

"Nothing in biology makes sense except in the light of evolution"

-- T. Dobzhansky, 1973

Phylogenetic Analysis of Haloalkane Dehalogenases

Eva Chovancová¹, Jan Kosinski², Janusz M. Bujnicki², Jiří Damborský¹

¹ *Loschmidt Laboratories, Protein Engineering Group, Masaryk University, Brno, CZ*

² *Laboratory of Bioinformatics and Protein Engineering, IIMCB, Warsaw, PL*

Why to study evolution?

- *evolution is underlying principle in biology, everything is built on that principle*
-

Why to study evolution?

- *evolution is underlying principle in biology, everything is built on that principle*
 - *crucial for understanding of life principles*
 - *evolution of drug resistance*
 - *evolutionary origins of diseases*
 - *identification of interesting genes, proteins*
-

Identification of haloalkane dehalogenases

- similarity searches – PSI-BLAST¹

NCBI *protein-protein* **BLAST**

Nucleotide Protein Translations Retrieve results for an RID

Search

MSE IGTGFPFDPHYVEVLGERMHYVDVGPRDGPVLF LHGNPTSSYLWRNIIPHVAPSHRCI.

Set subsequence From: To:

Choose database

or

PSI-BLAST results:
over **3400** protein sequences

¹ Altschul, et al. (1997). Nucleic Acids Research 25: 3389

Identification of haloalkane dehalogenases

- similarity searches – PSI-BLAST¹

NCBI *protein-protein* **BLAST**

Nucleotide Protein Translations Retrieve results for an RID

Search

MSEIGTGFPFDPHYVEVLGERMHYVDVGPRDGPVLF LHGNPTSSYLWRNIIPHVAPSHRCI.

Set subsequence From: To:

Choose database

Reset query Reset all

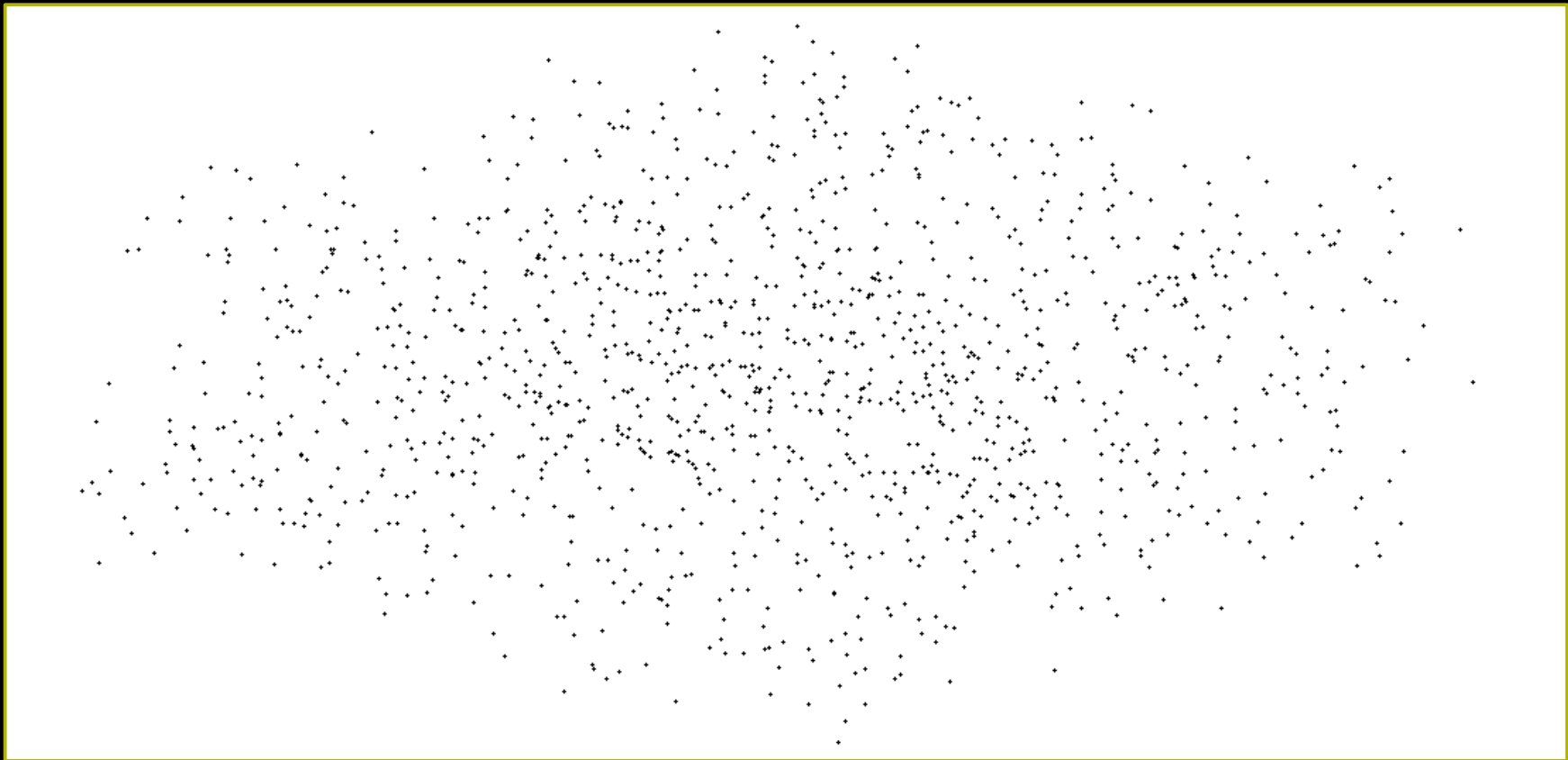
PSI-BLAST results:
over 3400 protein sequences

How to identify haloalkane
dehalogenases among them?

