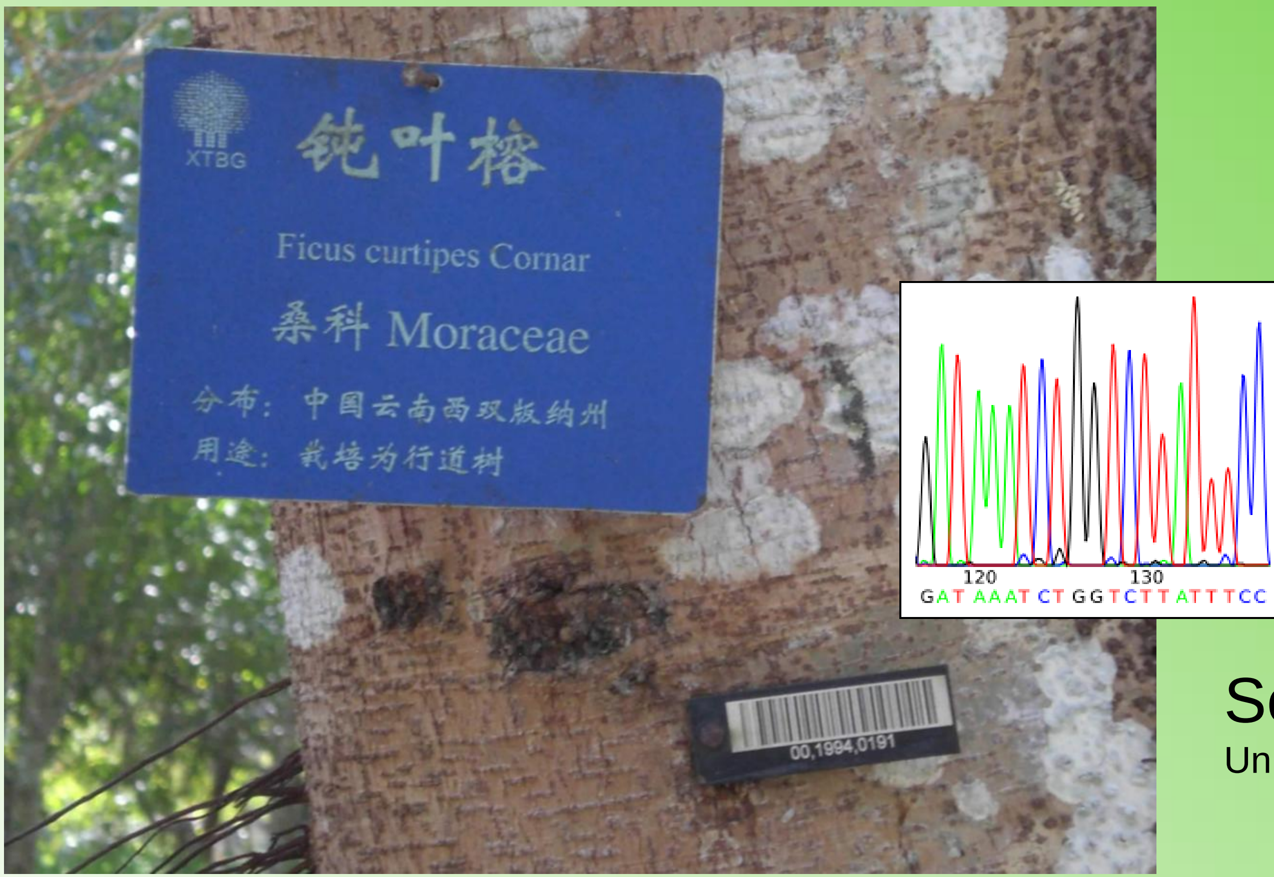
Plant DNA barcoding --
Advances, applications & limits



Sean Graham
University of British Columbia

Funding provided by:

GenomeCanada
Ontario Ministry of Research and Innovation
OntarioGenomicsInstitute
NSERC CRSNG
international BARCODE OF LIFE

ubcbotanicalgarden & centre for horticulture

Plant DNA barcoding --
Advances, applications & limits

CO-AUTHORS:

Prasad R. Kesanakurti; Aron J. Fazekas; Diana M. Percy; Kevin S. Burgess; Jeffery M Saarela; Steven G. Newmaster; Brian C. Husband; Mehrdad Hajibabaei; Spencer C. H. Barrett

Root project: Prasad R. Kesanakurti; Aron J. Fazekas; Kevin S. Burgess

Grasses & Willows: Diana Percy; Jeffery Saarela

Institutions: UBC; U. Guelph, U. Toronto, Canadian Museum of Nature

Overview

Choosing a multi-locus barcoding system

Are plants harder to barcode than animals?

