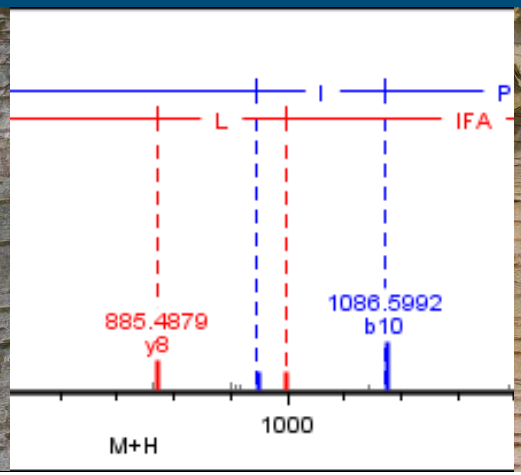
In search for hypoallergenic trees: Screening for genetic diversity in birch pollen allergens, a multigene family of Bet v 1 (PR-10) proteins

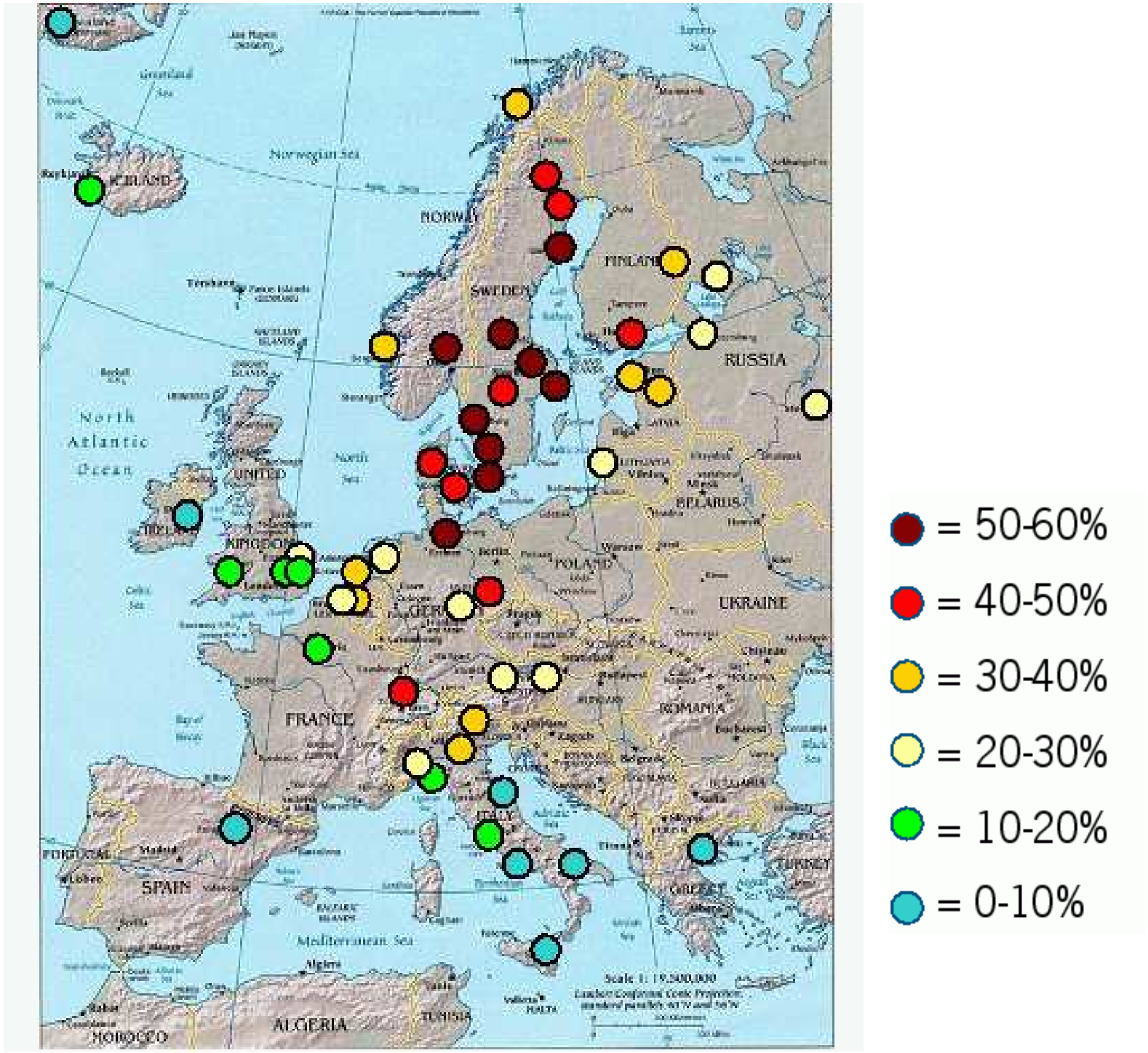
MJM Smulders, MF Schenk, LJWJ Gilissen



PLANT RESEARCH INTERNATIONAL
WAGENINGEN UR

Hay fever

- Hay fever affects 10-19% of the population in Western Europe
- Birch pollen is a major allergen source, especially in Central and North-European countries
- Allergy to birch is associated with a type of food allergy called Oral Allergy Syndrome (OAS)


Norwegian Sea
ICELAND
NORWAY
SWEDEN
FINLAND
RUSSIA
North Atlantic Ocean
North Sea
UNITED KINGDOM
IRELAND
BELARUS
POLAND
UKRAINE
GERMANY
FRANCE
ROMANIA
ITALY
PORTUGAL
SPAIN
Madrid
Mediterranean Sea
GREECE
TURKEY
ALGERIA
TUNISIA
MOROCCO
Scale 1: 19,500,000
= 50-60%
= 40-50%
= 30-40%
= 20-30%
= 10-20%
= 0-10%

Birch allergens

- Eight birch allergens are known: Bet v 1-8
- These allergens differ according to the number of birch sensitized patients that are allergic to them
- Bet v 1 is the only major birch allergen (>50% of the birch allergic patients react to this protein)

Bet v 1

- Bet v 1 is a pathogenesis-related (PR)-10 protein
- PR-10 proteins are present as a multigene family in many species
- Several Bet v 1 isoforms are found in birch pollen mixtures
- Bet v 1 isoforms differ in their IgE-binding activity and T-cell proliferation score

> Are multiple genes involved? Do several isoforms also exist within pollen of one tree?

