

CompositeSearch

Tutorial

JS Pathmanathan, P Lopez, F-J Lapointe and E Baptiste

STEP 1: SEQUENCE SIMILARITY

BLASTP parameters:

-seg yes

-soft_masking true

-max_target_seqs 5000

**-outfmt "6 qseqid sseqid evalue pident
bitscore qstart qend qlen sstart send
slen"**

STEP 2: CLEAN BLASTP OUTPUT

Your sequence name are:

1) integers :

`./cleanBlastp -i blastp.out`

2) not integers :

`./cleanBlastp -i blastp.out -n 1`

OUTPUT FILES:

- `blastp.out.cleanNetwork`**
- `blastp.out.cleanNetwork.genes`**
- `blastp.out.cleanNetwork.dico` (if you used `-n 1`)**

STEP 3: CompositeSearch

Find help: `./compositefinder -help`

STEP 3: CompositeSearch

(algorithm: composites)

Example:

```
./compositeSearch -i blastp.out.cleanNetwork -n blastp.out.cleanNetwork.genes -m composites  
-e 1e-05 -p 30 -c 80 -l 20 -t 20
```

composites



CompositeSearch

(algorithms: composites)

CompositeSearch give 3 files for the gene families detection.

-File 1 : family.nodes

-File 2 : family.edges

-File 3 : family.info

CompositeSearch

(algorithm: composites)

- CompositeSearch give 3 files for the gene families detection:

- family.nodes : gene families sequences ID (fasta format)

```
>F6
268103
201232
10
```

- family.edges : gene families sequences similarity (fasta format)

```
>F6
201232 268103
10      201232
10      268103
```