¹ Altschul, et al. (1997). Nucleic Acids Research 25: 3389

Identification of haloalkane dehalogenases

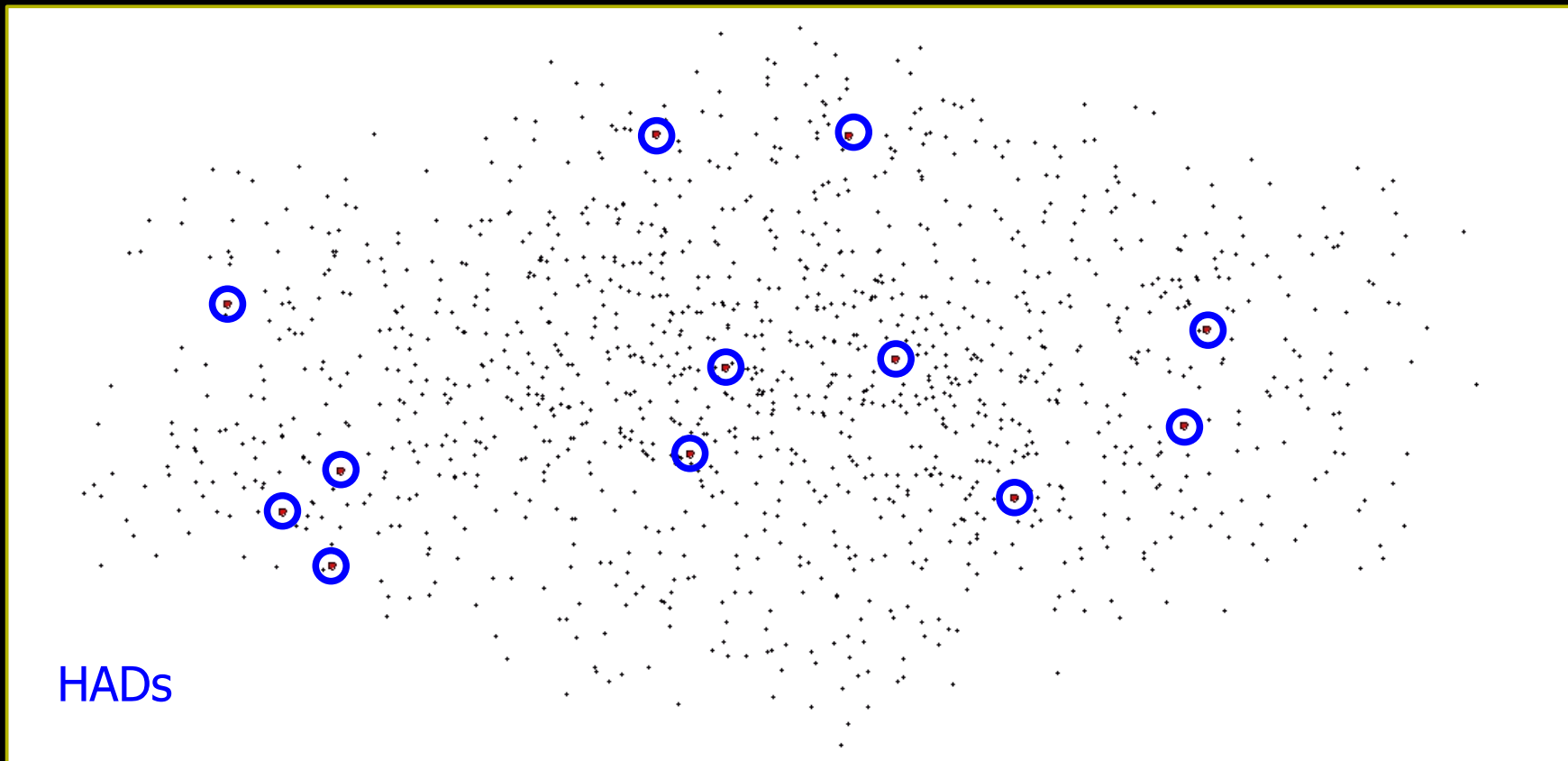
➤ *cluster analysis - CLANS¹*



¹ Frickey, et al. (2004). *Bioinformatics* 20: 3702

Identification of haloalkane dehalogenases

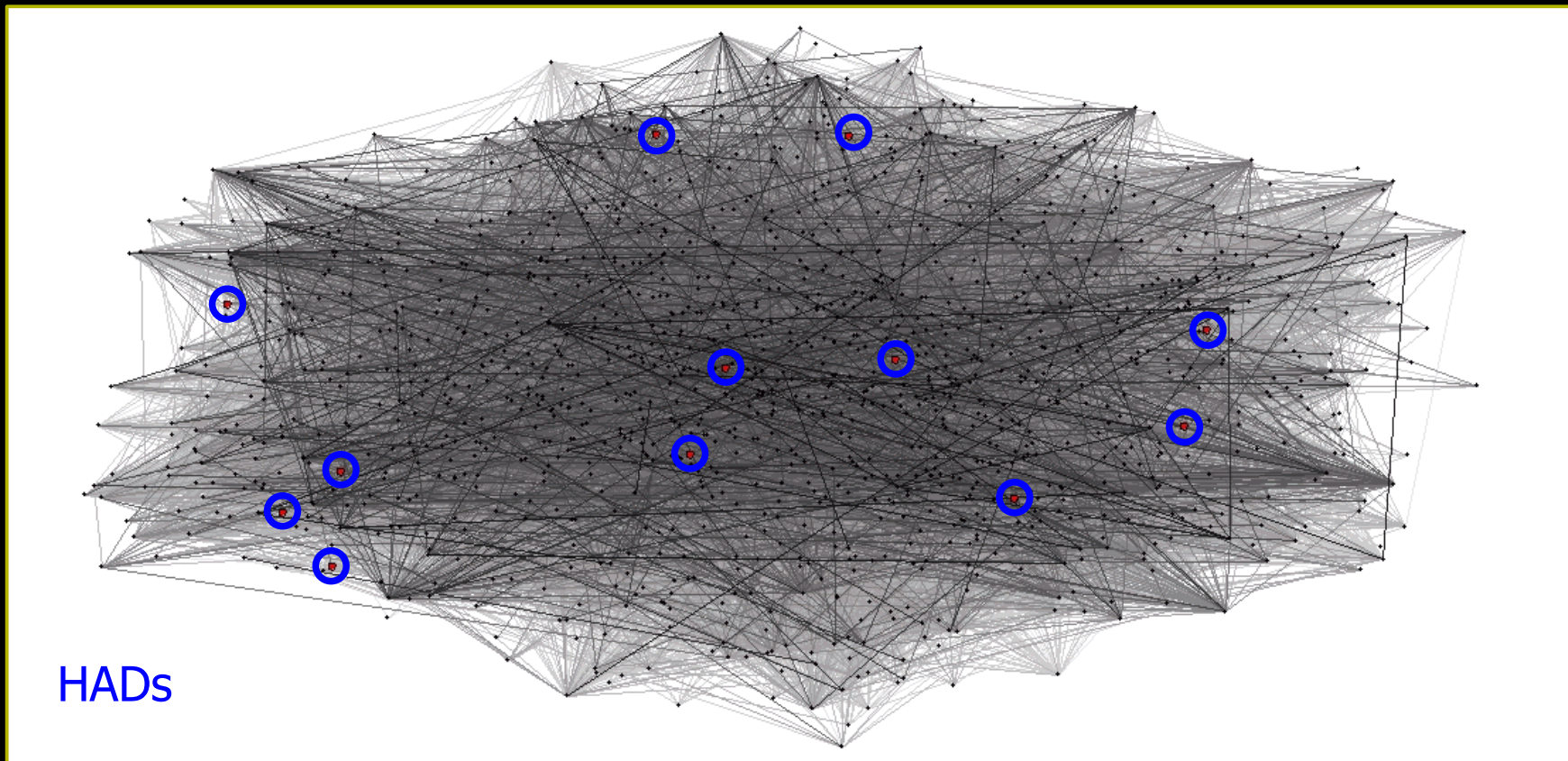
➤ *cluster analysis - CLANS¹*



¹ Frickey, et al. (2004). *Bioinformatics* 20: 3702

Identification of haloalkane dehalogenases

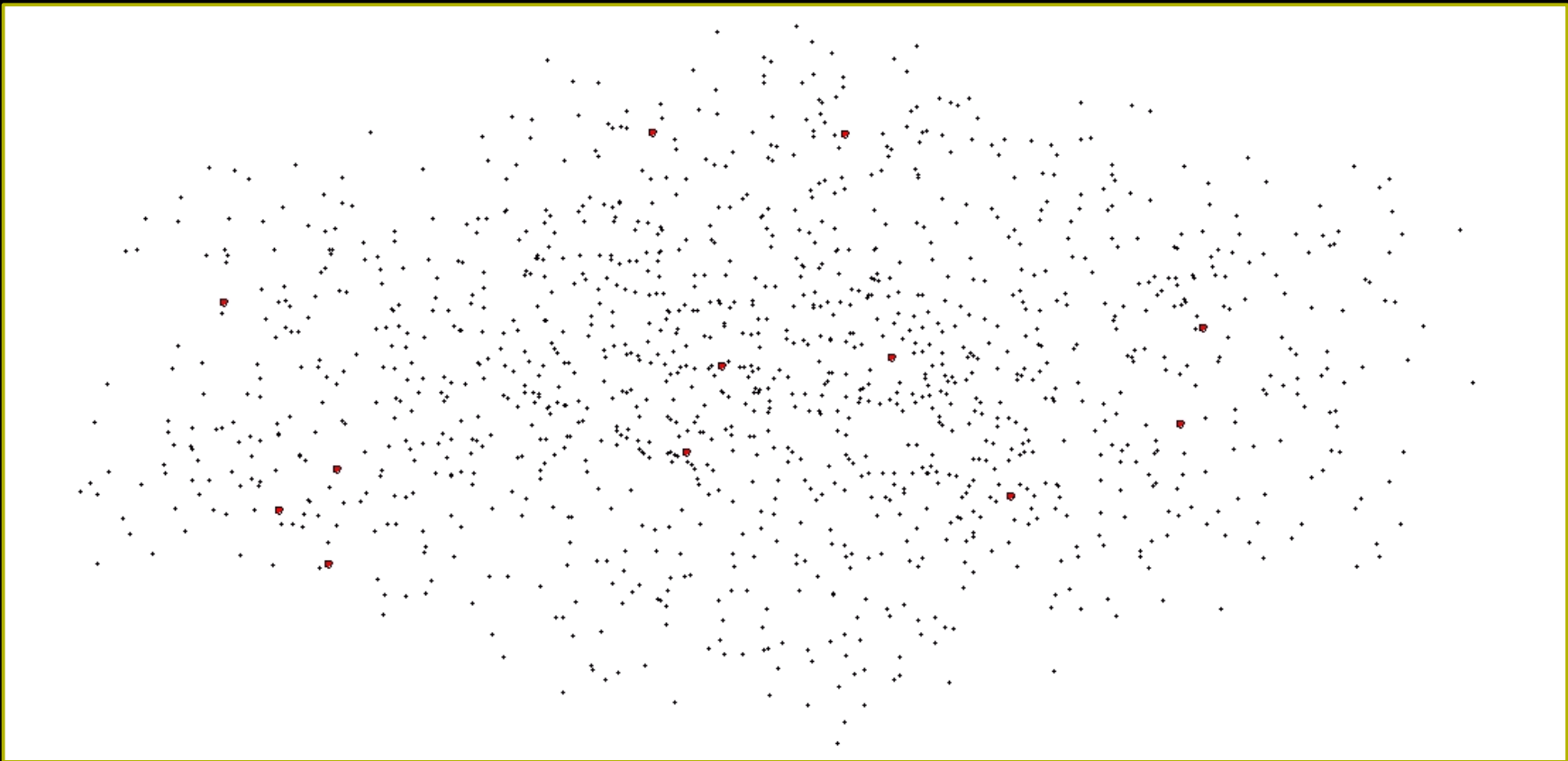
➤ *cluster analysis - CLANS¹*



¹ Frickey, et al. (2004). *Bioinformatics* 20: 3702

Identification of haloalkane dehalogenases

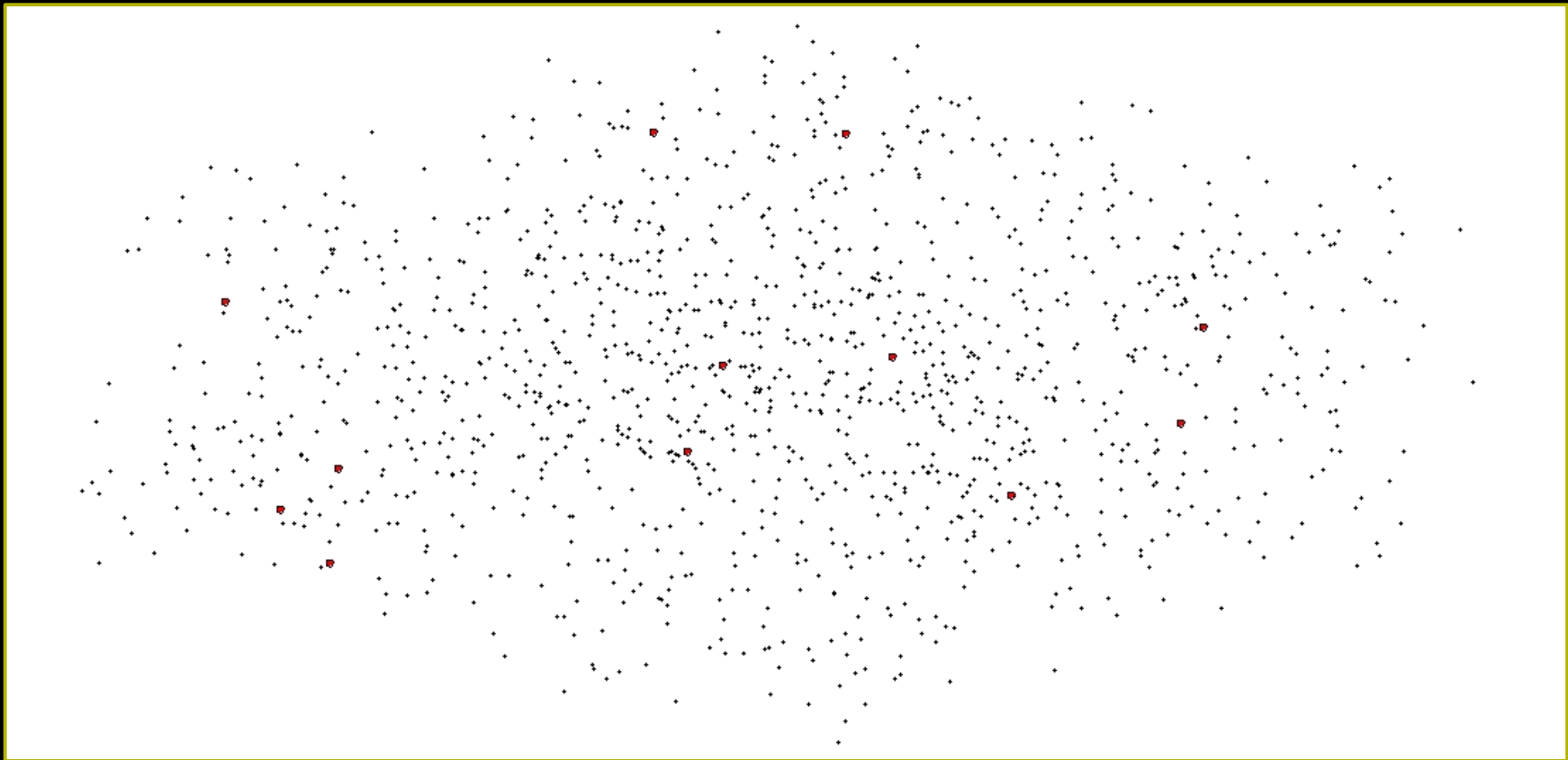
➤ *cluster analysis - CLANS¹*



¹ Frickey, et al. (2004). *Bioinformatics* 20: 3702

