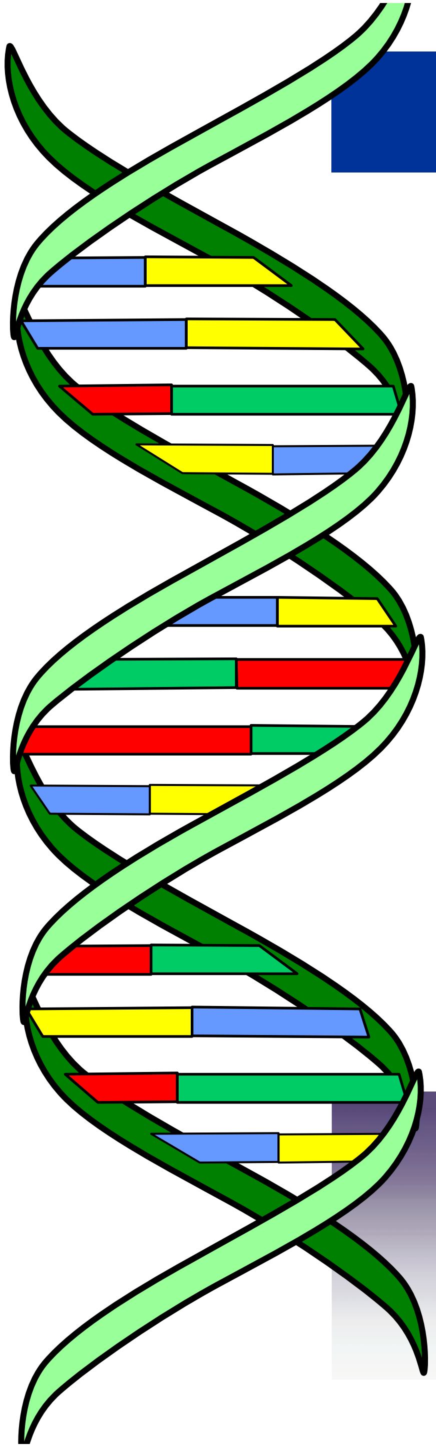


Computational Molecular Biology

Unique Probe Mapping met behulp van PQ-trees

Leiden, 6 april 2006

Hendrik Jan Hoogeboom
Algorithms / Fundamentele Informatica

www.liacs.nl/home/hoogeboo/

feb’01 - human genome

15 February 2001

nature

$10.00

www.nature.com

the **human** genome

Nuclear fission
Five-dimensional energy landscapes

Seafloor spreading
The view from under the Arctic ice

Career prospects
Sequence creates new opportunities