-- CBOL Plant Working Group (PNAS 2009)

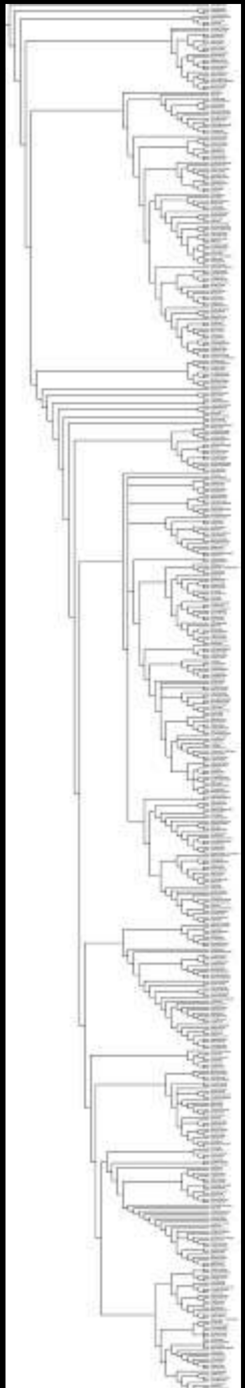
-- Fazekas & al., 2008 (*PLoS ONE* 2008)

-- Fazekas & al. 2009 (*Mol. Ecol. Resources 2009)*

Plant DNA barcoding studies

-- Poaceae & *Salix* of British Columbia, Canada

-- Eco-applications: Below-ground ecology



PLANT DIVERSITY

LAND PLANTS
- *c*. 400,000 species

VASCULAR PLANTS
- > 350,000 species
- c. 13,888 genera
- c. 511 families

TAXONOMY

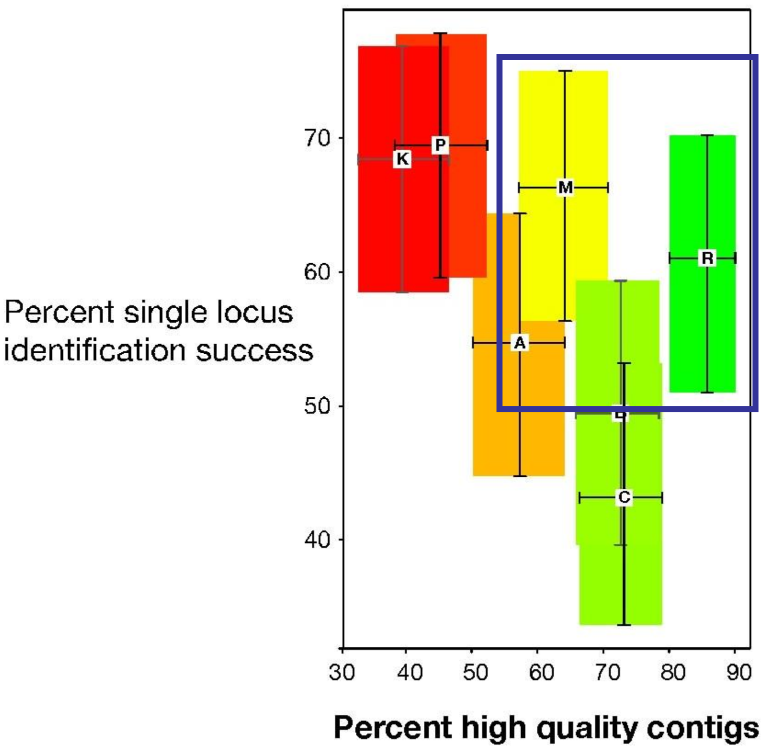
Wanted: A Barcode for Plants

Pennisi (2007)
Science 318: 190-191

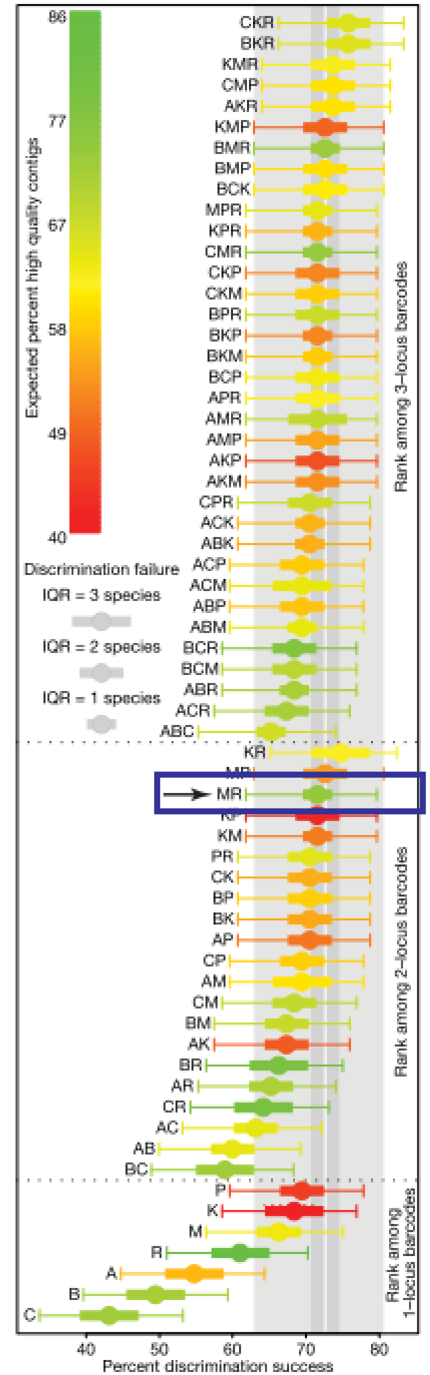
Plant working group
(Sept 2008)