Identification of haloalkane dehalogenases

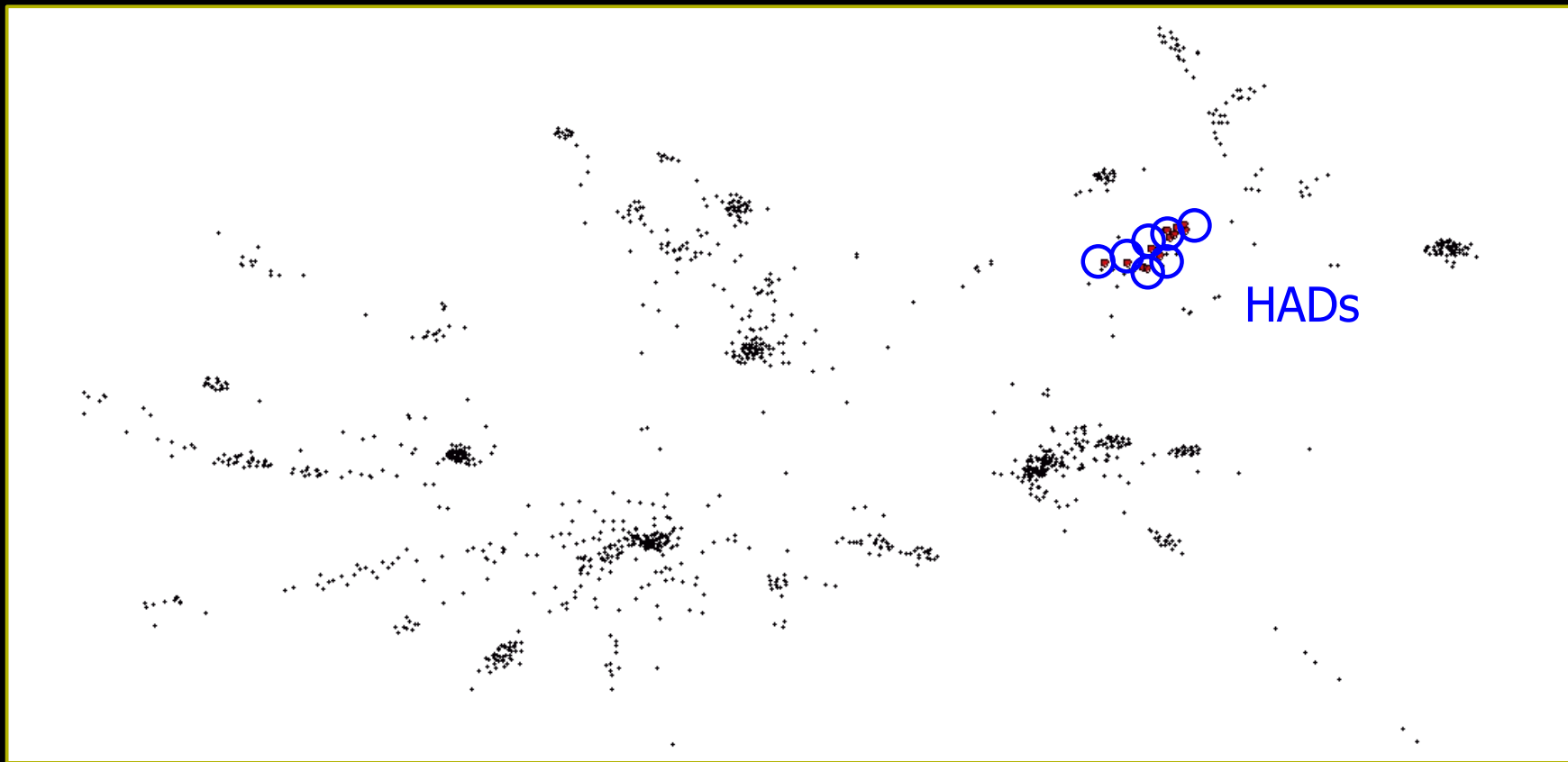
➤ *cluster analysis - CLANS¹*



¹ Frickey, et al. (2004). *Bioinformatics* 20: 3702

Identification of haloalkane dehalogenases

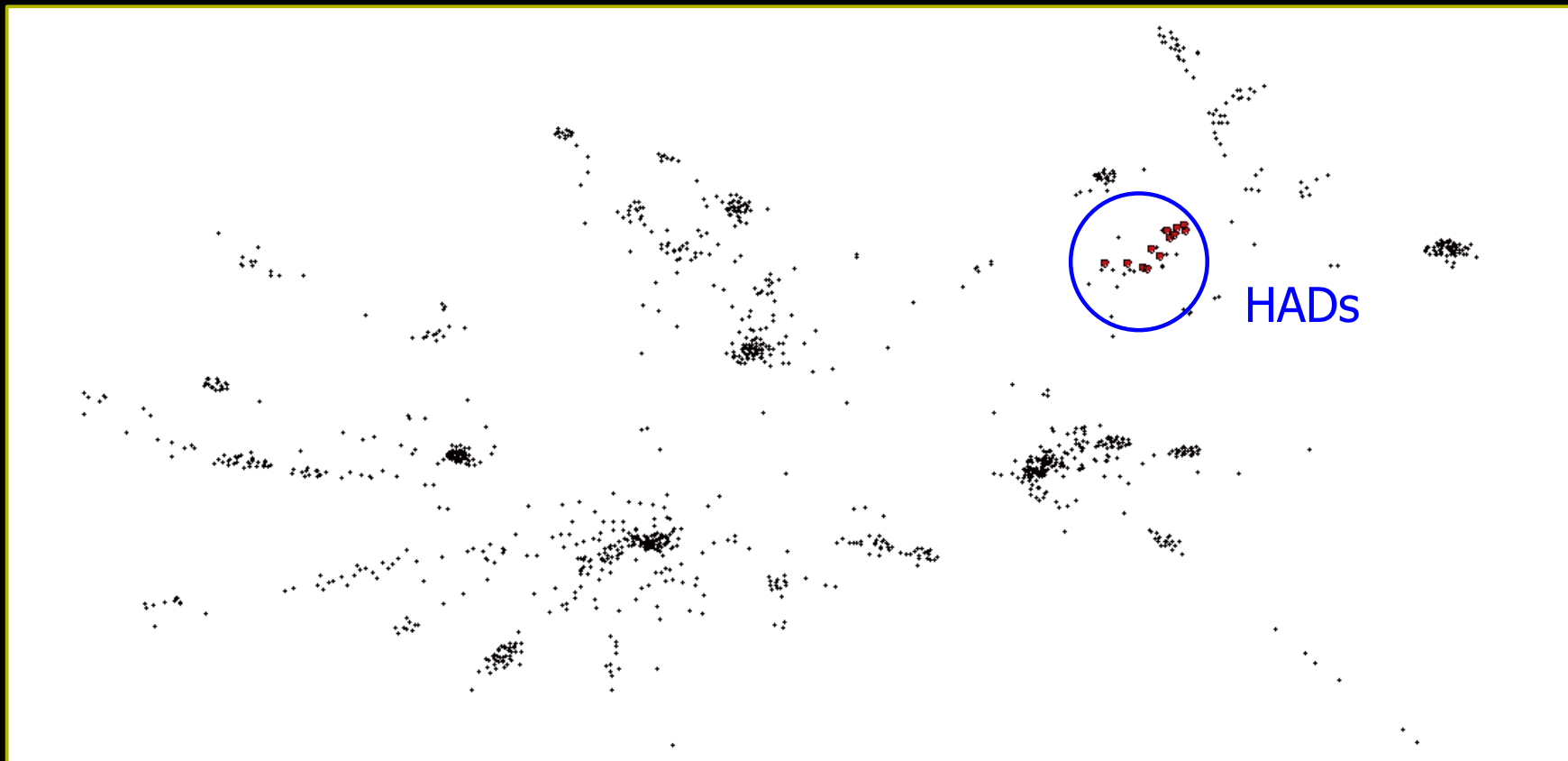
➤ *cluster analysis - CLANS¹*



¹ Frickey, et al. (2004). *Bioinformatics* 20: 3702

Identification of haloalkane dehalogenases

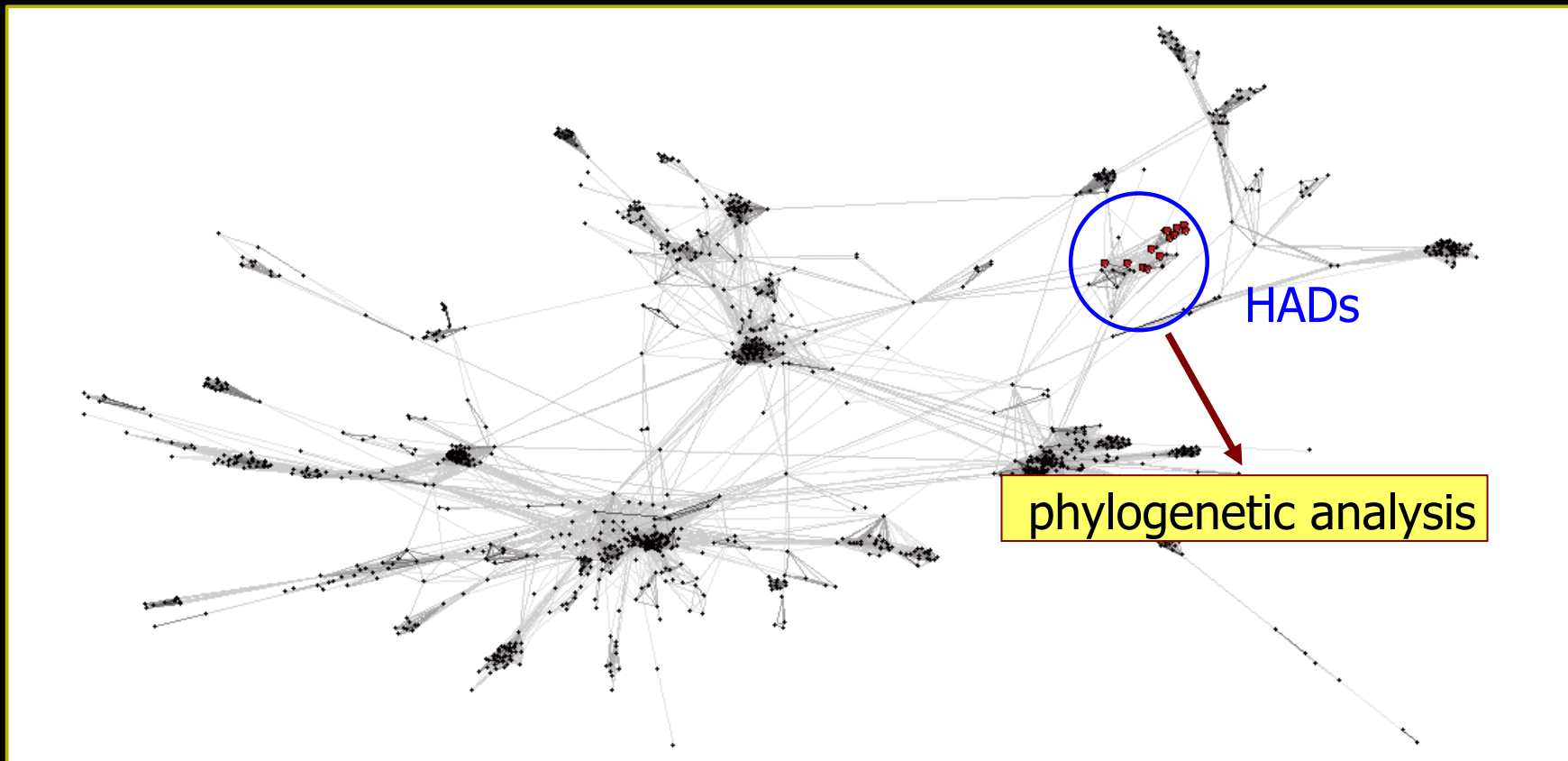
➤ *cluster analysis - CLANS¹*



¹ Frickey, et al. (2004). *Bioinformatics* 20: 3702

Identification of haloalkane dehalogenases

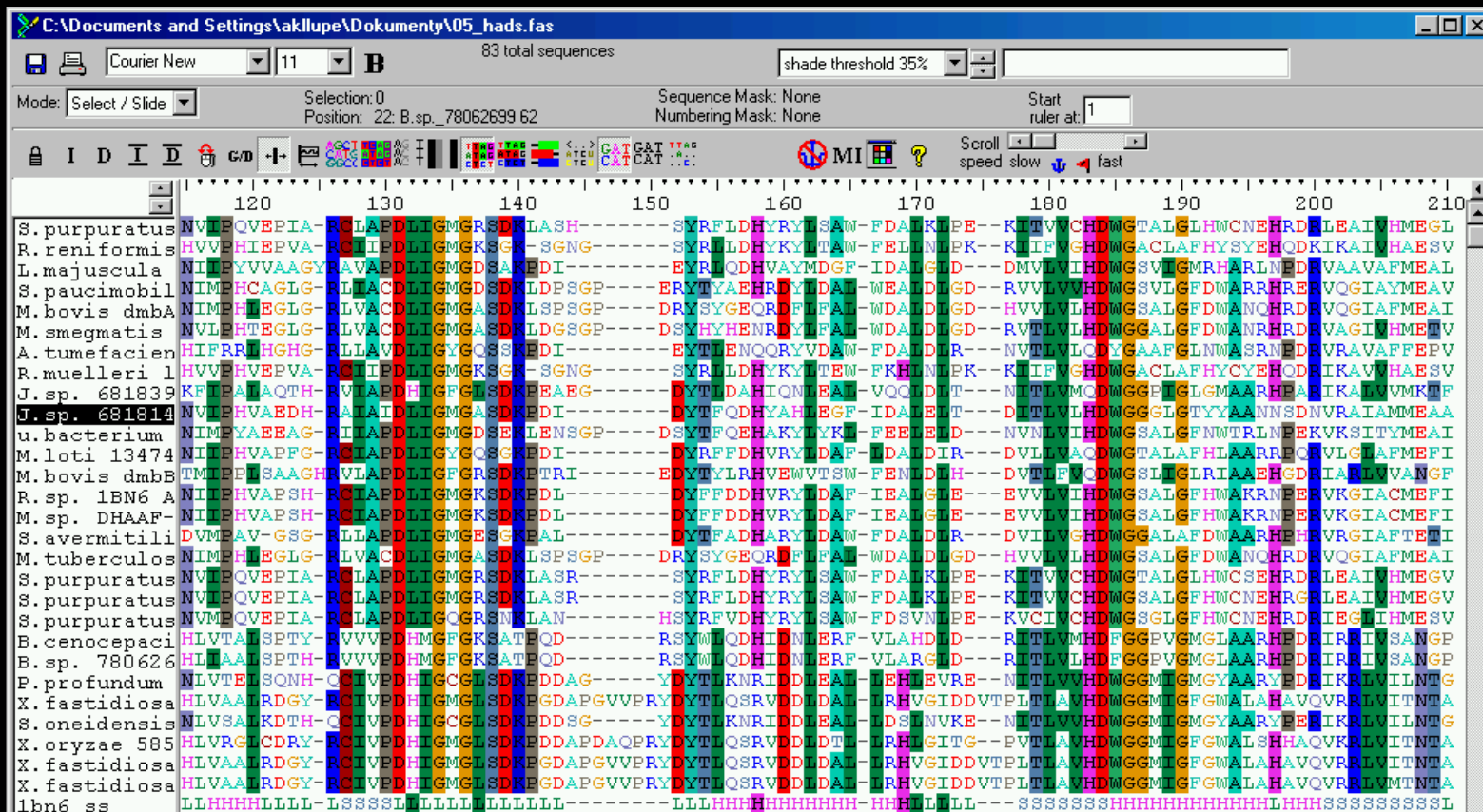
➤ *cluster analysis - CLANS¹*



¹ Frickey, et al. (2004). *Bioinformatics* 20: 3702

Phylogenetic analysis

- multiple sequence alignment – MUSCLE¹, BioEdit²