naturejobs
genomics special

16 February 2001

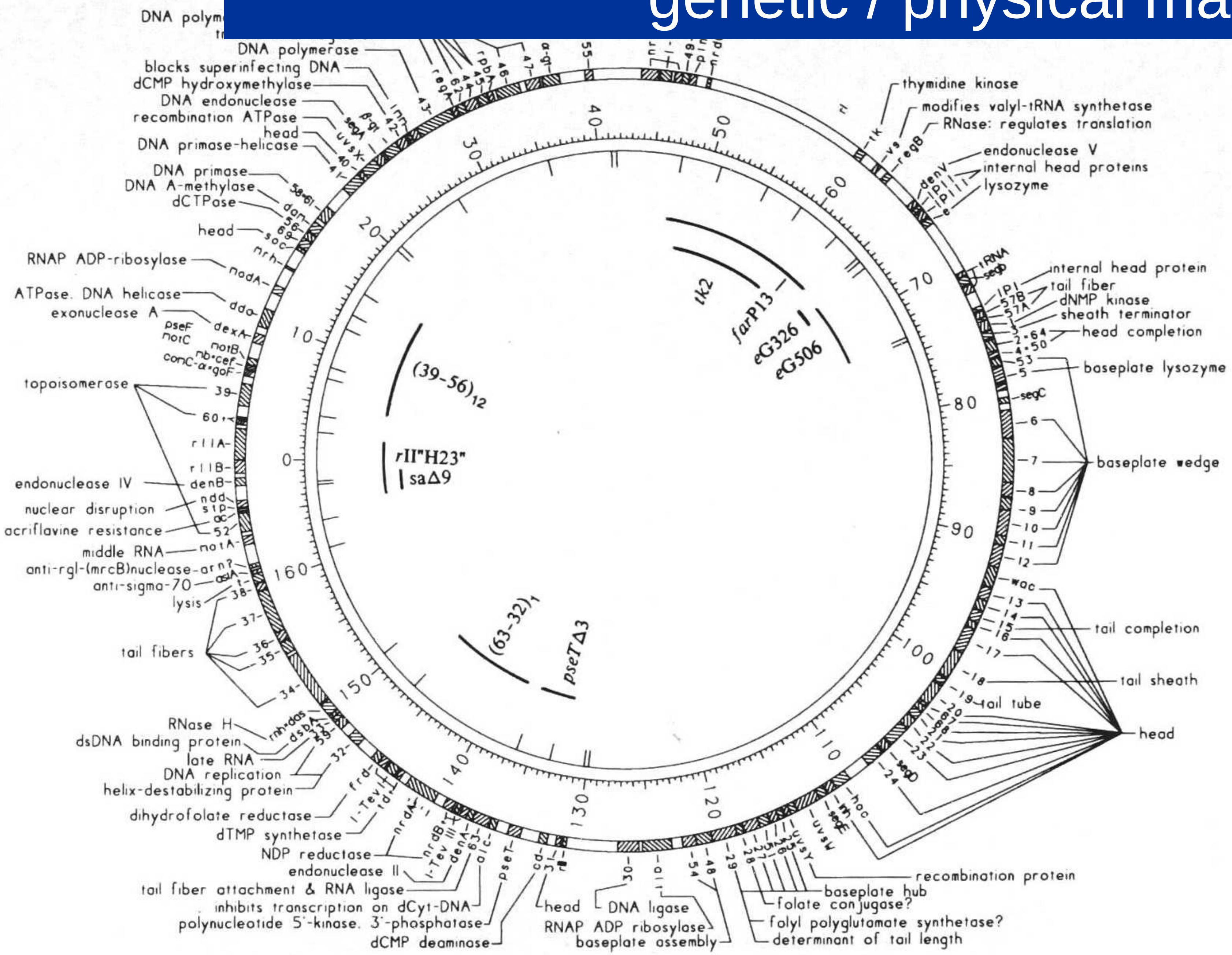
Science

Vol. 291 No. 5507
Pages 1145–1434 $9

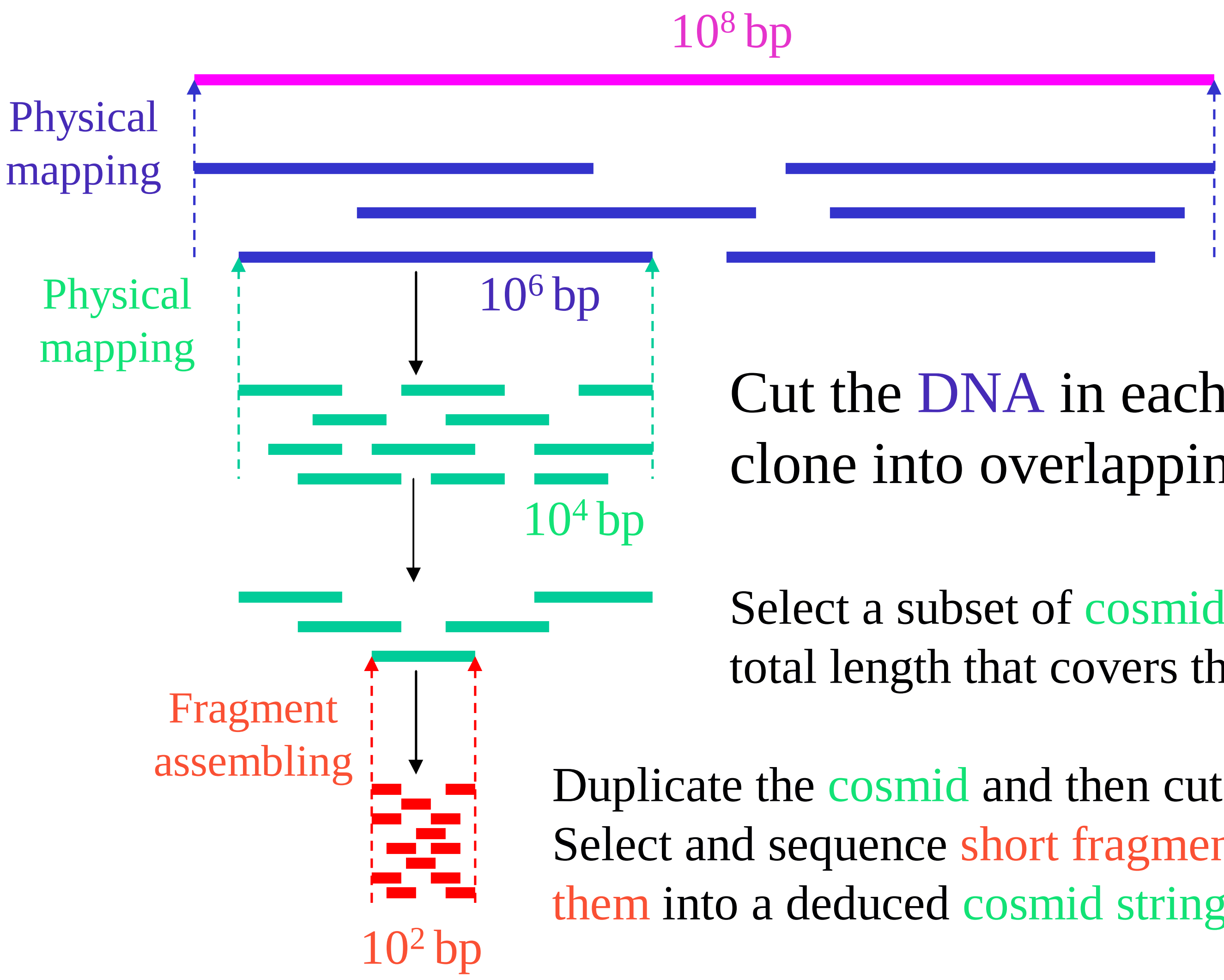
THE HUMAN GENOME

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

genetic / physical map



physical mapping



C: Full DNA

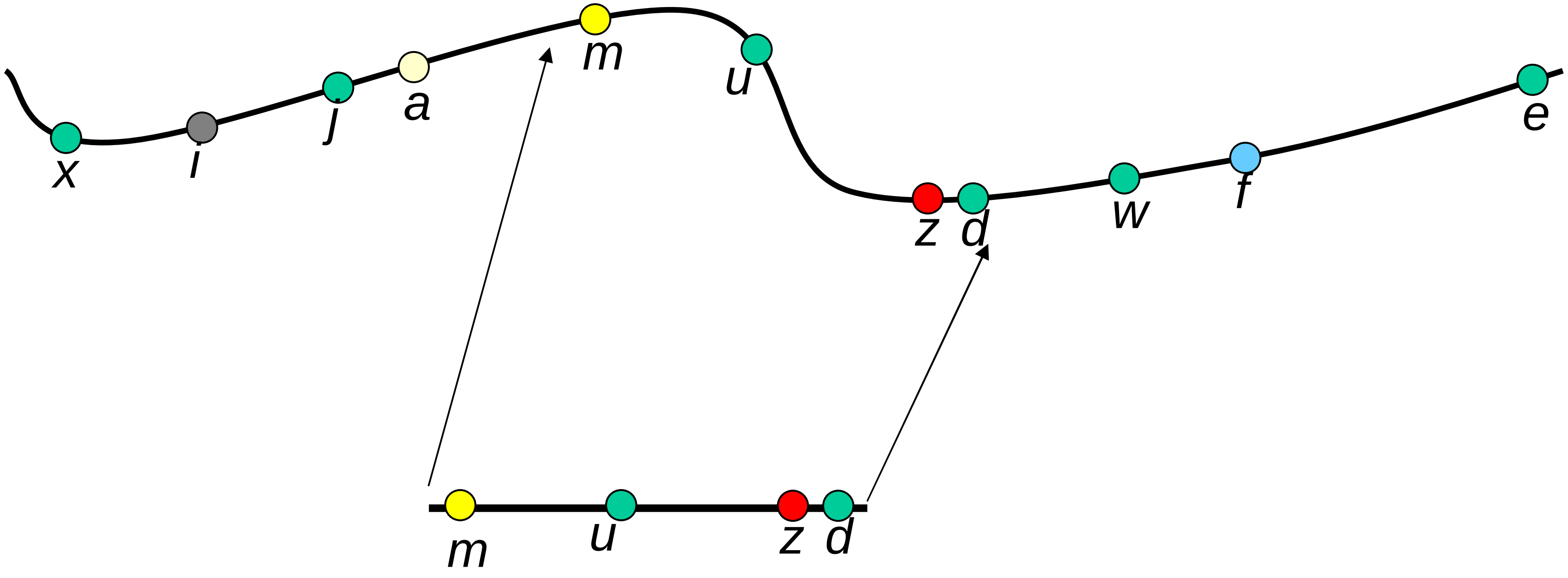
Cut *C* and clone into overlapping YAC clones.

Cut the DNA in each YAC clone and clone into overlapping cosmid clones.

