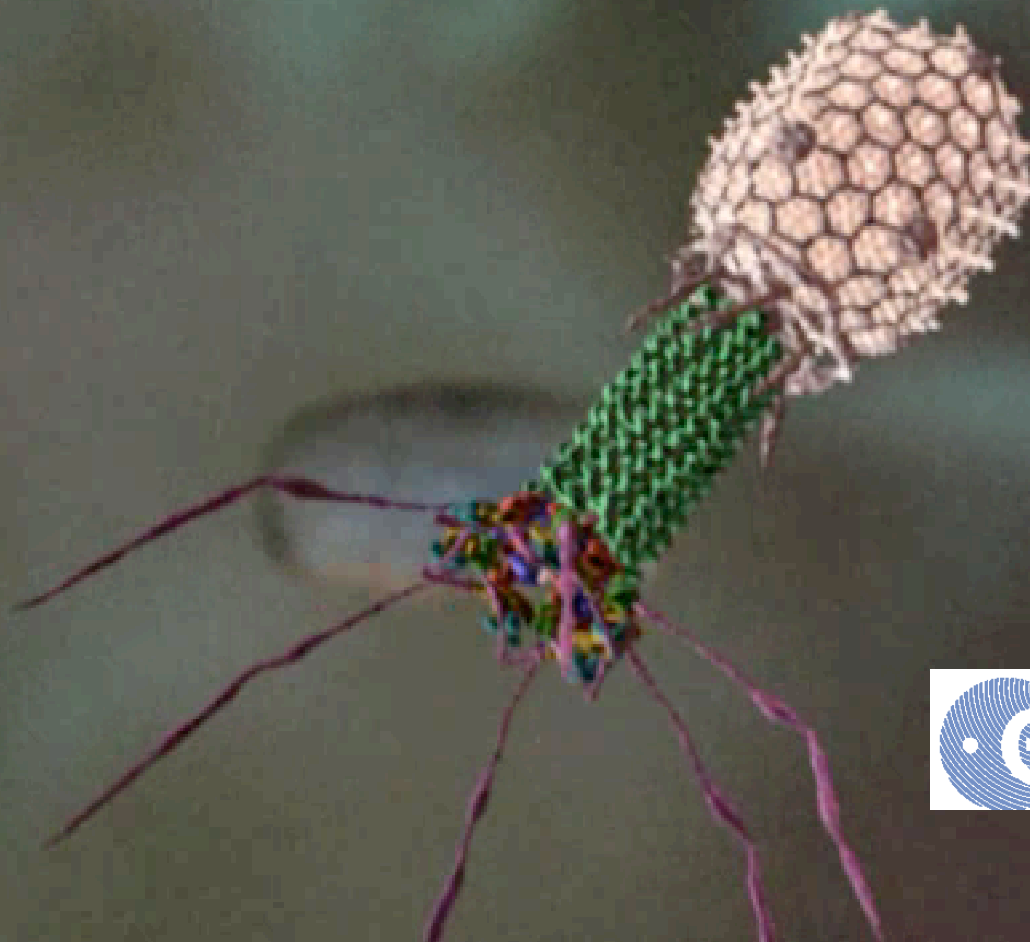


Phage and other MGEs ontologies

Ariane Toussaint, Gipsi Lima-Mendez, Raphaël Leplae
SCMBB,
Université Libre de Bruxelles