Core loci (CBOL): *rbc*L+*mat*K (2 locus)

(plus suppl. loci)



Noncoding	**Coding**
K = *psb*K-*psb*I	M = *mat*K
P = *psb*A-*trn*H	R = *rbc*L
A = *atp*F-*atp*H	B = *rpo*B
	C = *rpoC*1

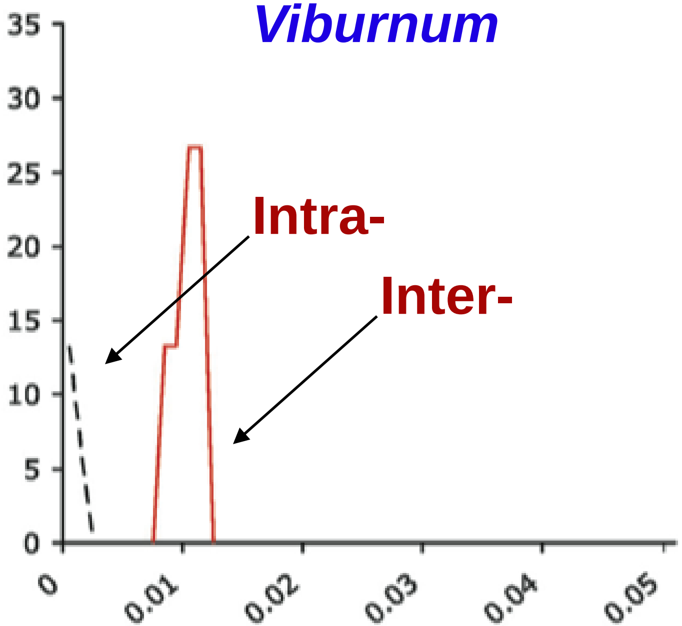


CBOL Plant working group (PNAS 2009)

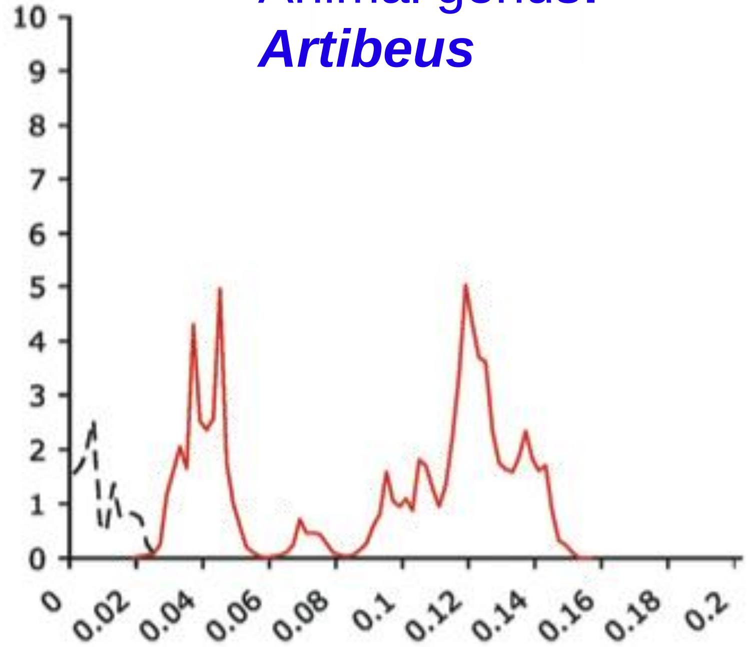
What about the barcoding gap?

-- Ideally: **Intra**specific < **Inter**specific variation

Plant genus: ***Viburnum***



Animal genus: ***Artibeus***

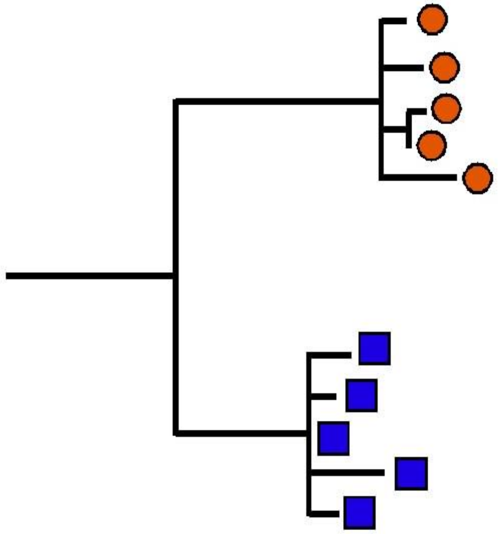


Genetic distance (K2P)