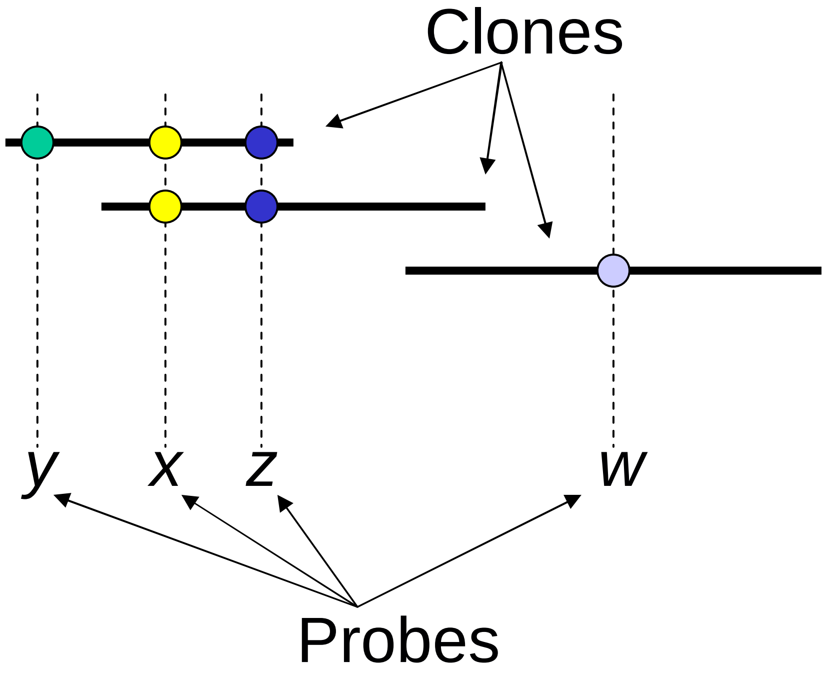
Select a subset of cosmid clones of minimum total length that covers the YAC DNA.

Duplicate the cosmid and then cut the copies randomly. Select and sequence short fragments and then reassemble them into a deduced cosmid string.

