"*Nothing in biology makes sense except in the light of evolution*" -- T. Dobzhansky, 1973

# Phylogenetic Analysis of Haloalkane Dehalogenases

Eva Chovancová<sup>1</sup>, Jan Kosinski<sup>2</sup>, Janusz M. Bujnicki<sup>2</sup>, Jiří Damborský<sup>1</sup>

<sup>1</sup> Loschmidt Laboratories, Protein Engineering Group, Masaryk University, Brno, CZ <sup>2</sup> 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

### similarity searches – PSI-BLAST<sup>1</sup>

0

S NC	СВІ	pr	otein-protein]	BLAST	
Nucleoti	ide Protei	in Trar	nslations Re	trieve results for an RID	
<u>Se</u>	MSEIGTGFPFDPHY	VEVLGERMHYVDVGP:	RDGTPVLFLHGNP'	ISSYLWRNIIPHVAPSHRC	I.
				]	
<u>Set subsequ</u>	ence From:	] To:			
Choose data	base nr 💌				
LAST results:					
400 protein seq	uences	eset query Reset all			
					-

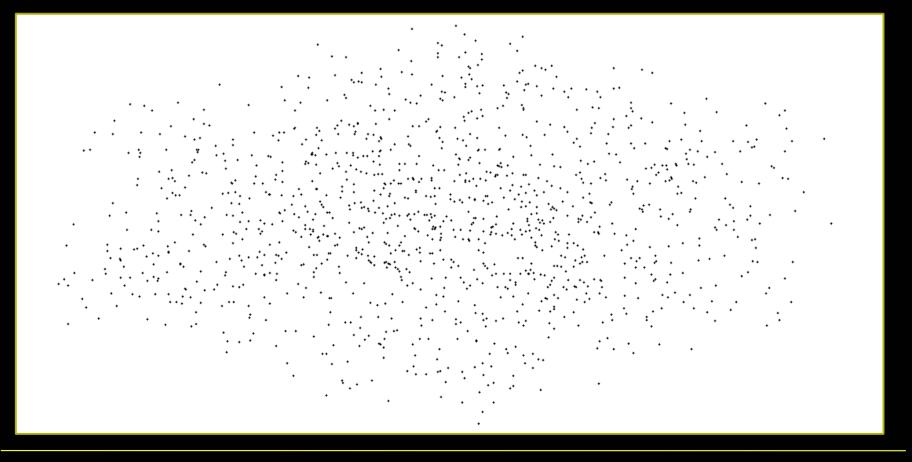
<sup>1</sup> Altschul, et al. (1997). Nucleic Acids Research 25: 3389

### similarity searches – PSI-BLAST<sup>1</sup>

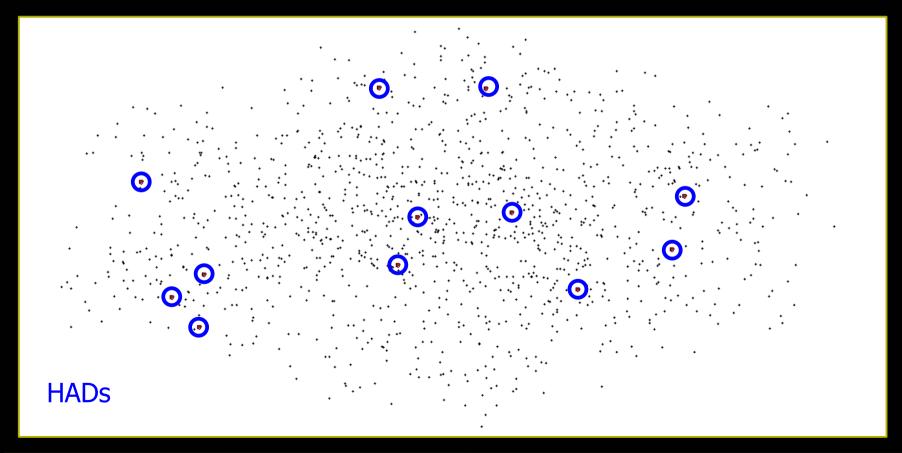
S NCBI	p	rotein-pro	tein <b>BLAST</b>	
Nucleotide	Protein Tra	anslations	Retrieve results for an RID	
	MSEIGTGFPFDPHYVEVLGERMHYVDVG	PRDGTPVLFL	HGNPTSSYLWRNIIPHVAPSHRCI.	
<u>Search</u>				
			Þ	
Set subsequence	From: To:			
	To.j			
Choose database	nr 💌			
BLAST results:		F	low to identify <b>h</b>	aloalkane
3400 protein seque			halogenases a	
3400 protein sequ	ences or Reset query Reset all	ue	naivychases a	

