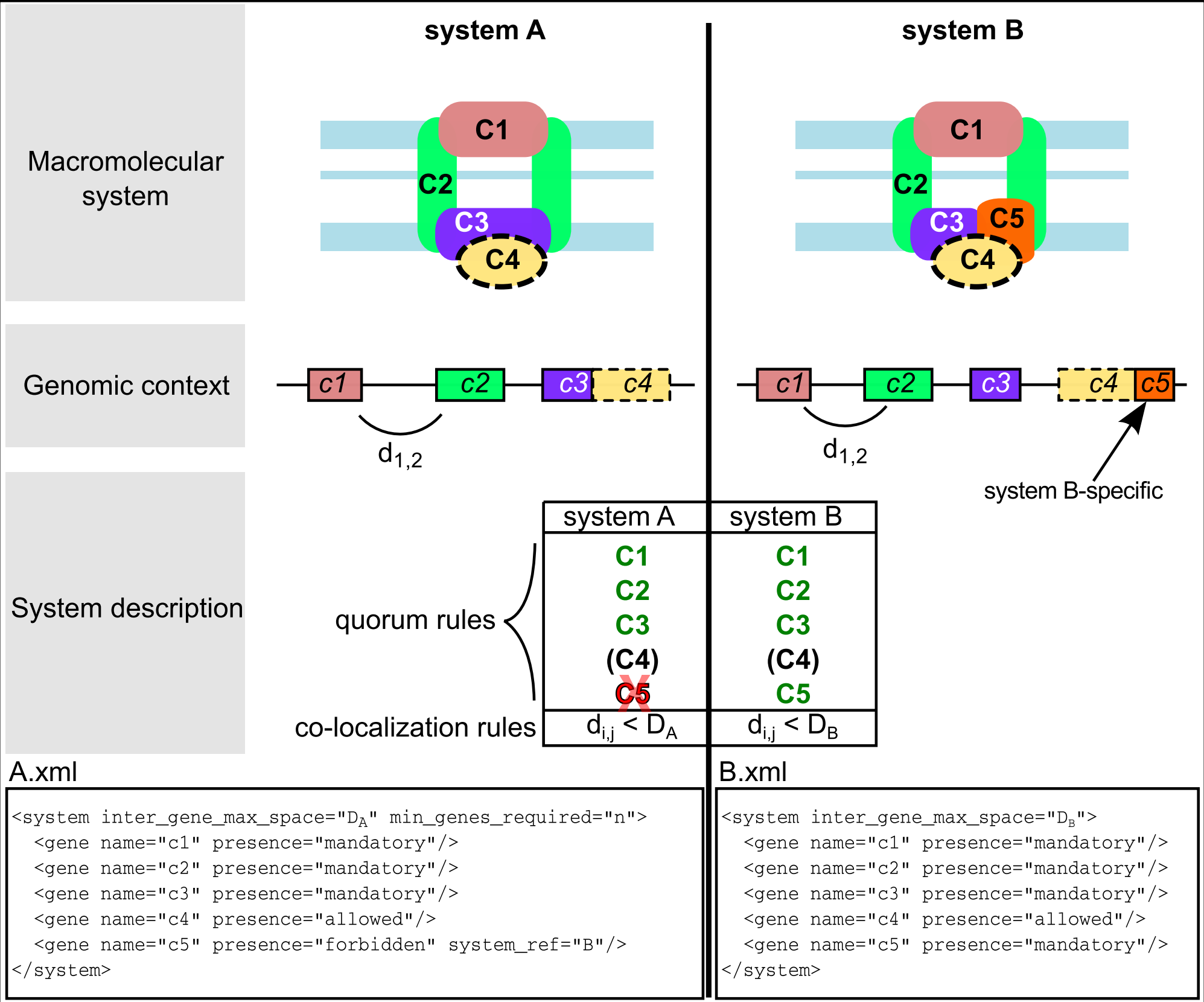
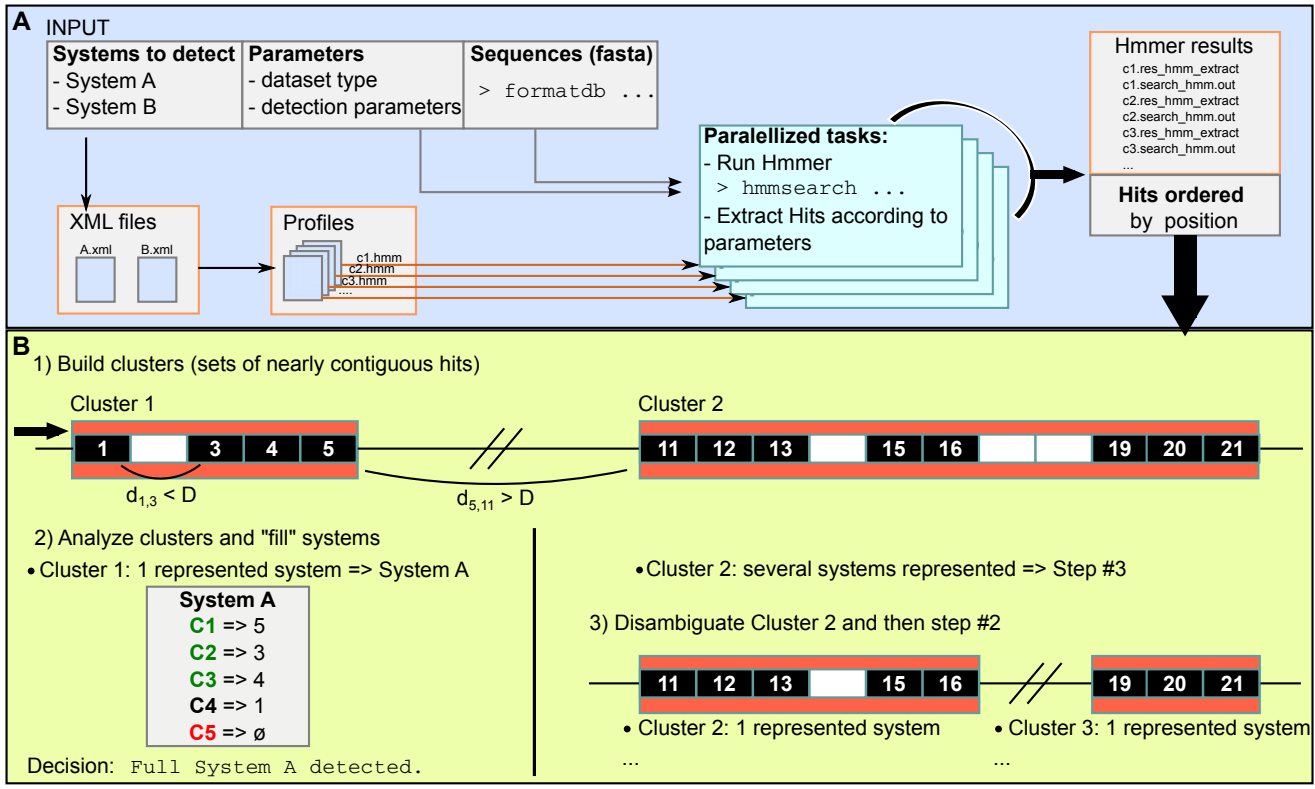
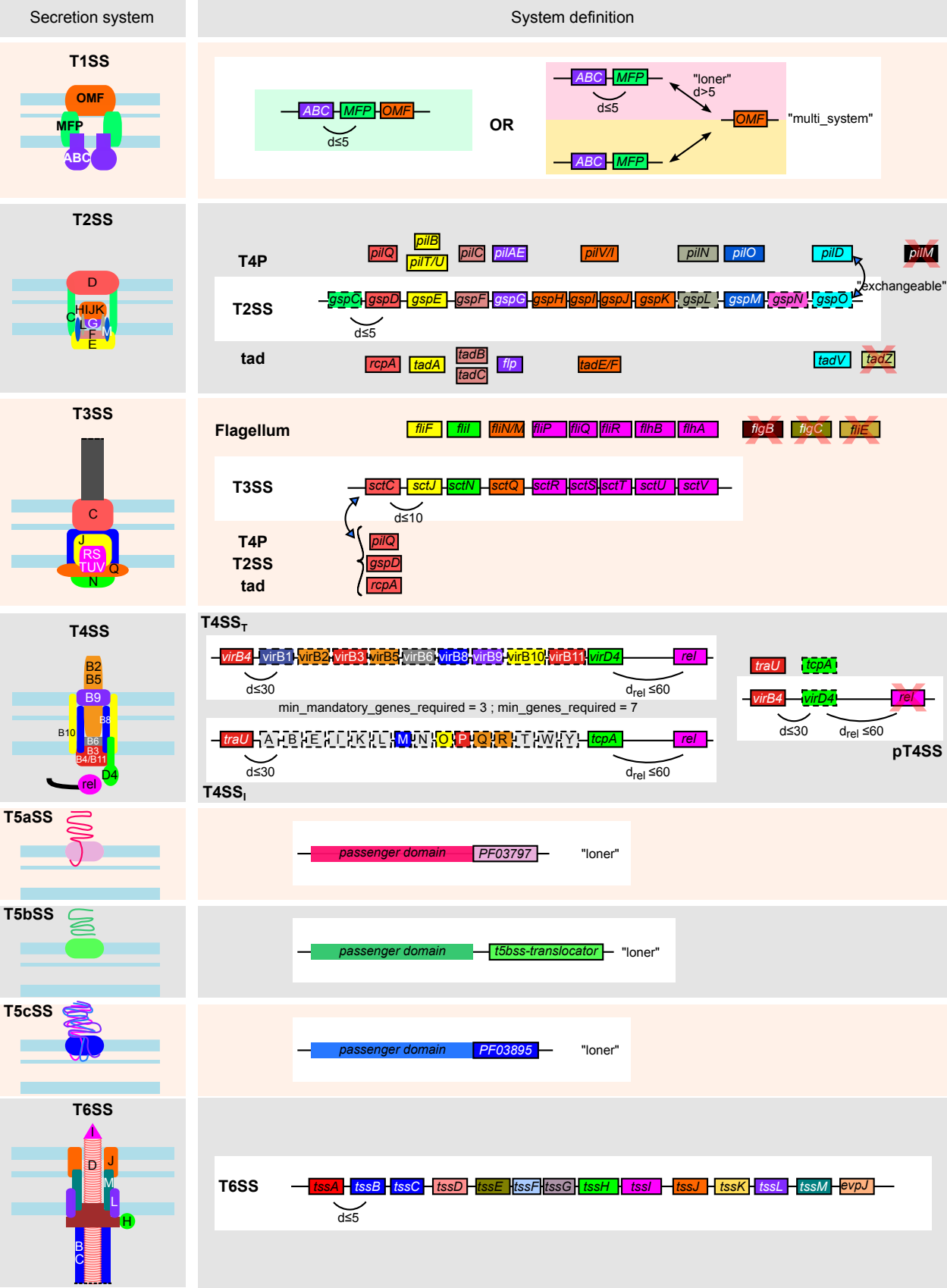
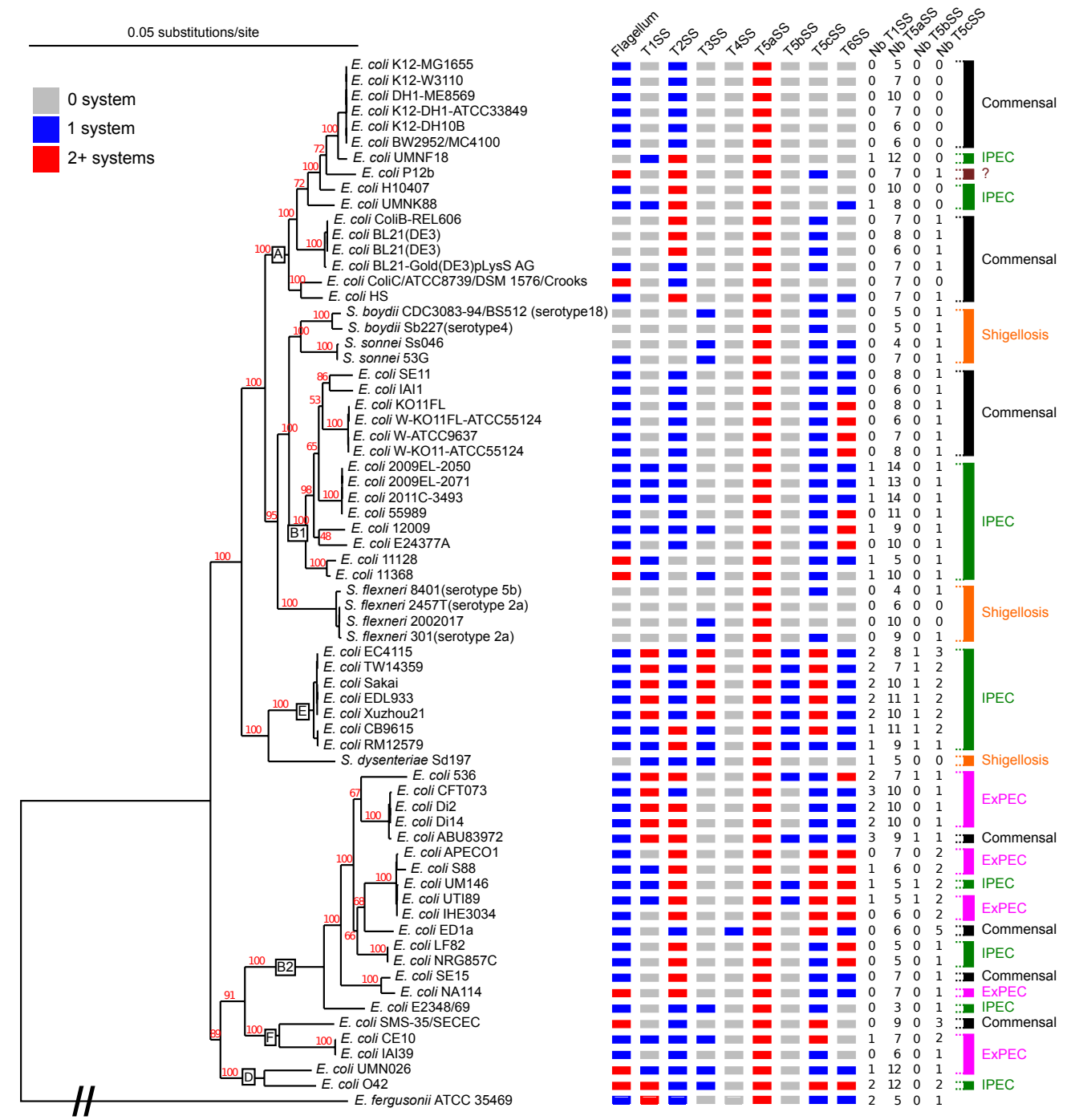