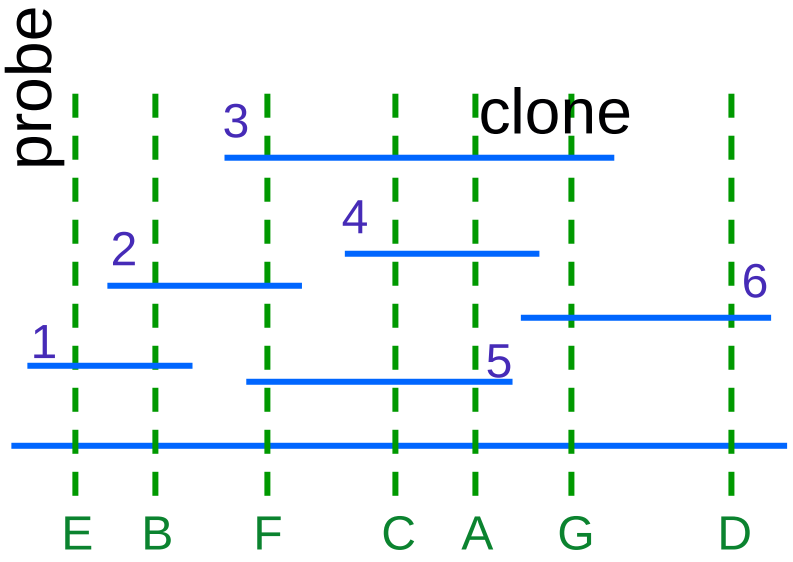
using a map of the genome



hybridization mapping



unique probe mapping

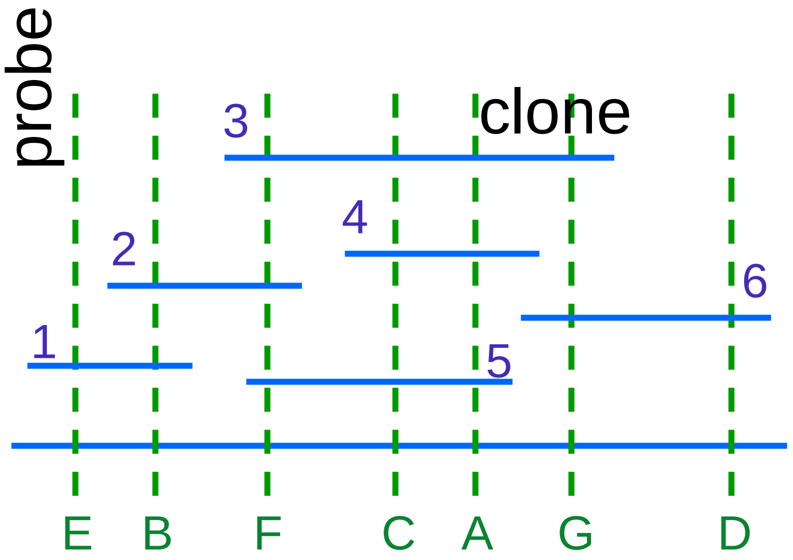


clones 1,2,…,6
probes A,B,…,G

matrix representation

	A	B	C	D	E	F	G
1		1			1		
2		1				1	
3	1		1			1	1
4	1		1				
5	1		1			1	
6				1			1

reordering of probes



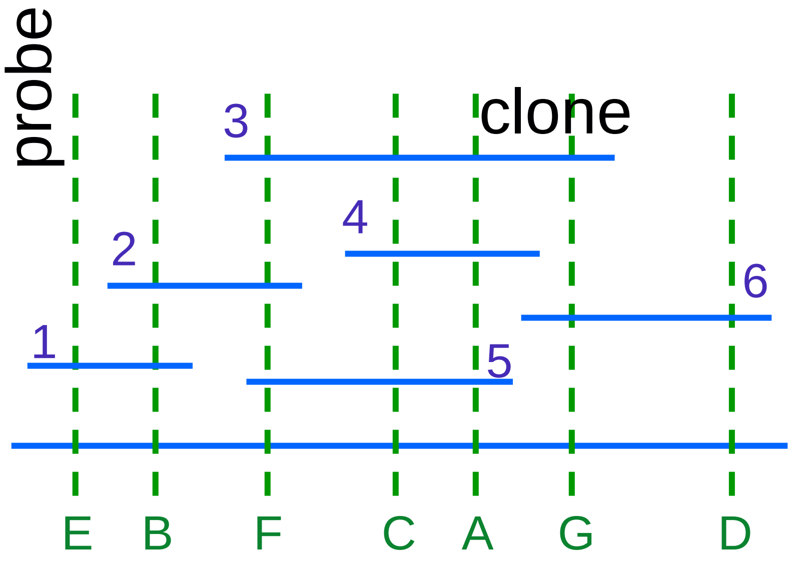
clones contain
consecutive probes

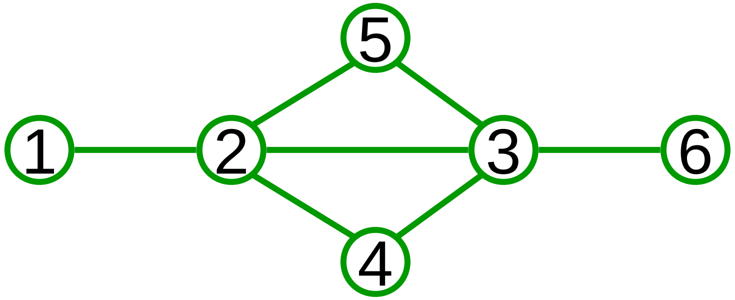
ordering

	D	G	C	A	F	B	E
1						1	1
2					1	1	
3		1	1	1	1		
4			1	1			
5			1	1	1		
6	1	1					

	A	B	C	D	E	F	G
1		1			1		
2		1				1	
3	1		1			1	1
4	1		1				
5	1		1			1	
6				1			1

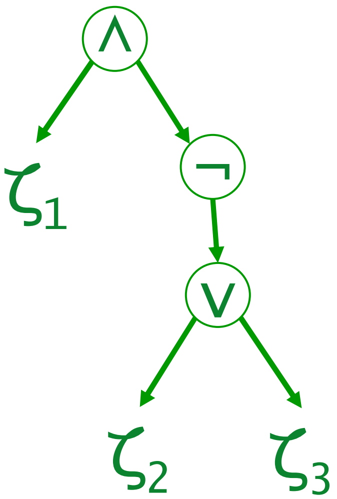
interval graphs



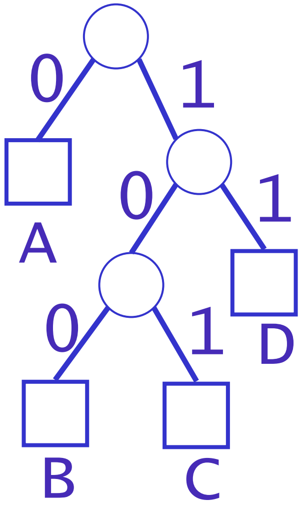


matrix representation

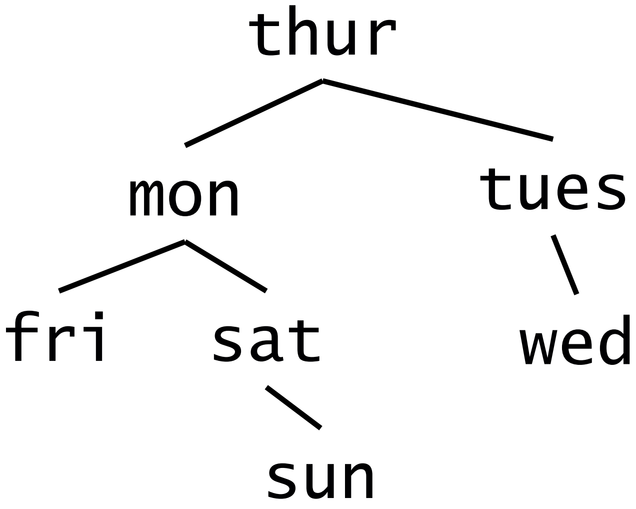
	A	B	C	D	E	F	G
1		1			1		
2		1				1	
3	1		1			1	1
4	1		1				
5	1		1			1	
6				1			1