This study

- Genetics:
  - How many Bet v 1 genes are there?
  - What differences exist among birch species?
  - Which are (similar to) hyperallergic isoforms?
- Proteomics:
  - Which isoforms are being expressed in pollen?
  - Can we predict (hypo)allergenicity?

Schenk et al. BMC Genomics 7:168 (2006)
Schenk et al. BMC Plant Biology (submitted)

Plant material

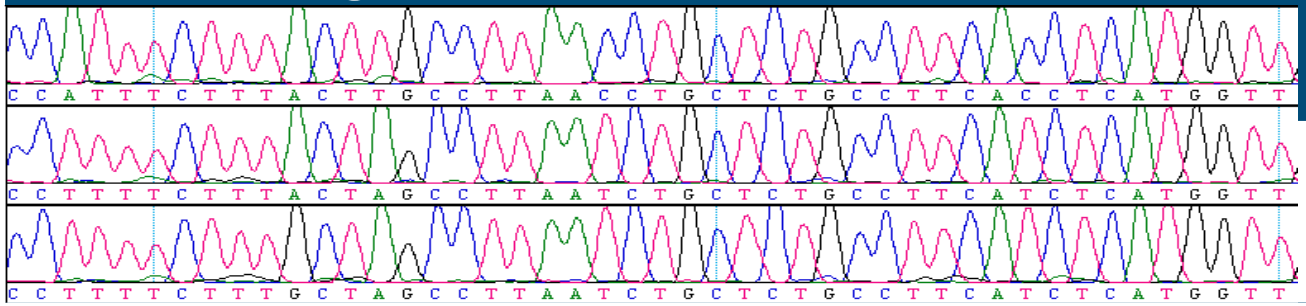
- *B. pendula*
  - *B. pendula* cultivar “Tristis”
  - *B. pendula* cultivar “Youngii”
  - *B. pendula* cultivar “Schneverdinger Goldbirke”
- Close relatives of *B. pendula*:
  - *B. plathyphylla*
  - *B. populifolia*
- Other *Betula* species:
  - *B. lenta*
  - *B. costata*
  - *B. schmidtii*
  - *B. chichibuensis*
  - *B. nigra*