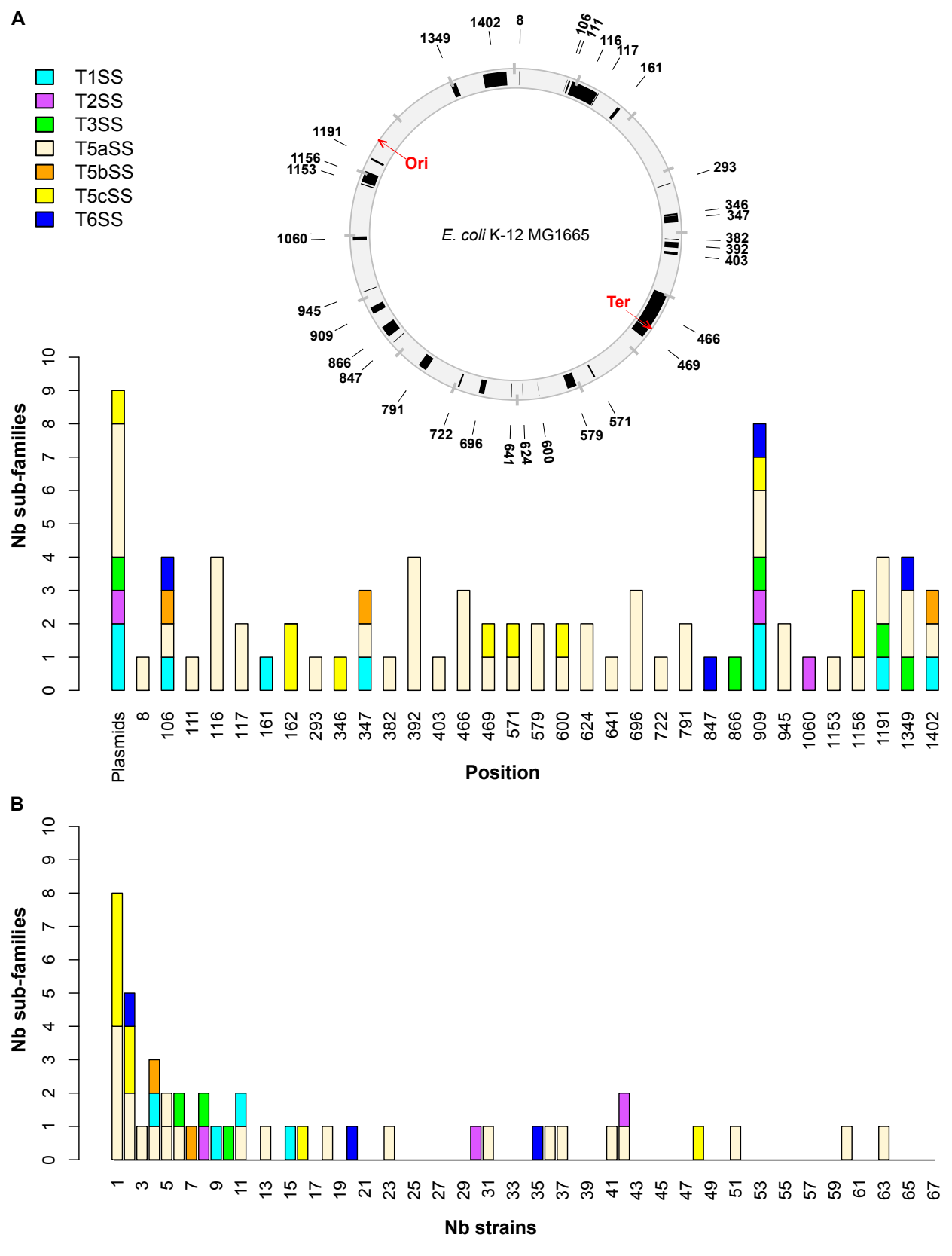

```
<system inter_gene_max_space="D_A" min_genes_required="n">
  <gene name="c1" presence="mandatory"/>
  <gene name="c2" presence="mandatory"/>
  <gene name="c3" presence="mandatory"/>
  <gene name="c4" presence="allowed"/>
  <gene name="c5" presence="forbidden" system_ref="B"/>
</system>
```

```
<system inter_gene_max_space="D_B">
  <gene name="c1" presence="mandatory"/>
  <gene name="c2" presence="mandatory"/>
  <gene name="c3" presence="mandatory"/>
  <gene name="c4" presence="allowed"/>
  <gene name="c5" presence="mandatory"/>
</system>
```


A
INPUT
Systems to detect
- System A
- System B
Parameters
- dataset type
- detection parameters
Sequences (fasta)
> formatdb ...