¹ Edgar (2004). Nucleic Acids Research 32: 1792

² <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>

Phylogenetic analysis

- multiple sequence alignment – MUSCLE, BioEdit
- selection of evolutionary model – PROTTEST¹

¹ Abascal, et al. (2005). *Bioinformatics* 21: 2104

Phylogenetic analysis

- multiple sequence alignment – MUSCLE, BioEdit
 - selection of evolutionary model – PROTTEST
 - construction of phylogenetic tree:
-

Phylogenetic analysis

- multiple sequence alignment – MUSCLE, BioEdit
- selection of evolutionary model – PROTTEST
- construction of phylogenetic tree:
 - *maximum likelihood method – PHYML¹*
 - *neighbour-joining method – MLDIST²*

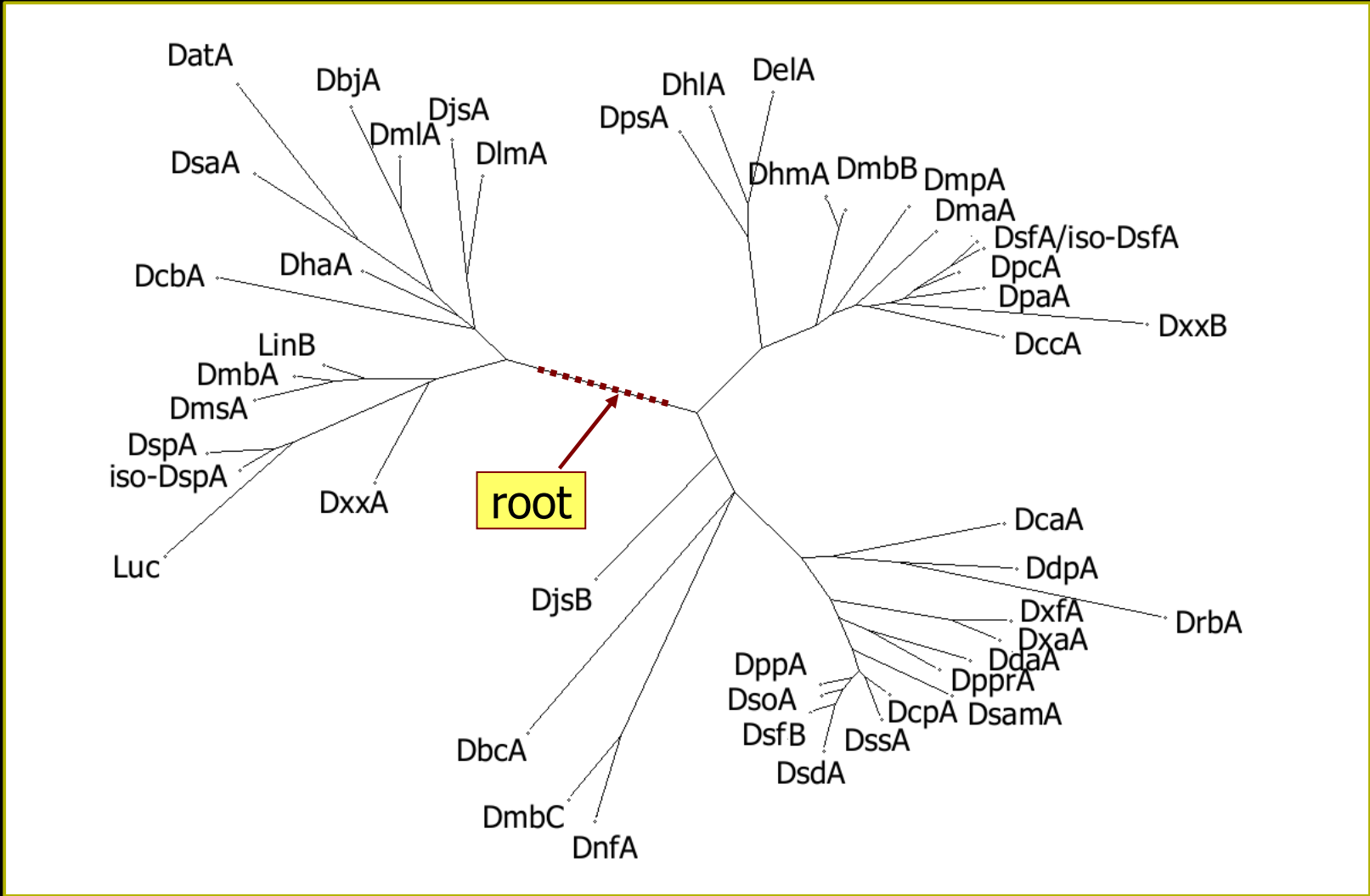
¹ Guindon, et al. (2003). Syst. Biol. 52: 696