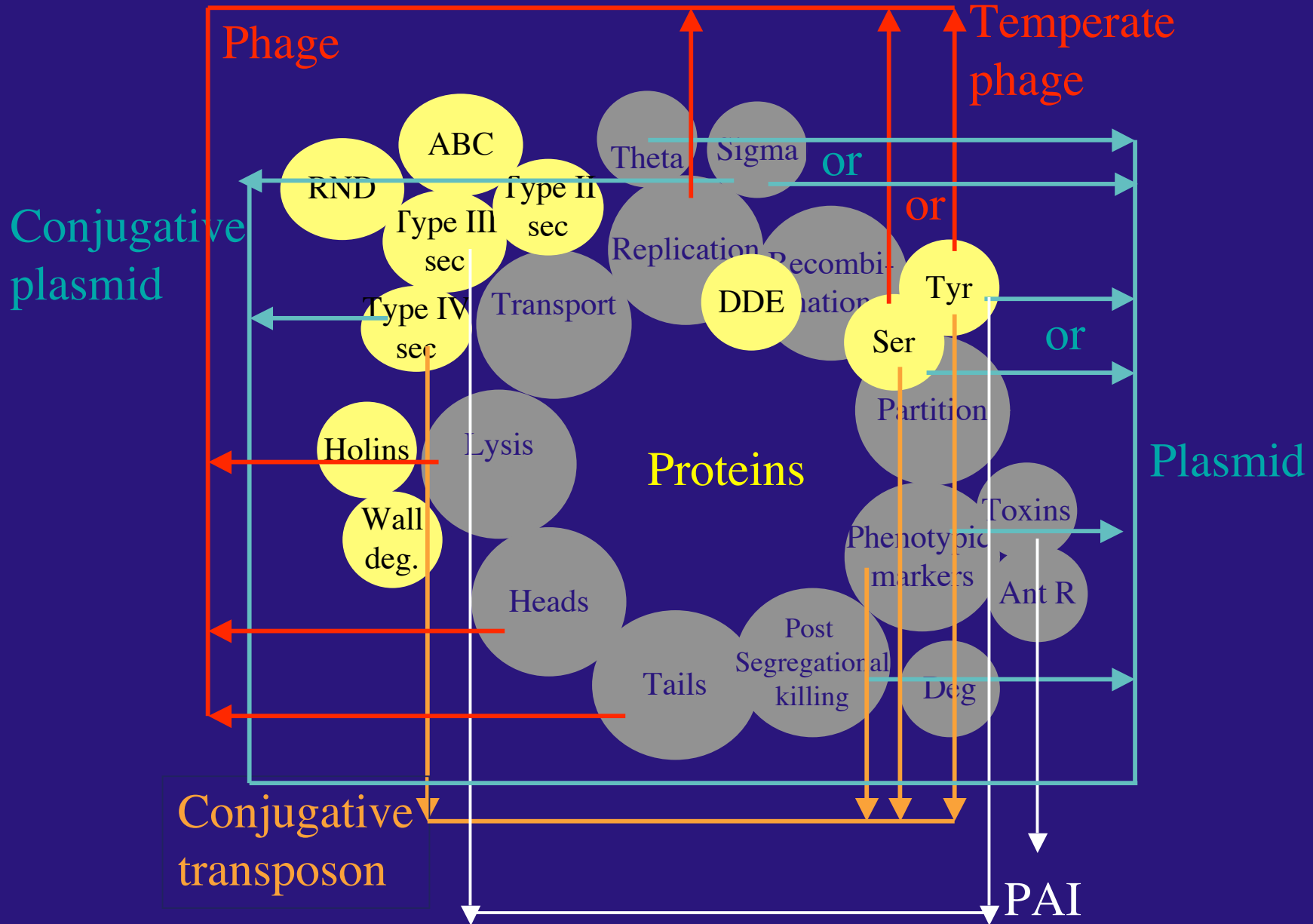
ULB
SCMBB



FNRS

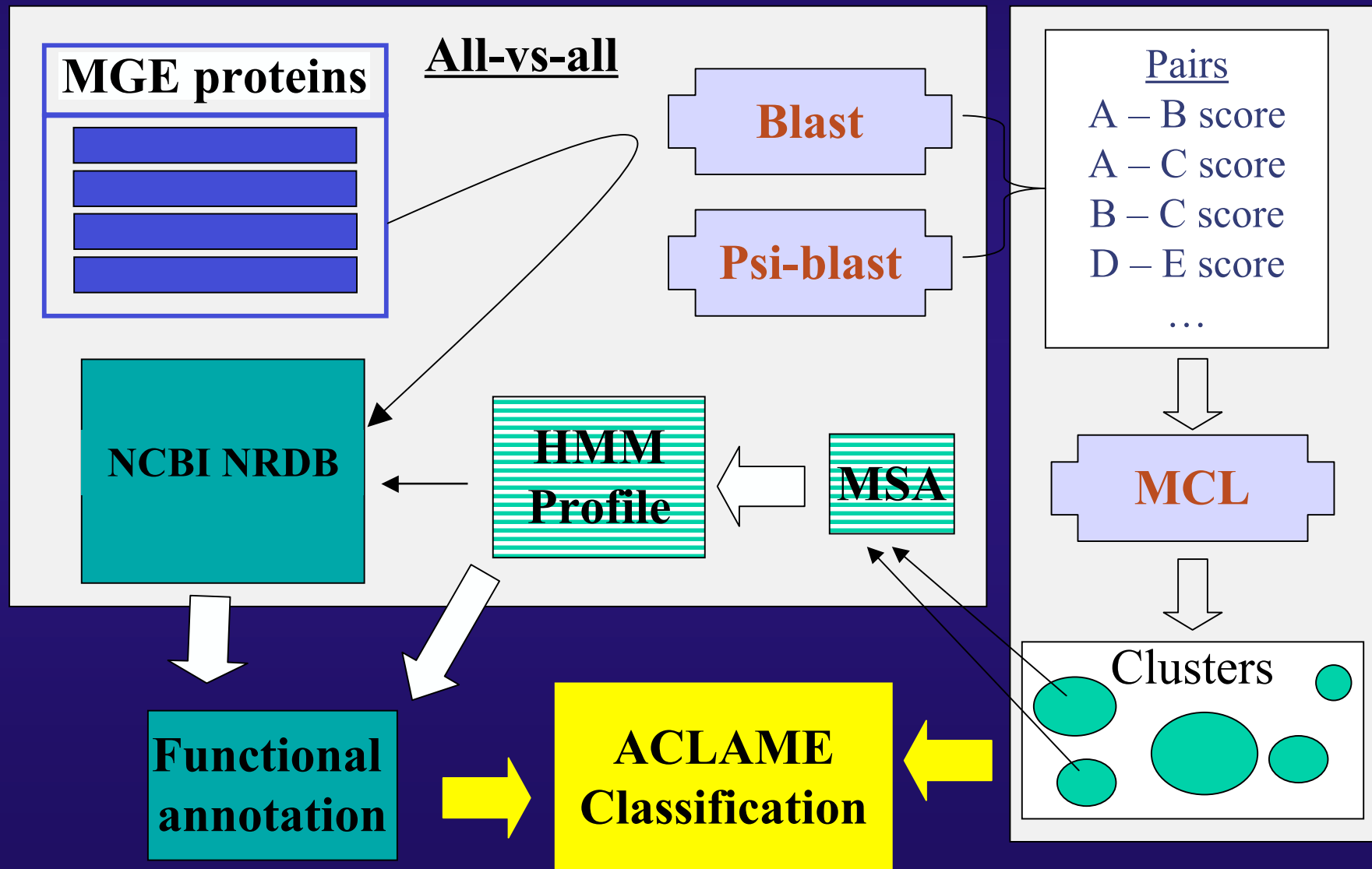
Basic ACLAME concept (Merlin et al. 2000)

Reconstruction of Various Bacterial MGE's



Generating protein families

Proteins clustering



Results for query sequence **gi|56692569|ref|YP_164037.1|**

Blast search results in Viruses

Hit vir:protein:1974 **[+]**

Description *Original annotation: Mor; ACLAME function(s): transcriptional activator Clusters: cluster:vir:997*

