

Concurrent Divergence in Coding and Promoter Regions of the Poplar Gene Family Encoding Xyloglucan Endotransglucosylase/Hydrolases

**Xia Ye¹, Qin Xu², Joshua Yuan¹, Hong Guo²,
Feng Chen¹, (Max) Zong-Ming Cheng***

¹Department of Plant Sciences and

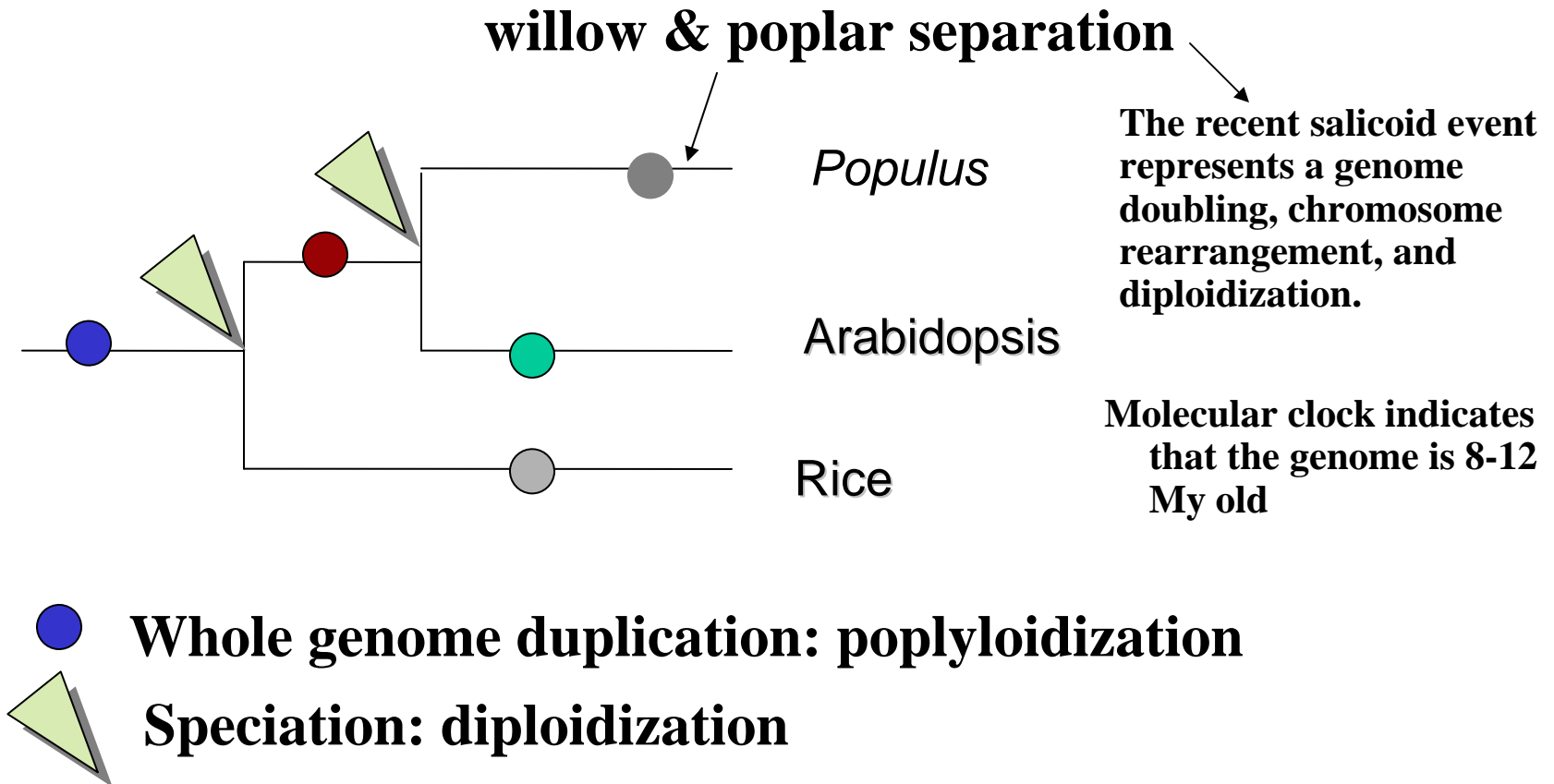
**²Department of Biochemistry, Cellular and
Molecular Biology, University of Tennessee,
Knoxville, TN 37996 *zcheng@utk.edu**



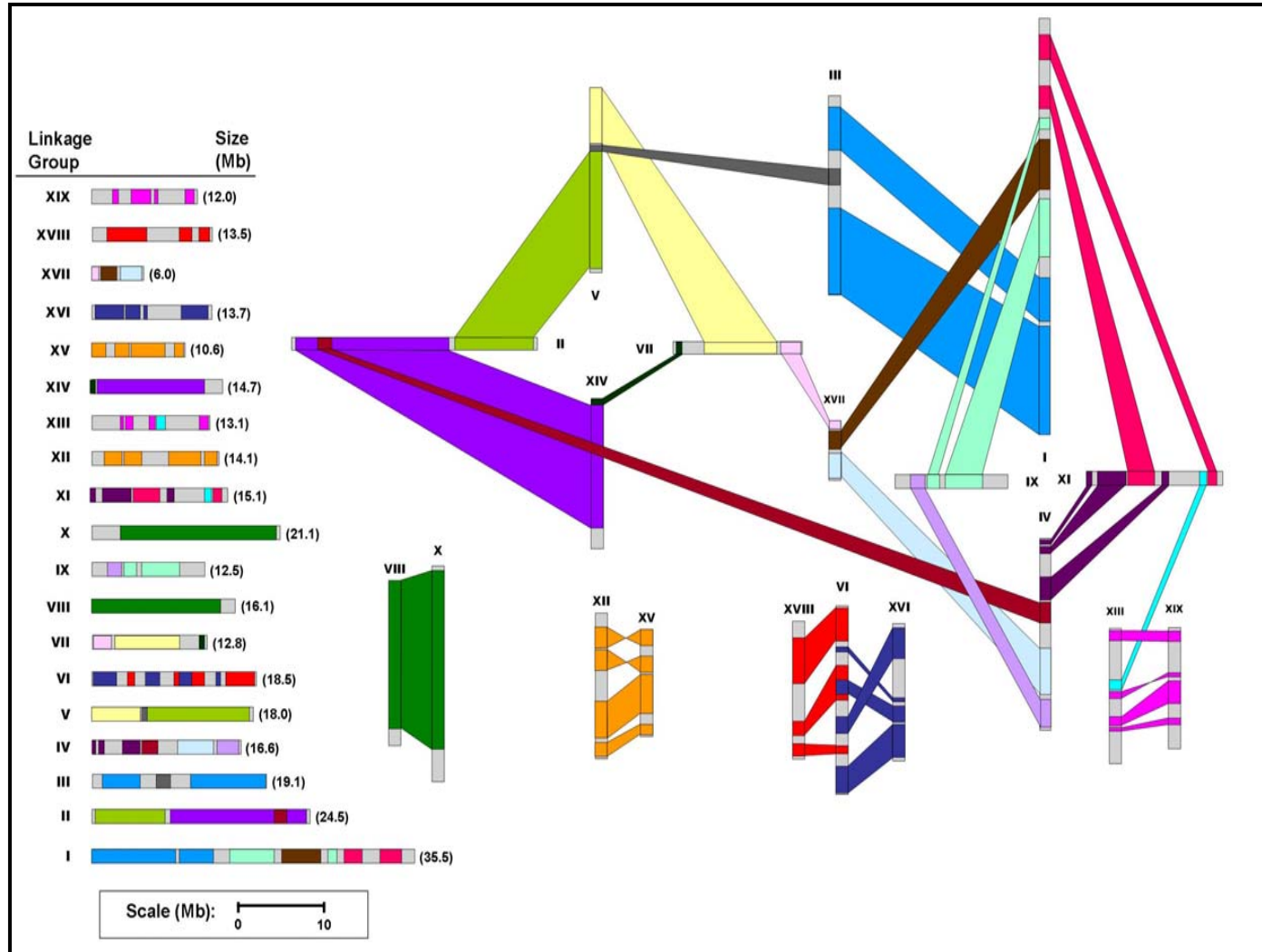
**Tuskan *et al.*, 2006 Science
313:1596-1604**

***Populus trichocarpa* genome
(485 Mb)**

At least three whole genome duplication events in *Arabidopsis* and poplar, and two in rice.