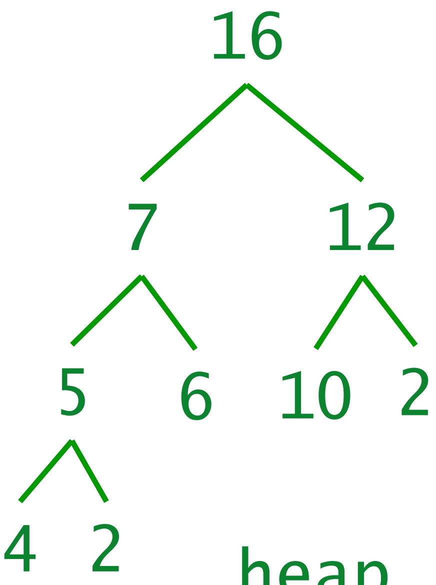
expressie



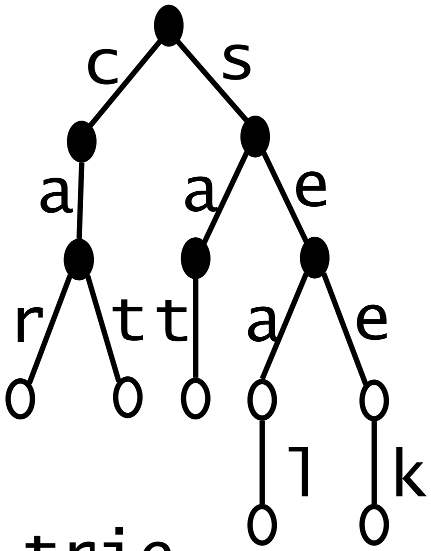
code



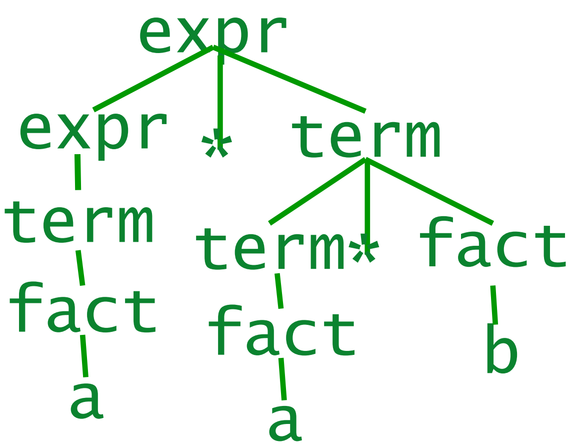
binaire zoek



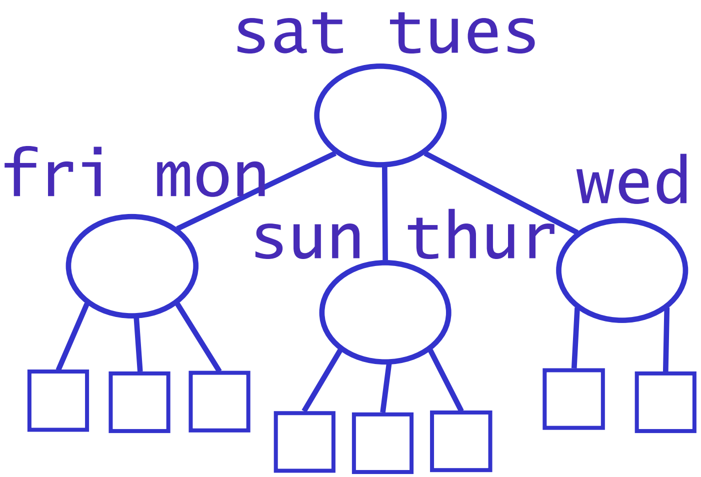
heap



trie



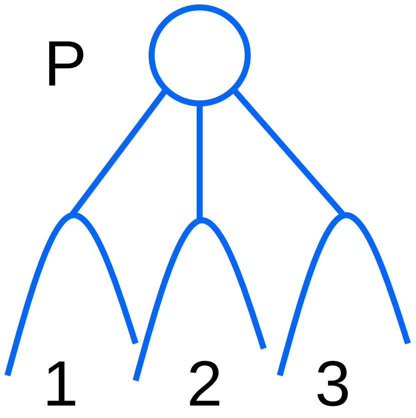
syntax



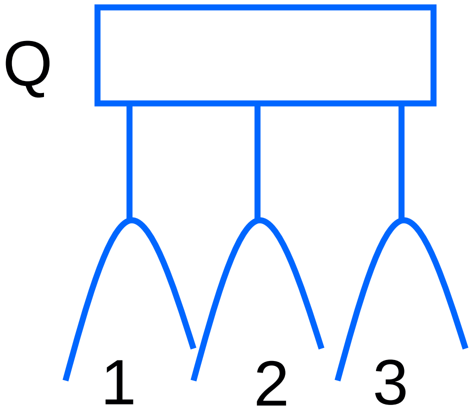
2,3 boom

PQ-trees

representation for permutations



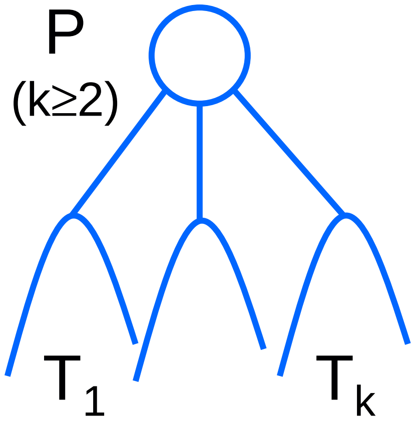
{ 123, 132, 213, 231, 312, 321 }



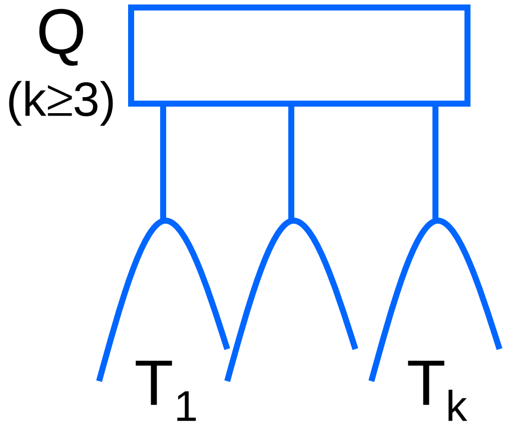
{ 123, 321 }

PQ-trees

datastructure to represent all possibilities



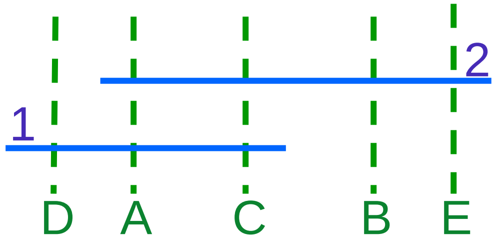
P permutation



Q linear order

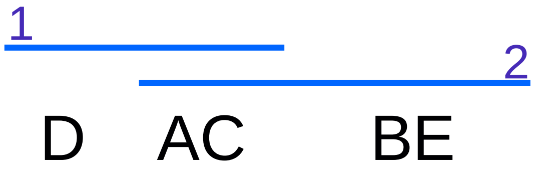
PQ trees
represent possible reorderings
(permutations of probes)