PLANT RESEARCH INTERNATIONAL
WAGENINGEN UR

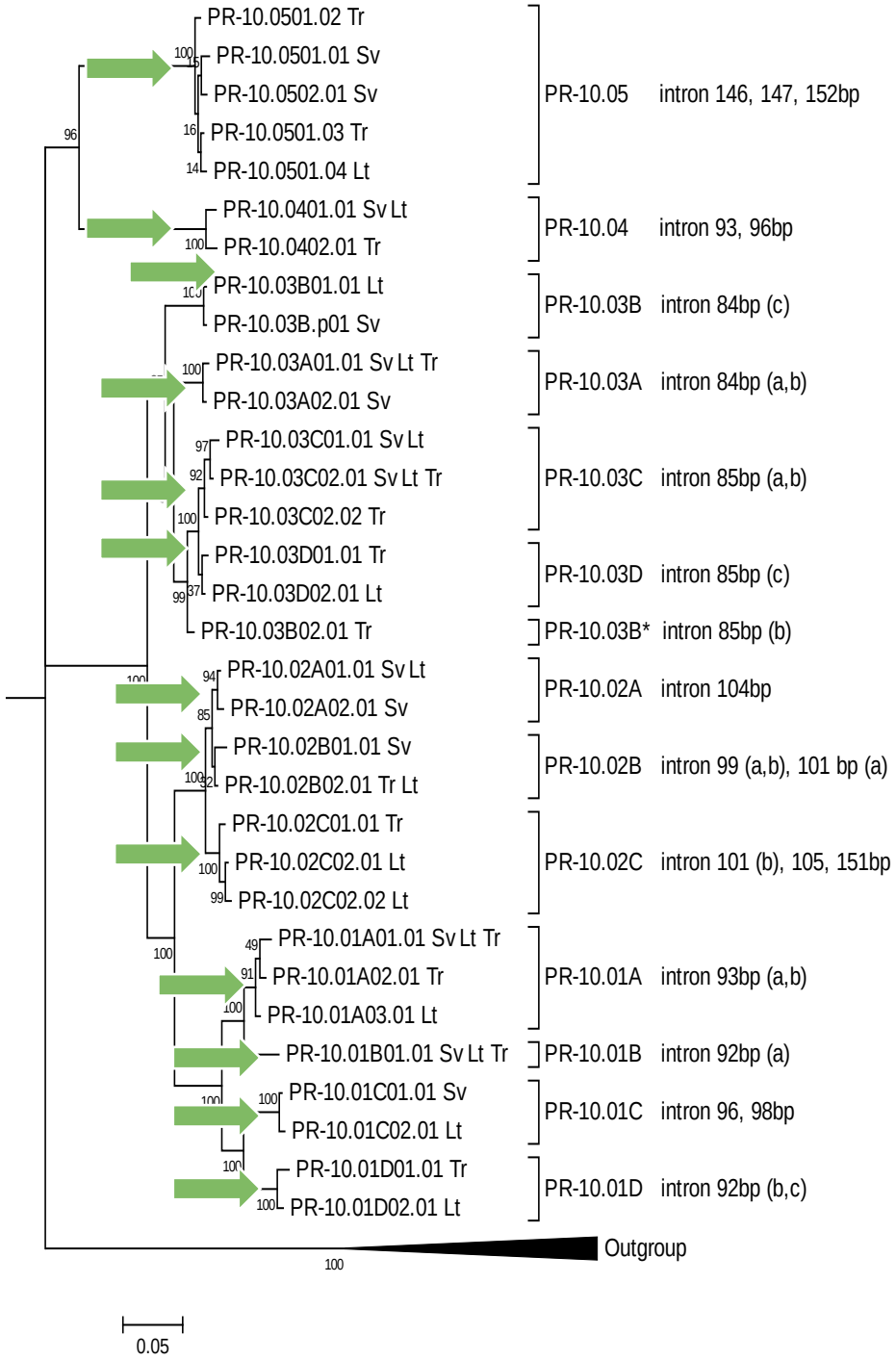
netherlands proteomics centre

Alleles cloned and sequenced

- **134 unique alleles**
  - *B. pendula* "Sch. Goldbirke" 14 alleles
  - *B. pendula* "Tristis" 15 alleles
  - *B. pendula* "Youngii" 16 alleles
  - *B. populifolia* 17 alleles
  - *B. plathyphylla* 20 alleles
  - *B. lenta* 14 alleles
  - *B. costata* 20 alleles
  - *B. schmidtii* 12 alleles
  - *B. chichibuensis* 22 alleles
  - *B. nigra* 25 alleles



netherlands proteomics centre

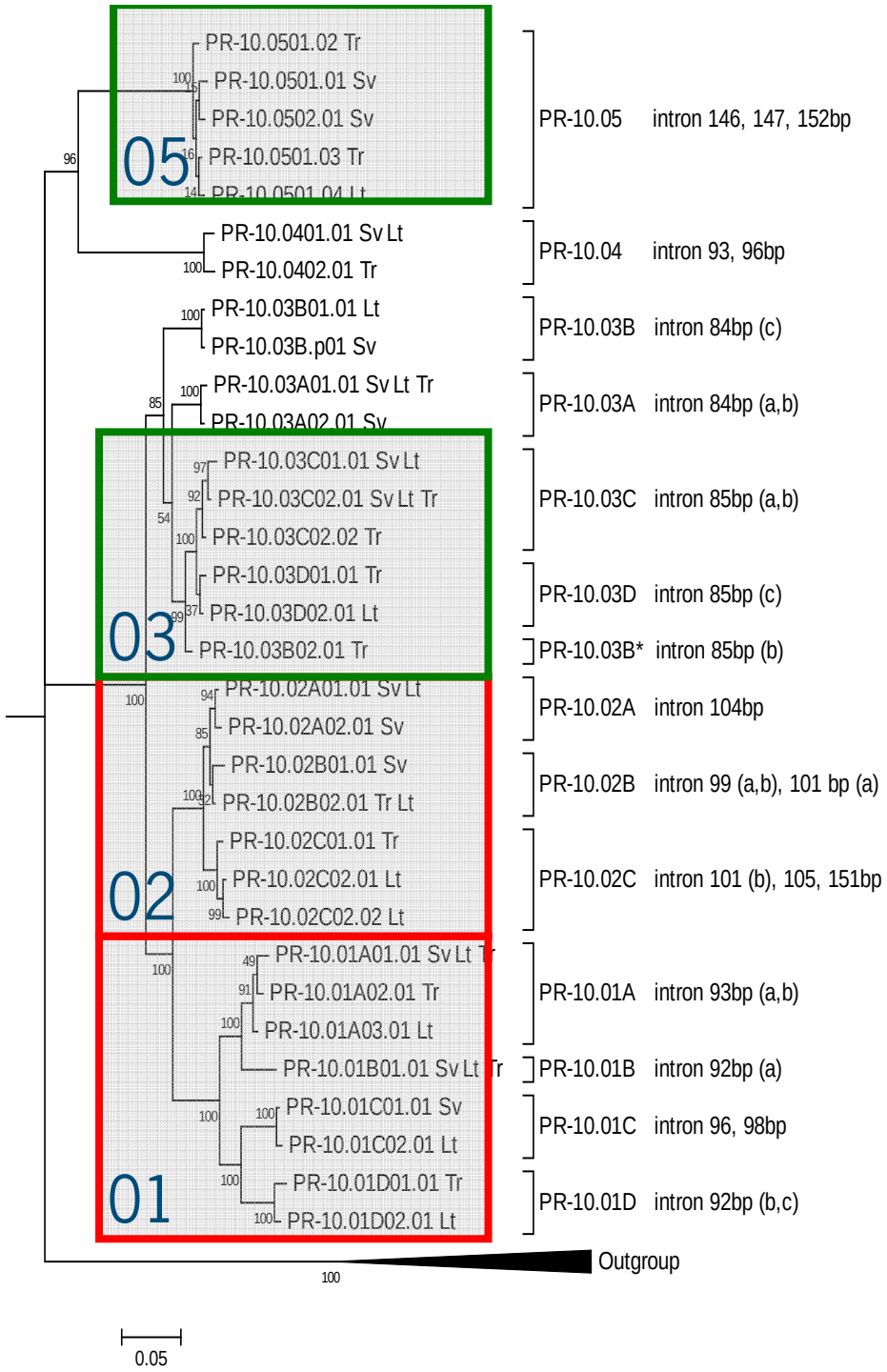


Results: *B. pendula*

- 13 distinct genes, based on:
  - Clustering of the alleles
  - Cut-off value of 98% similarity in exons