What about the barcoding gap? Ideally --

-- **Intra**specific < **Inter**specific variation

-- Gene-tree monophyly tracks **species boundaries**



= Mutual (reciprocal) **monophyly**

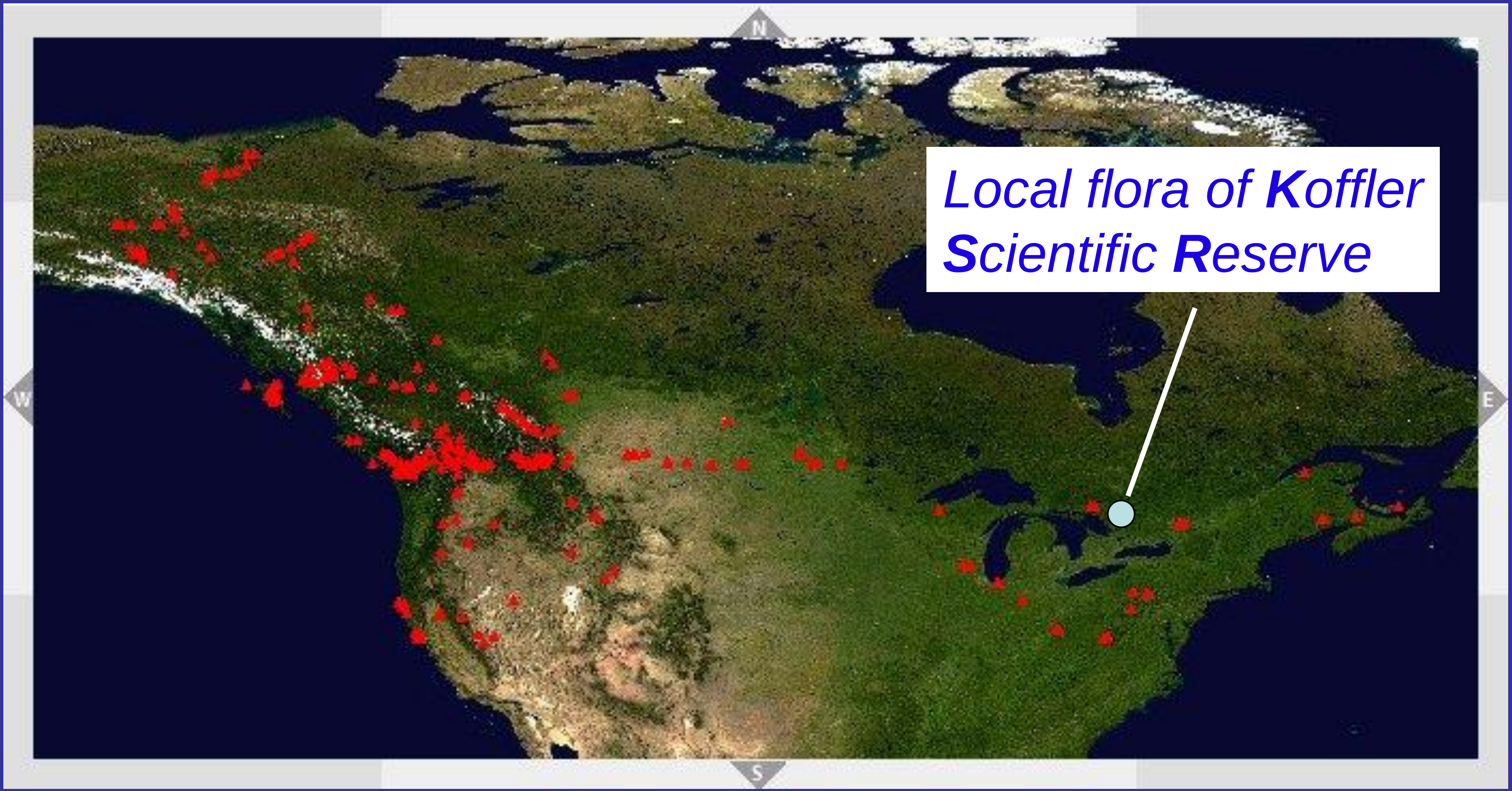
Samples of --

Species A

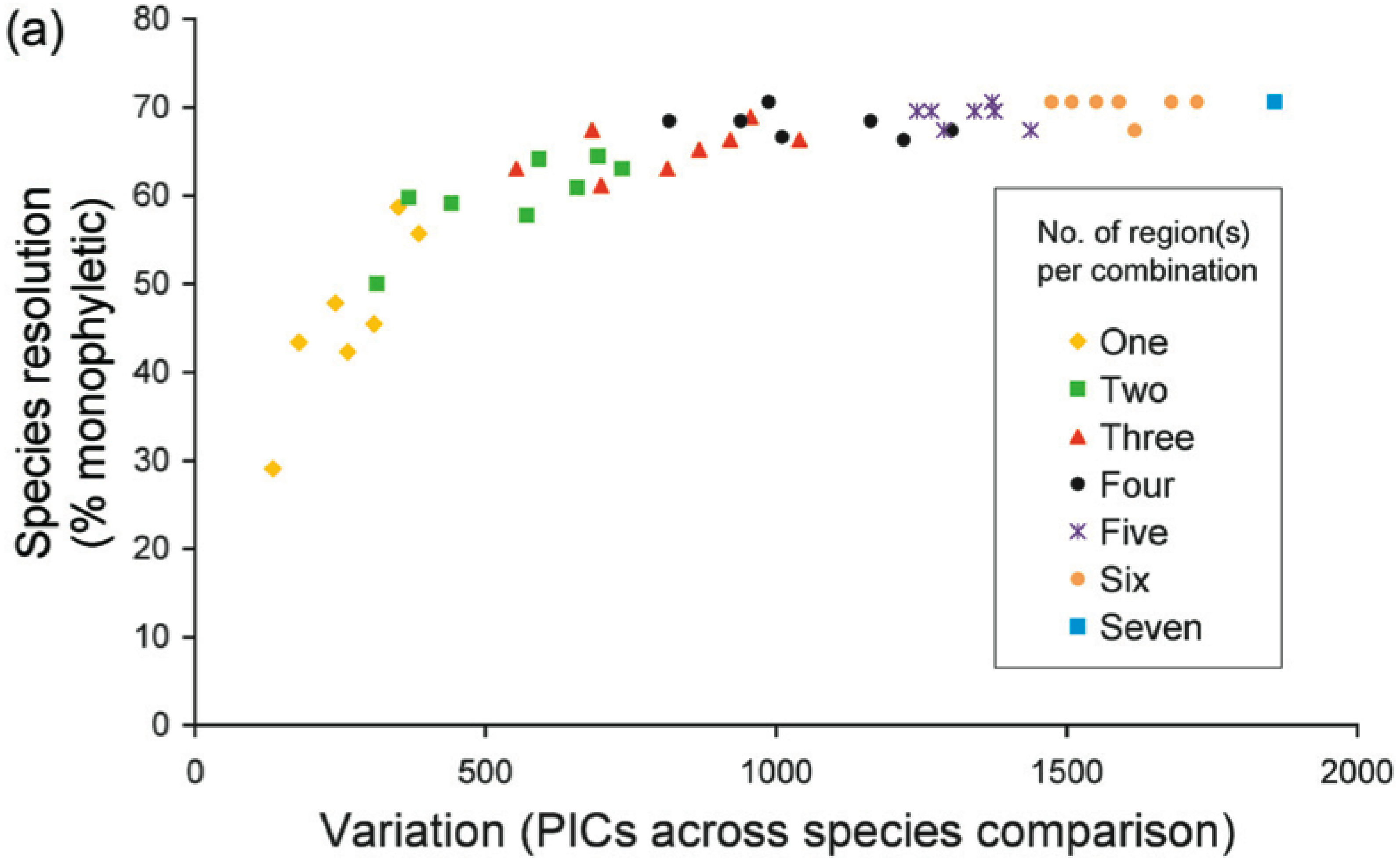