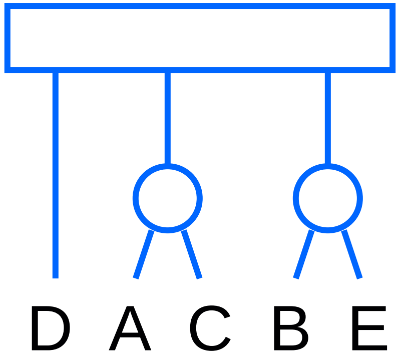
example



	A	B	C	D	E
1	1		1	1	
2	1	1	1		1

clones { A, C, D } { A, B, C, E }



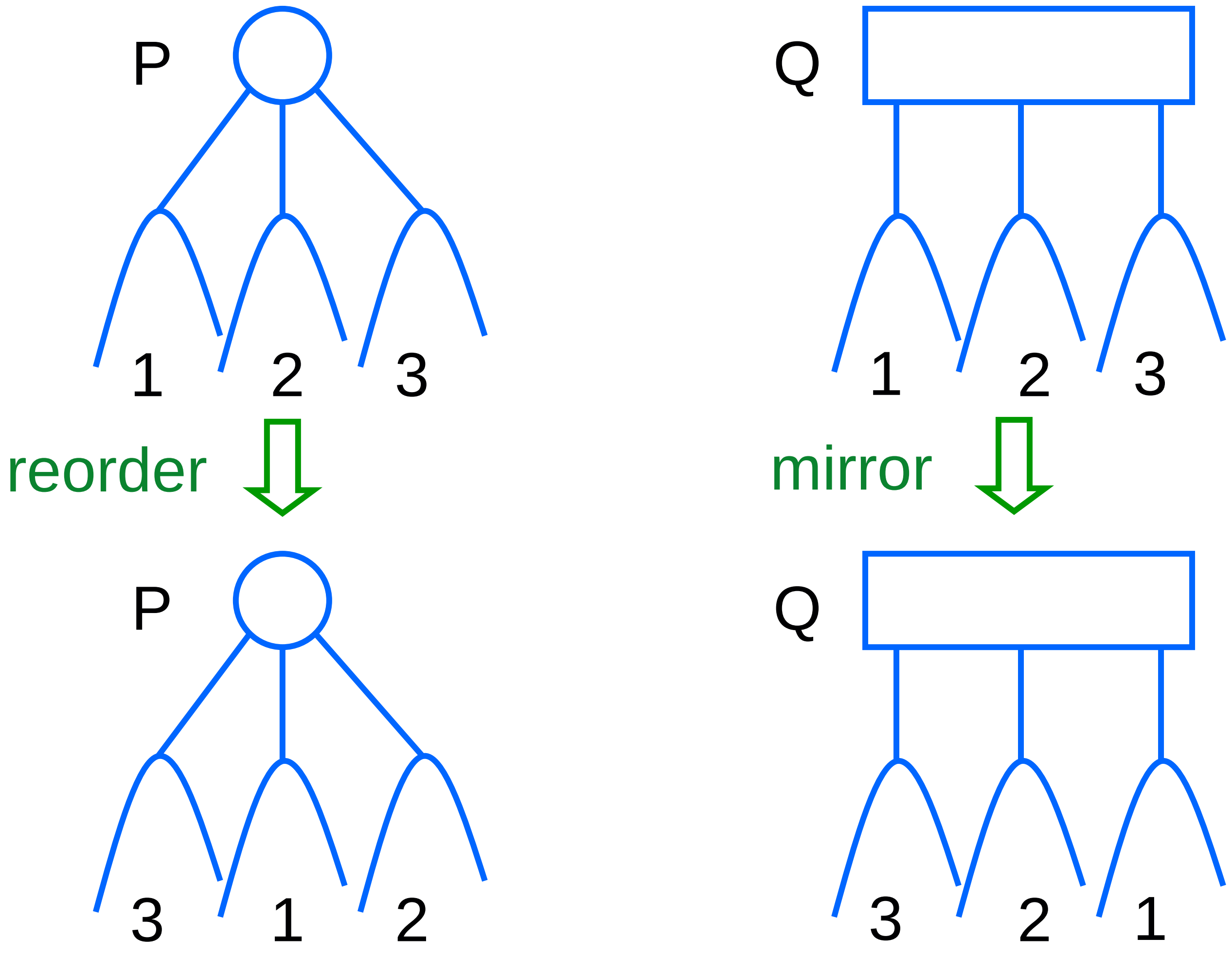


D AC BE
D CA BE
D AC EB
D CA EB

EB CA D
EB AC D
BE CA D
BE AC D

PQ-trees

equivalent representations



{ 123, 132, 213, 231, 312, 321 }

{ 123, 321 }

PQ-tree algorithm

reduce(T,S)

T PQ tree ~ set of permutations

S new clone ~ set of (consecutive) probes

add requirement S to tree T

‘keep S together’

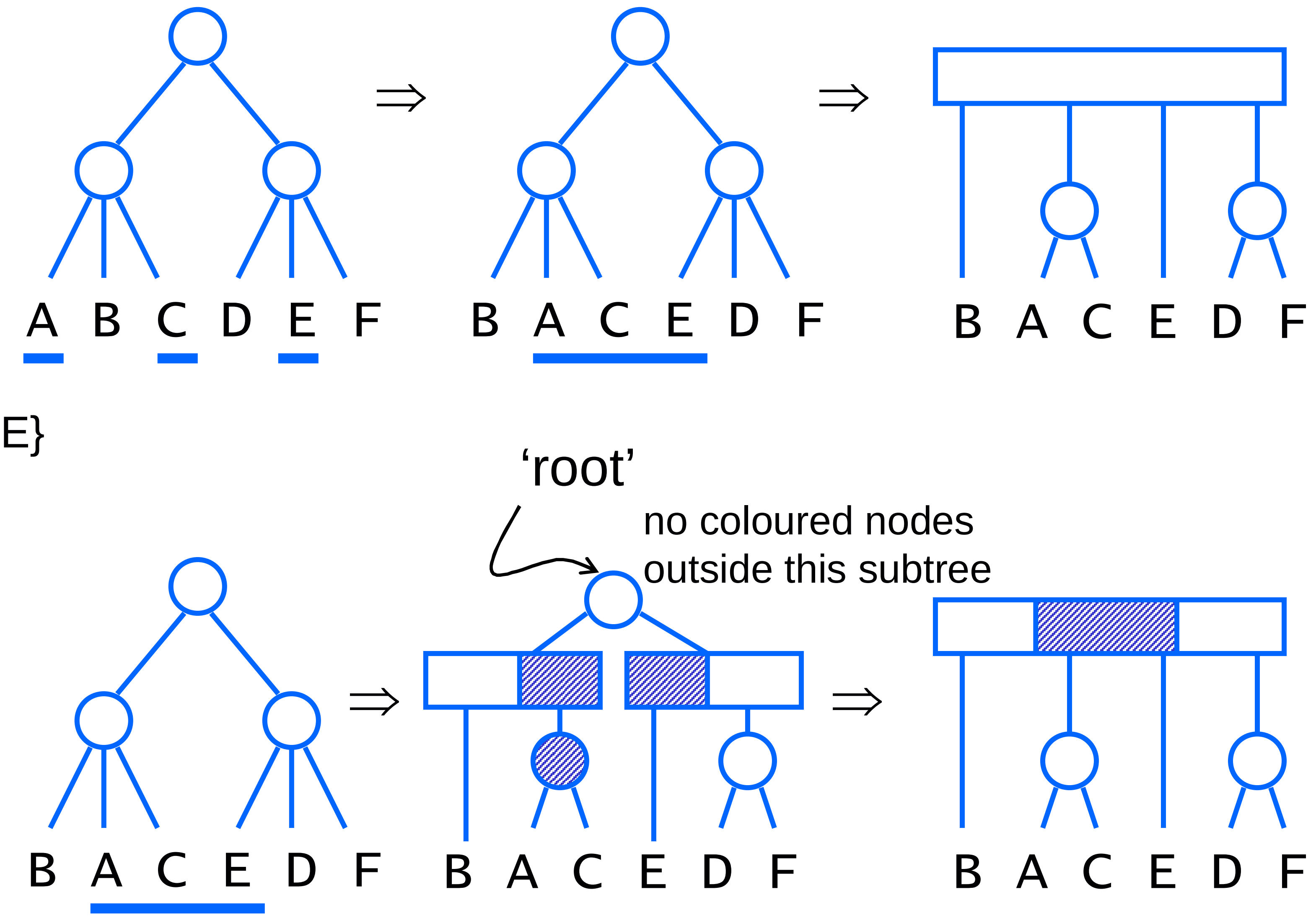
- colour leaves in S
- apply transformations
  to get consecutive leaves
- apply replacement rules
  to add new restriction to tree