Bits | 34

Eval | 3.e-03

%ID | 26 | 24 | 88 | 4 | 42 | 45 | 4 | 4

vir:protein:1974 **[+]** (Length: 129) - *Original annotation: Mor*

ACLAME function(s): transcriptional activator

Clusters: cluster:vir:997

Bits: 34 Eval: 3.e-03 %ID: 26 Query range: 15-106 Hit range: 31-118

gi 56692569 ref YP_164037.1	15	MLSELAEELIVEAFVRNGLSREKAVSESEELVFQLHRRWAGITFVFPVKDELARKRLELHILQRYDGSNADKLVREFGVTE
vir:protein:1974	31	LLAELNDLLRGELSRLGVDP----AHSLEIVVAICKHLGGGQVYIPRGQALDSLIRDRLRIWDFNGRNVSELTRYGVTF

gi 56692569 ref YP_164037.1	95	DWIYSVVRKHRR
vir:protein:1974	107	NTVYKAIRRMRR

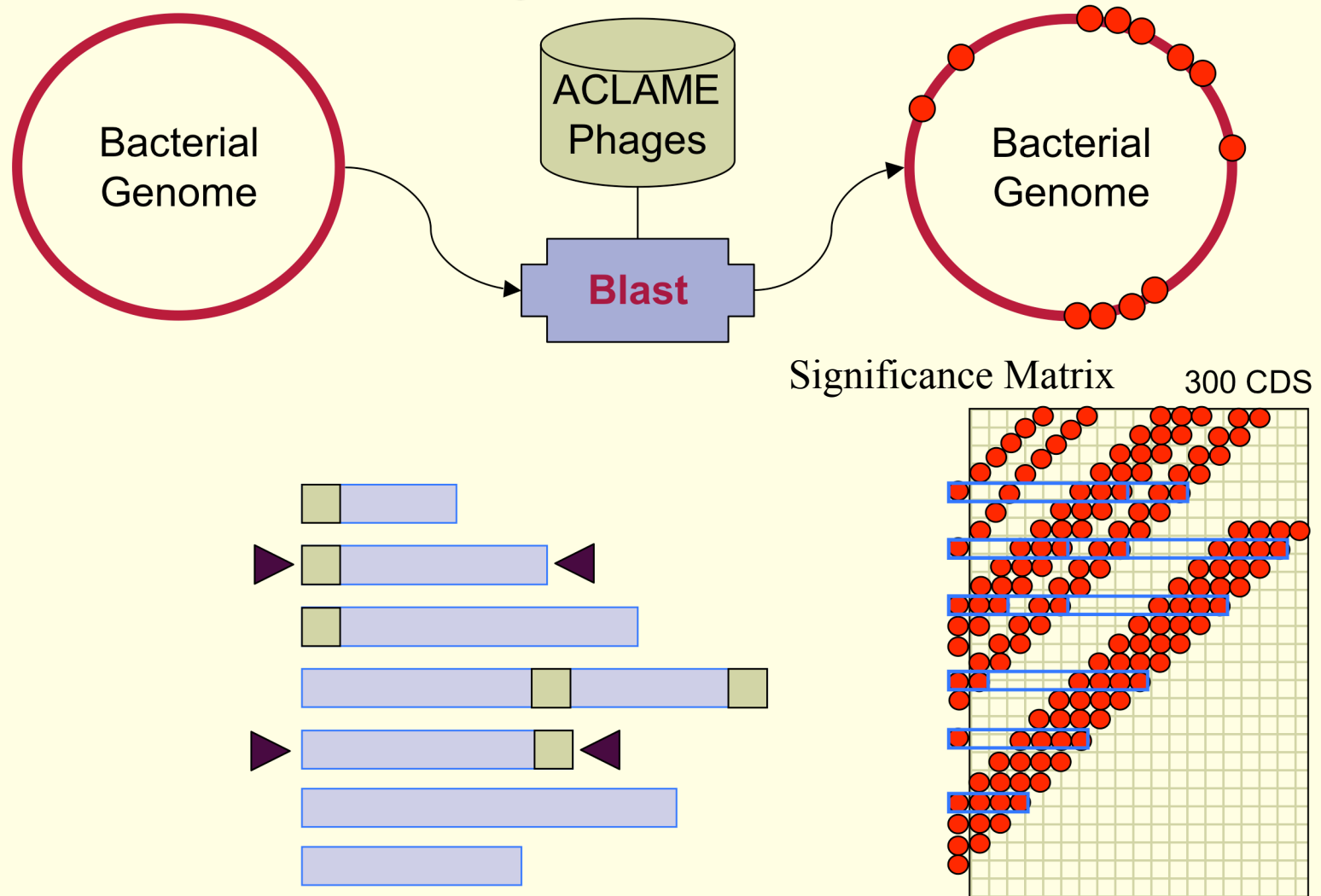
Results for query sequence **gi|56692571|ref|YP_164039.1|**

Blast search results in Viruses

No hits found!

Prophinder

- Using the information in ACLAME to detect other MGEs in bacterial genomes.



Prophage ID [NC_004347:300:2](#) [Size](#) 26.87 kb (26873 bp) [Return to genome view](#)
 Bacterial genome [NC_004347 - Shewanella oneidensis MR-1, complete genome](#). [Size](#) 4969.80 kb - (4969803bp) circular
 Nb CDS in prophage 300 With ACLAME hits 25 Number of DBs 0 [Info](#)

ACLAME: Clusters view

[NC_004347](#) 0.0kb
 300 CDS
[NC_004347:300:2](#) 279
 CDS

ACLAME

Login: Password:

[Classification](#) [Browse](#) [Description](#) [Proposal](#)

Control panel

- [Home](#)
- [Classification](#)
- [Browse](#)
- [Description](#)
- [Proposal](#)
- [Forums](#)
- [Ontology](#)
- [Resources](#)
- [People](#)
- [Community](#)
- [Annotation servers](#)
- [Ontologies](#)
- [Genomes](#)
- [Financial supports](#)
- [Publications](#)
- [Tools](#)
- [Annotators](#)
- [Blast](#)

Details (?)