- family.info : family ID, Nb nodes, Nb edges and connectivity

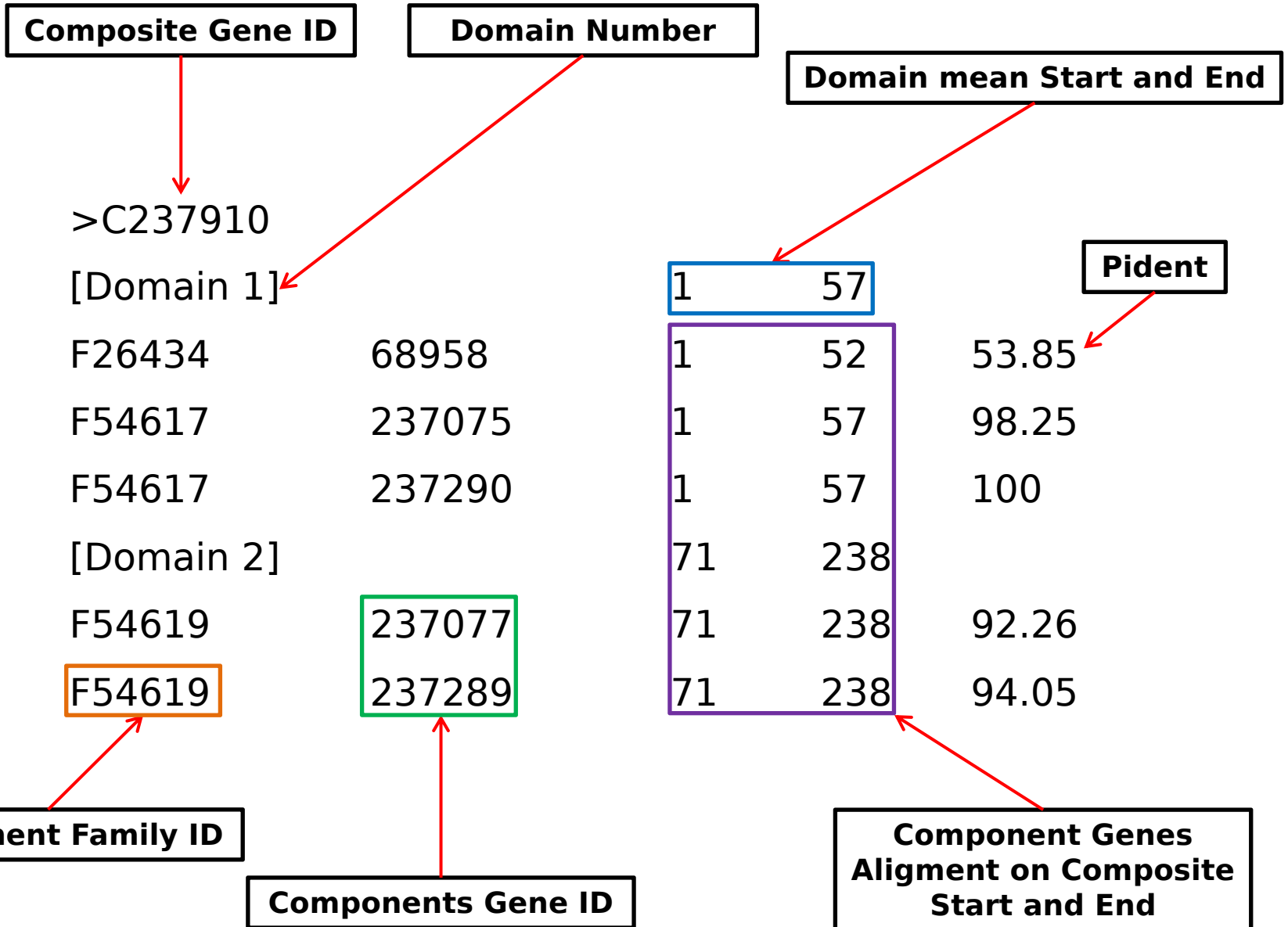
CompositeSearch