Species B

Observed **monophyly fraction** = one criterion for **species discrimination**

Barcoding a local flora (KSR, Ontario, Canada)



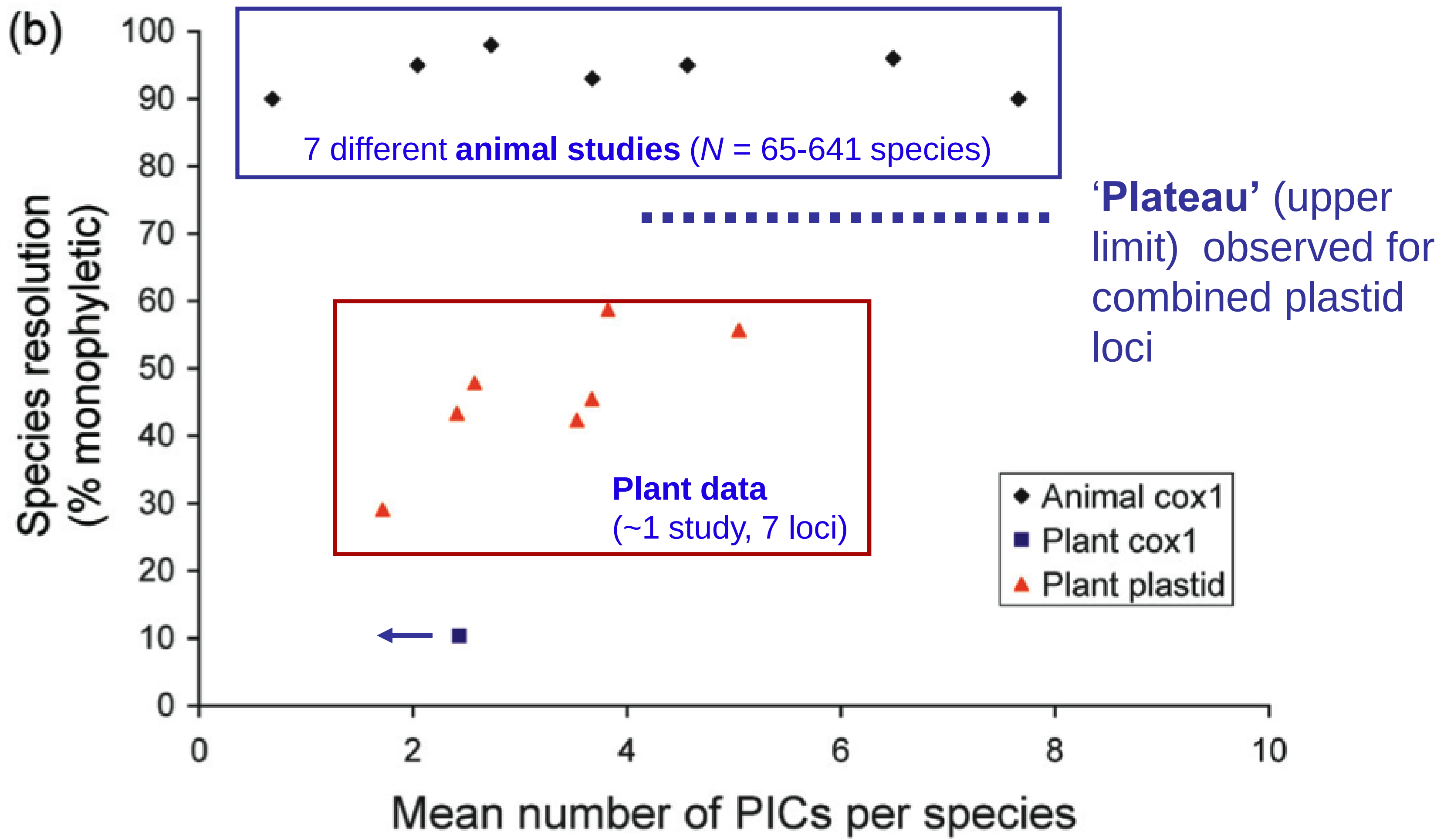
Relation between **sequence variation** & percent **species resolution**