<sup>1</sup> Altschul, et al. (1997). Nucleic Acids Research 25: 3389

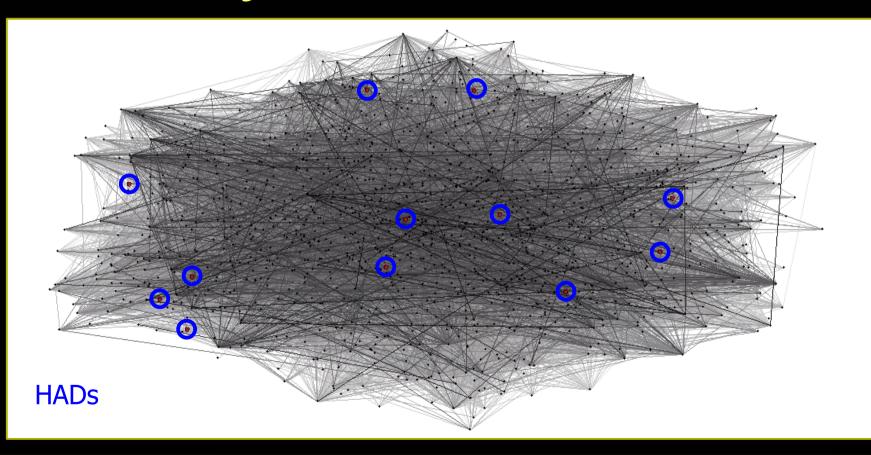
## cluster analysis - CLANS<sup>1</sup>



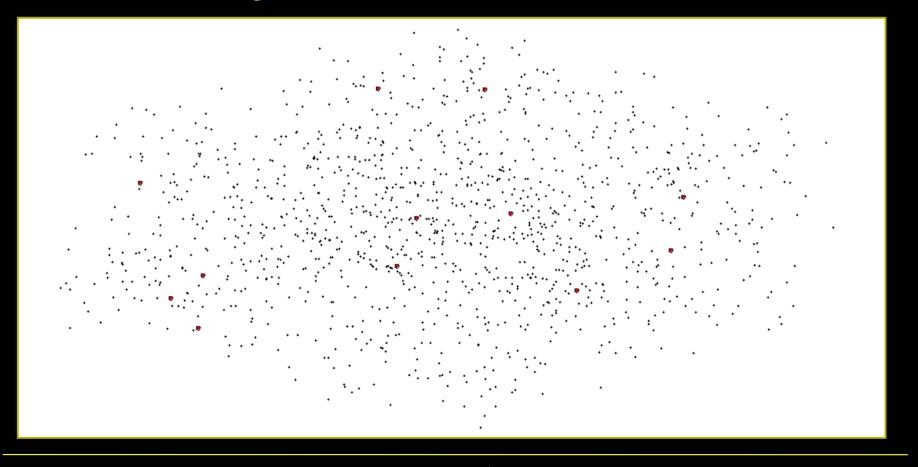
### cluster analysis - CLANS<sup>1</sup>



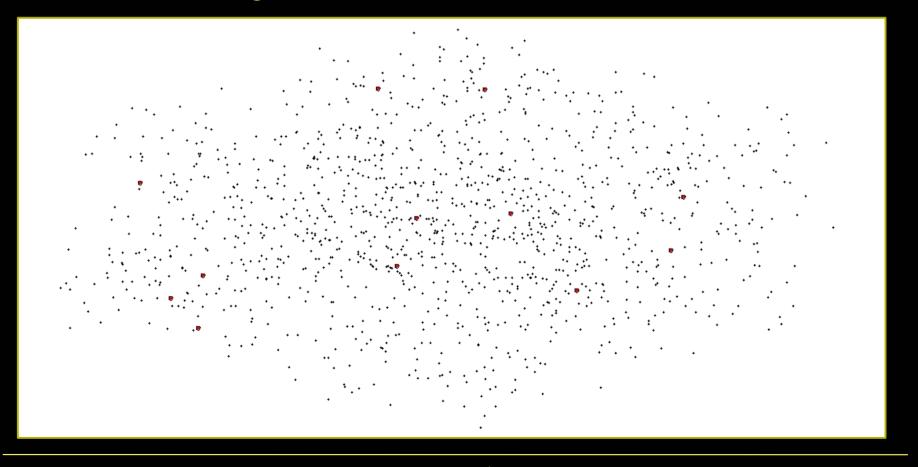
### cluster analysis - CLANS<sup>1</sup>