(algorithm: composites)

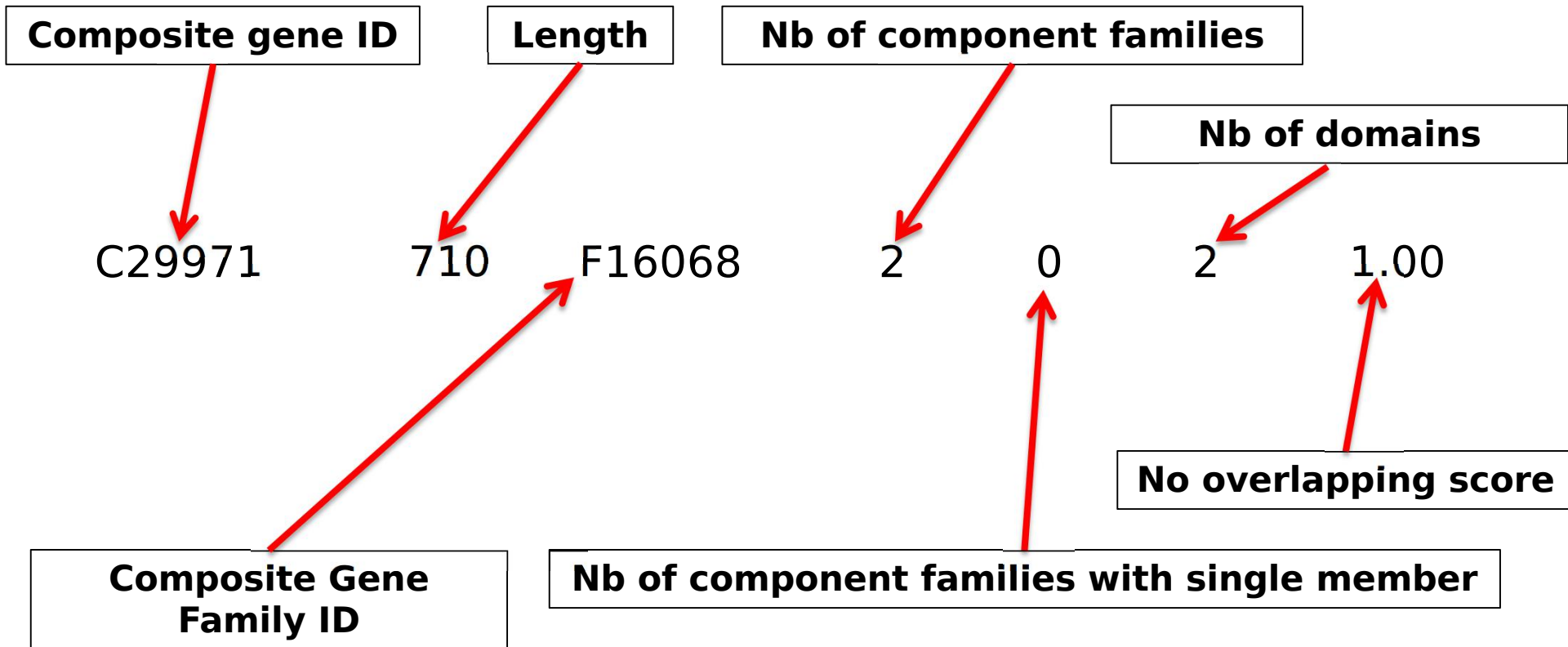
CompositeSearch give 2 files for the composite genes detection:

- File 1 : xxxxx.compoistes (detailed information)
- File 2 : xxxxx.compositesinfo (general information)

File 1 : xxxxx.composites



File 2 : xxxxx.compositesinfo



CompositeSearch

(algorithm: composites)

CompositeSearch give 2 files for the composite genes detection:

- File 1 : xxxxx.compoistefamilies (detailed information)
- File 2 : xxxxx.compositefamiliesinfo (general information)

File 1 : xxxxx.compositefamilies

Composite gene family ID

**Percentage of composite
genes composed with the
component family**

>CF65771

F4007

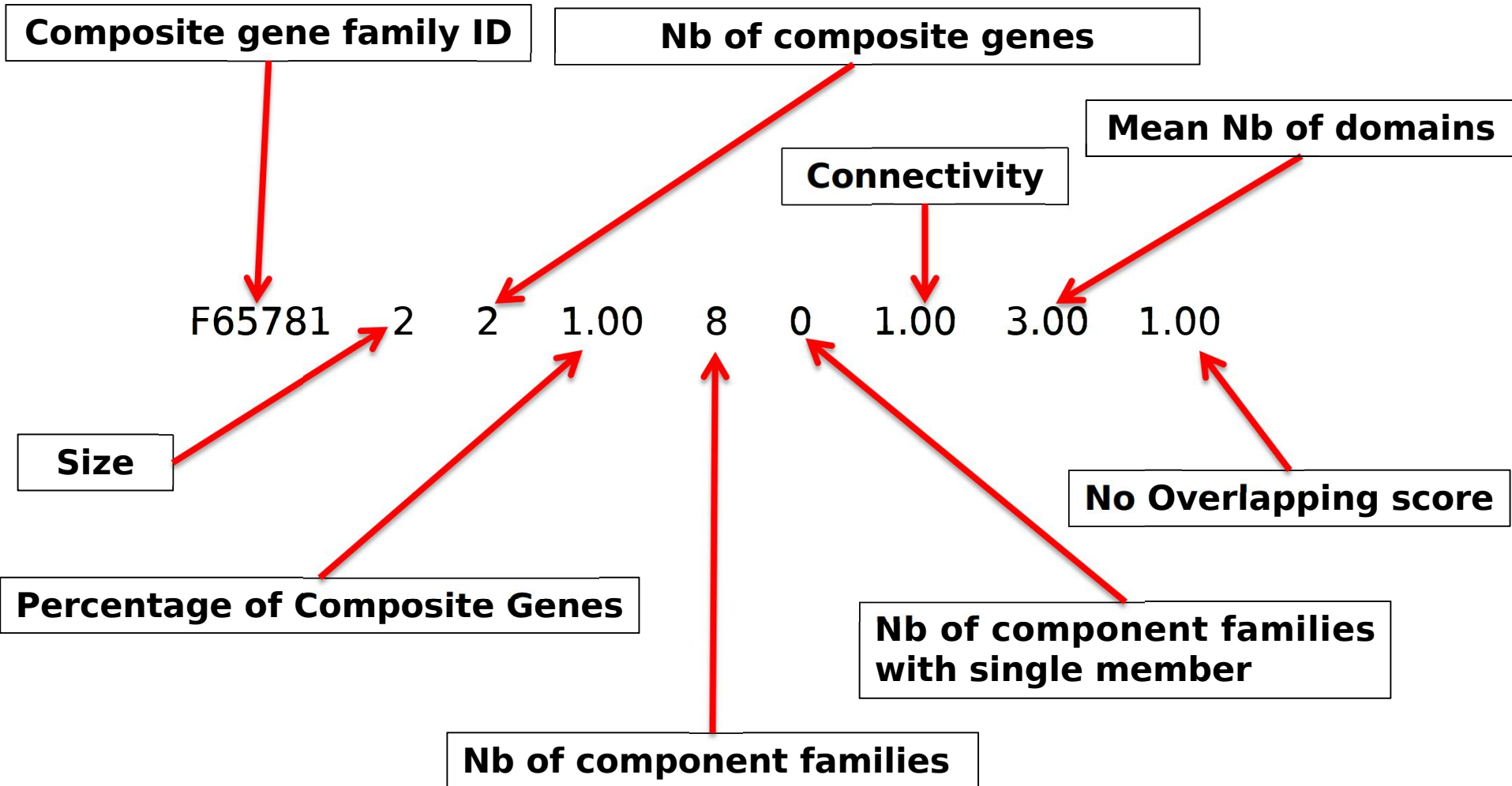
F50491

100.00

100.00

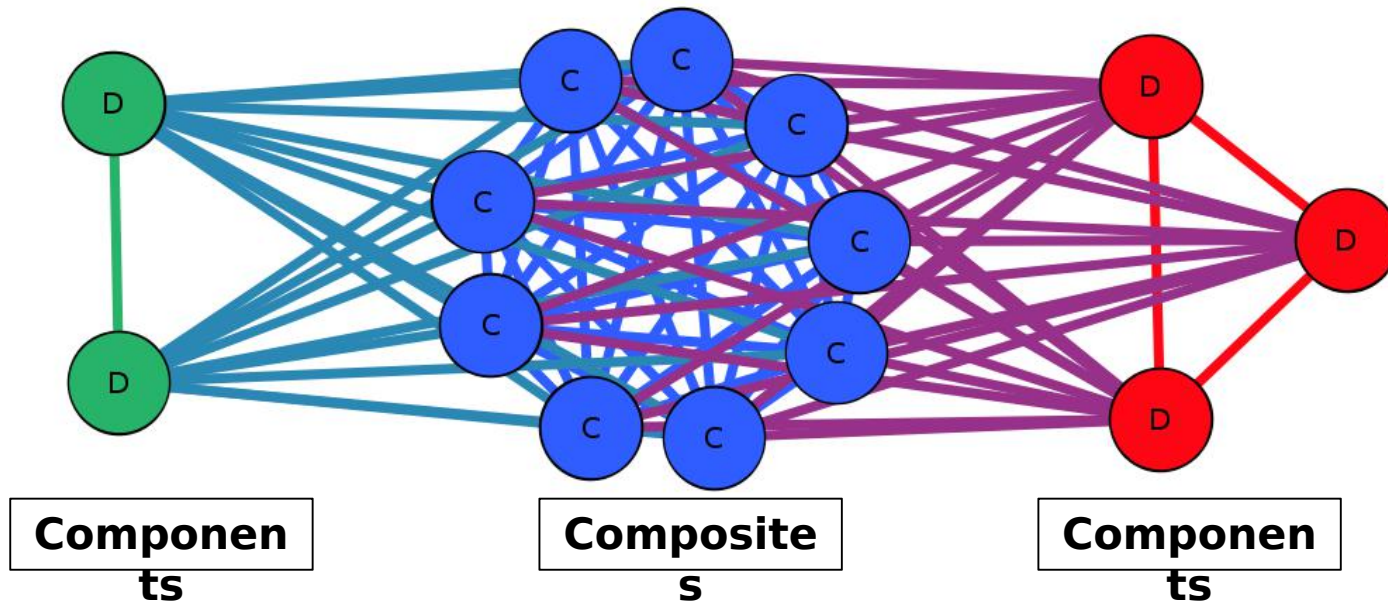
Component gene families ID

File 2 : xxxxx.compositefamiliesinfo

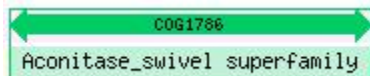


PLASMIDS

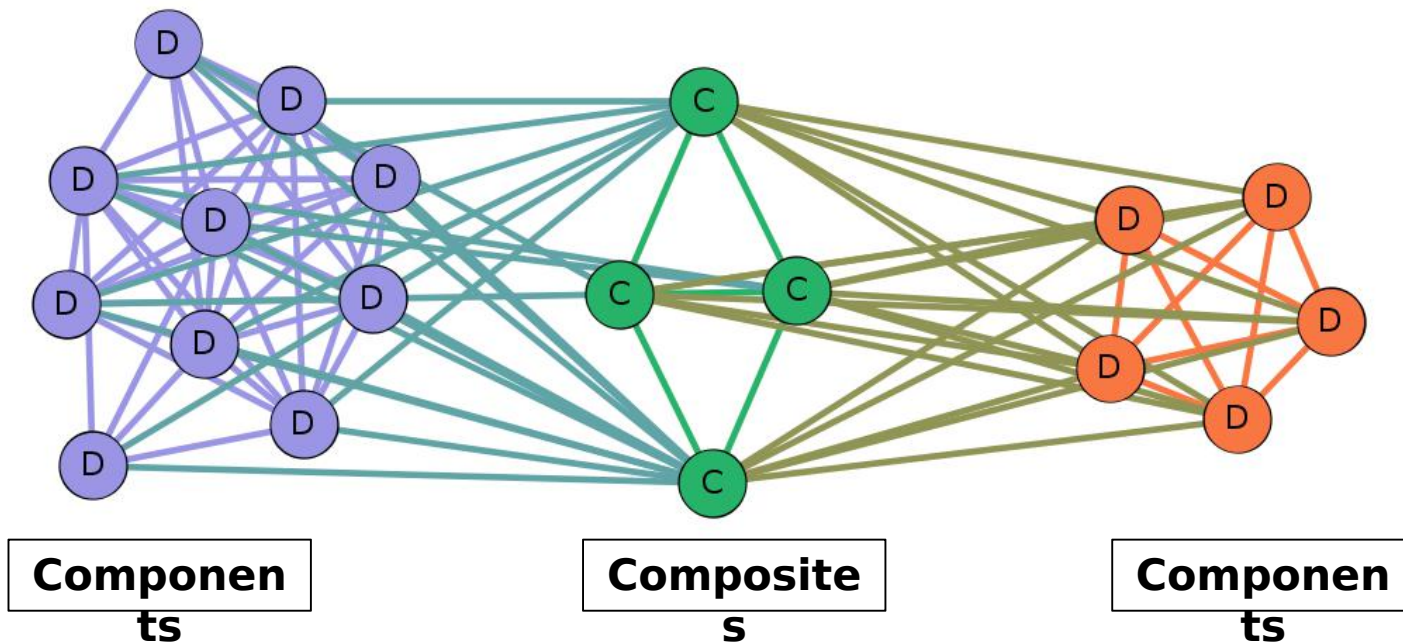
EXAMPLE N°1 (clique)



BLASTP



EXAMPLE N°2 (quasi clique)



BLASTP