  > Maximum of 2 alleles per locus in a diploid tree)

netherlands proteomics centre



Results: *B. pendula*

- Division into 5 subfamilies
  - Within a subfamily genes have >95 % sequence similarity
  - Between subfamilies sequence similarity is 80-90%
- Transcription
  - Information of 30 mRNA-derived (tissue-specific) GenBank sequences
    - Pollen
    - Root/Leaves

netherlands proteomics centre

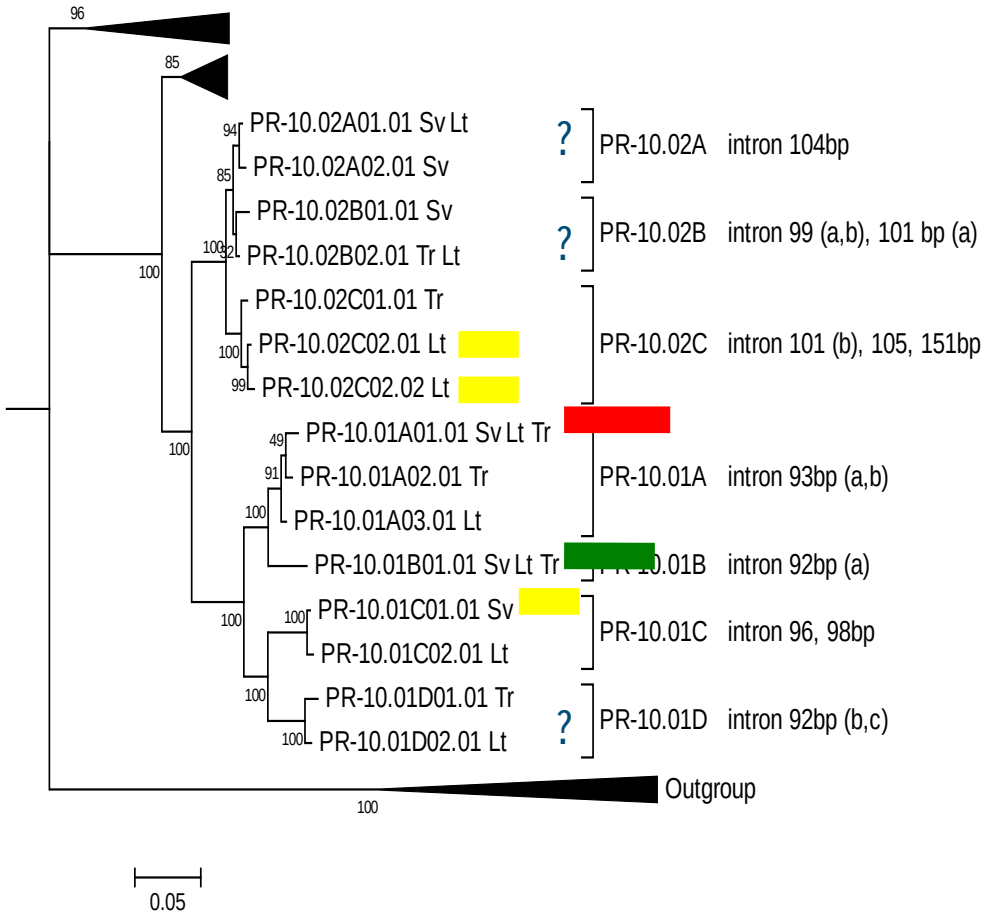
Results for other *Betula* species

- Genetic composition of PR-10 genes in *B. populifolia* and *B. plathyphylla* is similar to *B. pendula*
- Not possible to distinguish homologues from paralogues in the other *Betula* sp.
- Gene duplication, extinction, unequal crossing-over, and possibly gene conversion play a role

- Subfamily division is consistent across species
- Two new subfamilies were identified, but composed of few (pseudo) genes

IgE-reactivity at the genomic level

- The isoforms with varying IgE-reactivity are encoded by different genes
- In each of the cultivars and/or species protein variants are found that differ by 1 or 2 amino acids
  - However, we do not know whether these isoforms are actually expressed and in which quantities?