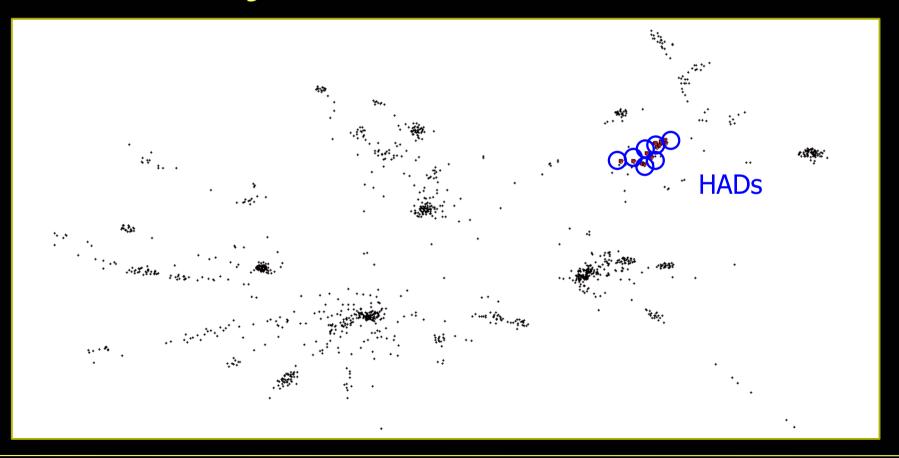
## cluster analysis - CLANS<sup>1</sup>



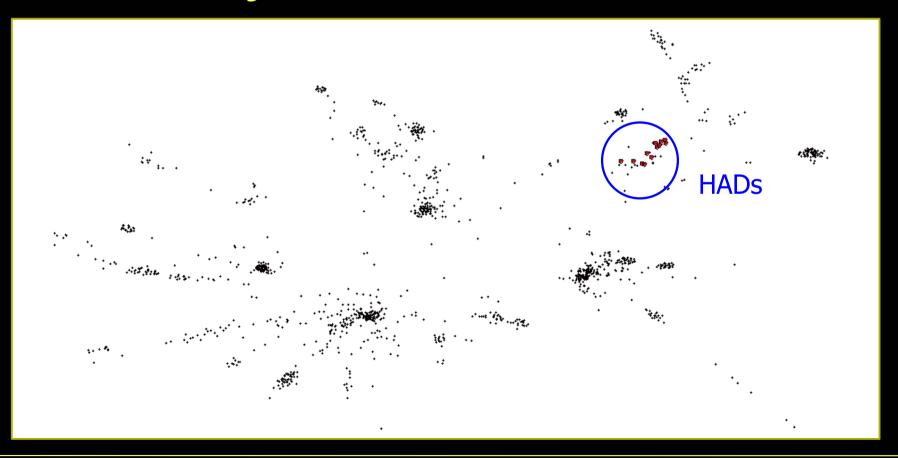
## cluster analysis - CLANS<sup>1</sup>



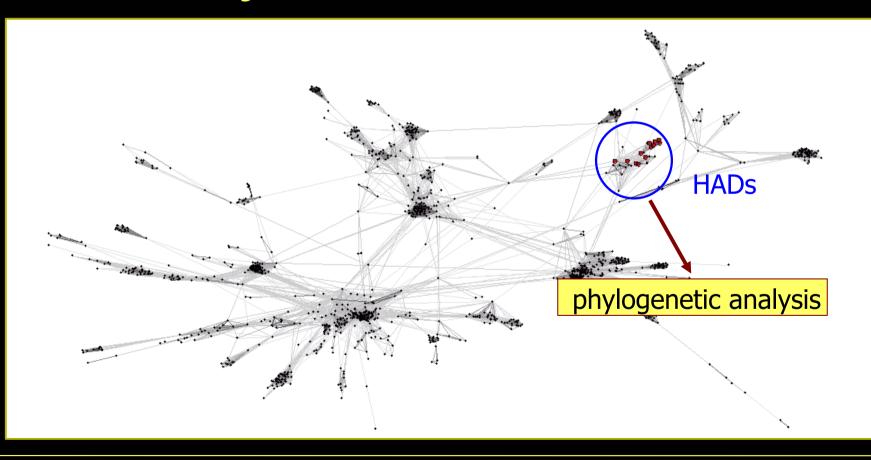
#### *cluster analysis -* CLANS<sup>1</sup>



### cluster analysis - CLANS<sup>1</sup>



### cluster analysis - CLANS<sup>1</sup>



### > multiple sequence alignment – MUSCLE<sup>1</sup>, BioEdit<sup>2</sup>

												١×							
🔒 📇 🛛 Courier Ne	ew 🗖	- 11	• B		83 total	sequences			shade t	hreshold	35%	▼⇒┌							
Mode: Select / Slide	<b>-</b>		ection: 0						k: None			9	Start ruler at: 1	_					
	_	Pos	ition: 22:	B.sp780	62699 62	2	Numbe	ring Mas	k: None										
	🔒 сл	+- ₩		Va 🖬 📗		ti î Cât	GAT TINS CAT :::::		- 🚯	MI 🚺	8	Scroll 💽 speed slov		 fast					
×	12		130 I	••••• • ]	140	••••	' '''' 150		' ''' 160		' '' 170	•••••	•   • • • 180		' '''' L90	•	200		0
S.purpuratus							_	FLDH	YRY	_		PEKI				NE <mark>H</mark> RI	BLEAI	HMEGI	
R.reniformis	HVVPHI	EPVA-		DLIGMG	KS <mark>GK</mark> -	SGNG	S <mark>Y</mark> F	LDH	YKY T	W-FEI	LLNL	PKKI	IFVG <mark>H</mark>	D <mark>WG</mark> ACI	LAFHYS	SYE <mark>H</mark> QI	KIKAI	VHAESV	, 💻
L.majuscula		VAAGY	AV <mark>A</mark> PI	DLIGMG	DSAKE	DI	E <mark>Y</mark> F	LODH	VAYMD	F-ID/	ALGL	DDM	VLVI <mark>H</mark>	Dw <mark>g</mark> sv	GMRH	RLNPI	RVAAU	AFMEAL	:
S.paucimobil	NIMPHC		RLIAC	DLIGMG		DPSGP	ERYI	YAE <mark>H</mark>	R <mark>D</mark> YLD	AL-WE/	ALDL	GDRV	VLVV <mark>H</mark>	D <mark>WG</mark> SVI	l <mark>g</mark> fdw <mark>/</mark>	RR <mark>H</mark> RI	R <mark>R</mark> VQGI	AYM <b>E</b> AV	7
M. bovis dmbA	NIMPHL	EGLG-	LVAC	DLIGMG	AS <mark>DK</mark> I	SPSGP	DR <mark>Y</mark> S	YGEQ	RDFLF	AL-WD2	ALDL	GDHV	VLVL <mark>H</mark>	DWG SAI	L <mark>G</mark> FDW <mark>Z</mark>	NQ <mark>H</mark> RI	RVQGI	AFMEAT	C 🗌
M. smegmatis	NVLPHT	EGLG-	LVAC	DLIGMG	ASDKI	DGSGP	DS <mark>Y</mark> F	<u>iy</u> hen	R <mark>D</mark> Y <b>L</b> F.	AL-WD2	ALDL	GDRV						VHMET V	
A.tumefacien	HIFRR	HGHG-	LLAV	DLI <mark>G</mark> YG	OSS <mark>K</mark> I	DI			QRYVD.	W-FD2	ALDL	R = -NV	TLVLQ	DY <mark>G</mark> AAI	7 <mark>G</mark> LNW <mark>∕</mark>	SRNPI	) <mark>R</mark> VRAV	/AFFEPV	7
R.muelleri l	HVVPHV	EPVA-	RCIIPI	DLIGMG	KSG <mark>K</mark> -	SGNG	S <mark>Y</mark> F	LDD <mark>H</mark>	YKY <mark>I</mark> TI	EW-FK	HLNL	PKKI	IF <mark>V</mark> G <mark>H</mark>	D <mark>WG</mark> AC]	LAFHY	'YE <mark>H</mark> QI	) <mark>r</mark> ikay	V <mark>HAE</mark> SV	7
J.sp. 681839	KFIPAL	AQTH-	RVIAPI	DHIGFG	LS <mark>DK</mark> E	EAEG	<mark>D</mark> YI	'LDA <mark>H</mark>	IQN <mark>IE</mark>	AL-VQO	<u> ID</u> I	T NT	TLVMQ	D <mark>WGG</mark> P	GLGM <mark>/</mark>	AR <mark>HP</mark> A	RIKA I	VVMKT F	r
J.sp. 681814	NVIPHV	AEDH-	ATAI	DLIGMG	AS <mark>DK</mark> E	DI	DYD	FQD <mark>H</mark>	YAH <mark>I</mark> E(	S <u>F</u> -IDA	A L E L							AMM <mark>E</mark> AA	
	NIMPYA	EEAG-	RITAPI	DLIGMG	DSEKI	ENSGP	DSY1	FQE <mark>H</mark>	AKY Y	K <mark>I – F</mark> EI	ELEL	D = -NV	NLVI <mark>H</mark>					TYMEAI	
M.loti 13474		APFG-		DLIGYG	03 <mark>G</mark> KI	DI	DYF	REFD <mark>H</mark>	VRY D	F-LD/	ALDI	RDV	LLV <mark>A</mark> Q					AFMEFI	
M.bovis dmbB	TMIPPL	SAAGH	NULAPI	DLIGFG	RS <mark>DK</mark> E	TRI	E <mark>D</mark> YI	YLR <mark>H</mark>	VEWVT:	3W-FEI	NLDL		TIFVQ					, V <mark>VA</mark> NGF	
R.sp. 1BN6 A	NITPHV	APSH-	RCIAPI	DLIGMG	KS <mark>DK</mark> E	DL	<b>DY</b> E	'FDD <mark>H</mark>	VRYLD.	F-IE/								ACMEFI	
M.sp. DHAAF-	NITPHV	APSH-	RCIAPI	DLIGMG	KS <mark>DK</mark> I	DL	DYE	FDDH	VRYLD.	F-IE/								ACMEFI	
S.avermitili	DVMPAV	-GSG-	RLL <mark>AP</mark>	DLIGMG	ESGKE	AL	<b>D</b> Y1	FAD <mark>H</mark>		NU−FD2		R = -DV						AFTETI	
M.tuberculos	NIMPHL	EGLG-	LVAC	DLIGMG	ASDKI	SPSGP	DRY 8	SYGE O	R <mark>D</mark> FLF.			GDHV						AFMEAI	
S.purpuratus	NVIPQV	EPIA-	RCLAPI	DLIGMG	RSDRI	ASR	S <mark>Y</mark> F	(FLD <mark>H</mark>	YRYT S	W-FD	ALKL	PEKI						VHMEGV	
S.purpuratus	NVIPQV	EPIA-	RCLAPI	DLIGMG		ASR	S <mark>A</mark> b	(FLD <mark>H</mark>	YRYT S	W-FD	ATKL	PEKI						VHMEGV	
S.purpuratus		EPIA-	RCLAPI	DTIGOG	RSN <mark>K</mark> I	AN	HS <mark>Y</mark> F	_		W-FD:	SVN	PEKV						IHMESV	
B.cenocepaci	HLVTAL	SPTY-	VVVPI	DHMGFC		-		ILQDH		RF-VL	AHDI							VSANGE	2
B.sp. 780626	HLTAAL	SPTH-	VVVPI	DHMGFG	KSATI		RS <mark>Y</mark> V				ARG			DFGGP					?
	NLVTEL				LS <mark>DK</mark> E			I KNR										'AITNIG	
X. fastidiosa	HLVAAL	RDGY-	RCIVPI		LS <mark>DK</mark> E			'LQS <mark>R</mark>				DDVTPL					_		
S.oneidensis	NLVSAL	KDTH-(	OCIVPI			DDSG		'I <mark>KNR</mark>		L-LD								'AITNIG	
X.oryzae 585	HLVRGL	CDRY-	RCIVPI	DHIGMG		DDAPDAC				ri-iri						_		'VITNT <mark>A</mark>	
X.fastidiosa	HLVAAL	RDGY-	RCIVPI	DHIGMG		GDAPGVV	'PRY <mark>DY</mark> '	'IQSR	VDDID.	AL-LR		DDVTPL		D <mark>W</mark> GGM		_		'VITNT <mark>A</mark>	
X.fastidiosa	HLVAA	RDGY-	RCIVPI		LS <mark>DK</mark> E		'PRY <mark>DY</mark> '		VDDDDD	AL-LR		DDVTPL			[ <mark>G</mark> FGW <mark>/</mark>	LA <mark>H</mark> AV	/QVR <mark>R</mark> I	, VMT NT <mark>A</mark>	1
1bn6 ss	LTHHHH	LLLL-I	LSSSSI	LLLI	LLLI	LL	LLI	HHH	ннннн	HH-HH	HLLL	LL S	33333	SHHHHI	ннннн	HHLHE	HSSSS	SSSSSSI	