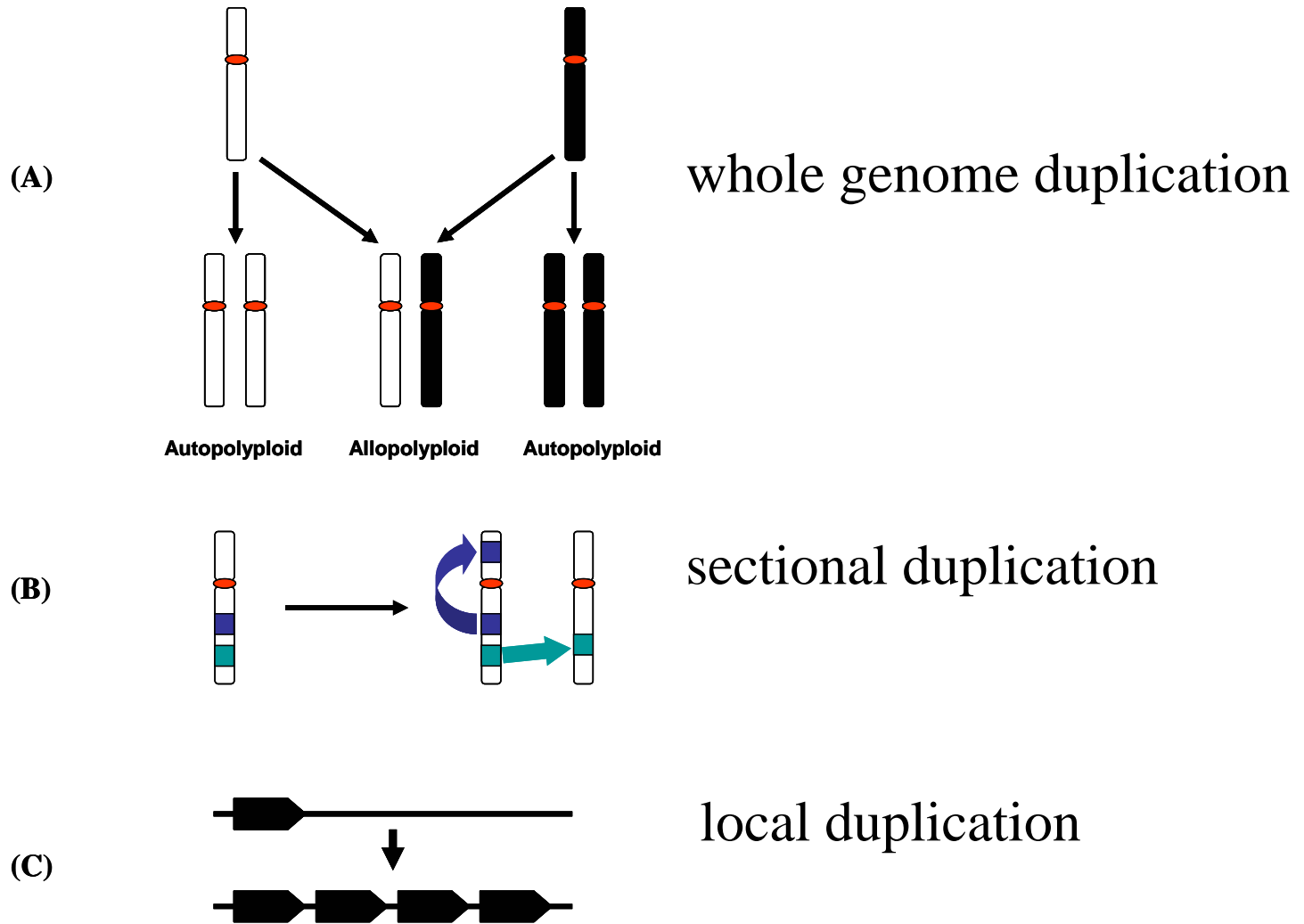
Diploidization after whole genome duplication



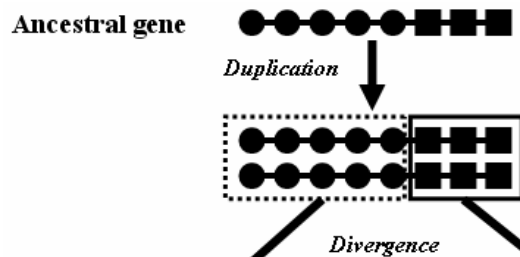
A large megabase blocks of the genome remain in complete co-linearity

17,380 genes in paralogous pairs, sharing ~ 94%+ identity

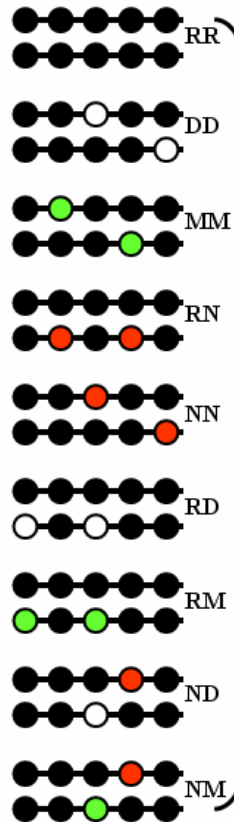
Modes of gene duplication



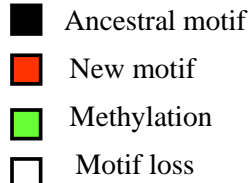
Possible evolution modes of duplicated paralogs



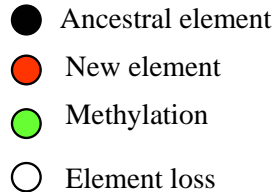
Gene fates



Protein motif



Cis-element



Category

Redundancy

Subfunctionalization

Neofunctionalization

Pseudogenization

X = D, M, N or R;

