches StringDecomposer (Dvorkina et al., 2020) to partition *Centromere* into monomer-blocks. After each launch of StringDecomposer, HORmon launches CentromereArchitect (Dvorkina et al., 2021) to cluster similar monomer-blocks into monomers, identify hybrid monomers (represented by a single hybrid D/E of monomers D and E), and transform *Centromere* into the monocentromere *Centromere**. Afterward, HORmon uses the generated monocentromere to construct a monomer-graph (red edges connect the hybrid monomer D/E with the rest of the monomer-graph). To comply with the Centromere Evolution Postulate, HORmon performs split/merge transformations and dehybridizations on the initial monomer-set. The orange dotted undirected edge connects similar monomers A and B to indicate that they represent candidates for merging. The breakable monomer D is shown as a dotted vertex to indicate that it is a candidate for splitting. The dehybridization substitutes the hybrid vertex D'/E by a single "red" edge that connects the prefix of D' with the suffix of E. Split, merge, and dehybridization operations result in a new monomer-set and transform *Centromere* into the monocentromere *Centromere***. The black cycle in the monomer-graph of *Centromere** represents the HOR; the purple edge connecting monomers G and C is a low-frequency chord in this cycle. HORmon uses this HOR to generate the HOR decomposition of *Centromere*** into the canonical (c_F, c_C), partial ($p_{(A+B)C}$, p_{FG}, p_{CE}), and auxiliary (the single block D'/E) HORs.

c_F and c_C refer to traversing the (canonical) HOR starting from monomers F and C, respectively. $p_{(A+B)C}$, p_{FG}, and p_{CE} refer to partial traversals of the HOR from monomer A+B to C, from F to G, and from C to E, respectively.