Prophage [NC_004347](#)
 List of CDS

GI number	C
24374206	2795
24374207	2796
24374208	2796
24374209	2797
24374210	2798
24374211	2798
24374212	2798
24374213	2799
24374214	2799
24374215	2800
24374216	2800
24374217	2800
24374218	2801
24374219	2801
24374220	2801
24374221	2803
24374222	2805
24374223	2806

Clusters view

Categories (?) : Jump (?) : Version 0.2 Change to (?) :

Clustered proteins in families filtered based on Viruses:

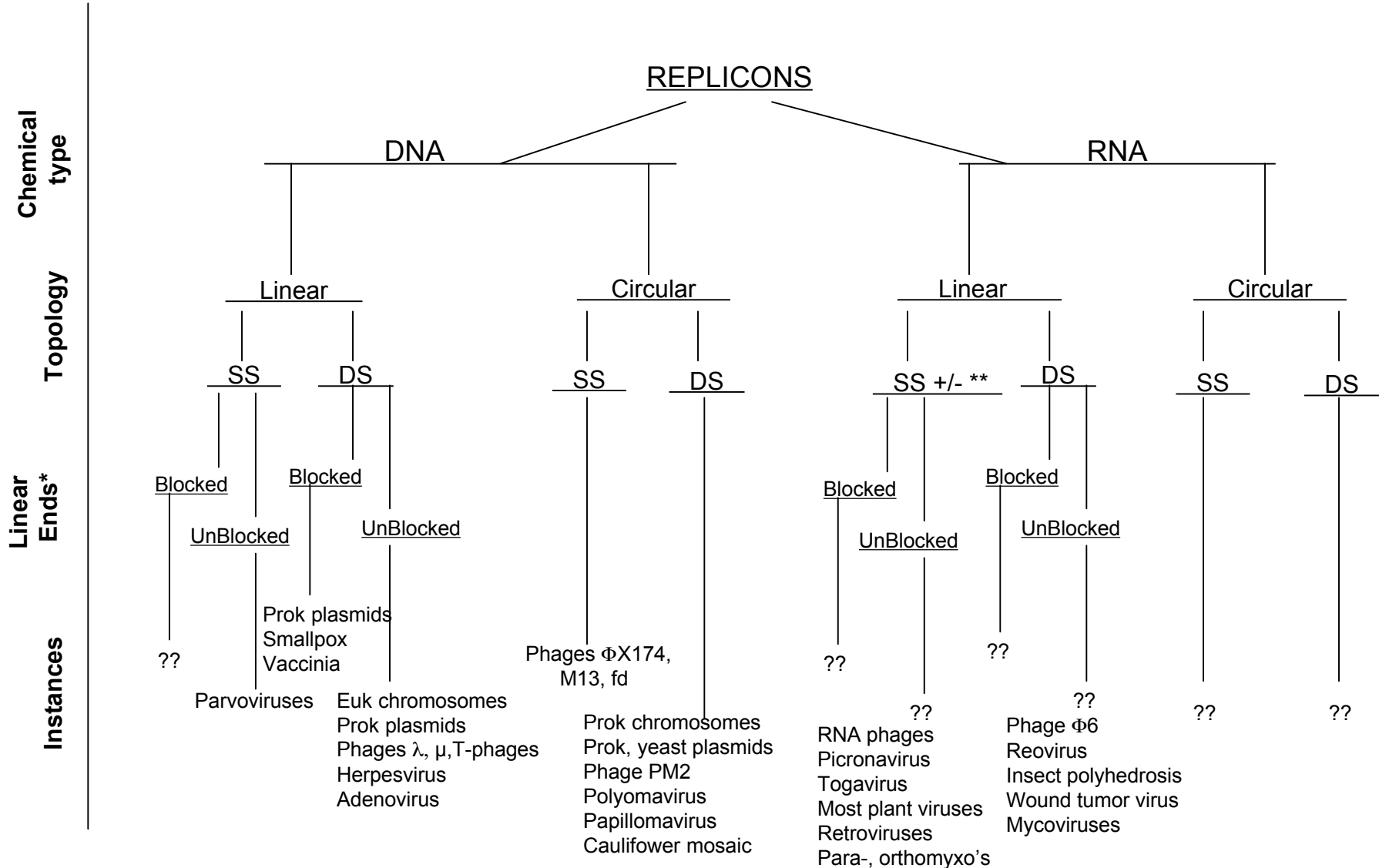
[Cluster ID](#) cluster:vir:997 [Nb of proteins](#) 2 [Cluster level](#) Families

[Function](#) function:182: transcriptional activator (

GO0016563

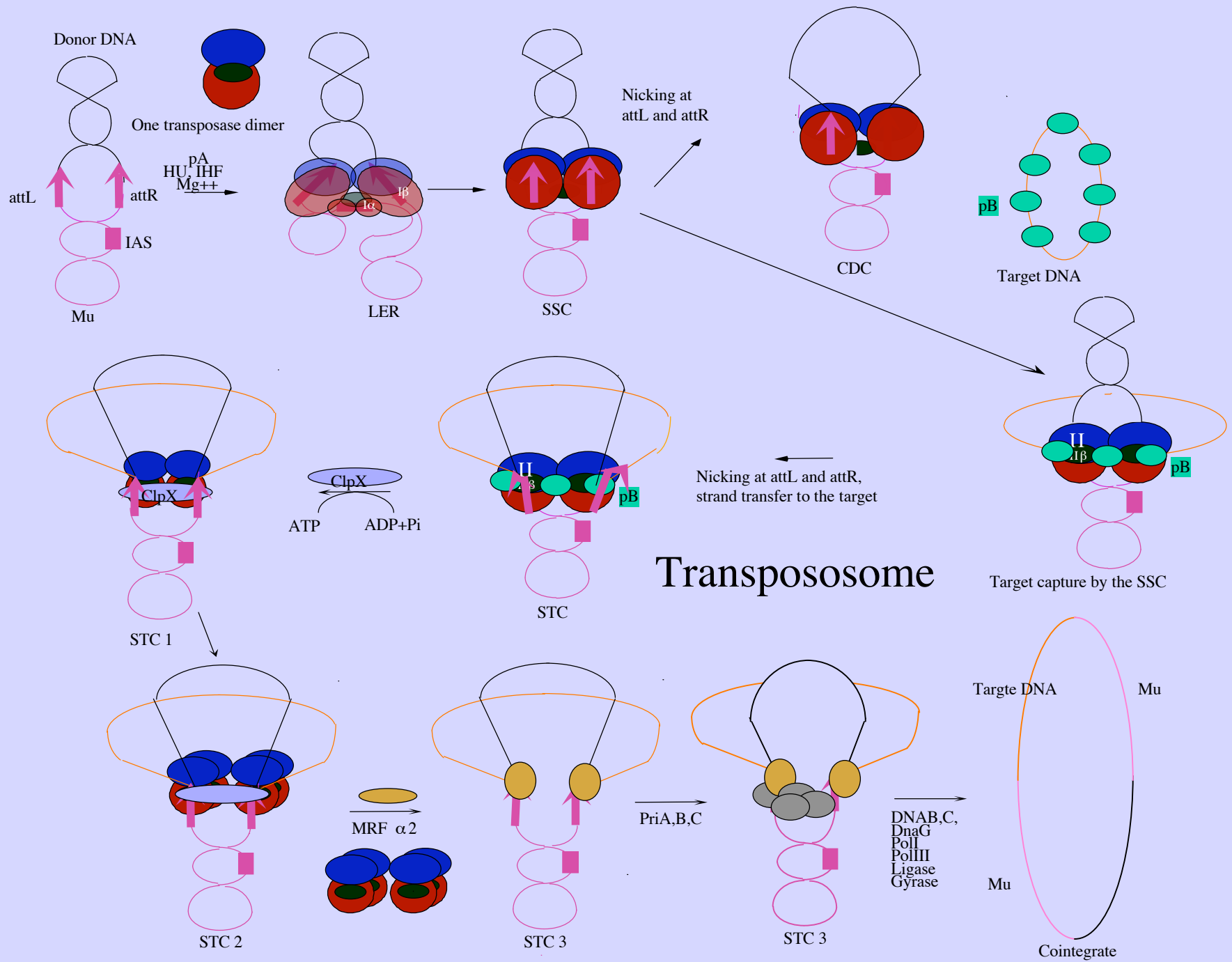
ID	[sw]	[nrdb]	[scop]	Description	Length	Genome and host(s)
vir:protein:1974	[+]	[4]	[36]	gpMor, middle operon regulator	129	Enterobacteria phage Mu Host(s): Escherichia coli
vir:protein:1978	[+]	[5]	[33]	gpC, late transcription activator	140	Enterobacteria phage Mu Host(s): Escherichia coli