PLANT RESEARCH INTERNATIONAL
WAGENINGEN UR

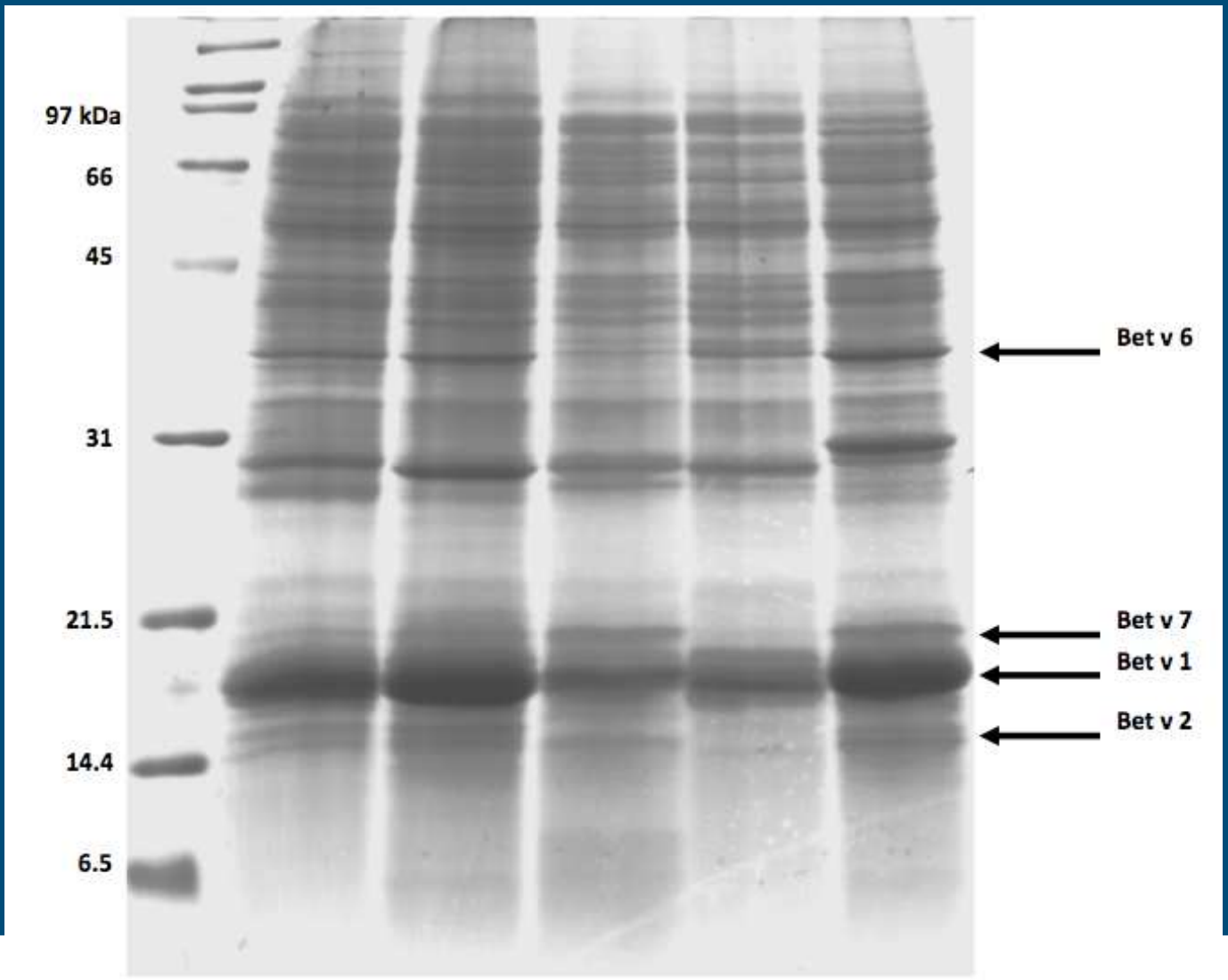
netherlands proteomics centre

Methods: protein extraction and sequencing

- Pollen collection (2004 -2005)
  - *B. nigra*
  - *B. lenta*
  - *B. chichibuensis*
  - *B. costata*
  - *B. nigra*
- Protein extraction in aqueous buffer
- Analyzed with SDS-PAGE

- 16-18 kDa band cut out
- In-gel protein digestion with trypsin
- Analyzed with Q-TOF LC-MS/MS and LC-MS^{E}

SDS-PAGE



1. *B. chichibuensis*
2. *B. costata*
3. *B. nigra*
4. *B. lenta*
5. *B. pendula*

PLANT RESEARCH INTERNATIONAL
WAGENINGEN UR

netherlands proteomics centre

Comparative LC MS approach

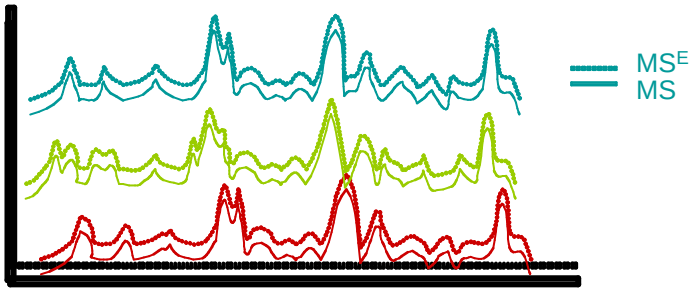
Protein extracts

1
2
3

Trypsin digest

Digest 1
Digest 2
Digest 3

replicate nanoLC-MS



Peak detection, Alignment, Quantitative Comparison, Peak selection & identification

Include list for LC-MS/MS

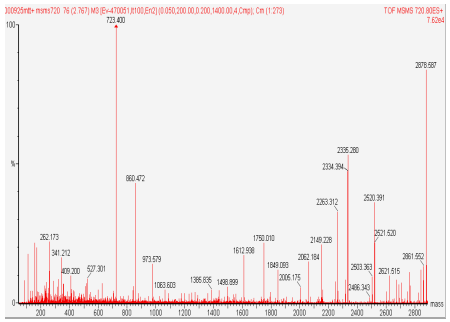
Quantification And pre-ID table

m/z	Ret.
546.543	12.3
645.597	13.4
873.345	25.9
763.563	31.8
845.254	32.2
977.223	33.1
É	É
É	É

Mw	Ret.	Peptide
851.522	18.4	SLLDFAR
1055.507	30.0	EDTSISLYK
1324.691	23.2	FGIAATGÉ
1308.677	31.8	GLTATMGÉ
1284.647	32.2	NQHGCVTÉ
1168.566	33.1	QDFANFÉ
É	É	É
É	É	É

DDA nanoLC-MS/MS

Selective Peptide MS/MS



Protein Identification and Quantification

Database Search of MS/MS data

Putative peptide fragments

Example:

Bet v 1 01A01 = PR-10 01A01

MGVFNYETETTSVIPAARLFKAFILDGDNLFPKV
APQAISSVENIEGNGGPGTIKKISFPEGFPFKYV
KDRVDEVDHTNFKYNYSVIEGGPIGDTLEKISNE
IKIVATPDGGSILKISNKYHTKGDHEVKAEQVKA
SKEMGETLLRAVESYLLAHSDAYN

Bet v 1 01A01 = PR-10 01A01

MGVFNYETETTSVIPAAR AFILDGDNLFPKV
APQAISSVENIEGNGGPGTIK(K)ISFPEGFPFK
(DR)VDEVDHTNFKYNYSVIEGGPIGDTLEK
IVATPDGGSILK
EMGETLLRAVESYLLAHSDAYN

MGVFNYETETTSVIPAAR
AFILDGDNLFPK
VAPQAISSVENIEGNGGPGTIK(K)
(K)ISFPEGFPFK
(DR)VDEVDHTNFK
YNYSVIEGGPIGDTLEK
IVATPDGGSILK
EMGETLLR
AVESYLLAHSDAYN