² Drummond, et al. (2001). Bioinformatics 17: 662

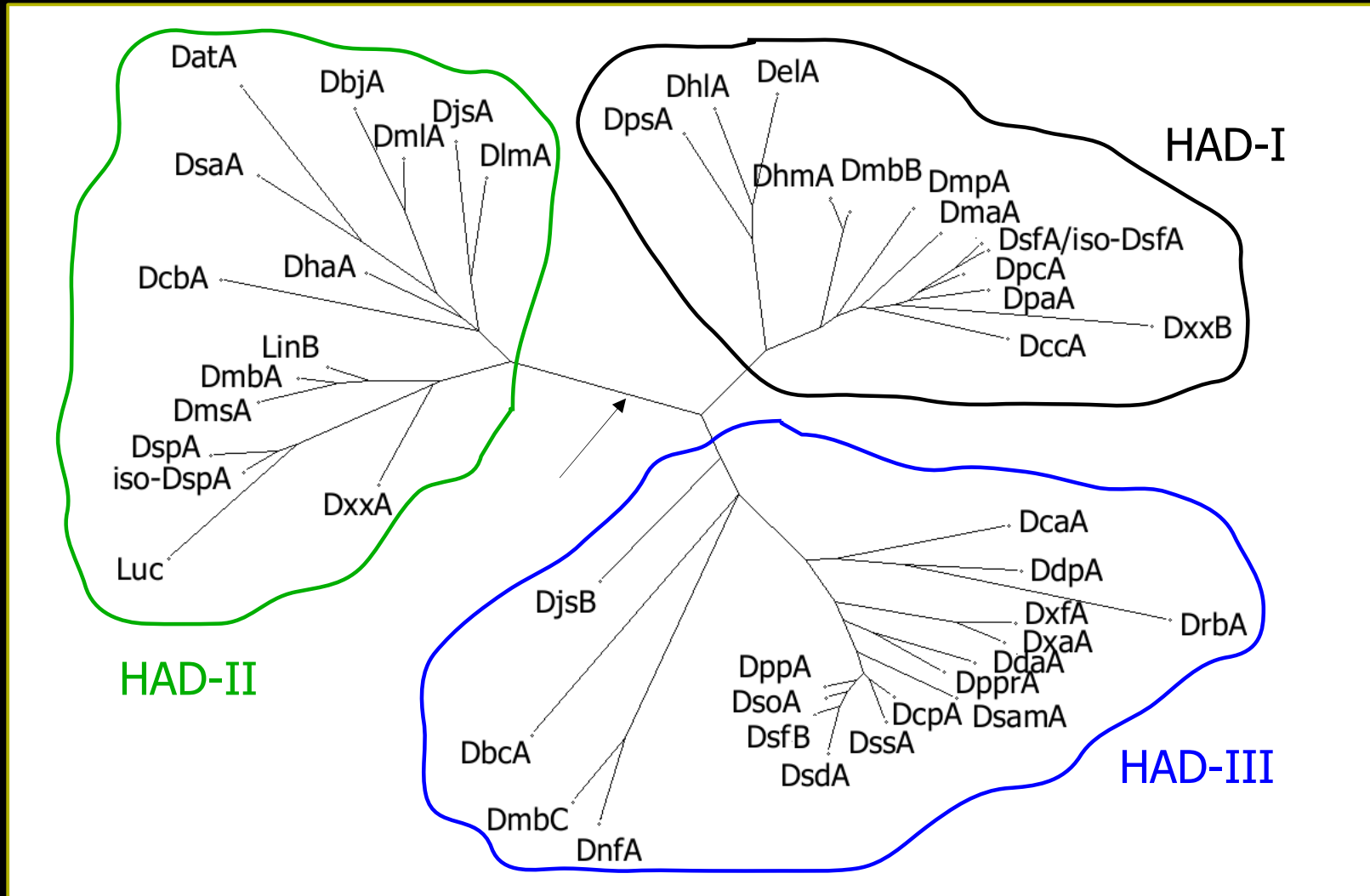
Phylogenetic analysis

- multiple sequence alignment – MUSCLE, BioEdit
 - selection of evolutionary model – PROTTEST
 - construction of phylogenetic tree:
 - *maximum likelihood method – PHYML*
 - *neighbour-joining method – MLDIST*
 - ***comparison of resulting trees – consensus tree***
-