<sup>1</sup> Edgar (2004). Nucleic Acids Research 32: 1792 <sup>2</sup> http://www.mbio.ncsu.edu/BioEdit/bioedit.html

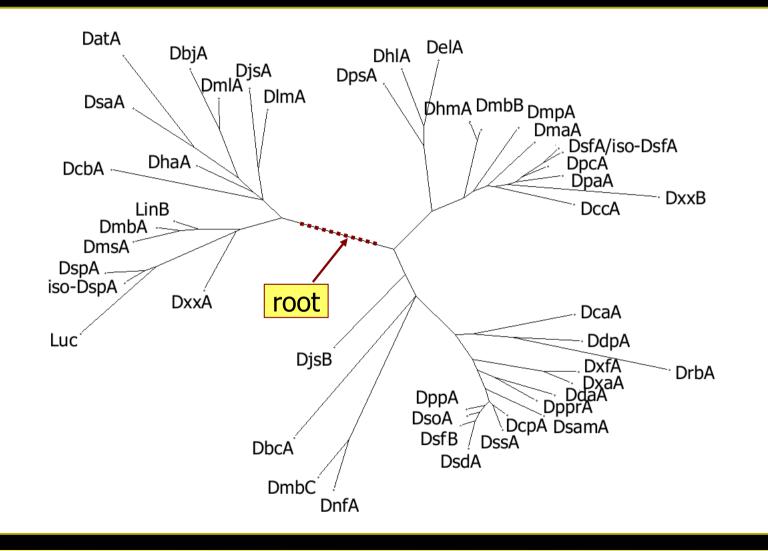
- > multiple sequence alignment MUSCLE, BioEdit
- selection of evolutionary model PROTTEST<sup>1</sup>