PLANT RESEARCH INTERNATIONAL
WAGENINGEN UR

netherlands proteomics centre

Putative peptide fragments

```
01A01 - MGVFNYETETTSVIPAAR AFILDGDNLFFK VAPQAISSVENIEGNGGPGTIK(K) (K)ISFPEGFPFK
01B01 - MGVFNYEIETTSVIPAAR AFILDGDNLFVK VAPQAISSVENIEGNGGPGTIK(K) (K)INFPEGFPFK
01C01 - MGVFNYEIEATSVIPAAR AFILDGDNLFFK VAPQAISSVENIEGNGGPGTIK(K) (K)ISFPEGFPFK
01D01 - MGVFNYEIEATSVIPAAR AFILDGDNLFFK VAPQAISSVENIEGNGGPGTIK(K) (K)ISFPEGIPFK
02A01 - MGVFNYESETTSVIPAAR AFILDGDNLFPK VAPQAISSVENIEGNGGPGTIK(K) (K)ITFPEGSPFK
02B01 - MGVFNYESETTSVIPAAR AFILDGDNLFPK VAPQAISSVENIEGNGGPGTIK(K) (K)ITFPEGSPFK
02C01 - MGVFNYESETTSVIPAAR AFILEGDNLFPK VAPQAISSVENIEGNGGPGTIK(K) (K)ITFPEGSPFK

03A01 - MGVFDYEGETTSVIPAAR AFILDGDNLFPK VAPQAVSCVENIEGNGGPGTIK(K) (K)ITFPEGSPFK
03B01 - MGVFNYEGETTSVIPAAR AFILDGDNLFPK VAPQAVSCVENIEGNGGPGTIK(K) (K)ITFPEGSPFK
03C01 - MGVFDYEGETTSVIPAAR AFILDGDNLFPK VAPQAVSCVENIEGNGGPGTIK(K) (K)ITFPEGSPFK
03D01 - MGVFDYEGETTSVIPAAR AFILDGDNLFPK VAPQAVSCVENIEGNGGPGTIK(K) (K)ITFPEGSPFK
04A01 - MGVFNDEAETTSVIPPAR SFILDADNILSK        SAENIEGNGGPGTIK(K) (K)ITFVEGSPFK
05A01 - MGVFNYEDEATSVIAPAR SFVLDADNLFPK VAPENVSSAENIEGNGGPGTIK(K) (K)ITFPEGSPFK
```

```
01A01 - (DR)VDEVDHTNFK YNYSVIEGGPIGDTLEK IVATPDGGSILK EMGETLLR    AVESYLLAHSDAYN
01B01 - (DR)VDEVDHTNFK YNYSVIEGGPVGDTLEK IVATPDGGCVLK EMGETLLR    AVESYLLAHSDAYN
01C01 - (DR)VDEVDHTNFK YNYSVIEGGPVGDTLEK IVATPNGGSILK EMGETLLR    AVESYLLAHSDAYN
01D01 - (DR)VDEVDHANFK YSYSLIEGGPVGDTLEK IVATPDGGSILK EMGETLLR    AVESYLLAHSDAYN
02A01 - (ER)VDEVDHANFK YSYSMIEGGALGDTLEK LVATPDGGSILK             AVESYLLAHSDAYN
02B01 - (ER)VDEVDHANFK YSYSMIEGGALGDTLEK LVATPDGGSILK             AVESYLLAHSDAYN
02C01 - (ER)VDEVDHANFK YSYSMIEGGALGDTLEK LVATPDGGSILK             AVESYLLAHSDAYN

03A01 - (ER)IDEVDHVNFK YSYSVIEGGAVGDTLEK IVAAPGGGSILK             AVESYLLAHSDAYN
03B01 - (ER)VDEVDHVNFK YSYSVIEGGVVGDTLEK IVAAPGGGSILK             AVESYLLAHSDAYN
03C01 -                YSYSVIEGGAVGDTLEK IVPAPGGGSILK             AVESYLLAHSDAYN
03D01 - (ER)VDEVDHENFK YSYSVIEGGAVGDTLEK IVPAPGGGSILK             AVESYLLAHSDAYN
04A01 - (QR)IDEIDHTNFK YSYSLIEGGPLGDTLEK IAAAPDGGSILK GNISINQEQIK AIEGYLLAHSDAYN
05A01 - (HR)VDEIDHTNFK YCYSIIEGGPLGDTLEK IVAAPGGGSILK GNISLNEEEIK AVENYLLAHSDAYN
```

Results: recovered fragments

01A01	-	GVFNYETETTSVIPAAR	AFILDGDNLFFK	VAPQAISSVENIEGNGGPGTIK(K)	(K)ISFPEGFPFK
01A06	-	GVFNYETETTSVIPAAR	AFILDGDNLFFK	VAPQAISSVENIEGNGGPGTIK(K)	(K)INFPEGFPFK
01B01	-	GVFNYEIETTSVIPAAR	AFILDGDNLFVK	VAPQAISSVENIEGNGGPGTIK(K)	(K)INFPEGFPFK
01C04	-	GVFNYEIEATSVIPAAR	AFILDGDNLFFK	VAPQAISSVENIEGNGGPGTIK(K)	(K)ISFPEGFPFK
01C05	-	GVFNYEIEATSVIPAAR	AFILDGDNLFFK	VAPQAISSVENIEGNGGPGTIK(K)	(K)ISFPEGFPFK
01D01	-	GVFNYEIEATSVIPAAR	AFILDGDNLFFK	VAPQAISSVENIEGNGGPGTIK(K)	(K)ISFPEGIPFK
02A01	-	GVFNYESETTSVIPAAR	AFILDGDNLFPK	VAPQAISSVENIEGNGGPGTIK(K)	(K)ITFPEGSPFK
02B01	-	GVFNYESETTSVIPAAR	AFILDGDNLFPK	VAPQAISSVENIEGNGGPGTIK(K)	(K)ITFPEGSPFK
02C01	-	GVFNYESETTSVIPAAR	AFILEGDNLFPK	VAPQAISSVENIEGNGGPGTIK(K)	(K)ITFPEGSPFK
03A01	-	GVFDYEGETTSVIPAAR	AFILDGDNLFPK	VAPQAVSCVENIEGNGGPGTIK(K)	(K)ITFPEGSPFK
04A01	-	GVFNDEAETTSVIPPAR	SFILDADNILSK	SAENIEGNGGPGTIK(K)	(K)ITFVEGSPFK
05A01	-	GVFNYEDEATSVIAPAR	SFVLDADNLFPK	VAPENVSSAENIEGNGGPGTIK(K)	(K)ITFPEGSPFK