A HIERARCHY OF REPLICONS BASED ON CHEMISTRY AND TOPOLOGY

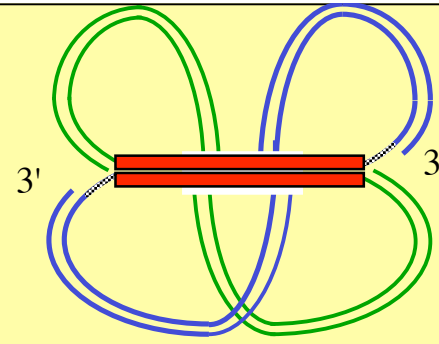


*Ends of linear replicons can either be free (unblocked) 5'phosphates and 3' hydroxyls or blocked in short hairpins (in SS) or covalently crosslinked (DS).
 ** SS RNA replicons can either be directly translatable mRNA's (+) or complementary to mRNA (-), needing to be copied before used as mRNA.

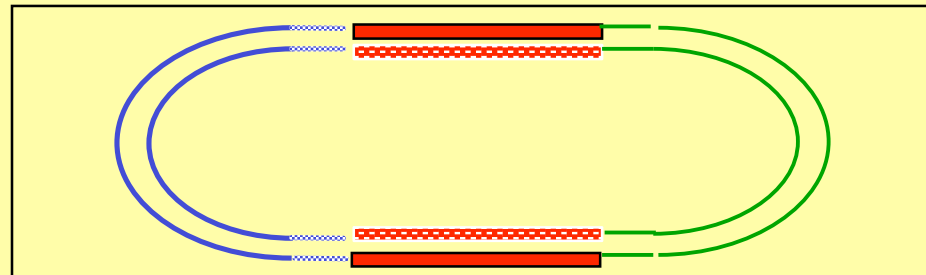
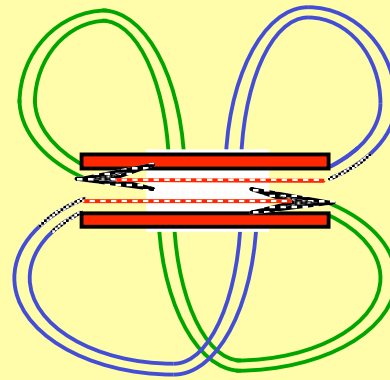
Replicative transposition