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

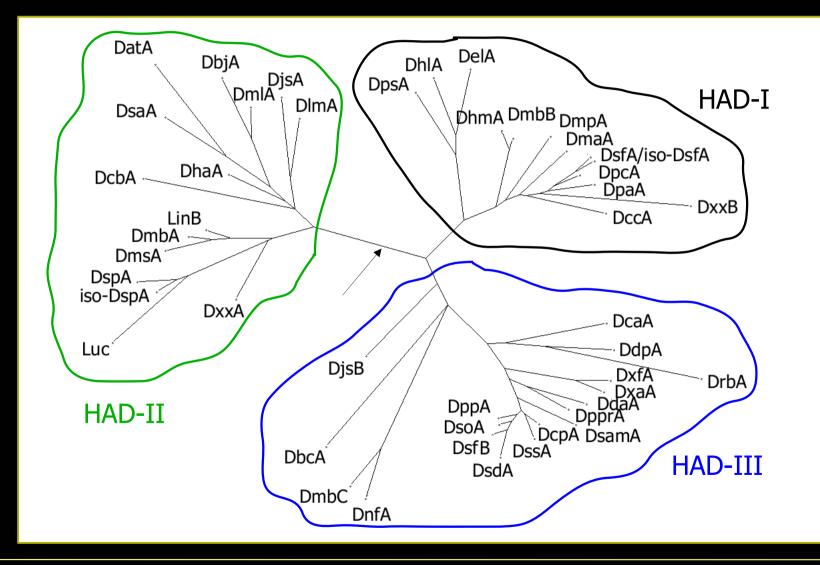
- > multiple sequence alignment MUSCLE, BioEdit
- » selection of evolutionary model PROTTEST
- > construction of phylogenetic tree:
- maximum likelihood method PHYML<sup>1</sup>
- » neighbour-joining method MLDIST<sup>2</sup>