P ▨ all leaves in S

Q □▨□ segment in S

example

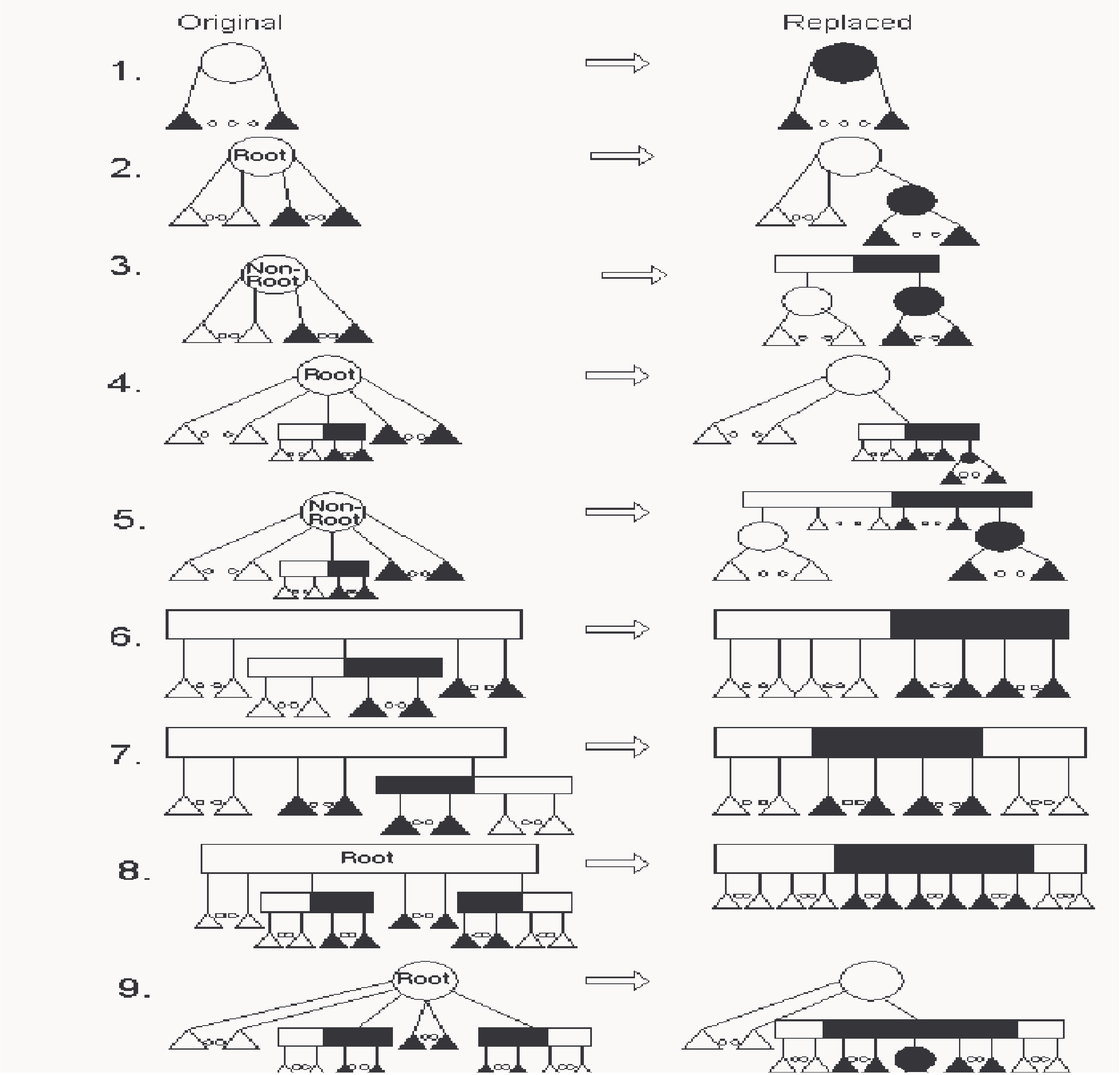
S = {A,C,E}



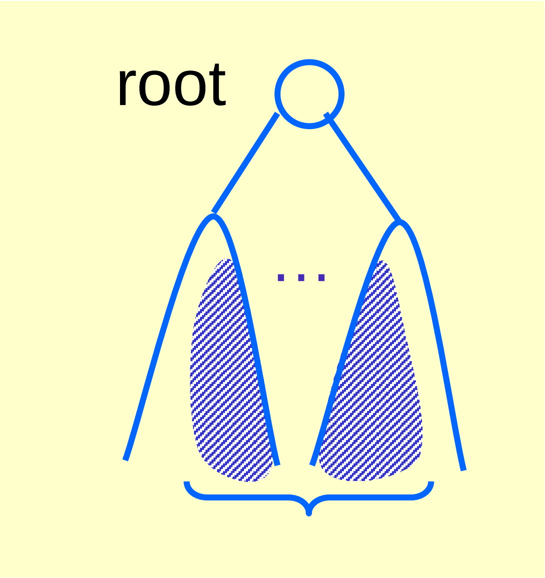
‘root’

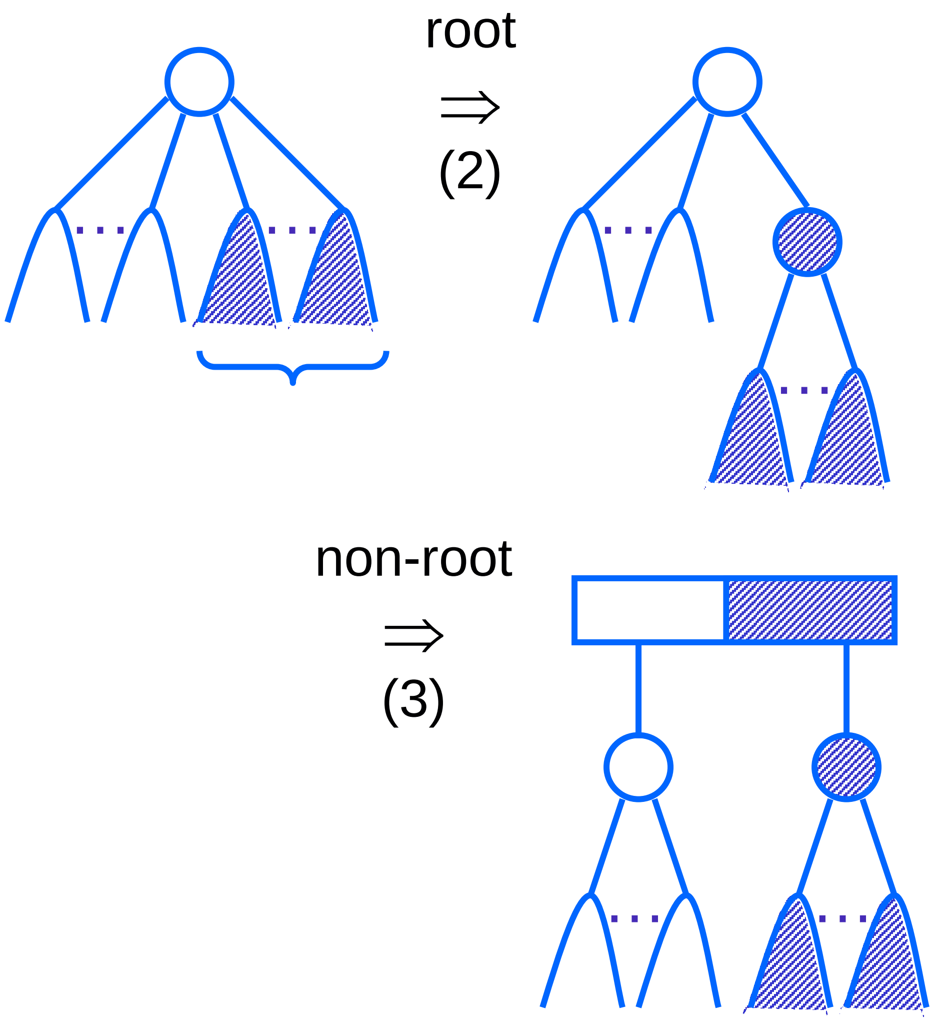
no coloured nodes outside this subtree

replacement rules



replacement rules (2,3)