XML files
A.xml
B.xml
Profiles
c1.hmm
c2.hmm
c3.hmm
....
Paralellized tasks:
- Run Hmmer
> hmmsearch ...
- Extract Hits according to parameters
Hmmer results
c1.res_hmm_extract
c1.search_hmm.out
c2.res_hmm_extract
c2.search_hmm.out
c3.res_hmm_extract
c3.search_hmm.out
...
Hits ordered by position
B
1) Build clusters (sets of nearly contiguous hits)
Cluster 1
1
3
4
5
Cluster 2
11
12
13
15
16
19
20
21
d1,3 < D
d5,11 > D
2) Analyze clusters and "fill" systems
• Cluster 1: 1 represented system => System A
System A
C1 => 5
C2 => 3
C3 => 4
C4 => 1
C5 => ø
Decision: Full System A detected.
• Cluster 2: several systems represented => Step #3
3) Disambiguate Cluster 2 and then step #2
11
12
13
15
16
19
20
21
• Cluster 2: 1 represented system
...
• Cluster 3: 1 represented system
...


Secretion system
System definition
T1SS
OMF
MFP
ABC
ABC MFP OMF
d≤5
OR
ABC MFP
"loner"
d>5
d≤5
OMF
"multi_system"
ABC MFP
T2SS
D
HIJK
C
LG
F
M
E
T4P
pilQ
pilB
pilT/U
pilC
pilAE
pilV/I
pilN
pilO
pilD
pilM
"exchangeable"
T2SS
gspC
gspD
gspE
gspF
gspG
gspH
gspI
gspJ
gspK
gspL
gspM
gspN
gspO
d≤5
tad
rcpA
tadA
tadB
tadC
flp
tadE/F
tadV
tadZ
T3SS
C
J
RS
TUV
Q
N
Flagellum
fliF
fliI
fliN/M
fliP
fliQ
fliR
flhB
flhA
figB
figC
fliE
T3SS
sctC
sctJ
sctN
sctQ
sctR
sctS
sctT
sctU
sctV
d≤10
T4P
T2SS
tad
pilQ
gspD
rcpA
T4SS
B2
B5
B9
B8
B10
B6
B3
B4/B11
D4
rel
T4SS_T
virB4
virB1
virB2
virB3
virB5
virB6
virB8
virB9
virB10
virB11
virD4
rel
d≤30
d_rel ≤60
min_mandatory_genes_required = 3 ; min_genes_required = 7
traU
A
B
E
I
K
L
M
N
O
P
Q
R
T
W
Y
tcpA
rel
d≤30
d_rel ≤60
T4SS_I
traU
tcpA
virB4
virD4
rel
d≤30
d_rel ≤60
pT4SS
T5aSS
passenger domain
PF03797
"loner"
T5bSS
passenger domain
t5bss-translocator
"loner"
T5cSS
passenger domain
PF03895
"loner"
T6SS
I
D
J
M
L
H
B
C
T6SS
tssA
tssB
tssC
tssD
tssE
tssF
tssG
tssH
tssI
tssJ
tssK
tssL
tssM
evpJ
d≤5