- » 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

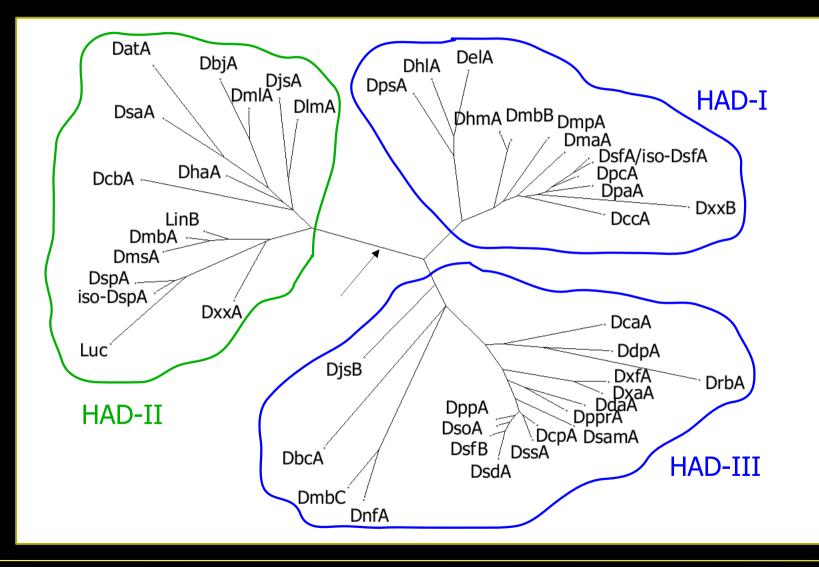
## **Phylogenetic tree**



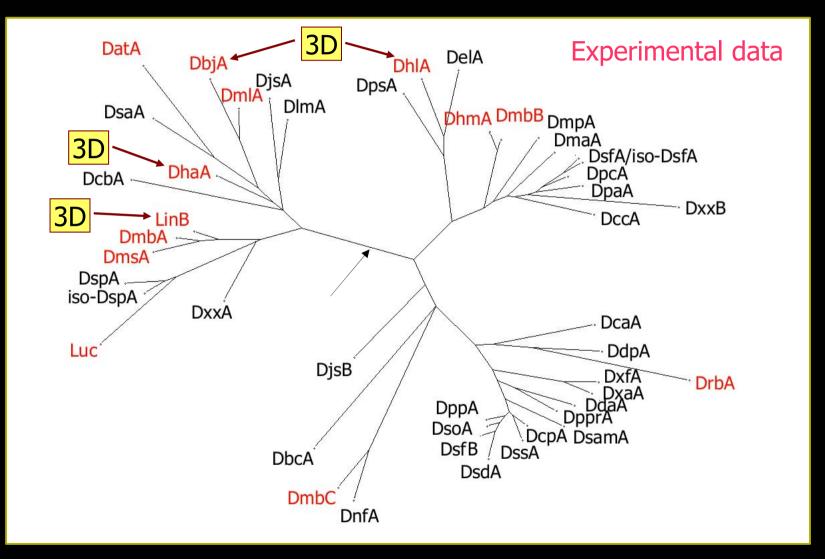
### **Classification**



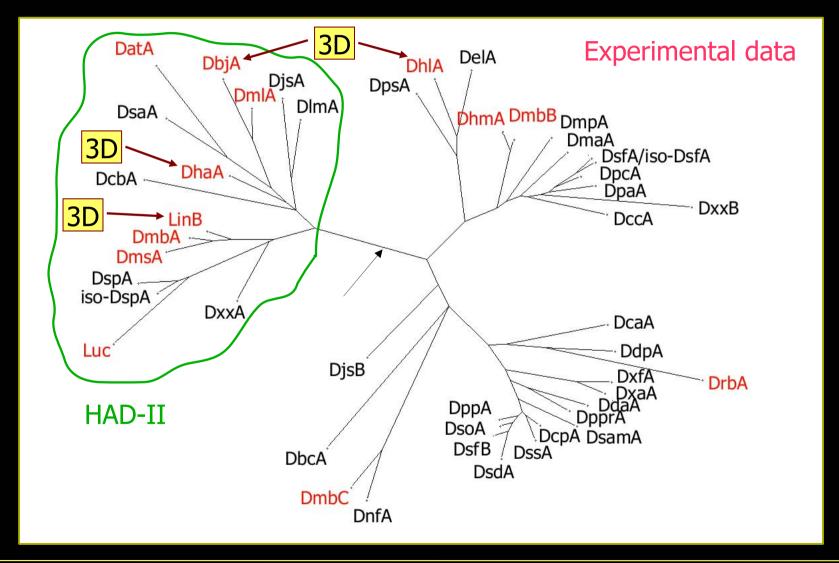
### **Relationships among subfamilies**

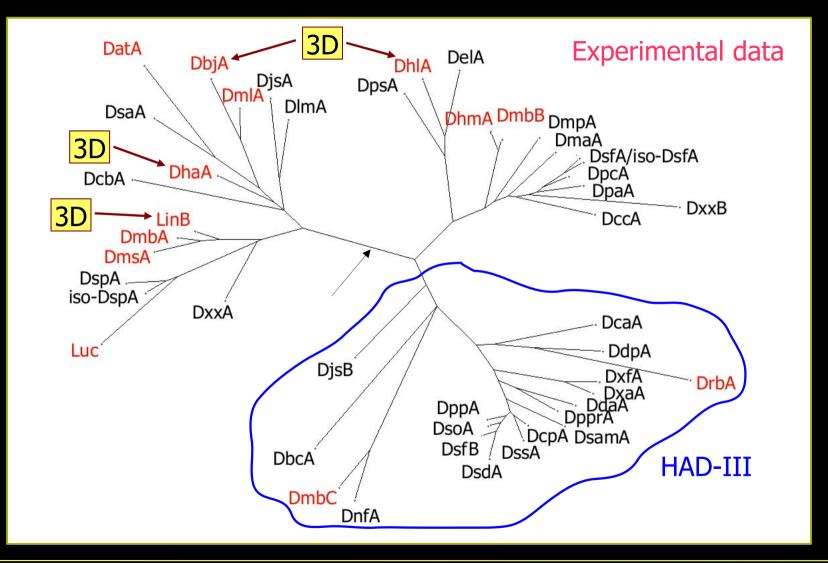


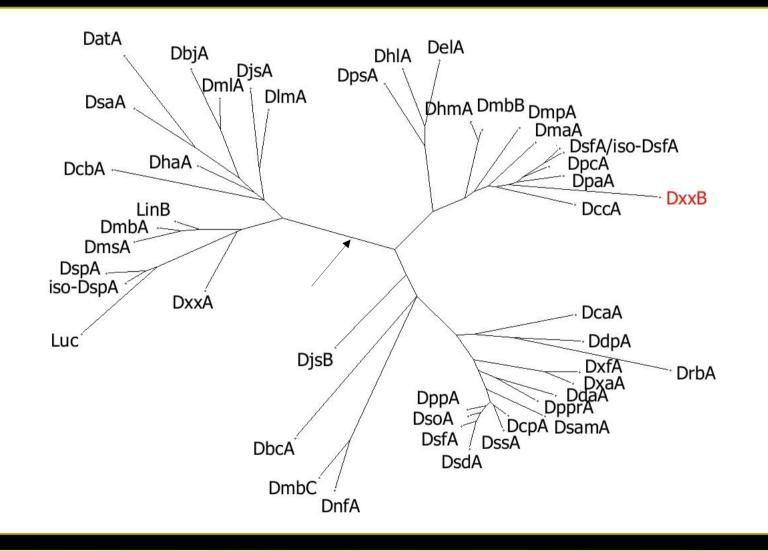
### Experimental data

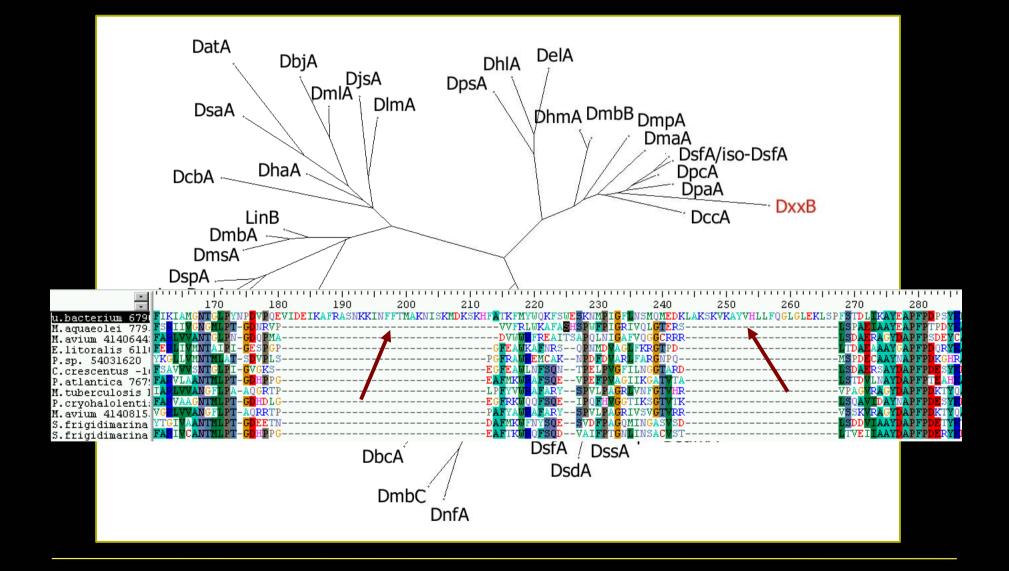


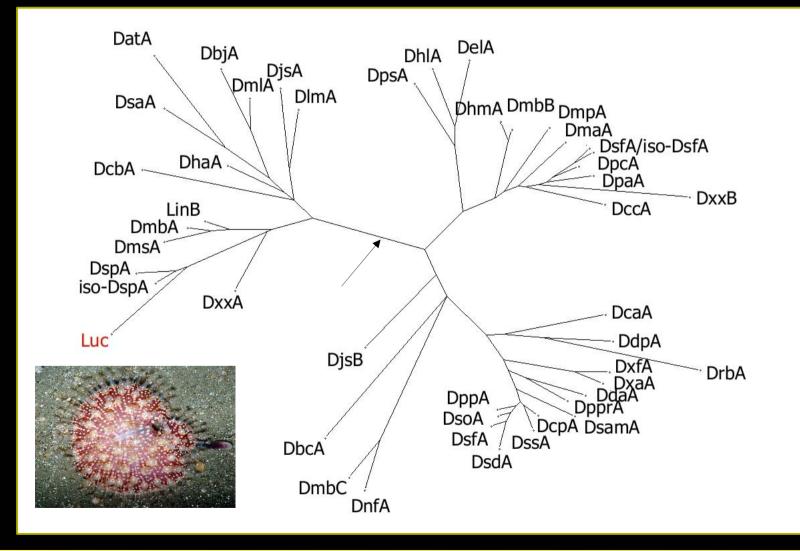
### Experimental data

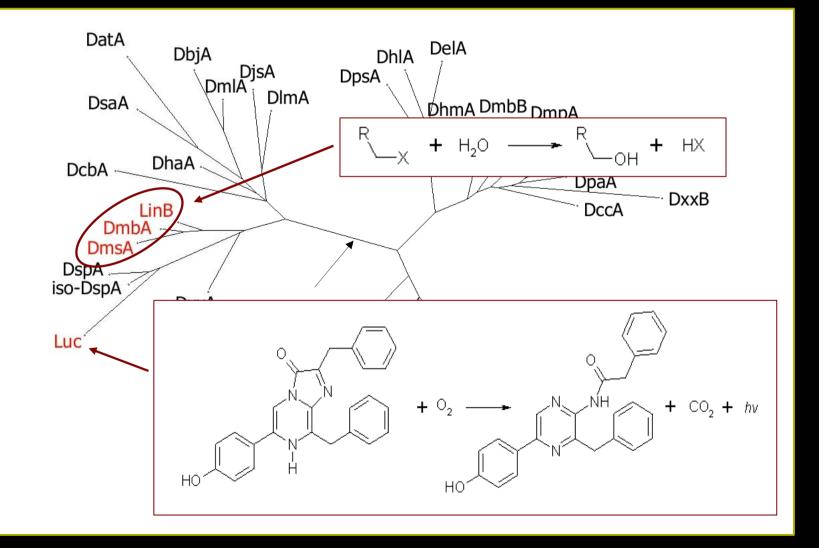




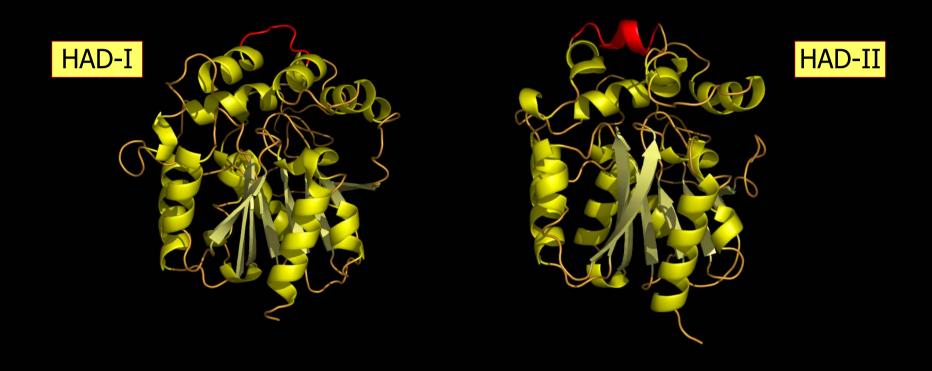






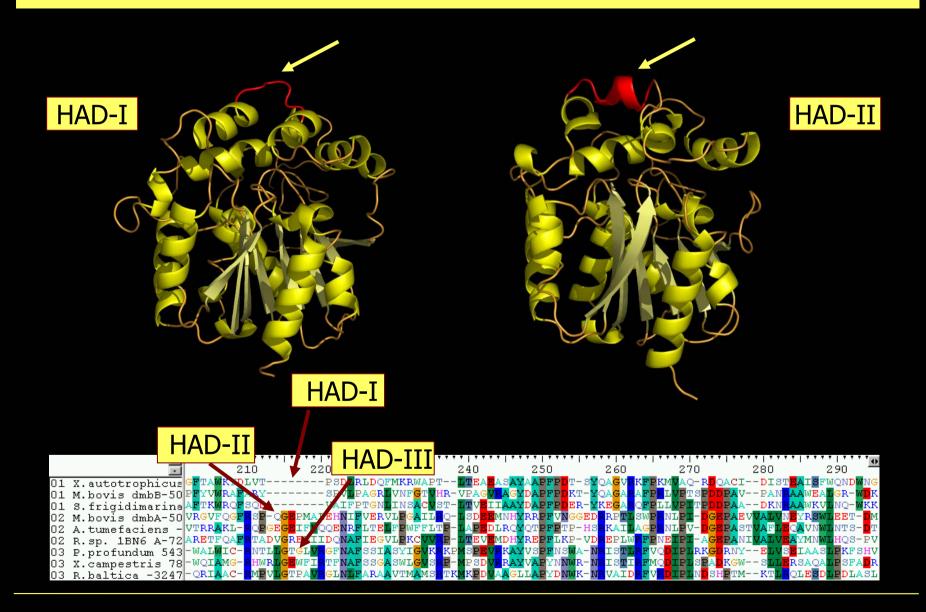


### **Evolution of enzyme elements**

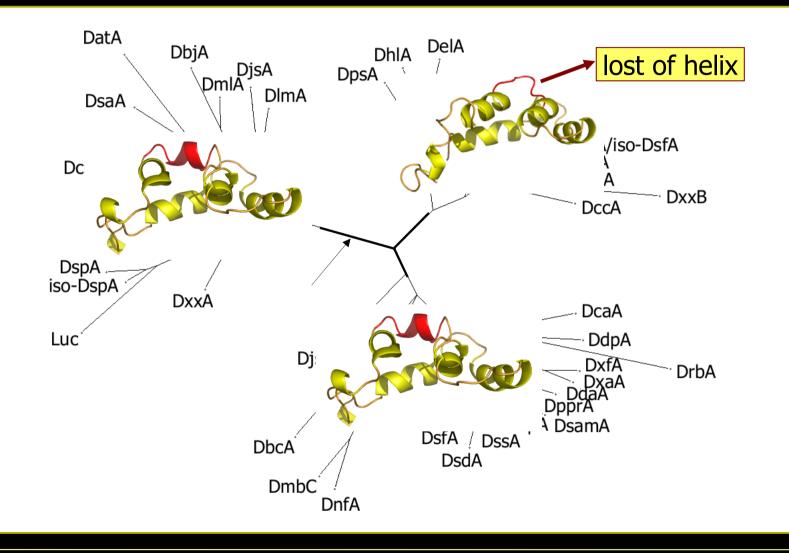


		210	220			250			280	290 -
01	X.autotrophicus <mark>GF</mark> TA <mark>W</mark> K	YDLVT	PSD <mark>L</mark> RLD	QFMKRWAPT-	-LTEA <mark>E</mark> AS <mark>A</mark>	YA <mark>APFPD</mark> T-S	YQA <mark>GVR</mark> KFPKI Yqagarafpri	M <mark>VAQ-R<mark>D</mark>QACI</mark>	IDIST <mark>E</mark> A	ISFWQNDWN0
01 1	M.bovis dmbB-50 P <b>F</b> YV <b>W</b> R		SPV <mark>I</mark> PAG	RLVNF <mark>G</mark> TVHR	<u>VPAG<mark>VRA</mark>G</u>	YDAPFP <mark>D</mark> KT-	YQA <mark>G</mark> A <mark>R</mark> AFP <mark>R</mark> i	L <mark>VPT</mark> SP <mark>DDPA</mark> V	VPAN <mark>R</mark> AA	WEA <mark>L</mark> GR-W <mark>D</mark> H