root

⇒

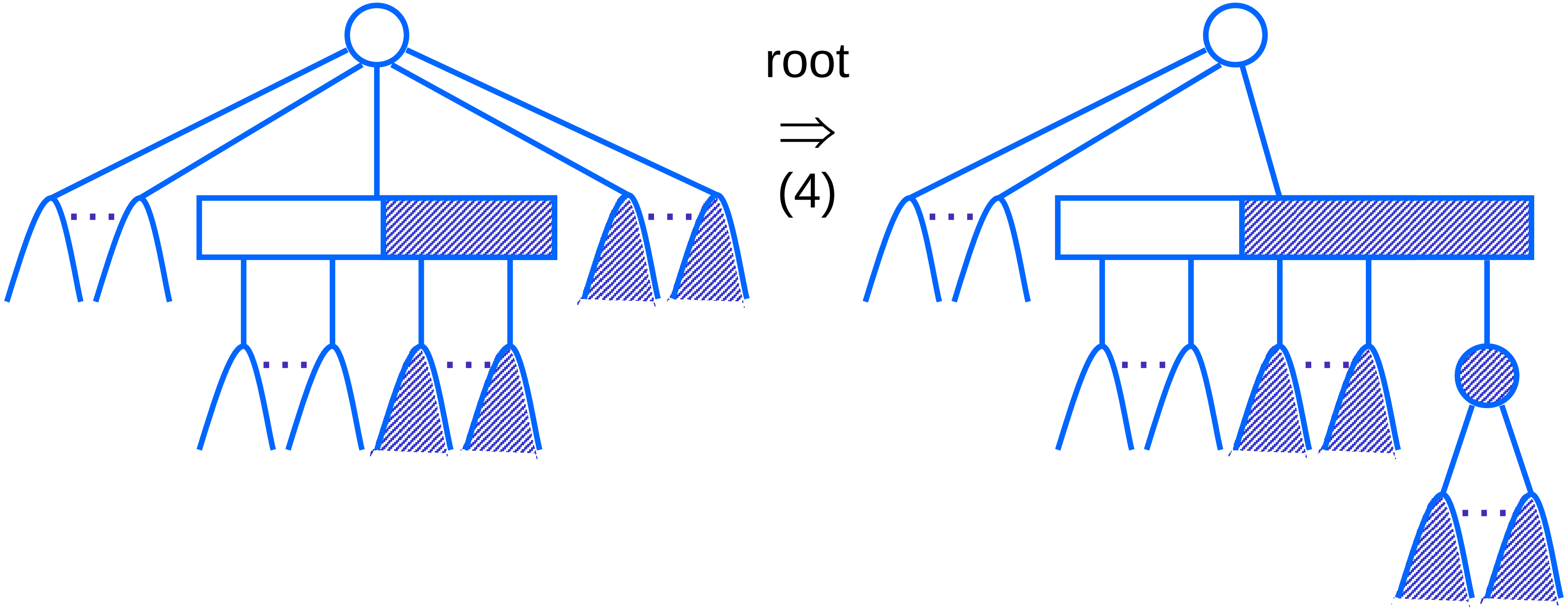
(2)

non-root

⇒

(3)

replacement rules (4,5)



root

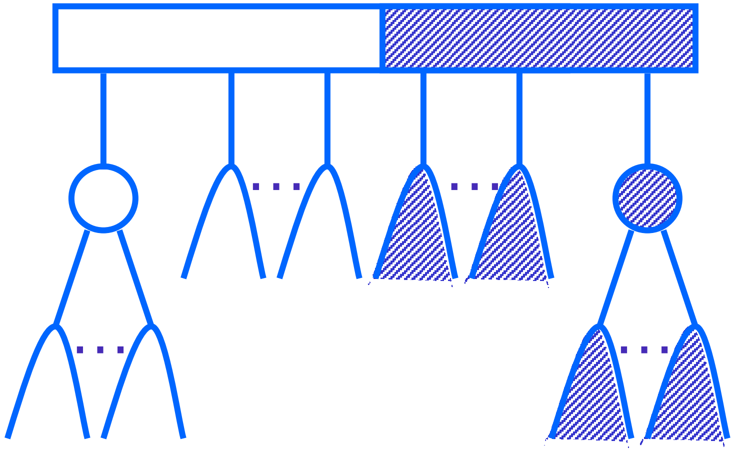
⇒

(4)

non-root

⇒

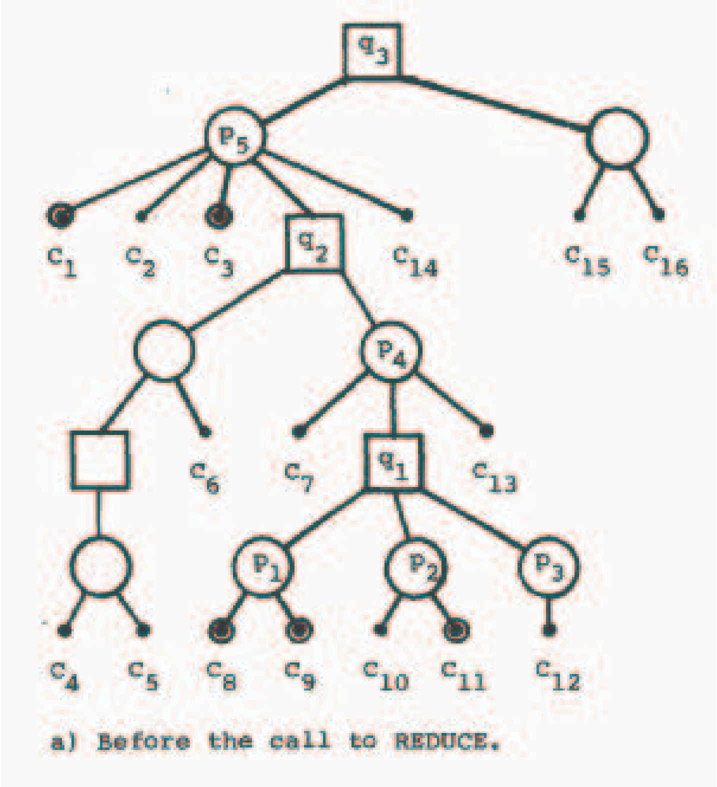
(5)



references

K.S. Booth and G.S. Leuker. Testing for the consecutive ones property, interval graphs, and graph planarity using PQ-tree algorithms. JCSS 13:335-379, 1976.

also 7th STOC, 1975.



JOURNAL OF COMPUTER AND SYSTEM SCIENCES

challenges

- find the right model (simplification)
  - noise, errors,
    too much / not enough data
  - heuristics & AI approach
  - is it data mining?
- interdisciplinary