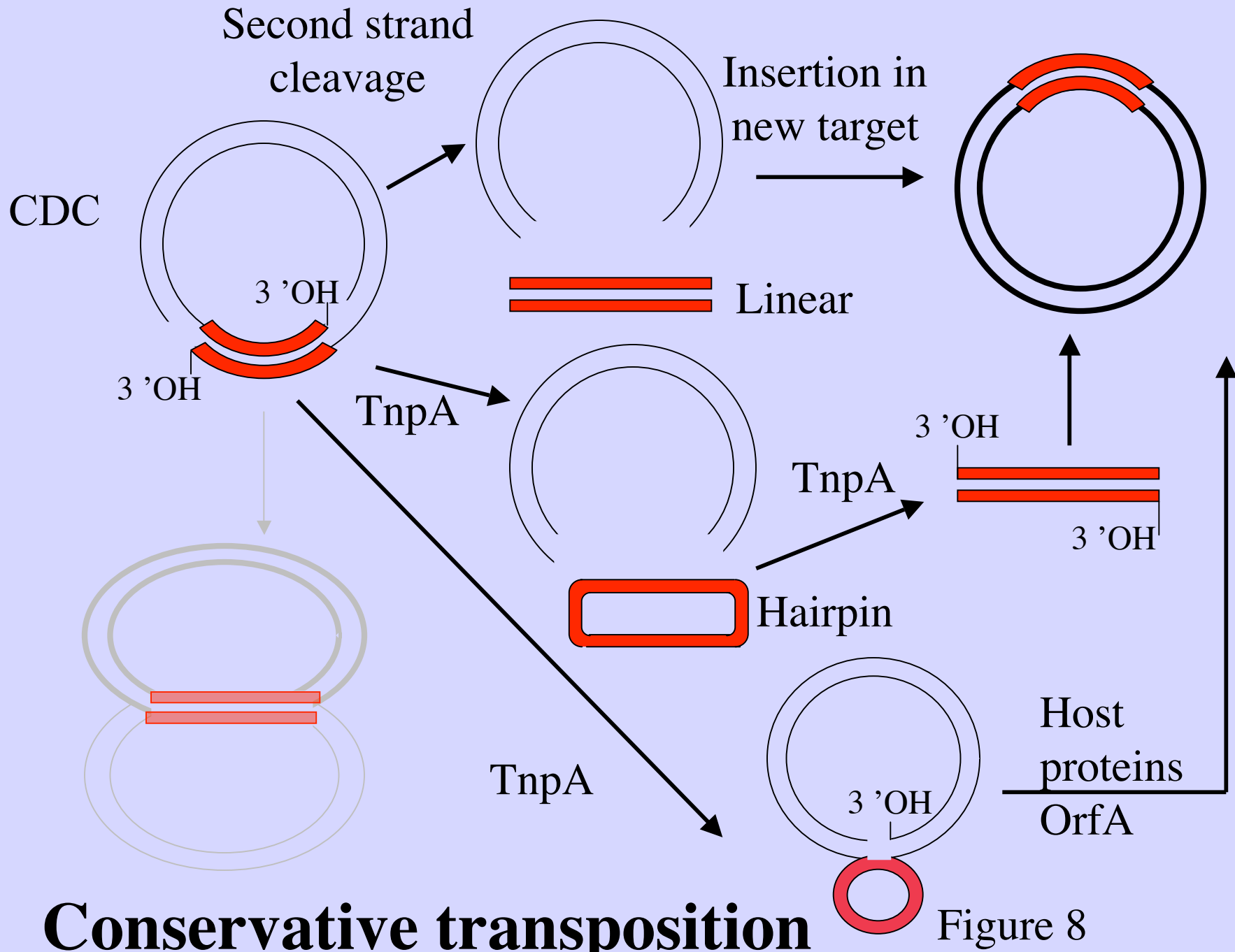
STC



**DNA replication
from 3'-OH target**



**Cointegrate
(replicative transposition)**



attL
attR are TnpA binding sites (transpososome)
IAS (internal activating sequence of transposition)
 IHF (integration host factor protein) binding site (specific)
 HU (histone-like protein) binding site (non specific usually)
 TnpA (transposase) nicking sites at the ends of *attR* and *attL*
 (concerted)
 DNA replication priming site? (primosome assembly site)

Except for *IAS* and, in many cases, priming assembly site, these are also valid for IS and Tn's, which use DDE transposases. Are usually called *IRL* and *IRR* and in composite transposons distinguish outside and inside *IRs*



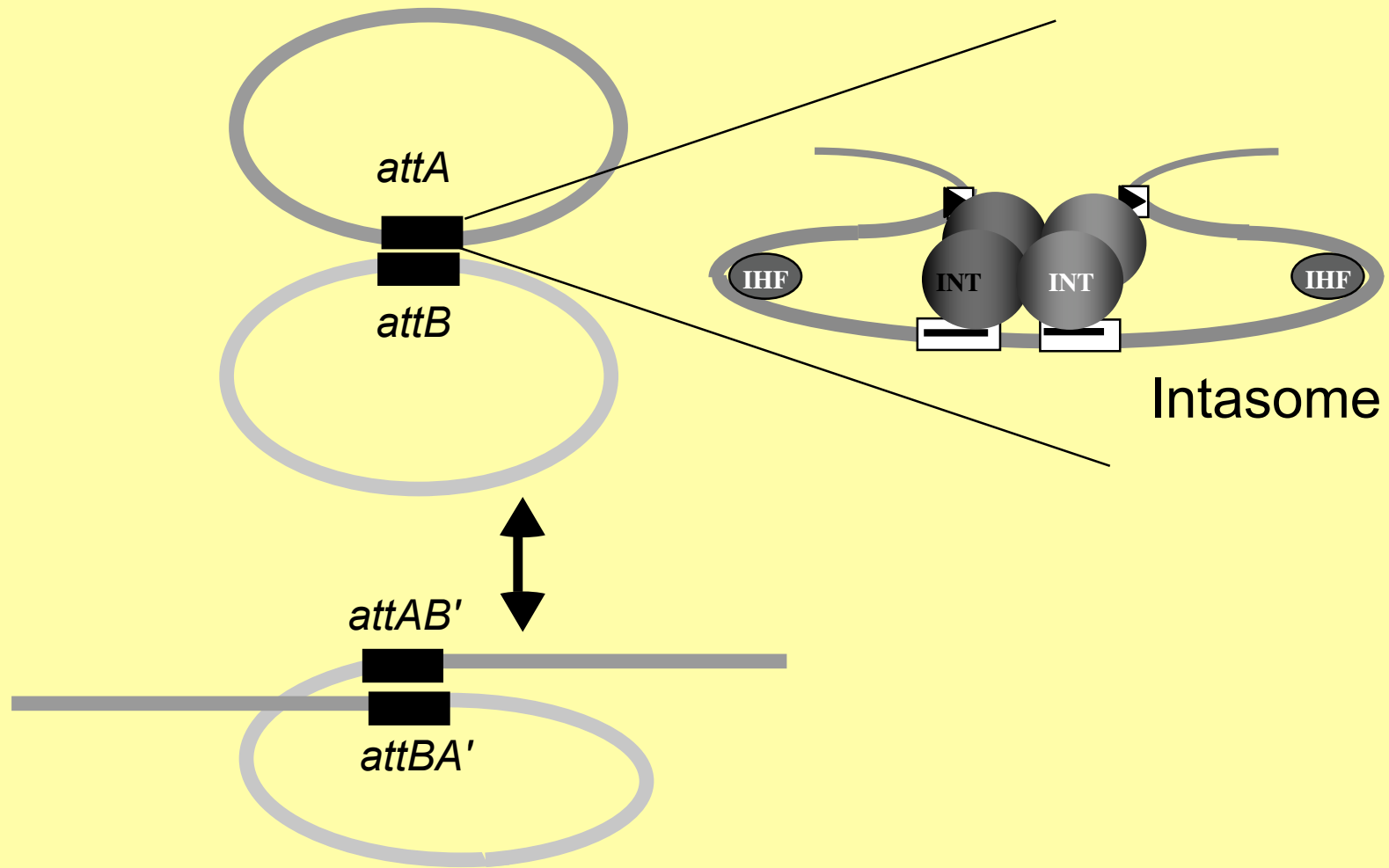
Frameshifting sequence (IS3 family at least)

Q: isn't it a process? Signals for frameshift are in RNA but process occurs during translation! Frameshift signals more OK for SO?