01A01	–	(DR)VDEVDHTNFK	YNYSVIEGGPIGDTLEK	IVATPDGGSILK	EMGETLLR	AVESYLLAHSDAYN
01A06	-	(DR)VDEVDHTNFK	YNYSVIEGGPIGDTLEK	IVATPDGGSILK	EMGETLLR	AVESYLLAHSDAYN
01B01	–	(DR)VDEVDHTNFK	YNYSVIEGGPVGDTLEK	IVATPDGGCVLK	EMGETLLR	AVESYLLAHSDAYN
01C04	–	(DR)VDEVDHTNFK	YSYSVIEGGPVGDTLEK	IVATPNGGSILK	EMGETLLR	AVESYLLAHSDAYN
01C05	-	(DR)VDEVDHTNFK	YSYSVIEGGPVGDTLEK	IVATPNGGSILK	EMGETLLR	AVESYLLAHSDAYN
01D01	–	(DR)VDEVDHANFK	YSYSLIEGGPVGDTLEK	IVATPDGGSILK	EMGETLLR	AVESYLLAHSDAYN
02A01	–	(ER)VDEVDHANFK	YSYSMIEGGALGDTLEK	LVATPDGGSILK		AVESYLLAHSDAYN
02B01	–	(ER)VDEVDHANFK	YSYSMIEGGALGDTLEK	LVATPDGGSILK		AVESYLLAHSDAYN
02C01	–	(ER)VDEVDHANFK	YSYSMIEGGALGDTLEK	LVATPDGGSILK		AVESYLLAHSDAYN
03A01	–	(ER)IDEVDHVNFK	YSYSVIEGGAVGDTLEK	IVAAPGGGSILK		AVESYLLAHSDAYN
04A01	–	(QR)IDEIDHTNFK	YSYSLIEGGPLGDTLEK	IAAAPDGGSILK	GNISINQEQIK	AIEGYLLAHSDAYN
05A01	–	(HR)VDEIDHTNFK	YCYSIIEGGPLGDTLEK	IVAAPGGGSILK	GNISLNEEEIK	AVENYLLAHSDAYN

Conclusions: Bet v 1 expression in *B. pendula*

- Only the genes in subfamily 01 and 02 are expressed in pollen
- The expression of **at least** 5 PR-10 genes was confirmed in *B. pendula* pollen
- Both isoforms with a high and low IgE reactivity can be detected in pollen from one birch tree

Isoform expression

- Unique peptides elucidate expression of isoforms/genes
- Examined 3 *B. pendula* cultivars:
  - 01A ~30%
  - 01B ~20%
  - 01C ~10%
  - 02A/B ~20%
  - 02C ~20%

Isoform expression

- For other *Betula* species, determining expression of individual isoforms is not useful because in most cases we do not know the allergenicity of all isoforms
- In stead we determined the relative quantity of those peptides that constitute isoforms with high, intermediate and low allergenicity in *B. pendula*

Peptide Sequence	Peptide fragment ID	High IgE-reactivity	Intermediate IgE-reactivity	Low IgE-reactivity	Not tested	*B. pendula 'Youngii'*	*B. pendula 'Youngii' (EDE)*	*B. pendula 'Tristis'*	*B. pendula 'Tristis' (EDE)*	*B. pendula 'Fastigiata'*	*B. sp*	*B. platyphylla*	*B. pubsecens (EDE)*	*B. papyrifera*	*B. davurica*	*B. nigra*	*B. chichibuensis*	*B. lenta*	*B. alleghaniensis*	*B. medwediewii*	*B. costata*
AFILDGDNLFPK	IIIa	X				51%	45%	44%	41%	48%	45%	18%	47%	56%	59%	39%	60%	6%	32%	40%	23%
AFILDGDNLVPK	IIIb			X		19%	12%	11%	9%	11%	19%	14%	14%	3%	4%	0%	0%	0%	0%	0%	0%
AFILDGDILFPK	IIId				X	0%	0%	0%	0%	0%	0%	3%	0%	0%	0%	0%	0%	19%	0%	0%	0%
AFILDGDNLIPK	IIIe				X	13%	27%	20%	31%	41%	25%	25%	38%	39%	29%	20%	23%	43%	14%	28%	60%
AFILEGDNLIPK	IIIf		X			17%	17%	25%	19%	0%	3%	22%	2%	3%	0%	0%	17%	10%	0%	15%	3%
AFILEGDTLIPK	IIIg		X			0%	0%	0%	0%	0%	8%	0%	0%	0%	8%	0%	0%	0%	6%	12%	14%
AFILDGDNVIPK	IIIh				X	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	22%	48%	5%	0%
AFILDGNNLIPK	IIIk				X	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	42%	0%	0%	0%	0%	0%
AFILGGDNLFPK	IIInew				X	0%	0%	0%	0%	0%	0%	18%	0%	0%	0%	0%	0%	0%	0%	0%	0%
(K)ISFPEGFPFK	Va	X				46%	39%	45%	41%	43%	41%	33%	31%	37%	33%	46%	62%	0%	52%	13%	40%
(K)INFPEGFPFK	Vb			X		23%	15%	11%	7%	14%	17%	12%	24%	12%	4%	16%	0%	0%	20%	10%	5%
(K)INFPEGIPFK	Vc				?	0%	0%	0%	0%	0%	0%	5%	0%	0%	11%	0%	0%	0%	0%	10%	0%
(K)ISFPEGFPFR	Vd				X	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	20%	0%	24%	0%
(K)ITFPEGSPFK	Ve		X			32%	46%	44%	52%	42%	42%	49%	45%	51%	52%	38%	39%	80%	28%	43%	56%
IVATPDGGSILK	Xa	X				82%	95%	80%	96%	87%	87%	86%	91%	95%	81%	86%	100%	100%	72%	89%	100%
LVATPDGGSILK	Xg		X																		
IVATPDGGCVLK	Xb			X		18%	5%	20%	4%	13%	9%	14%	9%	0%	0%	0%	0%	0%	21%	2%	0%
IVATPNGGSILK	Xc	X				?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
IVETPNGGSILK	Xf				X	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	14%	0%	0%	0%	0%	0%
IVATPDGGCILK	Xnew				X	0%	0%	0%	0%	0%	4%	0%	0%	5%	19%	0%	0%	0%	7%	9%	0%