x = d, m, n or r

Upper case for cis elements,
lowercase for protein motifs

Mode of divergence

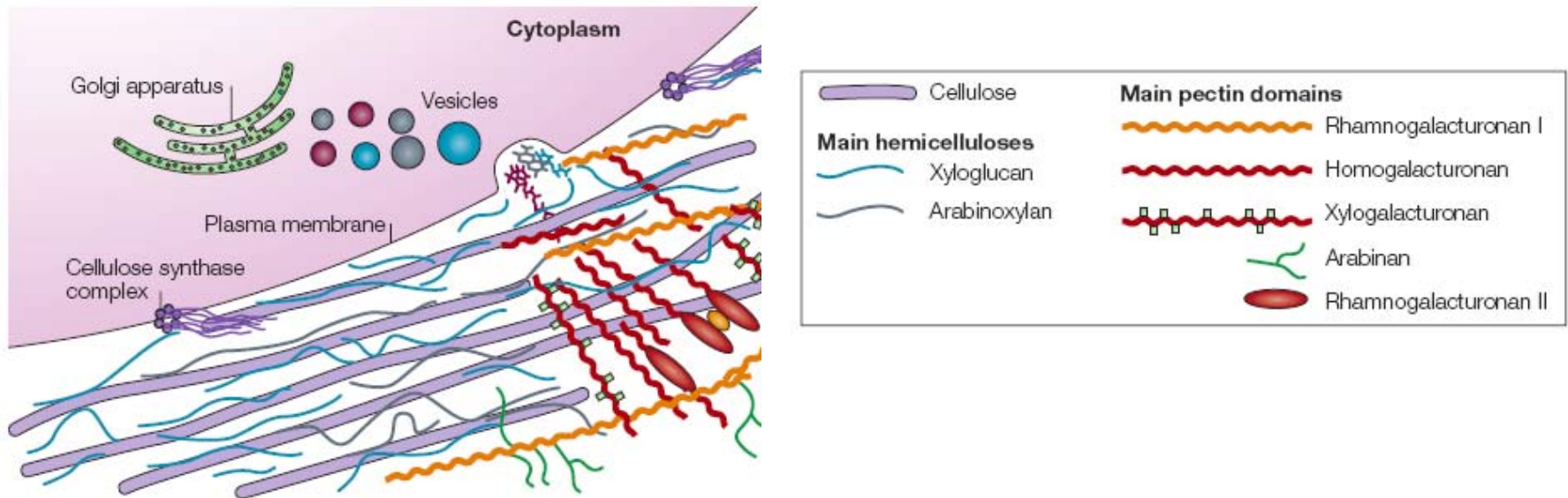
RRrr

DDrr RRdd

RNxx NNxx XXrn XXnn

RDxx RMxx NDxx NMxx XXrd XXrm XXnd XXnm

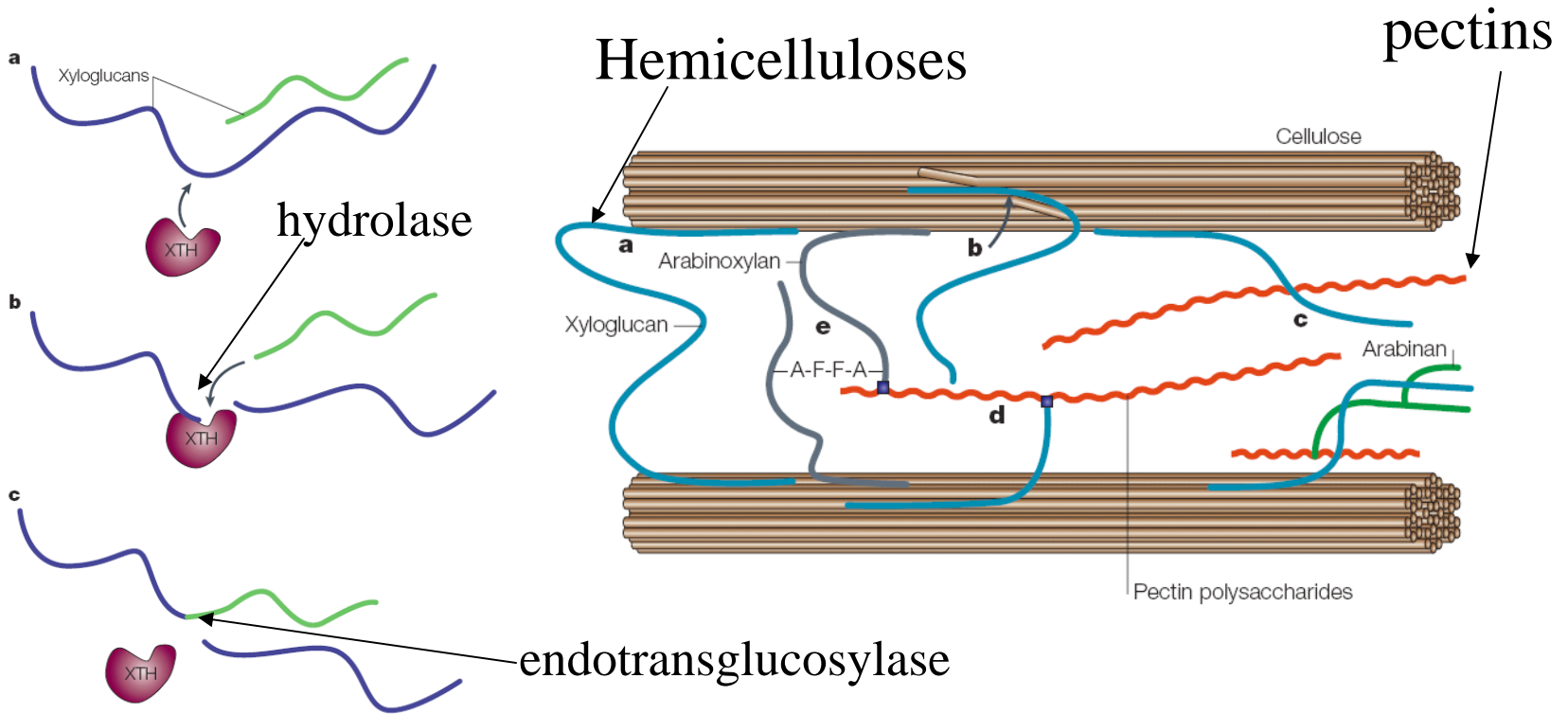
Xyloglucan endotransglucosylase/hydrolases (XTHs) gene family: cell wall formation: cellulosic bioenergy



Structure of the primary cell wall. Cellulose microfibrils are synthesized by large hexameric complexes in the plasma membrane. The hemicelluloses and pectins are synthesized in the golgi apparatus and are deposited to the wall surface by vesicles.

Functions of XTHs:

XTHs perform two different enzymatic reactions in two steps, cleavage of a cross-linking xyloglucan polymer to allow cellulose microfibrils to separate and the cell to expand (XEH), and the transfer of the newly generated end to another sugar polymer to restore stable cell wall structure (XET)



XTH gene family in plants

- **39 genes in poplar**
- **33 in Arabidopsis**
- **29 in rice**

The gene family can be divided into two or three subfamilies

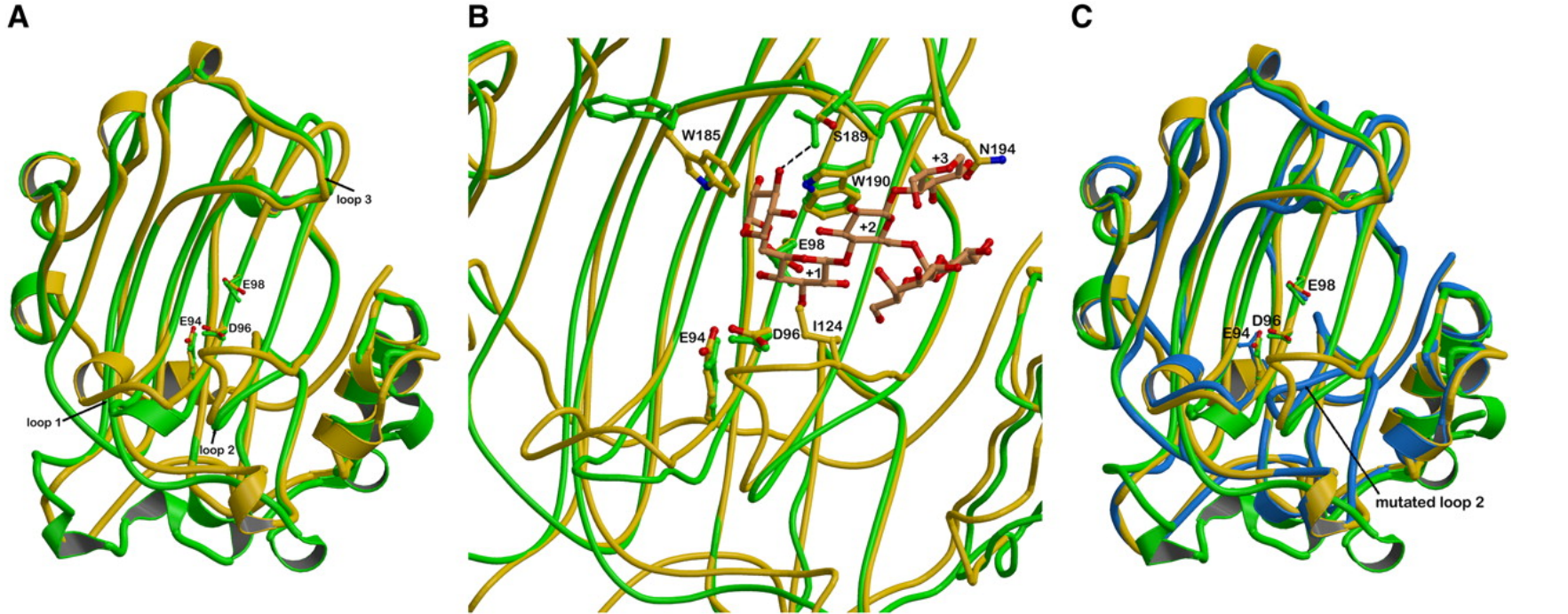
In rice: I/II and III

In Arabidopsis: I, II, and III

Previous major findings:

- **One XTH protein seems to be either predominantly XET or XEH, but functions may depend on substrate concentrations.**
- **SF I and II seem to be associated with XET and XEH appears to be associated with XEH.**

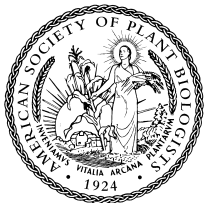
Structural Comparison of Ptt-XET16-34 (a subfamily I member) and Tm-NXG1 (a subfamily III member)



(A) Ribbon representation of the superimposition of Ptt-XET16-34 (green) onto Tm-NXG1 (yellow) illustrating the **structural differences that appear mainly in loops 1 through 3 and in the C-terminal region**. The catalytic machinery is labeled.