01	3.friqidimarina A <mark>F</mark> TK <mark>M</mark> R	Q <mark>F</mark> SQD <u></u> -	VAIFPTG	NLINS <mark>ACV</mark> ST	-LTV <mark>E</mark> II <mark>AA</mark>	YDAPFP <mark>D</mark> ER-	YKE <mark>GAR</mark> OFPL:	L <mark>VPI</mark> TP <mark>DDPA</mark> /	ADKN <mark>RAA</mark>	WKVLNQ- <mark>W</mark> KF
02 0	M.bovis dmbA-50 VRGVFQ	G <mark>FR</mark> SP-Q <mark>GE</mark> P	MALEHNIFVE	RVLPG <mark>A</mark> IL <mark>R</mark> Q	- <mark>I</mark> SD <mark>EE</mark> MNH	YRRPFVNGGE	D <mark>r</mark> rptiswpr	NLP <b>I-<mark>DGEPA</mark>I</b>	EV <mark>VA</mark> LVN <mark>E</mark> Y.	RSWLEET- <mark>D</mark> N
02.	A.tumefaciens - VTRR <mark>A</mark> K	L- <mark>R</mark> OPGE <mark>GE</mark> I	F <mark>VQQENRF</mark> LT	ELFPWF <u>FLT</u> P	- <mark>LAPE</mark> DLRQ	YQTPFPTP-H	IS <mark>R</mark> KAT <mark>LAG</mark> P <mark>R</mark> I	NLPV- <mark>DGEPA</mark> :	ST <mark>VA</mark> FL <mark>E</mark> QA	VN <mark>WL</mark> NTS- <mark>D</mark> I
02	R.sp. 1BN6 A-72 ARETFO	A <mark>FR</mark> TADV <mark>G</mark> RE	LIIDO <mark>NAF</mark> IE	GVLPKC <mark>VVR</mark> P	-LTEV <mark>E</mark> MDH	YREPFLKP-V	D <mark>replwrfp</mark> ni	EIPI-A <mark>ge</mark> pa	NI <mark>VA</mark> LV <mark>EA</mark> YI	MN <mark>WL</mark> HQS-PV