(*PIC = Parsimony Informative Character)

Fazekas & al., 2008
(*PLoS ONE* e2802)

Plants are harder to barcode precisely
(controlling for amount of variation per species)

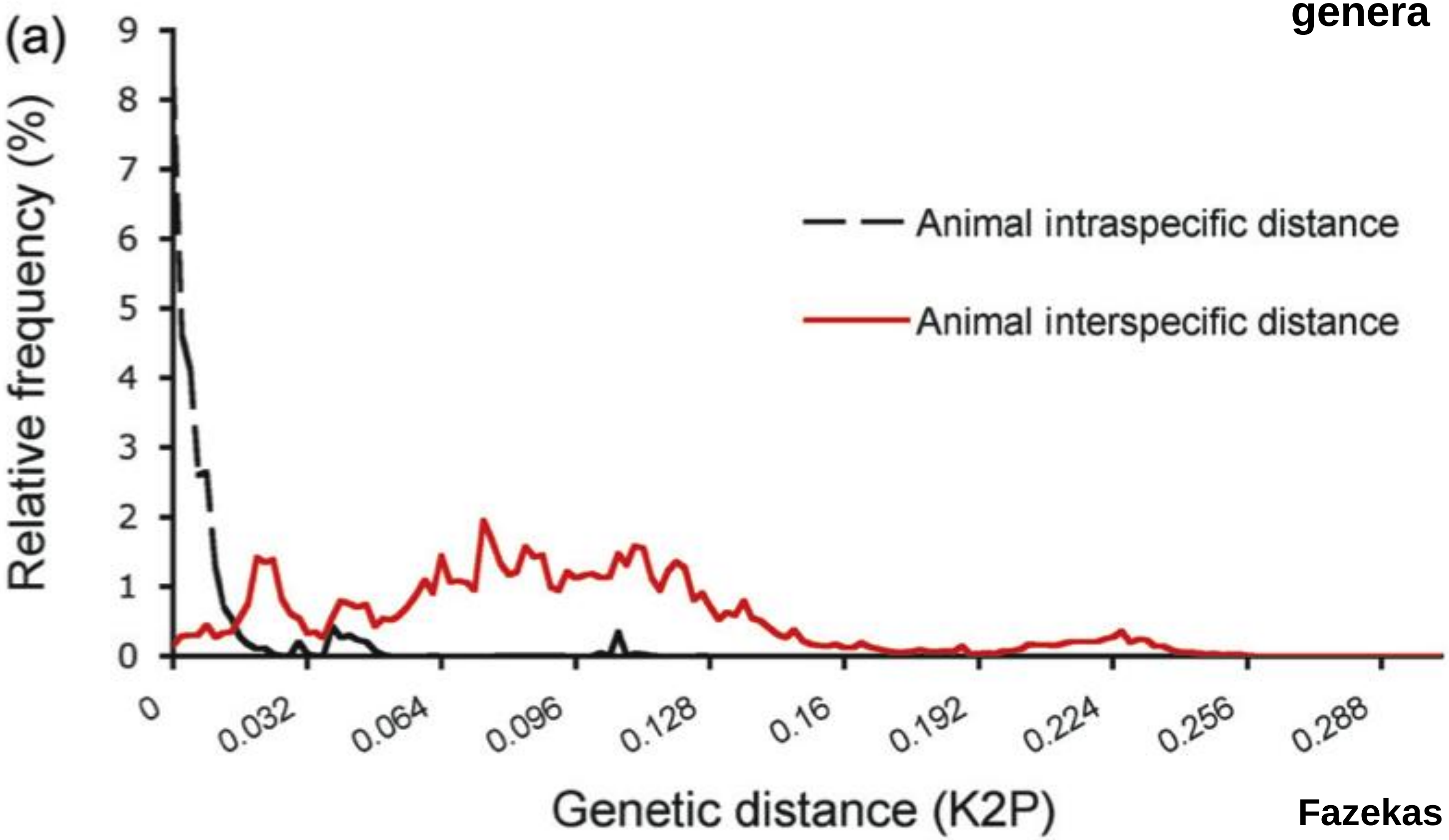


(*PIC = Parsimony Informative Character)