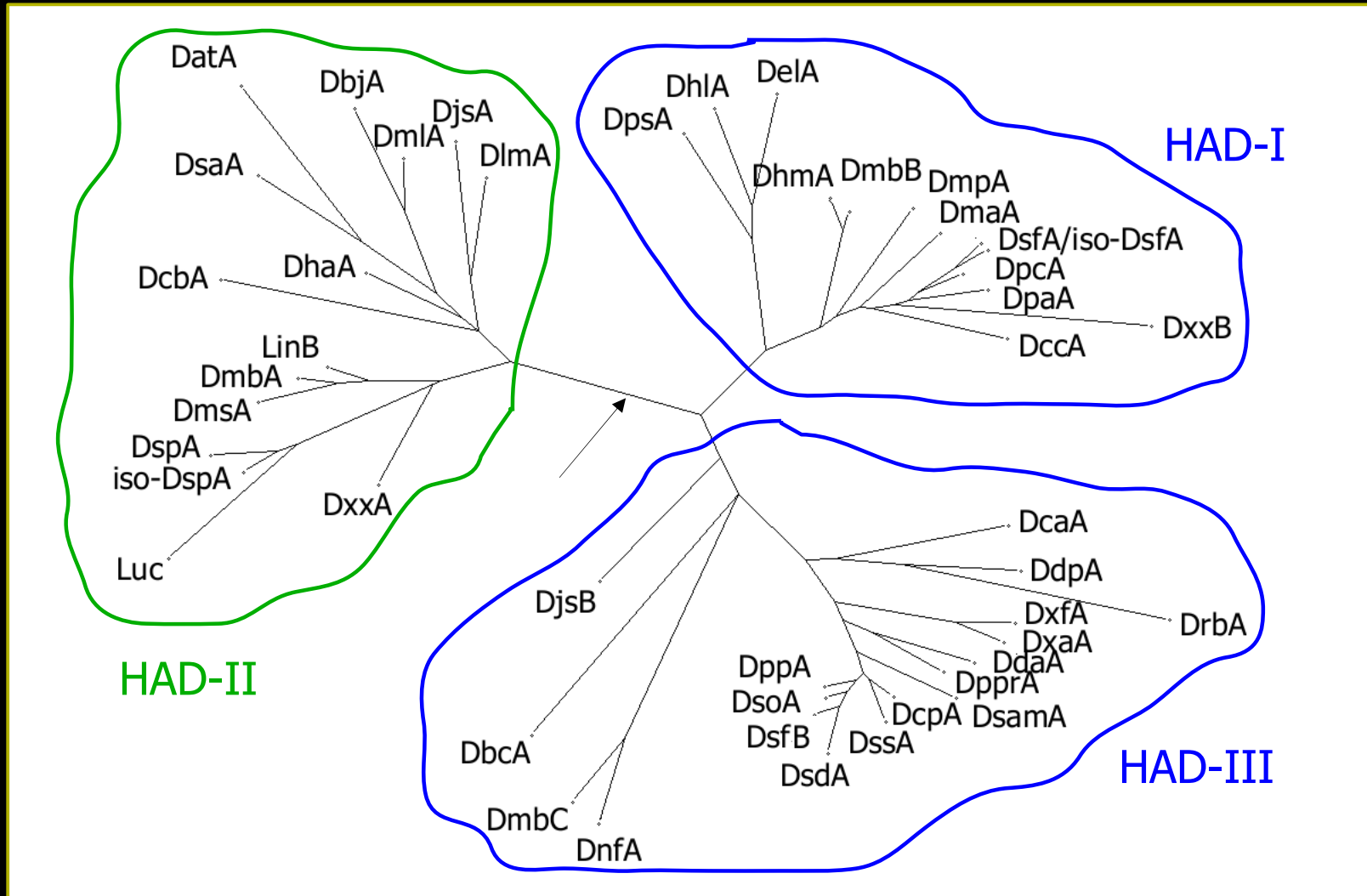
Phylogenetic tree



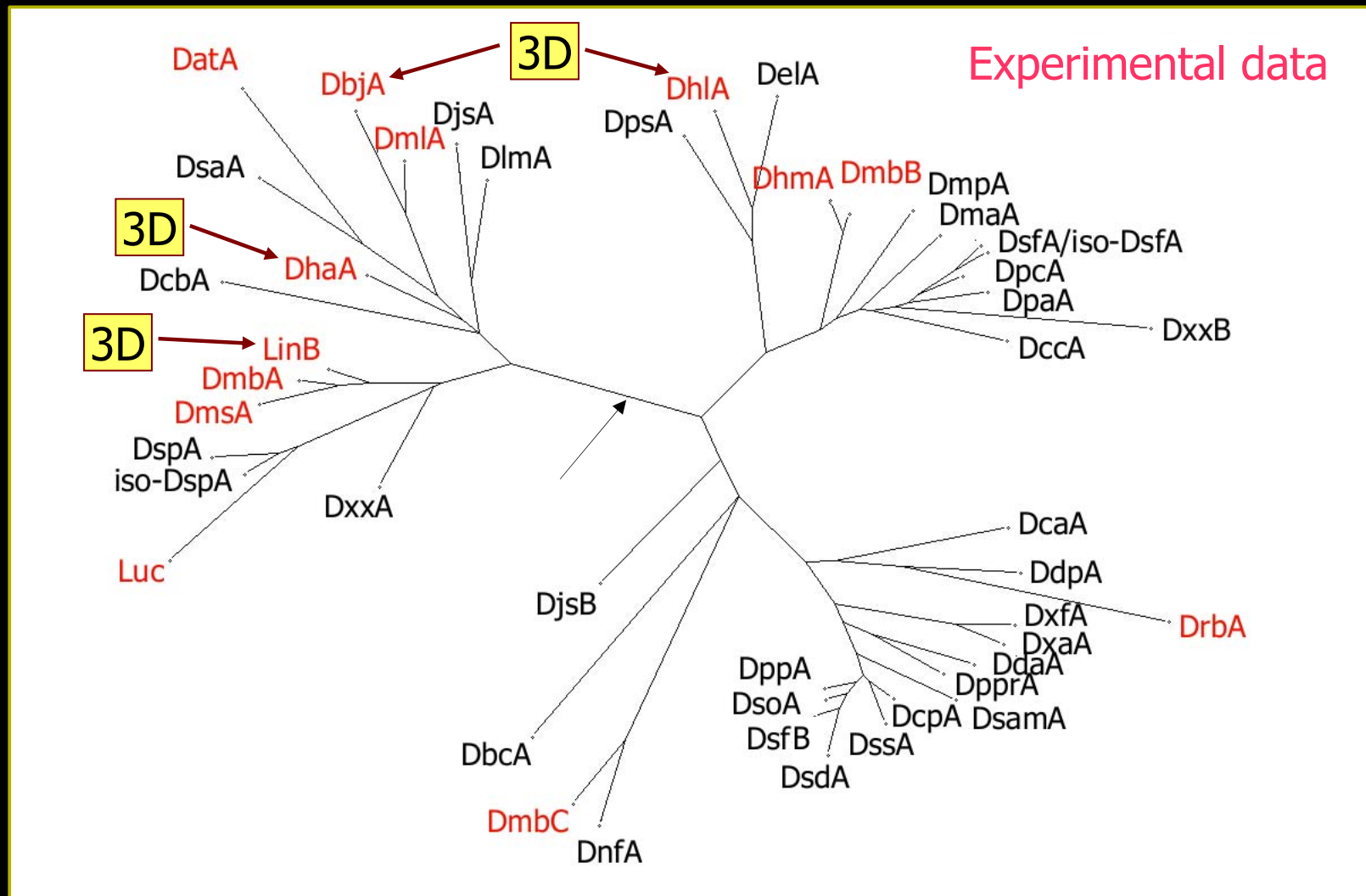
Classification



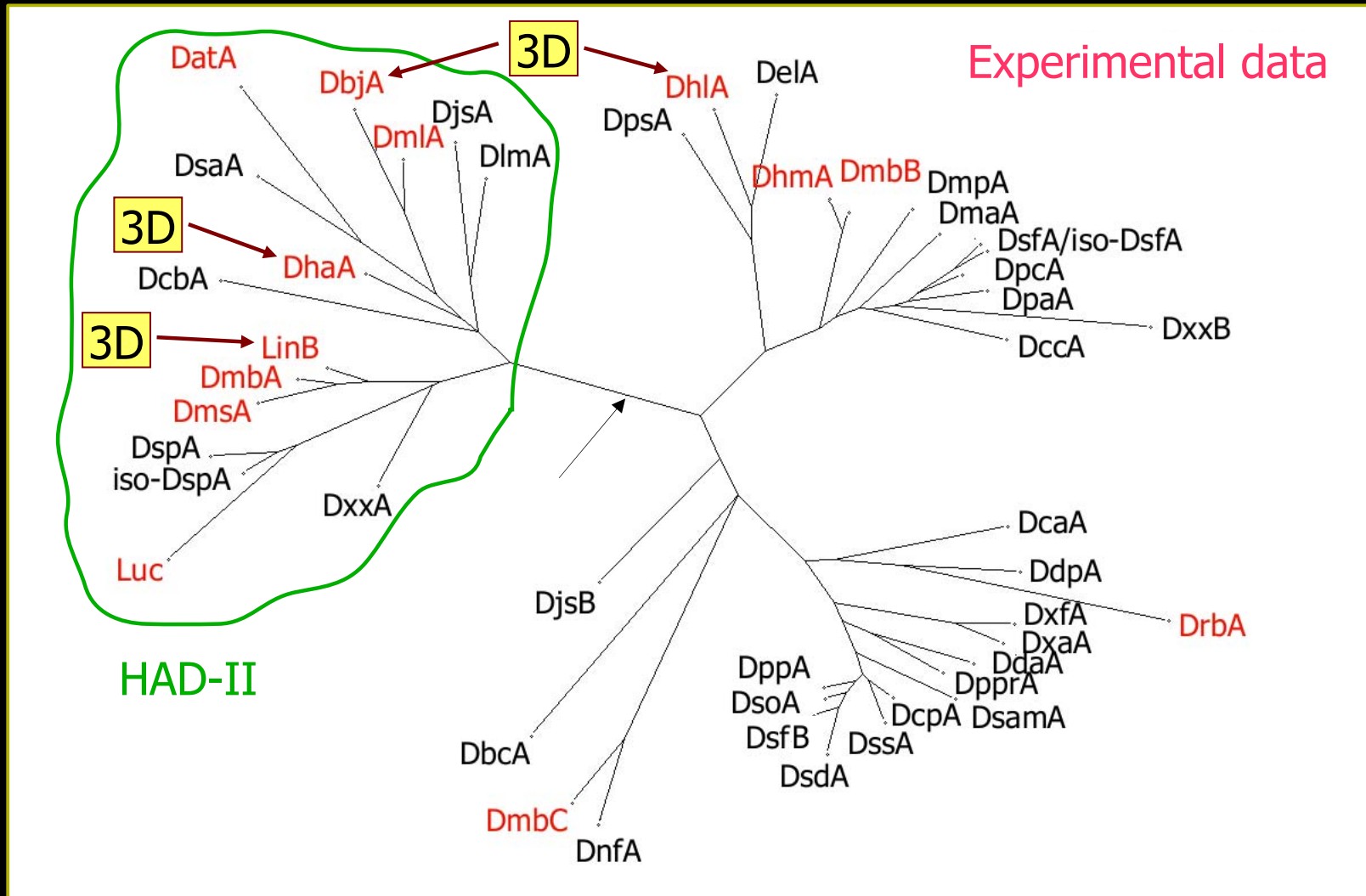
Relationships among subfamilies



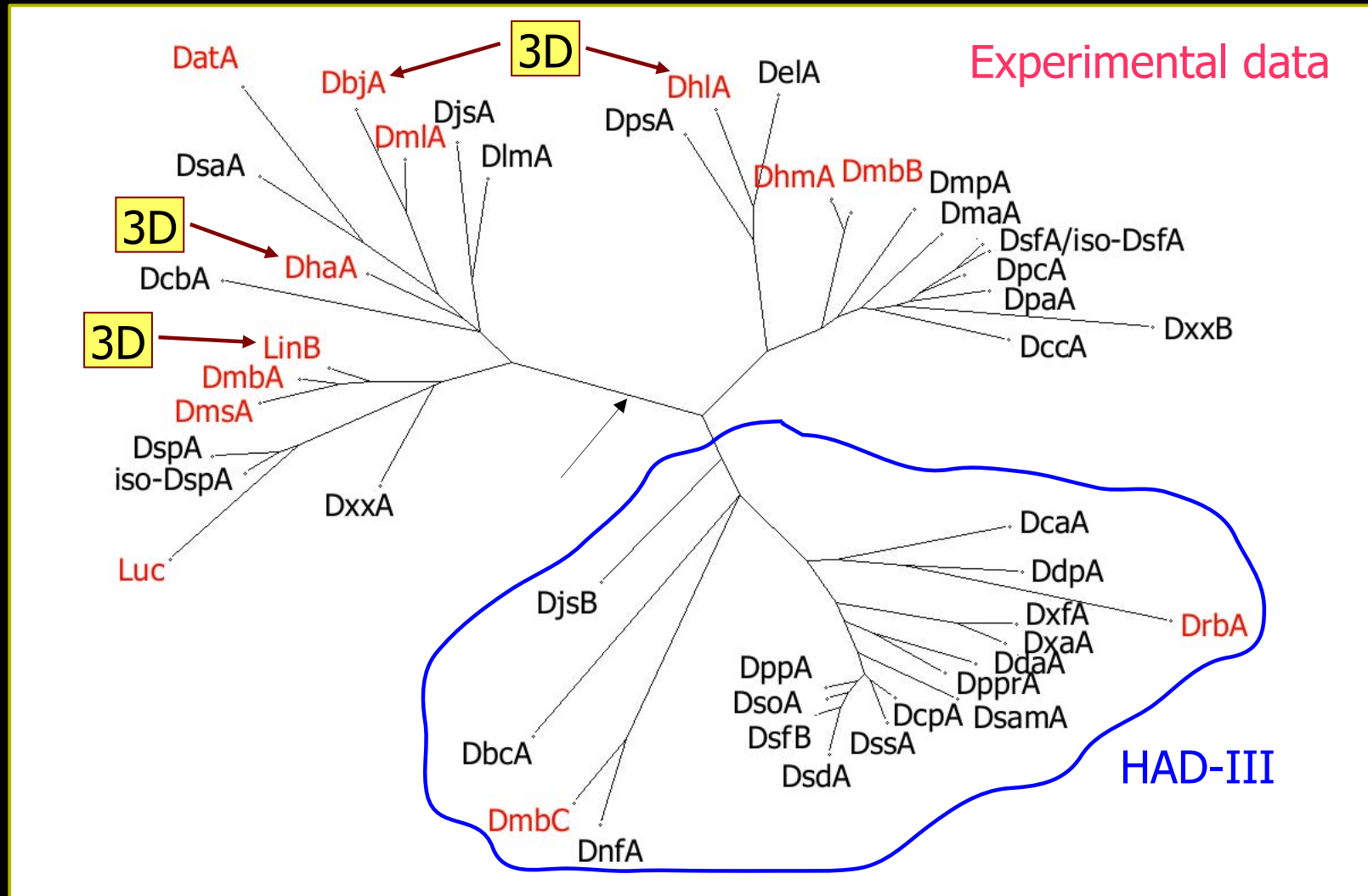