03	P.profundum 543-WALWI X.campestris 78-WQIAM	C- <mark>R</mark> NTL <mark>LG</mark> TG	l <mark>vr</mark> gf <mark>naf</mark> ss	I <mark>A</mark> SYI <mark>GV</mark> K <mark>R</mark> K	PMSPEVRKA	YVSPFNSWA-	NRISTLRFVQ	DIPLRK <mark>GD</mark> RNY	ZELVS <mark>E</mark> I.	AAS <mark>LP</mark> KFS <mark>H</mark> V
03	X.campestris 78-WQI <mark>A</mark> M	3- <mark>R</mark> HWR <mark>LGE</mark> W	FI <mark>R</mark> TF <mark>NAF</mark> SS	G <mark>a</mark> swl <mark>gv</mark> s <mark>rp</mark>	-MPSDVR <mark>R</mark> A	YV <mark>A</mark> PYNNWR-	NRISTIRFMO	DIPL SPA <mark>D</mark> KGI	VS <mark>LL<mark>E</mark>RS.</mark>	AQA <mark>LP</mark> SFA <mark>D</mark> F
03	R.baltica -3247-QRIAA	C- <mark>R</mark> MEVLGTP	A <mark>VR</mark> GL <mark>NLF</mark> AR	A <mark>A</mark> VTM <mark>A</mark> MS <mark>R</mark> T	KMKPDVA <mark>A</mark> G	LL <mark>APYDNW</mark> K-	NRVAID <mark>RF</mark> V <mark>R</mark>	dipln <mark>d</mark> shpti	MKT <mark>LR</mark> QL	ESDLEDLASI

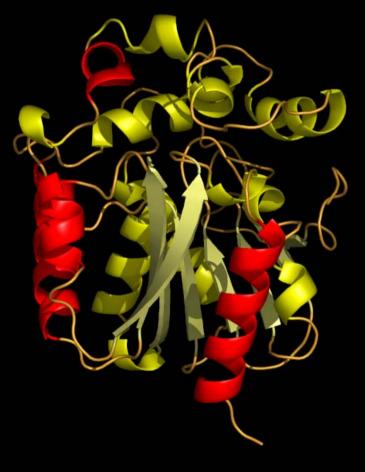
#### **Evolution of enzyme elements**



## **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