(B) Close-up view into the positive subsites of Ptt-XET16-34. The superimposition was performed using the complex structure of Ptt-XET16-34 with bound XLLG (PDB identifier 1UMZ). Asp-178 in Ptt-XET16-34 forms a hydrogen bond to a xylosyl branch of the bound oligosaccharide. Ser-189 in Tm-NXG1 (or Tm-NXG2) is too distant to be able to form an equivalent bond. Gly-183 is replaced by Asn-194 in Tm-NXG1, which would collide with bound sugar in this loop conformation. Ile-124, which is part of the critical loop insertion in Tm-NXG1, would collide with the glucose unit bound in subsite +1 of the acceptor binding cleft.

(C) Ribbon representation of the superimposition of Tm-NXG1- YNIIG (blue) onto Ptt-XET16-34 (green) and Tm-NXG1 (yellow) illustrating that the C trace of the truncated loop 2 in Tm-NXG1- YNIIG is now closer to that of Ptt-XET16-34. The catalytic machinery is labeled.



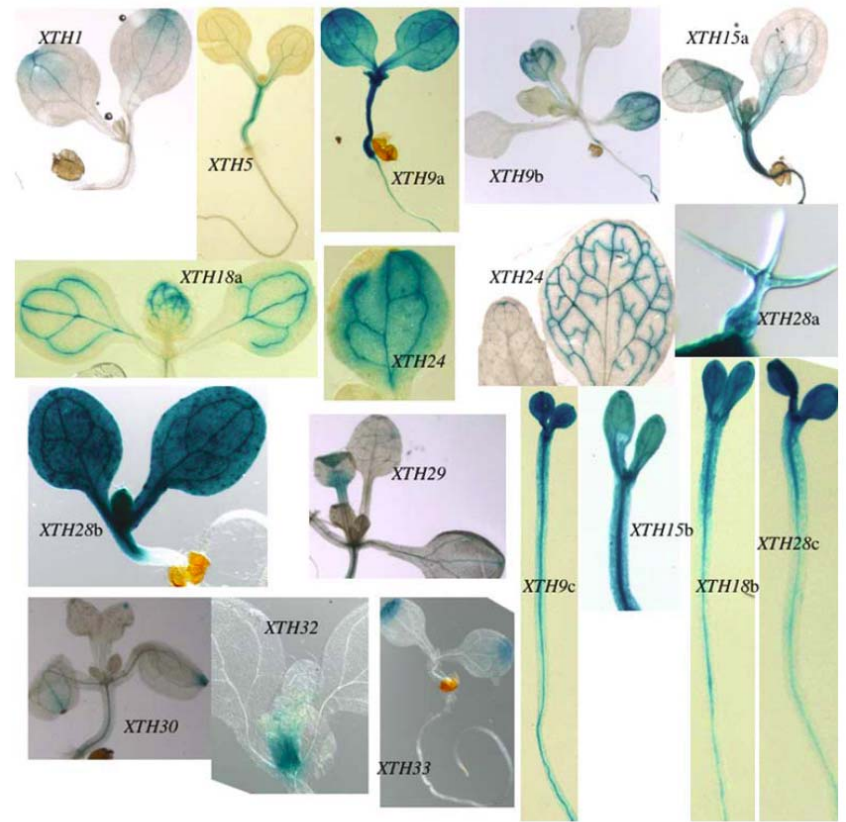
Baumann, M. J., et al. *Plant Cell* 2007;19:1947-1963

Previous major findings:

The Arabidopsis XTH genes are expressed in tissue-specific manner and respond to various hormones and stresses and other environmental factors.



re 4. Localization of *XTH::GUS* expression in Arabidopsis floral organs. Photo labeling is described in Fig



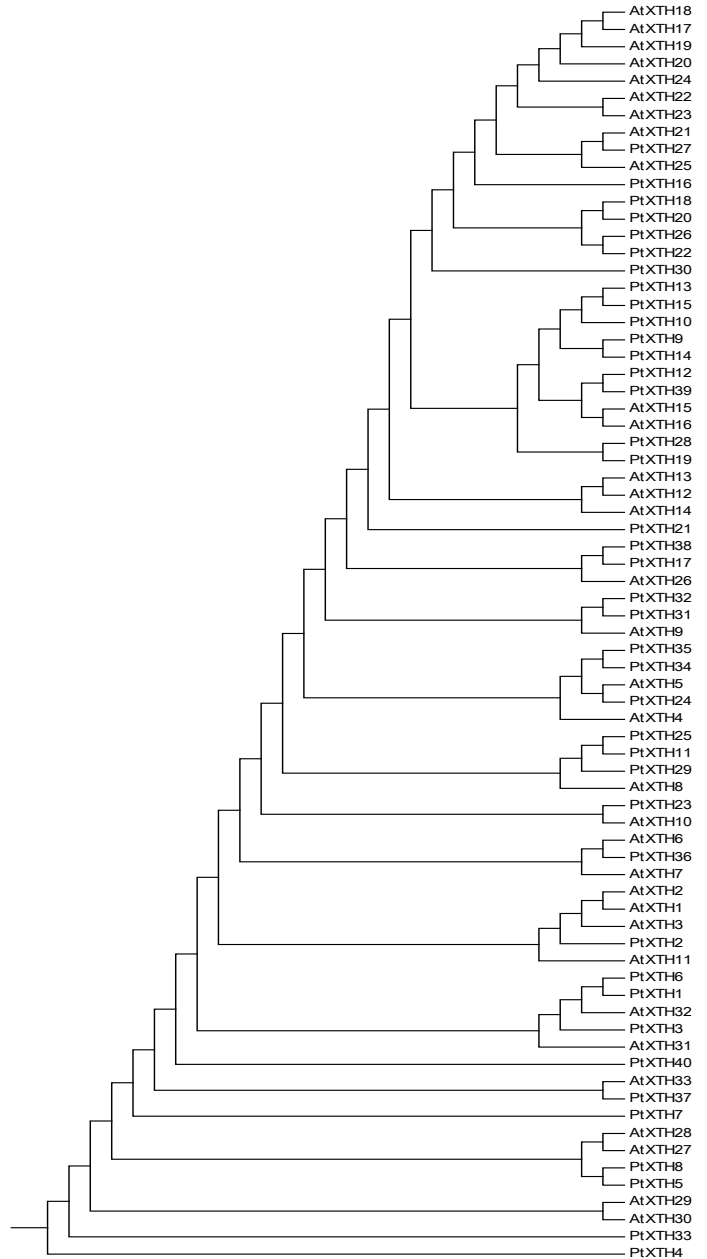
Localization *XTH::GUS* expression in light- and/or dark-grown Arabidopsis seedlings. Photo labeling i

To characterize the XTH genes in poplar

Specific Questions:

- Does one enzyme have two very different functions or only one function (XET or XEH)?
- Are there any correlations between sub-families and physiological or biochemical functions?
- Is there any functional diversification of enzymes in each subfamily?
- What is the function of each XTH gene?