X_XXN ou X_XXZ_ZZN and SD 9-15 bp upstream of previous motif or hairpin (IR GC rich) 5-9 bp long after the previous motif.

! Looking at VDJ definitions we will need every gene on the planet!

Site-specific recombination



For temperate phages:

attP (and also *attB* for bacterial chromosomes and, *attBP'*, *attPB'* when inserted prophage).

Int binding sites (distinction to be made between Int N-term and C-term binding site)

Int nicking site

(*Int* stands for Tyr or Ser recombinases here!)

When Ser or Tyr recombinases are resolvases:

res site to be subdivided into Res binding sites and Res cleavage site.

These binding and cleavage sites are also valid for other islands, which use Tyr or Ser recombinases to integrate and excise, *attI* from integrons and *attC* from gene cassettes
- Do they need to be distinguished and if yes how? *Sensu...*

For Tn7-like make sure to also consider
TnsA and C (?) binding (cleavage) sites and
preferential chromosomal *att* site *attTn* and
its TnsD binding site.

And ???

PROCESSES

Under GO:0006310: DNA recombination.

site-specific recombination:

- integration

- resolution (transposon-sensu)

- resolution (plasmid-sensu)

transpositional recombination:

- replicative transpositional recombination

- conservative transpositional recombination

 - hairpin intermediate

 - figure 8 intermediate

 - linear intermediate

illegitimate recombination ??

more??

ACTIVITIES

In GO:

integrase ?

tyrosine-based site-specific recombinase activity

serine-based site-specific recombinase activity

relaxase

PROPOSAL:

site-specific recombinases

integrase (sensu retrovirus): see DDE transposase

integrase (sensu phage-prophage, genomic island and conjugative transposon)

tyrosine-based site-specific recombinase activity

serine-based site-specific recombinase activity

excisionase

resolvase ?

DDE transposase

relaxase

Also need to deal with

- Tn916 family transposases that are Tyr-Rec integrases
- IS607 transposases which may be Ser recombinases
- transposases that are 'relaxases'
 - IS605-608, Y transposases
 - YII IS91 RC replication
 - ...MIMIVIRUS

And : IS91 is a helitron NO rather helitrons are similar to IS91 family (see IS-Finder DB) now called ISCR elements (Toleman et al. MMBR 70, 296 (2006). PNAS ref in OBO only refers to fly.

COMPONENTS (primosome and replisome exist in GO)

intasome

intermediates??

excisome?

intermediates??

relaxosome

intermediates??

transpososome

LER complex

stable synaptic complex (or type 0)

cleaved donor complex (or type 1)

diverse strand transfer complex(es) (including type 2)

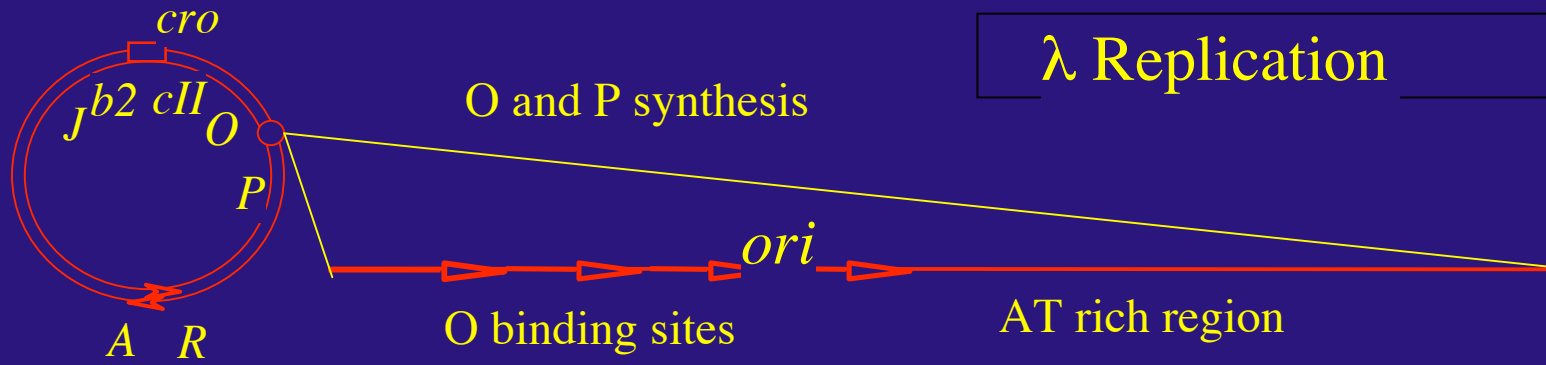
others?

BACK to PHAGES...

Infection and lytic
cycle...