0.05 substitutions/site
0 system
1 system
2+ systems
Flagellum
T1SS
T2SS
T3SS
T4SS
T5aSS
T5bSS
T5cSS
T6SS
Nb T1SS
Nb T5aSS
Nb T5bSS
Nb T5cSS
E. coli K12-MG1655 0 5 0 0 Commensal
E. coli K12-W3110 0 7 0 0
E. coli DH1-ME8569 0 10 0 0
E. coli K12-DH1-ATCC33849 0 7 0 0
E. coli K12-DH10B 0 6 0 0
E. coli BW2952/MC4100 0 6 0 0
E. coli UMNF18 1 12 0 0 IPEC
E. coli P12b 0 7 0 1 ?
E. coli H10407 0 10 0 0 IPEC
E. coli UMNK88 1 8 0 0
E. coli ColiB-REL606 0 7 0 1 Commensal
E. coli BL21(DE3) 0 8 0 1
E. coli BL21(DE3) 0 6 0 1
E. coli BL21-Gold(DE3)pLysS AG 0 7 0 1
E. coli ColiC/ATCC8739/DSM 1576/Crooks 0 7 0 0
E. coli HS 0 7 0 1
S. boydii CDC3083-94/BS512 (serotype18) 0 5 0 1 Shigellosis
S. boydii Sb227(serotype4) 0 5 0 1
S. sonnei Ss046 0 4 0 1
S. sonnei 53G 0 7 0 1
E. coli SE11 0 8 0 1 Commensal
E. coli IAI1 0 6 0 1
E. coli KO11FL 0 8 0 1
E. coli W-KO11FL-ATCC55124 0 6 0 1
E. coli W-ATCC9637 0 7 0 1
E. coli W-KO11-ATCC55124 0 8 0 1
E. coli 2009EL-2050 1 14 0 1 IPEC
E. coli 2009EL-2071 1 13 0 1
E. coli 2011C-3493 1 14 0 1
E. coli 55989 0 11 0 1
E. coli 12009 1 9 0 1
E. coli E24377A 0 10 0 1
E. coli 11128 1 5 0 1
E. coli 11368 1 10 0 1
S. flexneri 8401(serotype 5b) 0 4 0 1 Shigellosis
S. flexneri 2457T(serotype 2a) 0 6 0 0
S. flexneri 2002017 0 10 0 0
S. flexneri 301(serotype 2a) 0 9 0 1
E. coli EC4115 2 8 1 3 IPEC
E. coli TW14359 2 7 1 2
E. coli Sakai 2 10 1 2
E. coli EDL933 2 11 1 2
E. coli Xuzhou21 2 10 1 2
E. coli CB9615 1 11 1 2
E. coli RM12579 1 9 1 1
S. dysenteriae Sd197 1 5 0 0 Shigellosis
E. coli 536 2 7 1 1 ExPEC
E. coli CFT073 3 10 0 1
E. coli Di2 2 10 0 1
E. coli Di14 2 10 0 1
E. coli ABU83972 3 9 1 1 Commensal
E. coli APECO1 0 7 0 2 ExPEC
E. coli S88 1 6 0 2
E. coli UM146 1 5 1 2 IPEC
E. coli UTI89 1 5 1 2 ExPEC
E. coli IHE3034 0 6 0 2
E. coli ED1a 0 6 0 5 Commensal
E. coli LF82 0 5 0 1 IPEC
E. coli NRG857C 0 5 0 1
E. coli SE15 0 7 0 1 Commensal
E. coli NA114 0 7 0 1 ExPEC
E. coli E2348/69 0 3 0 1 IPEC
E. coli SMS-35/SECEC 0 9 0 3 Commensal
E. coli CE10 1 7 0 2 ExPEC
E. coli IAI39 0 6 0 1
E. coli UMN026 1 12 0 1
E. coli O42 2 12 0 2 IPEC
E. fergusonii ATCC 35469 2 5 0 1
A
B1
B2
D
E
F


A
T1SS
T2SS
T3SS
T5aSS
T5bSS
T5cSS
T6SS
E. coli K-12 MG1665
Ori
Ter
1402
8
106
111
116
117
161
293
346
347
382
392
403
466
469
571
579
600
624
641
696
722
791
847
866
909
945
1060
1153
1156
1191
1349
1402
Nb sub-families
Position
Plasmids
B
Nb sub-families
Nb strains