Phylogenetic tree of poplar and Arabidopsis XTHs



Subfamily 2
Pt: 20; At: 15

Subfamily 1
Pt: 9; At: 7

Subfamily 4
Pt: 1, At: 4

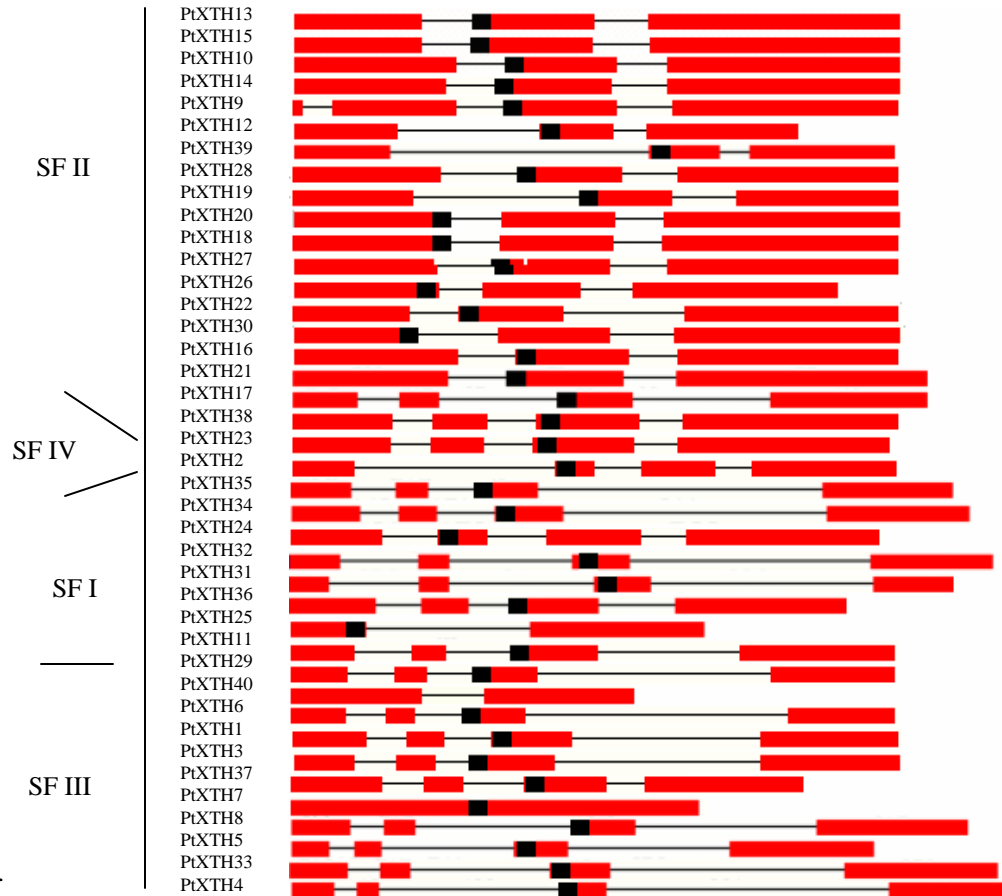
Subfamily 3
Pt: 10, At: 7

The structures of the poplar XTH genes

**The active site
is mostly in
the first and
second exon**



**The active site is
mostly in the third
exon**

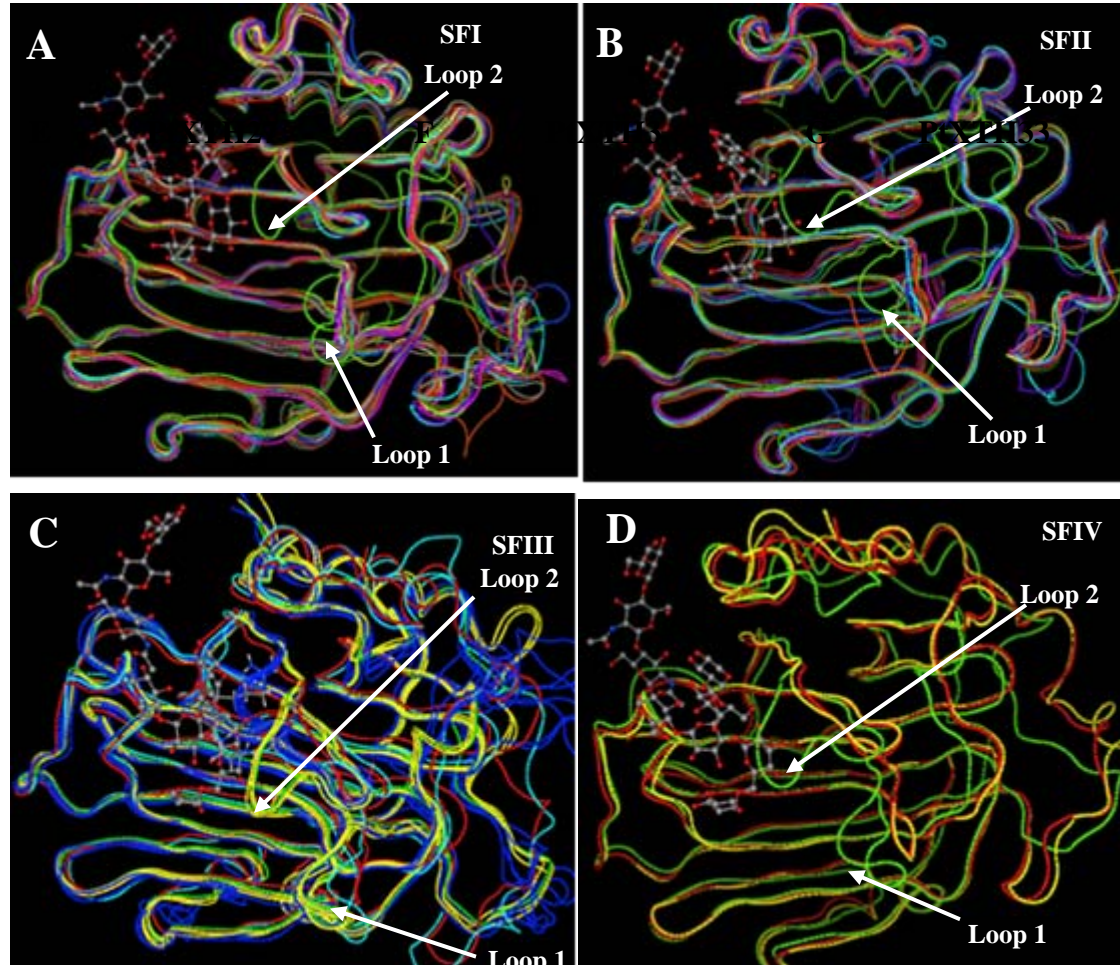


Partial sequence alignments for the regions between and near loop 1 and loop 2. The catalytic triad EXDXE (colored in blue) is totally conserved. The three templates used in homology modelings are the X-ray structure of Ptt-XET16-34 or 1UWZ (Chain 1), Tm-NXG1 or 2UWA (Chain 2), and lichenase or 1GBG (chain 43). The black bars indicate the regions near loop 1 and 2.