Experimental data



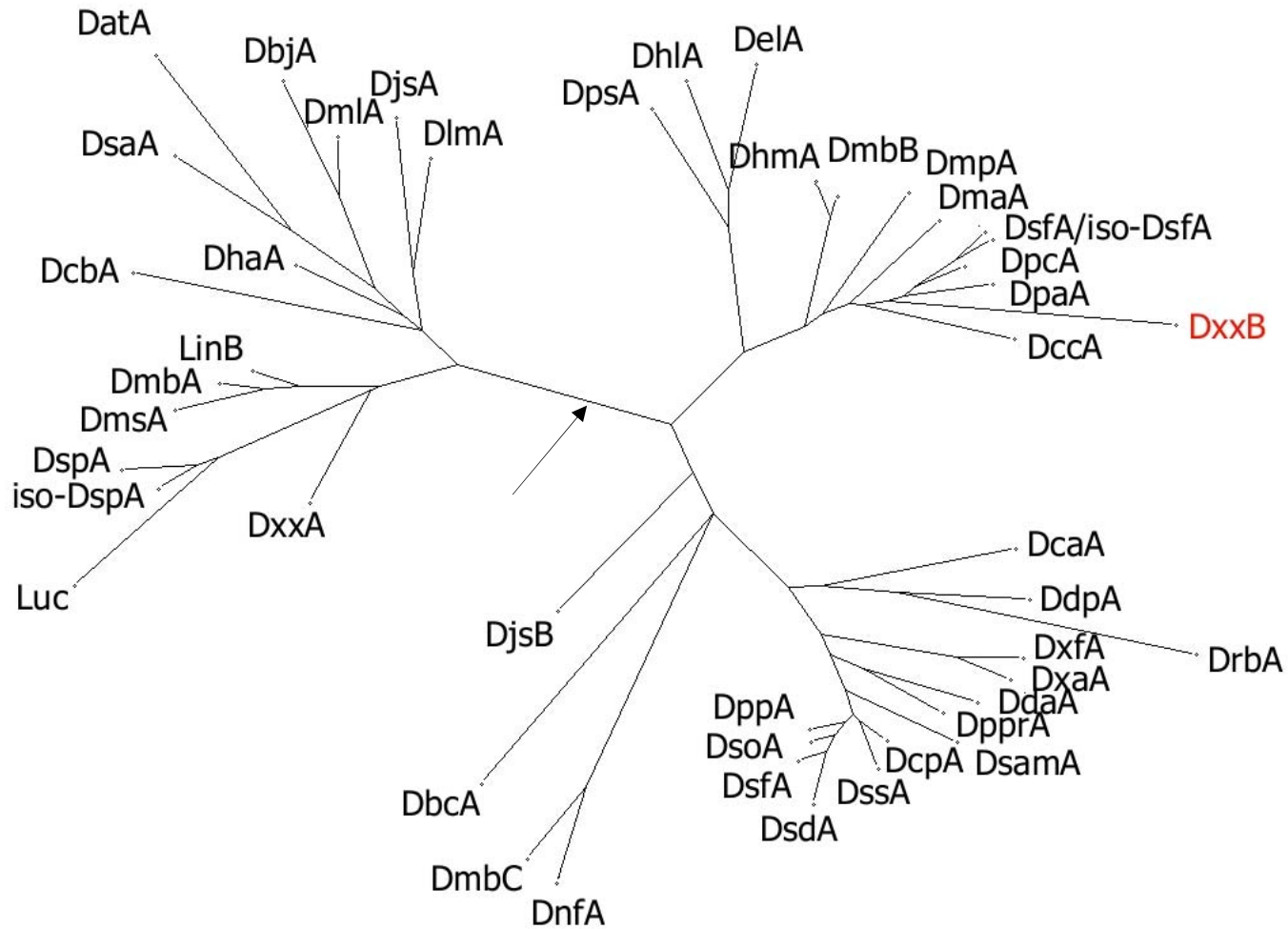
Experimental data



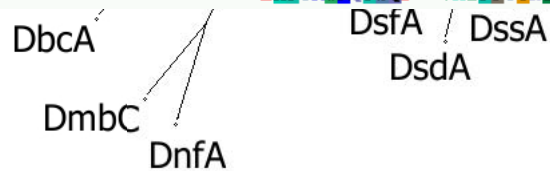
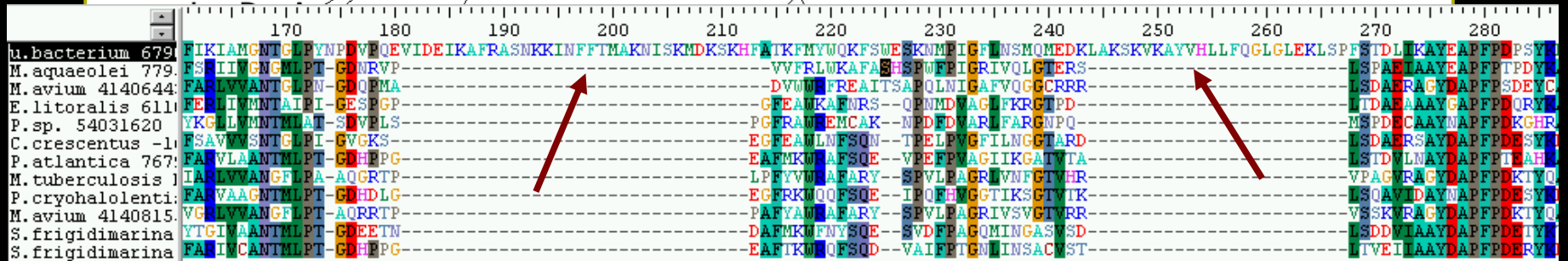
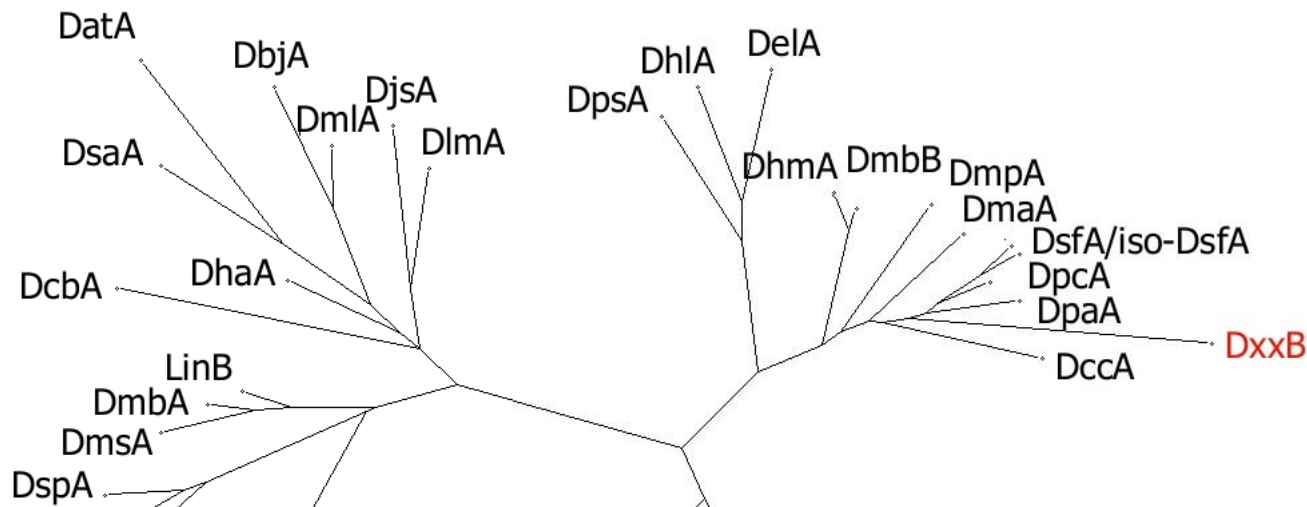
Interesting proteins



