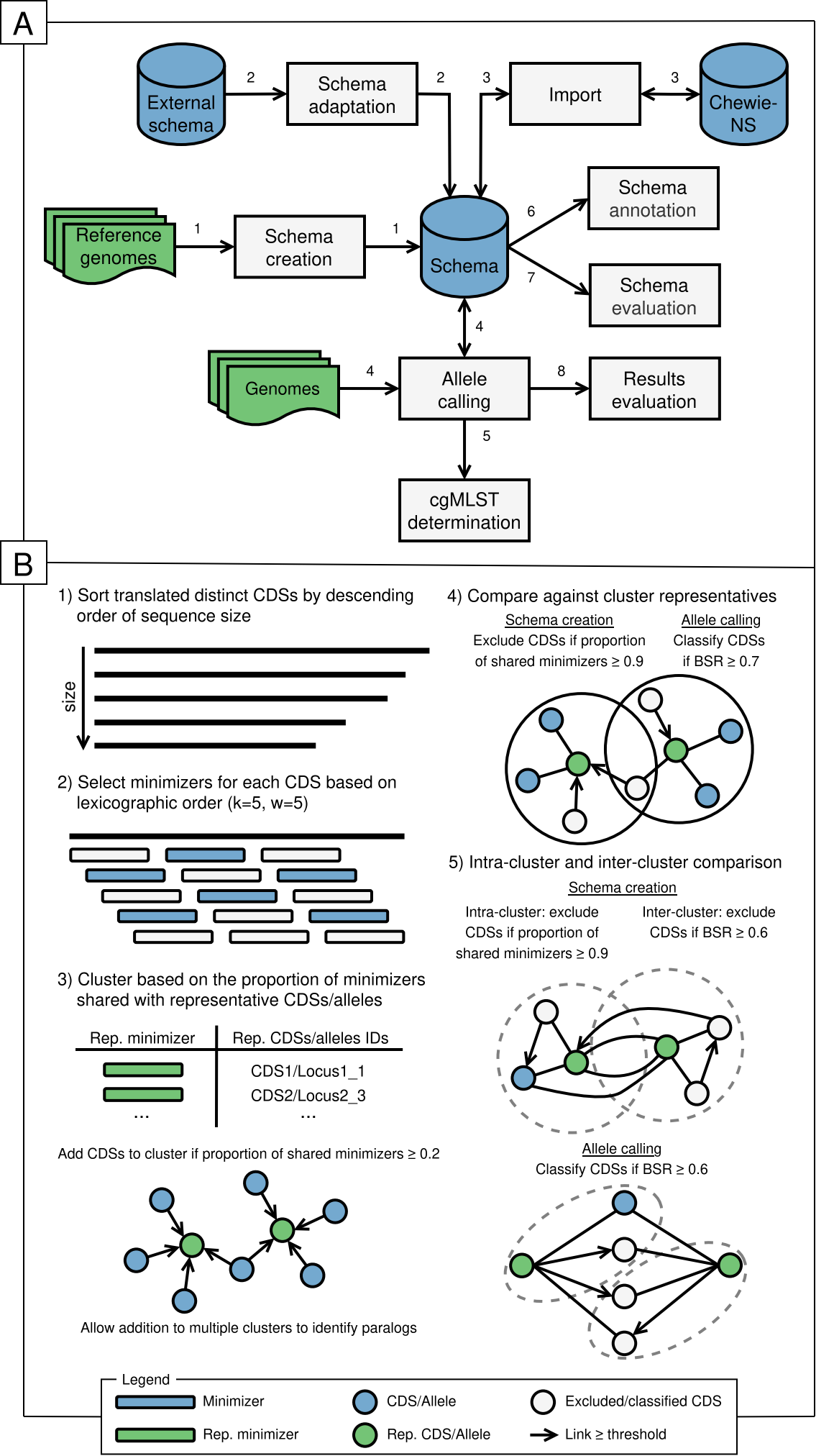
A
External schema
2
Schema adaptation
2
3
Import
3
Chewie-NS
Reference genomes
1
Schema creation
1
Schema
6
Schema annotation
7
Schema evaluation
4
Genomes
4
Allele calling
8
Results evaluation
5
cgMLST determination
B
1) Sort translated distinct CDSs by descending order of sequence size
size
2) Select minimizers for each CDS based on lexicographic order (k=5, w=5)
3) Cluster based on the proportion of minimizers shared with representative CDSs/alleles
Rep. minimizer
Rep. CDSs/alleles IDs
CDS1/Locus1_1
CDS2/Locus2_3
...
...
Add CDSs to cluster if proportion of shared minimizers ≥ 0.2
Allow addition to multiple clusters to identify paralogs
4) Compare against cluster representatives
Schema creation
Exclude CDSs if proportion of shared minimizers ≥ 0.9
Allele calling
Classify CDSs if BSR ≥ 0.7
5) Intra-cluster and inter-cluster comparison
Schema creation
Intra-cluster: exclude CDSs if proportion of shared minimizers ≥ 0.9
Inter-cluster: exclude CDSs if BSR ≥ 0.6
Allele calling
Classify CDSs if BSR ≥ 0.6
Legend
Minimizer
Rep. minimizer
CDS/Allele
Rep. CDS/Allele
Excluded/classified CDS
Link ≥ threshold