	Chain	95	100	105	110	115	120	125	130	135	140	145	150
SF II	1UMZ	1	YLSSQ	· · ·	NSEH	EDDF	EFLGN	RRTG	QPYIL	QTNV	FTGG	KGD	· · · · ·
	2UWA	2	YLSNNQ	DYPG	KHDEID	IE	EFLGT	IPGK	PYTL	QTNV	FI	EGSGD	YNIIG
	PtXTH10	3	YLSSQ	· · ·	GPTH	EDDF	EFLGN	LTTG	EPYTL	LMTN	VFSQ	GKGN	· · · · ·
	PtXTH12	4	YLSSQ	· · ·	GPTH	EDDF	EFLGN	LTTG	EPYTL	LMTN	VFTQ	GKGN	· · · · ·
	PtXTH13	5	YLSSQ	· · ·	GPTH	EDDF	EFLGN	VTTG	EPYTL	LMTN	VFSQ	GKGN	· · · · ·
	PtXTH14	6	YLSSQ	· · ·	GPTH	EDDF	EFLGN	ATG	EPYV	LMTN	VFSQ	GKGD	· · · · ·
	PtXTH15	7	YLSSQ	· · ·	GPTH	EDDF	EFLGN	VTTG	EPYTL	LMTN	VFSQ	GKGN	· · · · ·
	PtXTH16	8	YLKSS	· · ·	GNSW	EDDF	EFLGN	LTTG	DPYTL	LMTN	VYSQ	GKGD	· · · · ·
	PtXTH17	9	YVSSS	· · ·	GDRH	EDDF	EFLGN	TSG	QPYTI	MTNI	YTTQ	GNGS	· · · · ·
	PtXTH18	10	YLSK	· · ·	GSAW	EDDF	EFLGN	LTTG	DPYIL	LMTN	VFSQ	GKGN	· · · · ·
PtXTH19	11	YLT	SQ	· · ·	GNKH	EDDF	EFLGN	QSGN	PYTL	LMTN	VYTT	QGGN	· · · · ·
PtXTH20	12	YLSK	· · ·	GSAW	EDDF	EFLGN	LTTG	DPYIL	LMTN	VFSQ	GKGN	· · · · ·	
PtXTH21	13	YLSL	· · ·	GPYH	EDDF	EFLGN	TSG	QPYTL	LMTN	VFSQ	GKGN	· · · · ·	
PtXTH22	14	YLQSK	· · ·	GSAW	EDDF	EFLGN	LTTG	DPYIL	LMTN	VYSQ	GKGD	· · · · ·	
PtXTH23	15	YSSD	· · ·	QPKR	EDDF	EFLGN	VSG	QPYIL	QTNV	YADG	ND	· · · · ·	
PtXTH26	16	YLSK	· · ·	GSTW	EDDF	EFLGN	LTTG	DPYIL	LMTN	VFSQ	GKGN	· · · · ·	
PtXTH27	17	YLSQ	· · ·	GSAW	EDDF	EFLGN	LTTG	DPYLV	HTNV	YTQCKG	D	· · · · ·	
PtXTH28	18	YLSSE	· · ·	GPYH	EDDF	EFLGN	LTTG	EPYTV	MNTN	VYTT	QGGD	· · · · ·	
PtXTH30	19	YLSL	· · ·	GSAH	EDDF	EFLGN	LTTG	DPYIL	LMTN	VFTQ	GKGN	· · · · ·	
PtXTH38	20	YVSSS	· · ·	GDRH	EDDF	EFLGN	ASG	QPYTI	MTNI	YTTQ	GNGS	· · · · ·	
PtXTH39	21	YLSQ	· · ·	GATH	EDDF	EFLGN	LTTG	DPYIL	LMTN	VFTQ	GKGN	· · · · ·	
PtXTH9	22	YLSQ	· · ·	GSTH	EDDF	EFLGN	ATG	EPYIL	LMTN	VFSQ	GKGN	· · · · ·	
SF I	PtXTH11	23	·	NGA	· · ·	GPTR	DELDF	EFLGN	RRTG	EPYLI	QTNV	YTKN	NGE
	PtXTH24	24	YLSQ	· · ·	TNEH	EDDF	EFLGN	RRTG	QPYIL	QTNV	YTTG	GKGD	· · · · ·
	PtXTH25	25	·	AGPT	RELD	DF	EFLGN	RRTG	EPYLI	QTNV	YKNG	TGN	· · · · ·
	PtXTH29	26	YMCT	ENG	AGPT	RELD	DF	EFLGN	RRTG	EPYLI	QTNV	YKNG	TGN
	PtXTH31	27	YMSSE	· · ·	GPYH	NEDF	EFLGN	TTTG	EPYLV	QTNV	VYV	NGVGN	· · · · ·
	PtXTH32	28	YMSSE	· · ·	GTNH	NEDF	EFLGN	TTTG	EPYLV	QTNV	VYV	NGVGN	· · · · ·
	PtXTH34	29	YLSQ	· · ·	NSEH	EDDF	EFLGN	RRTG	QPYIL	QTNV	FTGG	KGD	· · · · ·
	PtXTH35	30	YLSQ	· · ·	NSEH	EDDF	EFLGN	RRTG	QPYIL	QTNV	FTGG	KGD	· · · · ·
	PtXTH36	31	YMSDT	· · ·	DAVR	DELDF	EFLGN	RRTG	QPYTV	QTNV	IFAH	GKGD	· · · · ·
	SF IV	PtXTH2	32	YLSL	· · ·	GPYH	EDDF	EFLGN	TSG	QPYTL	LMTN	VFSQ	GKGN
PtXTH1		33	YLSN	NEA	HPGF	HDEVD	IE	EFLGT	TPDK	KPYTL	QTNV	YIRG	SGDGR
PtXTH3		34	YLSN	NEA	HPGD	HDEVD	IE	EFLGT	TPDK	KPYTL	QTNV	YIRG	SGDR
PtXTH6		35	YLSN	SEA	HPGY	HDEVD	IE	EFLGT	TPDK	KPYTL	QTNV	YIRG	SGDGR
PtXTH4		36	YTSNG	DV	FKKTH	DELDF	EFLGN	TKGP	WR	FQTN	LYGN	GSTSH	· · ·
PtXTH5		37	YMSNG	DIFE	KNH	EDDF	EFLGN	IRG	KDWRI	QTNV	YGN	GSTSV	· · ·
PtXTH7		38	YTSNA	DEHT	TNH	EDDF	EFLGN	TGK	PWT	LQTN	LYGN	GSTGR	· · ·
PtXTH33		39	YTSNG	DV	FEKTH	DELDF	EFLGN	TGK	WR	QTNLY	GN	GSTSR	· · ·
PtXTH37		40	YLSN	ADN	YPHN	HDEVD	IE	ELG	HDMR	NDW	VLQTN	VYANG	STGT
PtXTH8		41	YMSNG	DIFE	KNH	EDDF	EFLGN	IRG	KDWRI	QTNV	YGN	GSTSA	· · ·
SF III	PtXTH40	42	YSSLEG	·	DKSQ	EDDF	EFLGK	· · ·	DKTIV	QTNV	YAS	SGTGN	· · · · ·
	1GBG	43	FTYTG	P	DGTP	WDIED	IE	EFLG	KDTTK	· · ·	VQFN	YNTNG	VGN

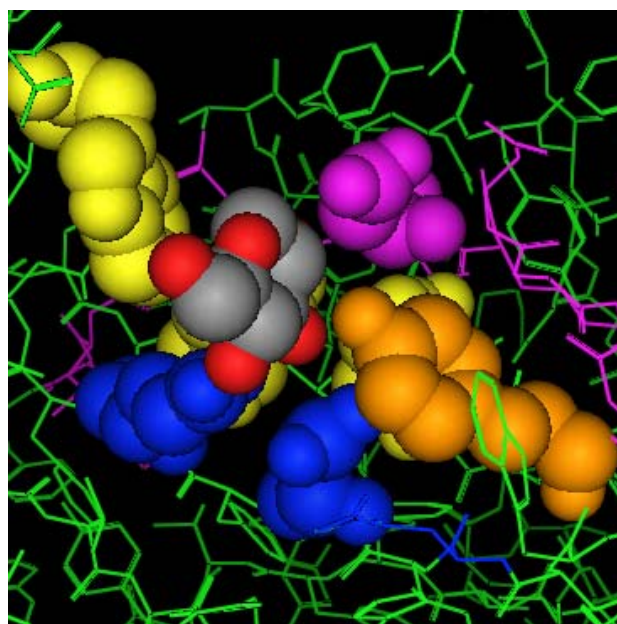
Coding region divergence:

Superposition of the homology models for each SF with (1UWZ, red) and (2UWA, green): SF I (A), SF II (B), SF III (C) and SF IV (D).

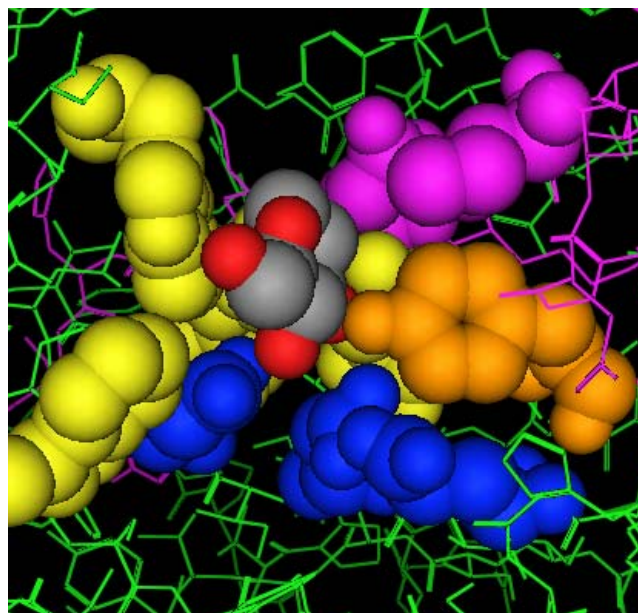


The active-site structure of PtXTH27, 3 and 33

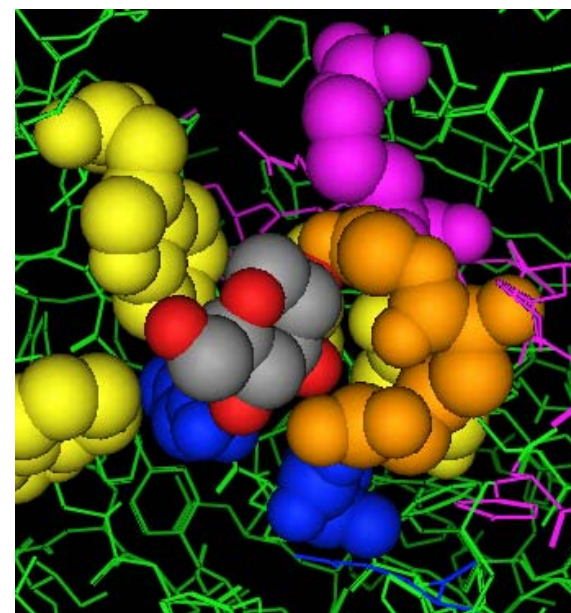
E: PtXTH27
(a rep of SFI&II)