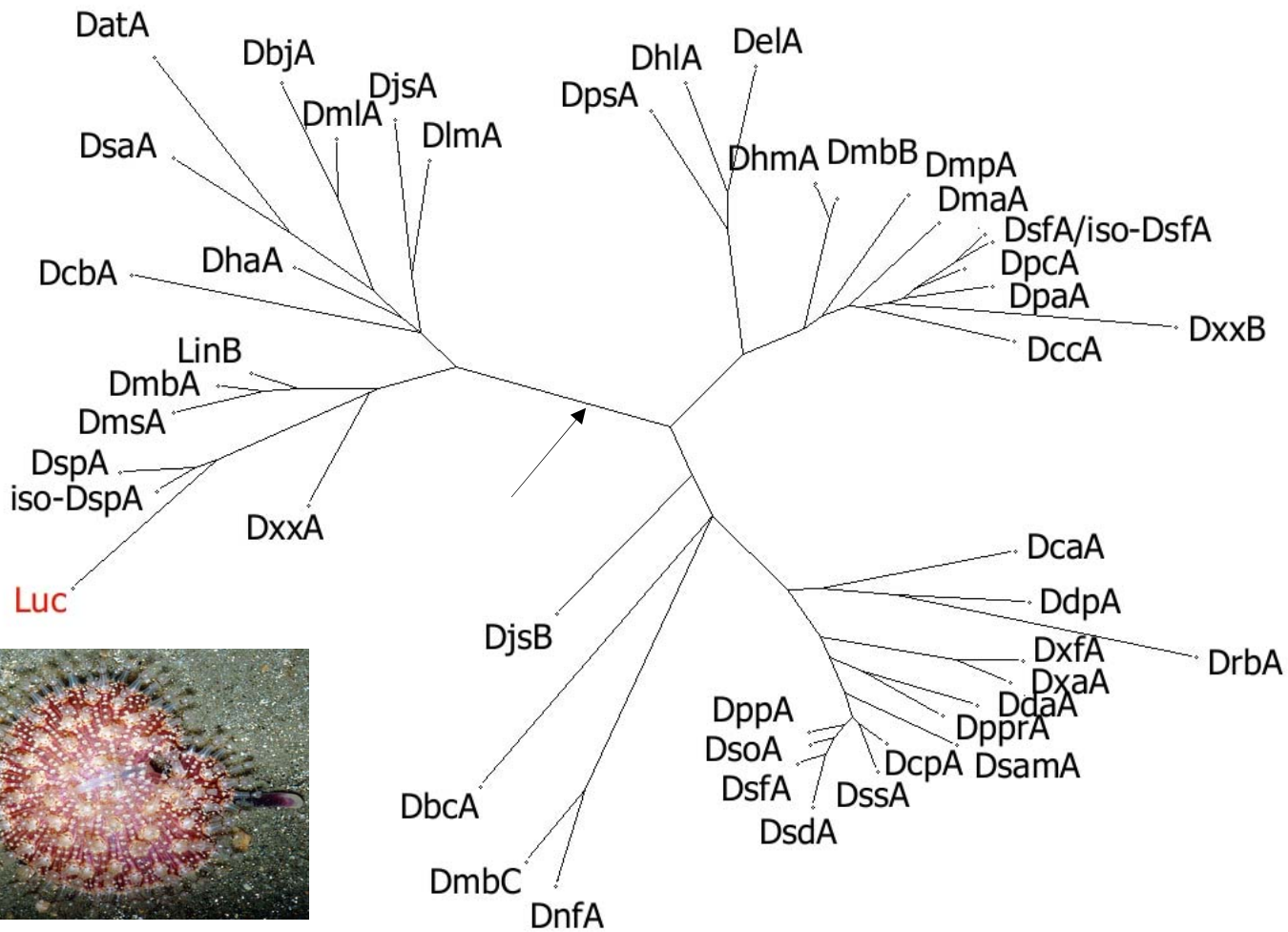
Interesting proteins



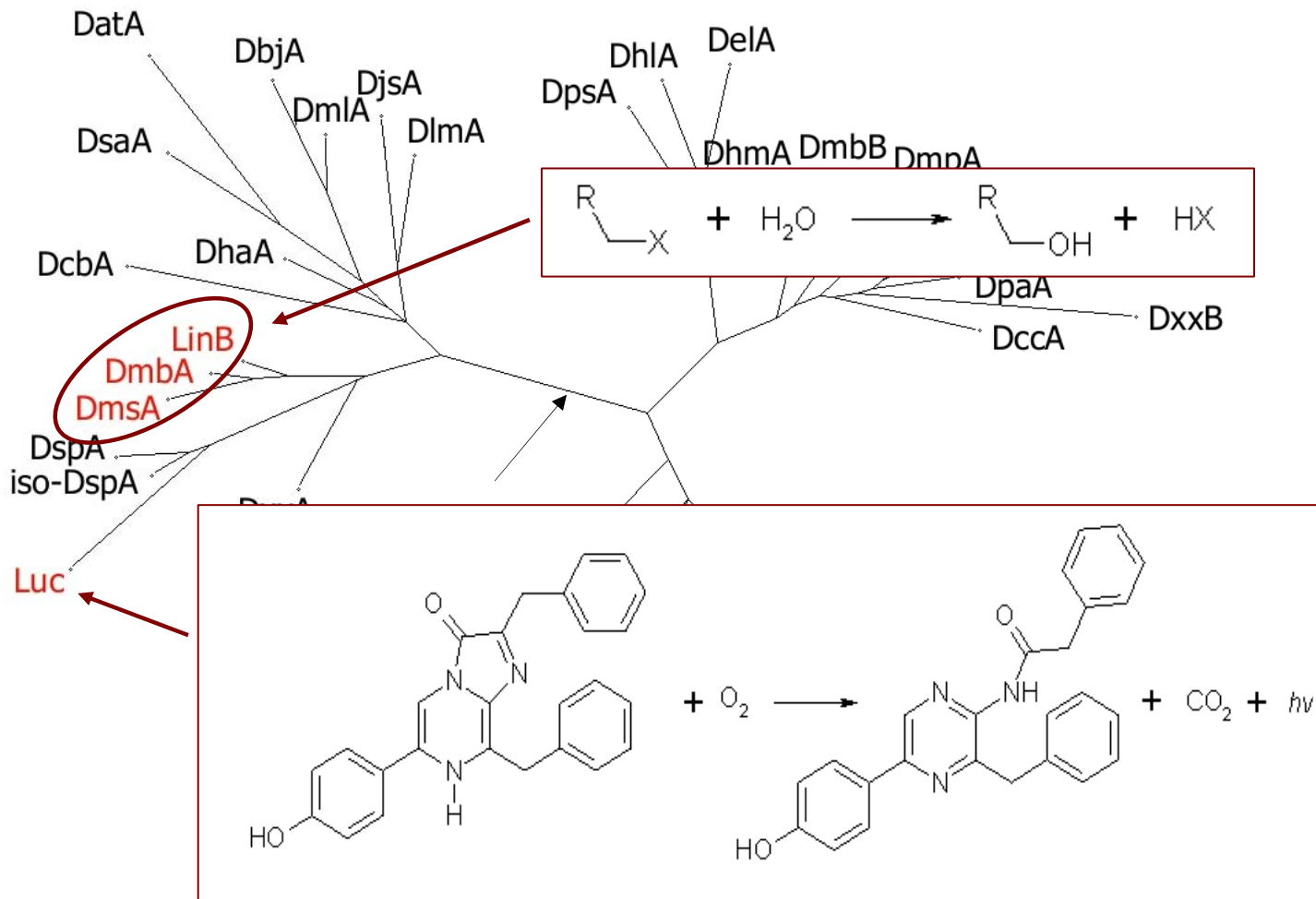
Interesting proteins



Interesting proteins



Interesting proteins

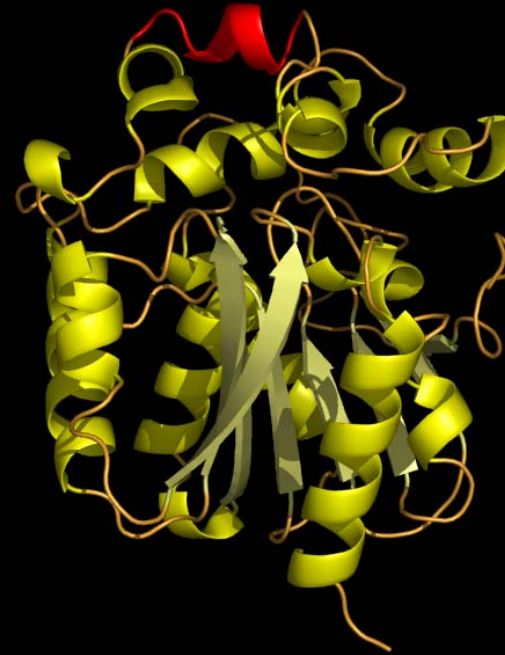


Evolution of enzyme elements

HAD-I



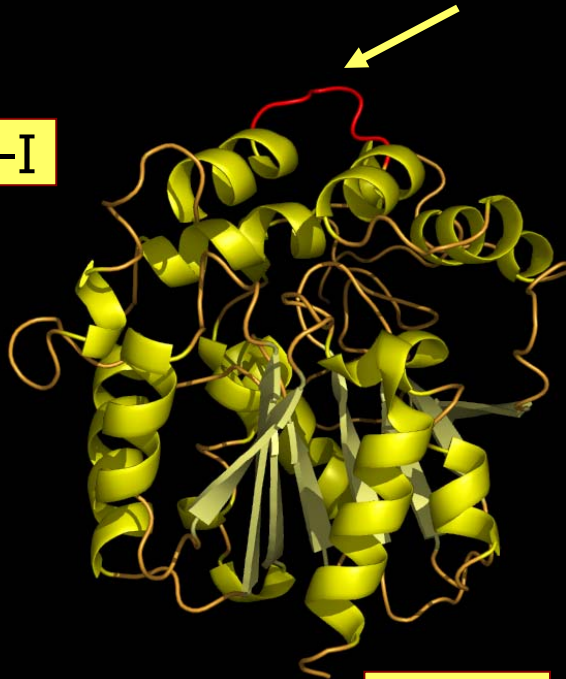
HAD-II



	210	220	230	240	250	260	270	280	290	
01 X.autotrophicus	GFTAMKYDLVT	-----PSD	RLDQFMKRWAPT	--LTEA	ASAYAAPFDDT	-SYQAGV	RRKFERKMAQ	-RDQACTI	--DISTEA	ISFWQND
01 M.bovis dmbB-50	PFYVMRAFARY	-----SPVL	PAGRIVNF	GTVHR	-VPAGV	RAGYDAPFDKT	-YQAGAR	AFPELVETS	PDDEPAV	--PANRAA
01 S.frigidimarina	ATTKWRQFSQD	-----VAIF	PTGNLINS	ACVST	-LTV	ELIAAYDAPF	DER	-YKEGAR	QFPELLV	PTDDPAA
02 M.bovis dmbA-50	VRGVFQGF	FRSE-CQ	EPMALEH	NIFVER	VLEGA	ILRQ	-LSD	EMNH	YRRPFV	NGGED
02 A.tumefaciens	VTRRAKL	-RQEG	EGEIF	VQENR	FLTE	FFWFFL	TF	-LAF	EDLR	QYQTF
02 R.sp. 1BN6 A-72	ARETFQAF	FTADV	GRELI	IDQNA	FTIE	GVLEK	CVVRE	-LTE	VE	MDHYR
03 P.profundum 543	-WALM	IC-RNT	LLGT	GLV	RGFN	AFSSIA	SYIG	VKR	KPMSE	EV
03 X.campestris 78	-WQIAMG	-RHW	RLGE	WFIR	TNF	AFSSG	ASWLG	VSR	-MPSD	V
03 R.baltica -3247	-QR	IAC	-RME	VLT	PA	V	GLN	IF	RAA	V

Evolution of enzyme elements

HAD-I



HAD-I

HAD-II

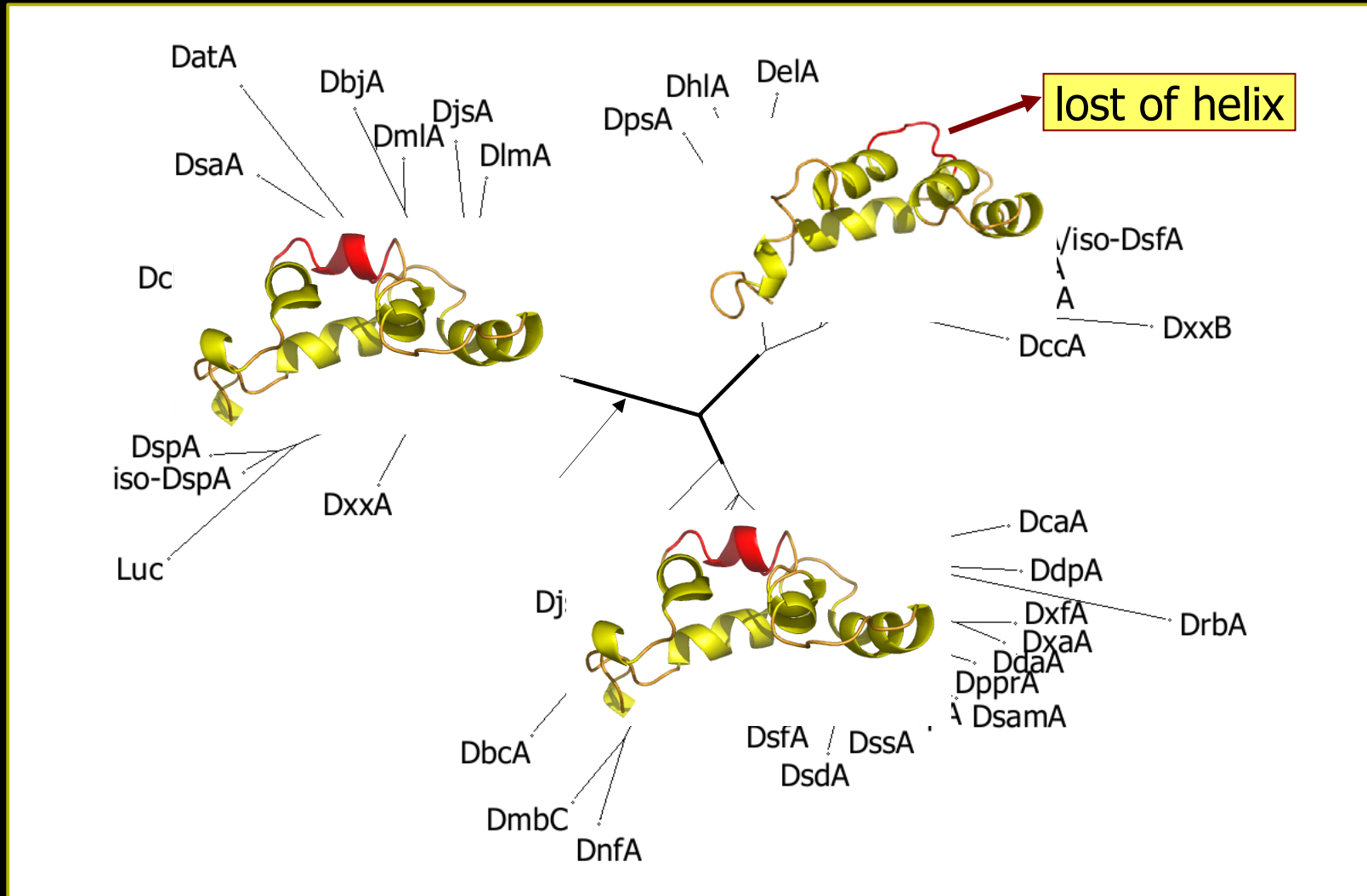
HAD-III

HAD-II



	210	220	240	250	260	270	280	290																																																																																
01 X. autotrophicus	GFTAMKIDLV	-----P	SDRLDQFMKRW	APT--L	EAASAYA	APF	PT-SYQAGV	RKFERM	AAQ-RDQ	ACT--D	ISTE	AI	IS	FW	QND	W	NG																																																																							
01 M. bovis dmbB-50	PFYVMRAF	RY-----	SEVLPAGR	VNF	GT	VHR-	VPAGV	RAGYD	APF	PKT-YQAG	ARAF	EL	V	ET	S	PD	DP	AV--																																																																						
01 S. frigidimarina	AFTKWRQFSQ	Y-----	VAIFPTGN	LINS	AC	VST-L	TV	ELIA	AYD	APF	DER-	YKEG	AR	QF	EL	LV	PT	DD	PA--																																																																					
02 M. bovis dmbA-50	VRGVFQGF	RSF-C	HEPMA	EH	NIF	VER	VLE	G	A	IL	Q-	L	S	D	E	M	N	H	Y	R	R	F	V	N	G	G	E	D	R	P	T	I	S	W	P	N	L	P	I	-	D	G	E	P	A	E	V	V	A	L	V	N	E	Y	R	S	W	L	E	E	T	-	D	M																								
02 A. tumefaciens	VTRRAKL	ROE	GE	GE	IF	Q	EN	R	F	L	T	E	F	F	W	F	L	T	E	-	L	A	P	E	D	L	R	C	Y	Q	T	F	F	T	P	-	H	S	R	K	A	I	L	A	G	E	R	N	L	E	V	-	D	G	E	P	A	S	T	V	A	F	L	E	C	A	V	N	W	L	N	T	S	-	D	T												
02 R. sp. 1BN6 A-72	ARETFQAF	FTAD	V	G	R	E	I	I	D	Q	N	A	F	T	E	G	V	L	K	C	V	V	R	E	-	L	T	E	V	E	M	D	H	Y	R	E	P	L	K	P	-	V	D	R	E	P	L	W	R	F	E	N	E	T	P	T	-	A	G	E	P	A	N	I	V	A	L	V	E	A	Y	M	N	W	L	H	Q	S	-	P								
03 P. profundum 543	-WALM	I	C	-R	N	T	L	G	T	G	L	V	R	G	F	N	A	F	S	S	I	A	S	Y	I	G	V	R	K	K	P	M	S	E	V	R	K	A	Y	V	S	P	E	N	S	W	A	-	N	R	I	S	T	L	R	F	V	C	D	I	P	L	R	K	G	D	R	N	Y	--	E	L	V	S	E	I	A	A	S	L	E	K	F	S	H	V		
03 X. campestris 78	-WQIAMG	-R	H	W	R	L	G	E	M	F	I	R	T	F	N	A	F	S	S	G	A	S	W	L	G	V	S	R	E	-	M	P	S	D	V	R	R	A	Y	V	A	P	Y	N	N	W	R	-	N	R	I	S	T	I	R	F	M	C	D	I	P	L	S	E	A	D	K	G	W	--	S	L	L	E	R	S	A	Q	A	L	E	S	F	A	L	R		
03 R. baltica -3247	-QR	I	A	C	-R	M	E	V	L	G	T	P	A	V	R	G	L	N	I	F	A	R	A	A	V	T	M	A	M	S	E	T	K	M	K	E	D	N	A	A	G	L	L	A	F	Y	D	N	M	K	-	N	R	V	A	I	D	R	E	V	R	D	I	P	L	N	S	H	E	T	M	--	K	T	L	R	Q	L	E	S	D	L	E	D	L	A	S	L

Evolution of enzyme elements



Identification of conserved/variable residues



Summary

- phylogenetic analysis of haloalkane dehalogenases
 - evolutionary history
 - identification of new family members
 - classification of haloalkane dehalogenases
 - selection of potentially interesting proteins
-

Acknowledgements

➤ Protein Engineering Group



Acknowledgements

- Protein Engineering Group
- Laboratory of Bioinformatics and Protein Engineering