Fazekas & al., 2009
(*Mol Ecol Res*)

Overview of barcoding gap (across multiple **animal genera**)

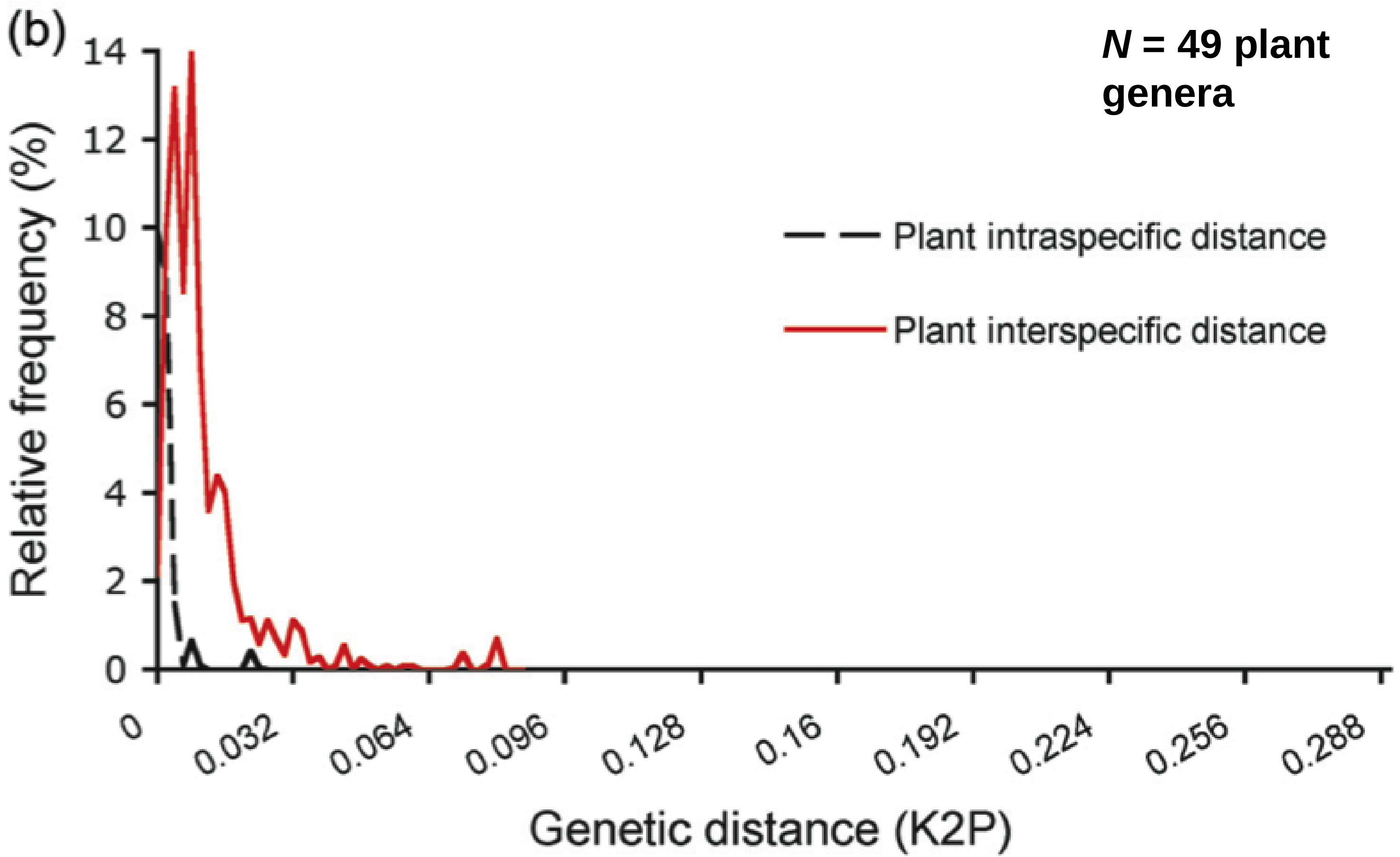
***N* = 326 animal genera**



Fazekas & al., 2009
(*Mol Ecol Res*)

Plants typically have **more overlap** in inter- vs. intraspecific distances

(& **smaller scale**: plastid genome is slower than animal mt)