F: PtXTH3
(a rep of the SF III-A)

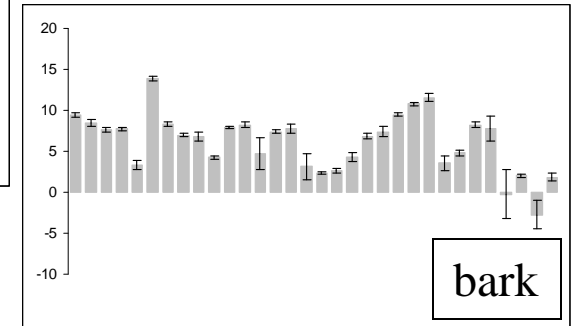
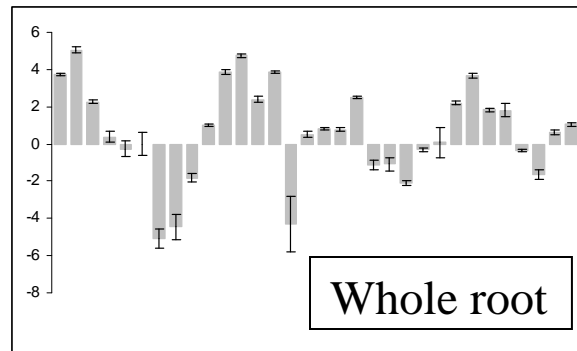
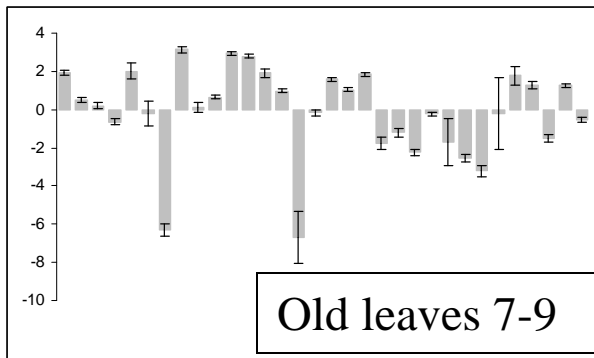
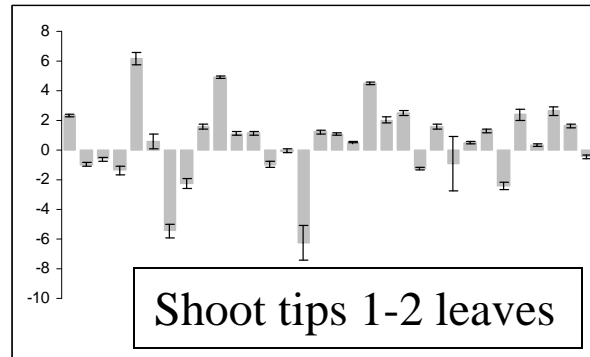
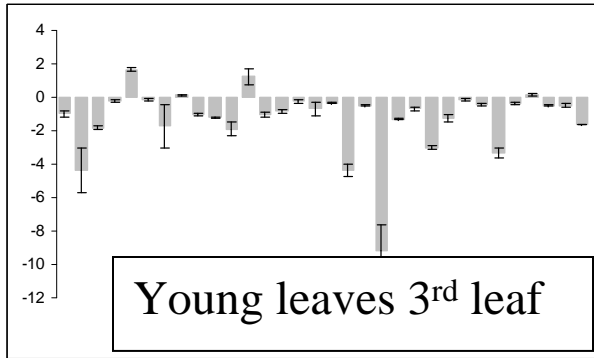


G PtXTH33
(a rep of the SF III-B)



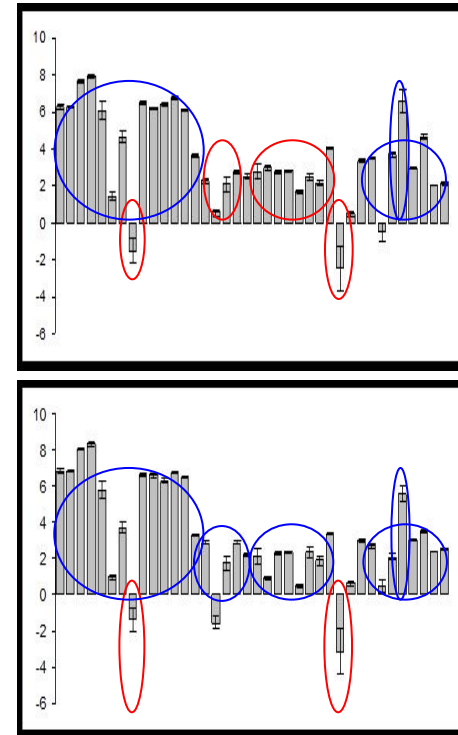
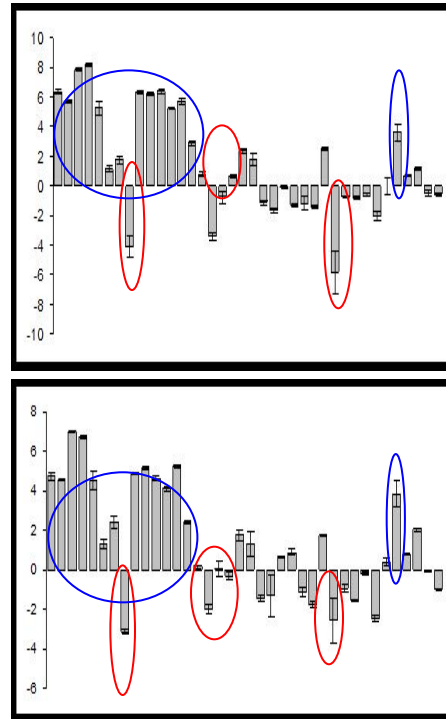
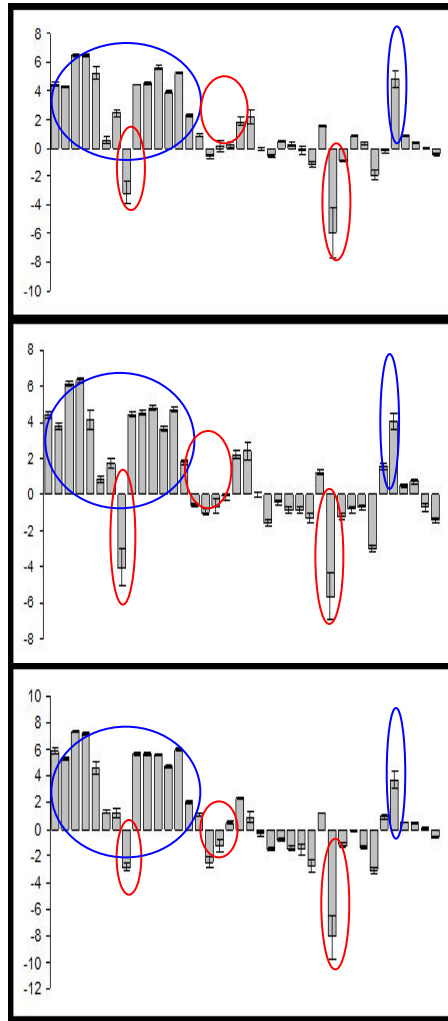
Promoter region divergence:

XTH gene family expression in different organs



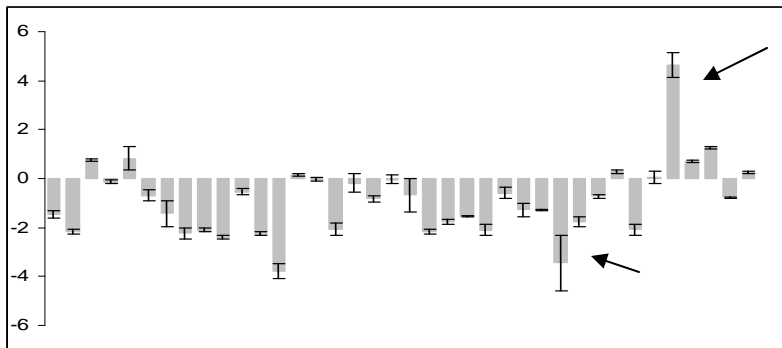
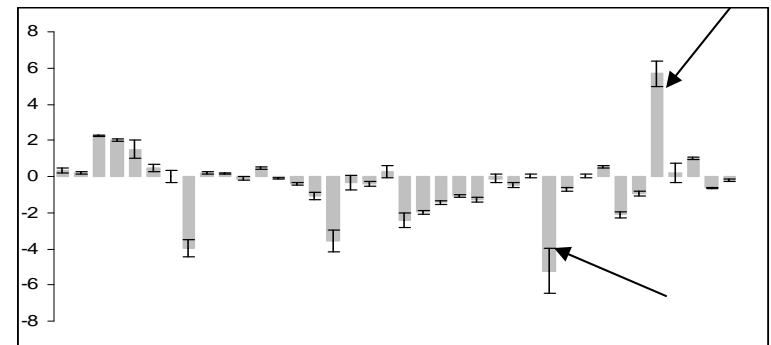
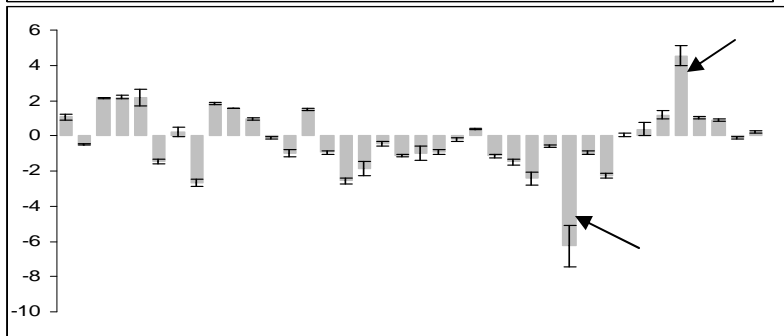
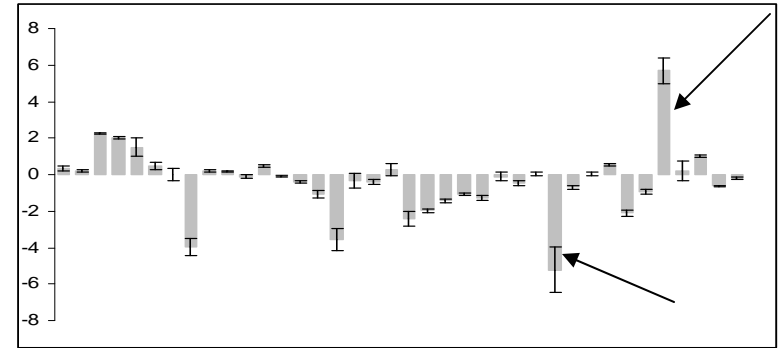
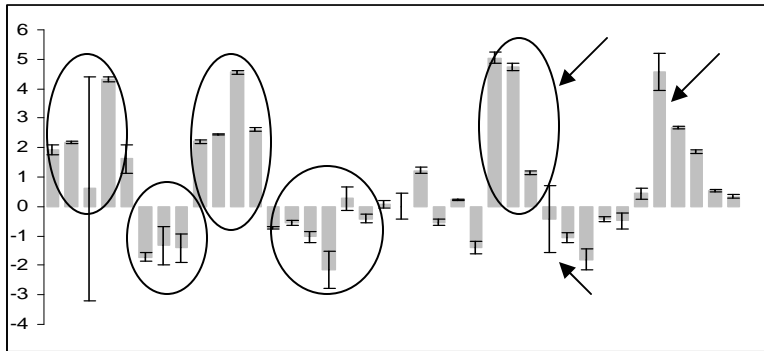
Promoter region divergence:

XTH gene family expression: responses to hormones: BA, IAA, SA, GA, BR, JA, ABA



Promoter region divergence:

XTH gene family expression: responses to stresses:
high temp, low temp, drought, flood, salt



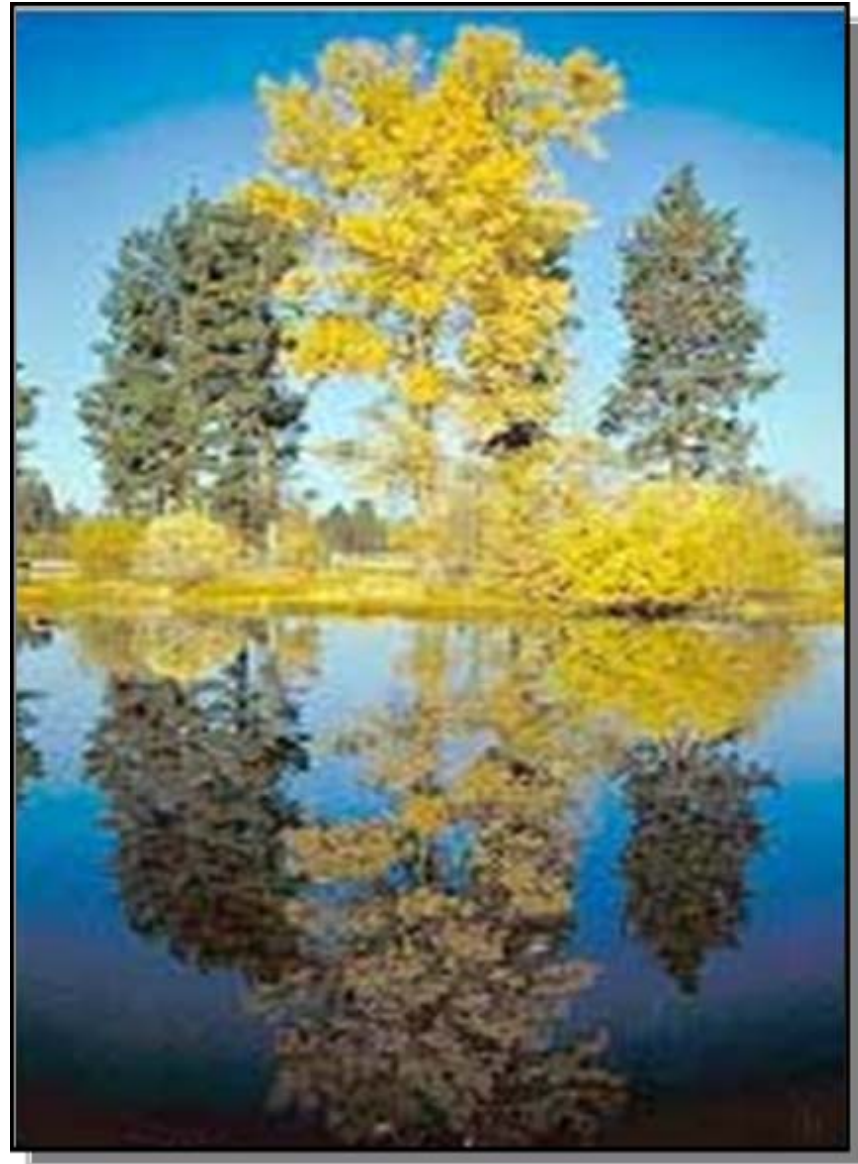
Conclusions—take home messages

- The poplar XTH gene family has undergone considerable expansion and divergence after duplication
- The proteins (coding regions) have been diverged into either XET or XEH → one-gene-one-function model
- There is a clear association between subfamilies and biochemical functions: SF I, II, IV: XET, SF III: XEH
- Promoters also diverged by responding to physiological stimuli to guide the functions
- Evolution is correlated with development
- Several candidate genes/promoters have been cloned and are under characterization

Acknowledgements

Funding:

- DOE: Bioenergy Science Center
- Tennessee Agricultural Experimental Station



Thank
you!