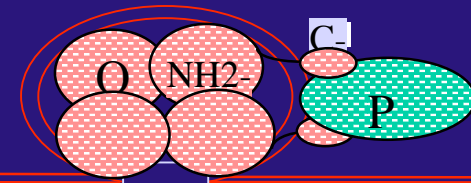


λ Replication

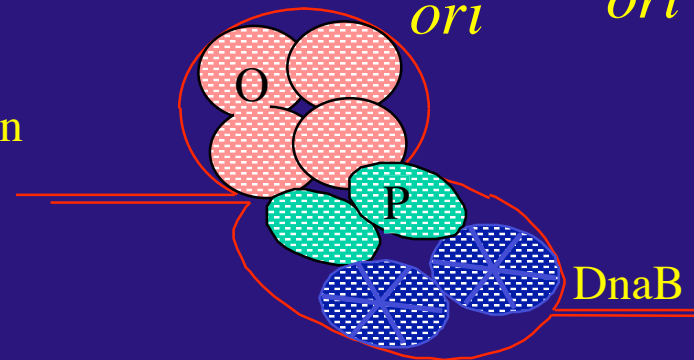


O binds DNA through N-terminus, P through C-terminus

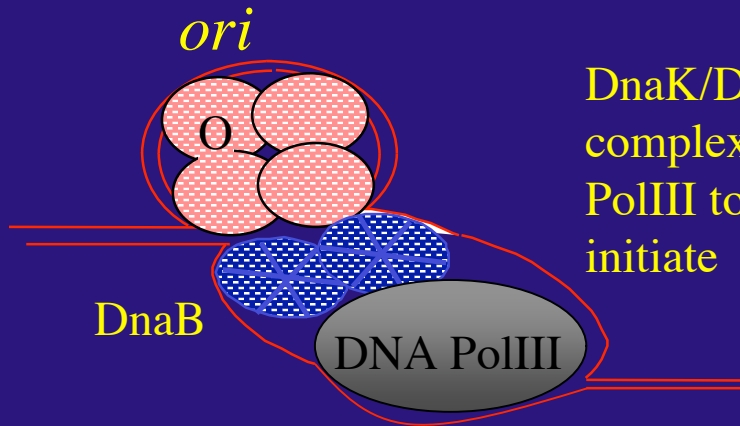
AT rich region



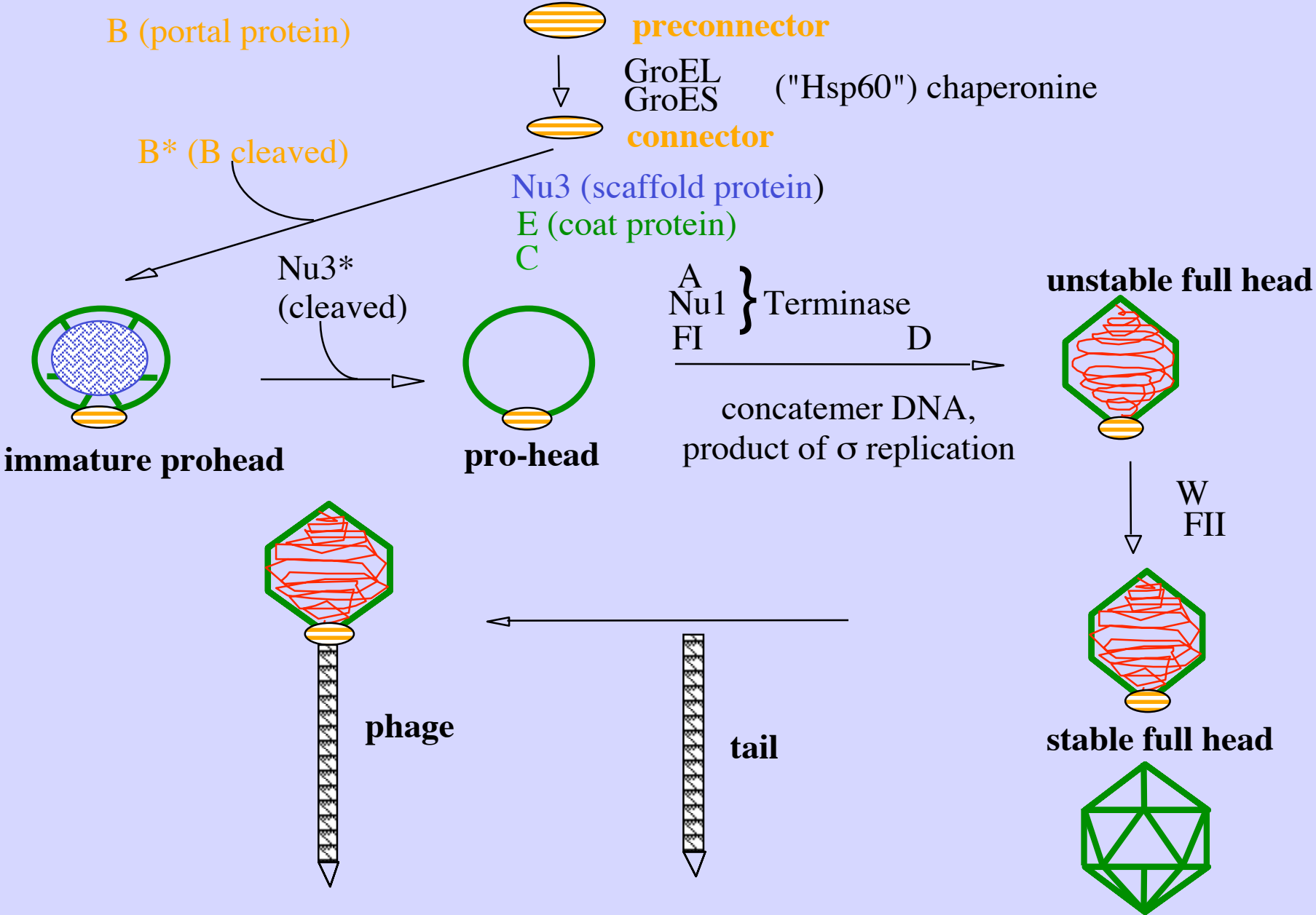
P binds host DnaB helicase
Loop formation at AT rich region
Very stable complex



DnaK/DnaJ, GrpE chaperon machine recognizes the complex. P leaves, DnaB remains which allows DNA PolIII to be loaded and bidirectional replication to initiate



λ morphogenesis



That means for all phages but transposable ones:
oriV, replication proteins binding and nicking sites
(primosome, replisome)

For all phages:

regulatory sequences (same as for bacteria including
transcription antitermination sites)

terminase binding sites

pac

cos (3' and 5' extended) site,
terminase nicking site

*Note: RCR to be distinguished between phages and e.g.
helitrons?*

Various questions:

Where do we enter the phage lytic cycle, the related subprocesses, phage components and related molecular functions (more GO than SO problem)?

Intra vs inter chromosomal transposition. Translocation?

SO:0000182 etc. transposition is via a RNA/DNA intermediate: all have DNA, some have RNA intermediates!

DNA region: gene vs cistron (polycistronic and not polygenic mRNA!)

gene array

gene cassette



Raphaël: ACLAME , developer,
Prophinder on the web



Gipsi: Prophinder,
graph analysis