Conclusion

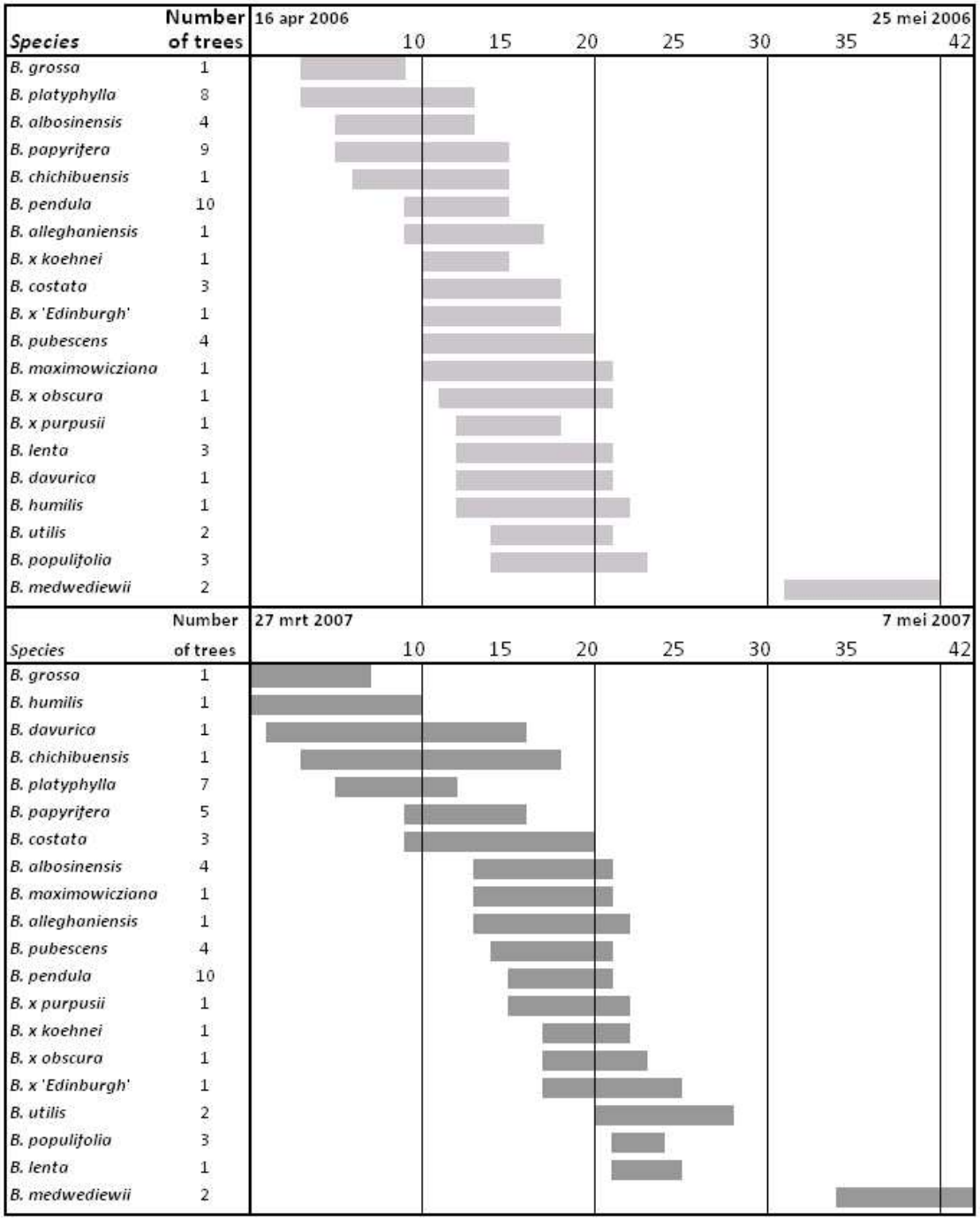
- Q-TOF LC-MS^E data provide:
  - Presence/absence of peptides/isoforms
  - Quantitative data
- All (examined) *Betula pendula* cultivars are likely to be highly allergenic
- Allergens that are similar to the most allergenic isoforms are probably present in all species, except possibly *B. lenta*

Hypoallergenic birch trees?

- Perhaps in *B. lenta*

or

- Use male sterile hybrid cultivars
- Decrease length of the pollen season by limiting the number of different species planted in a city


Species
Number of trees
16 apr 2006
25 mei 2006
10 15 20 25 30 35 42
B. grossa 1
B. platyphylla 8
B. albosinensis 4
B. papyrifera 9
B. chichibuensis 1
B. pendula 10
B. alleghaniensis 1
B. x koehnei 1
B. costata 3
B. x 'Edinburgh' 1
B. pubescens 4
B. maximowicziana 1
B. x obscura 1
B. x purpusii 1
B. lenta 3
B. davurica 1
B. humilis 1
B. utilis 2
B. populifolia 3
B. medwediewii 2
Species
Number of trees
27 mrt 2007
7 mei 2007
10 15 20 25 30 35 42
B. grossa 1
B. humilis 1
B. davurica 1
B. chichibuensis 1
B. platyphylla 7
B. papyrifera 5
B. costata 3
B. albosinensis 4
B. maximowicziana 1
B. alleghaniensis 1
B. pubescens 4
B. pendula 10
B. x purpusii 1
B. x koehnei 1
B. x obscura 1
B. x 'Edinburgh' 1
B. utilis 2
B. populifolia 3
B. lenta 1
B. medwediewii 2

2006

2007

Acknowledgements

Genomics

- D. Esselink
- W. Van 't Westende

Proteomics

- A.H.P. America
- J.H.G. Cordewener

PLANT RESEARCH INTERNATIONAL
WAGENINGEN UR

netherlands proteomics centre