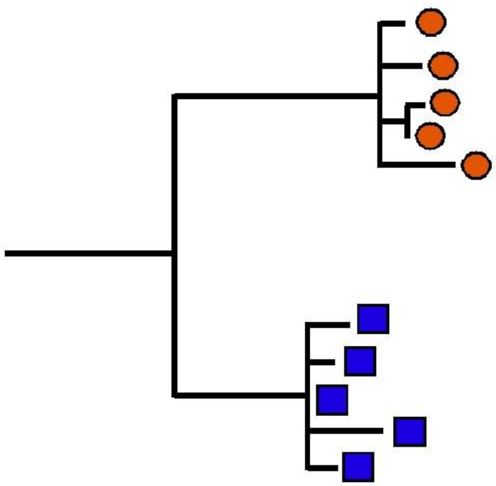


***N* = 49 plant genera**

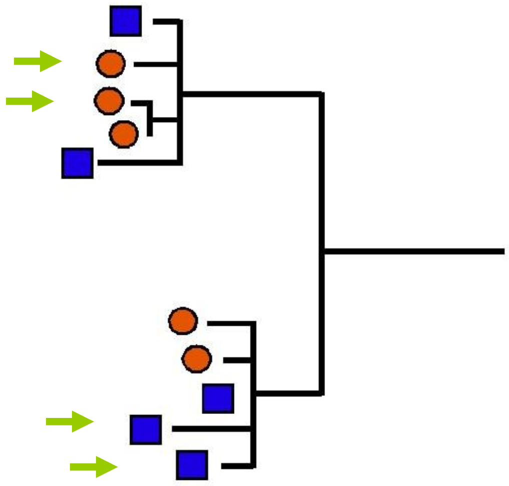
Fazekas & al., 2009
(*Mol Ecol Res*)

Is gene-tree paraphyly more extensive in plants?

Why? More introgression? Shallower gene-tree coalescences? “Poorer” species-level circumscriptions?



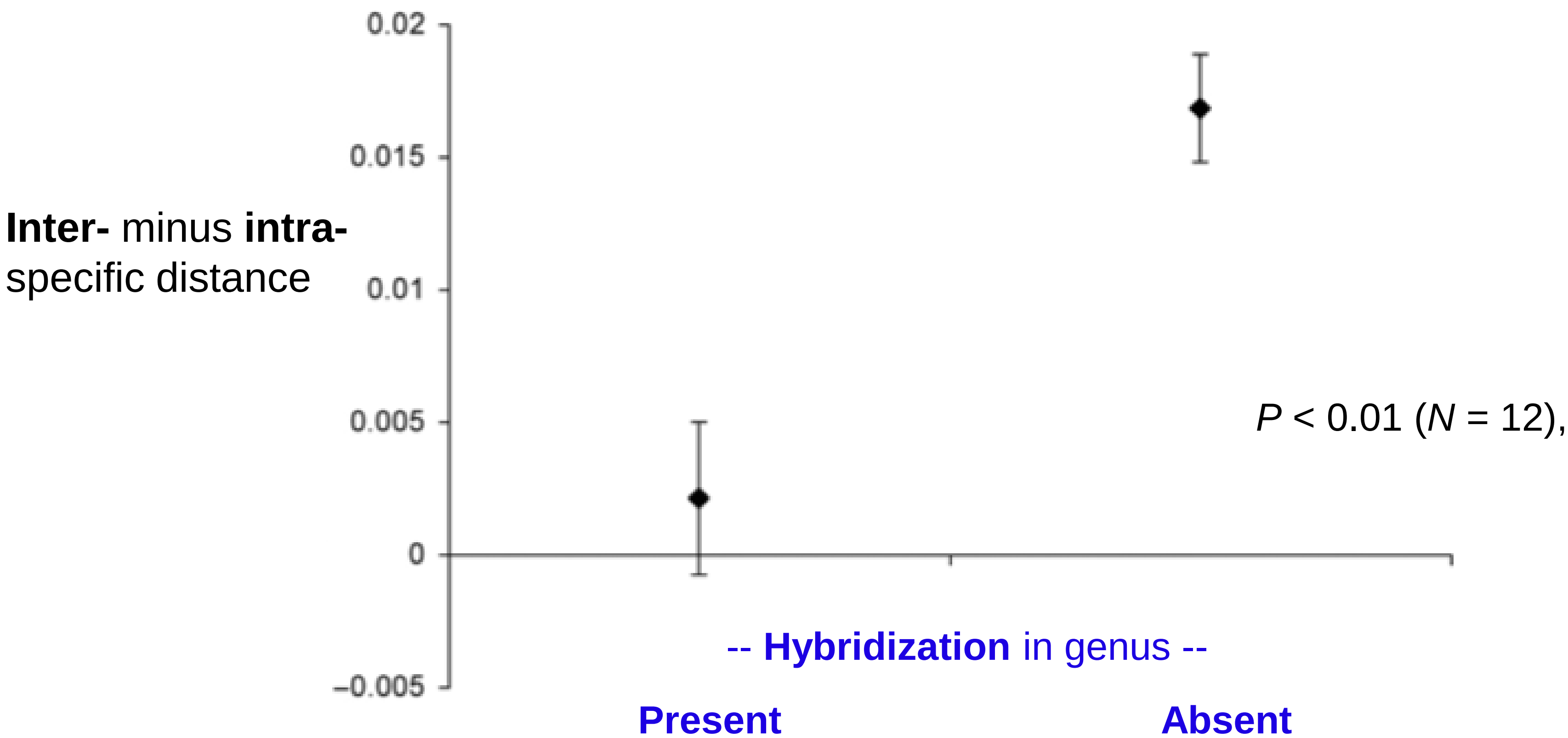
vs.



Sister species
mutually monophyletic

Gene-tree paraphyly
(introgression, lineage-sorting, etc)

Lack of ‘barcoding gap’ in plants is associated with presence of **hybridization** (but **not** polyploidy)



Fazekas & al., 2009
(*Mol Ecol Res*)

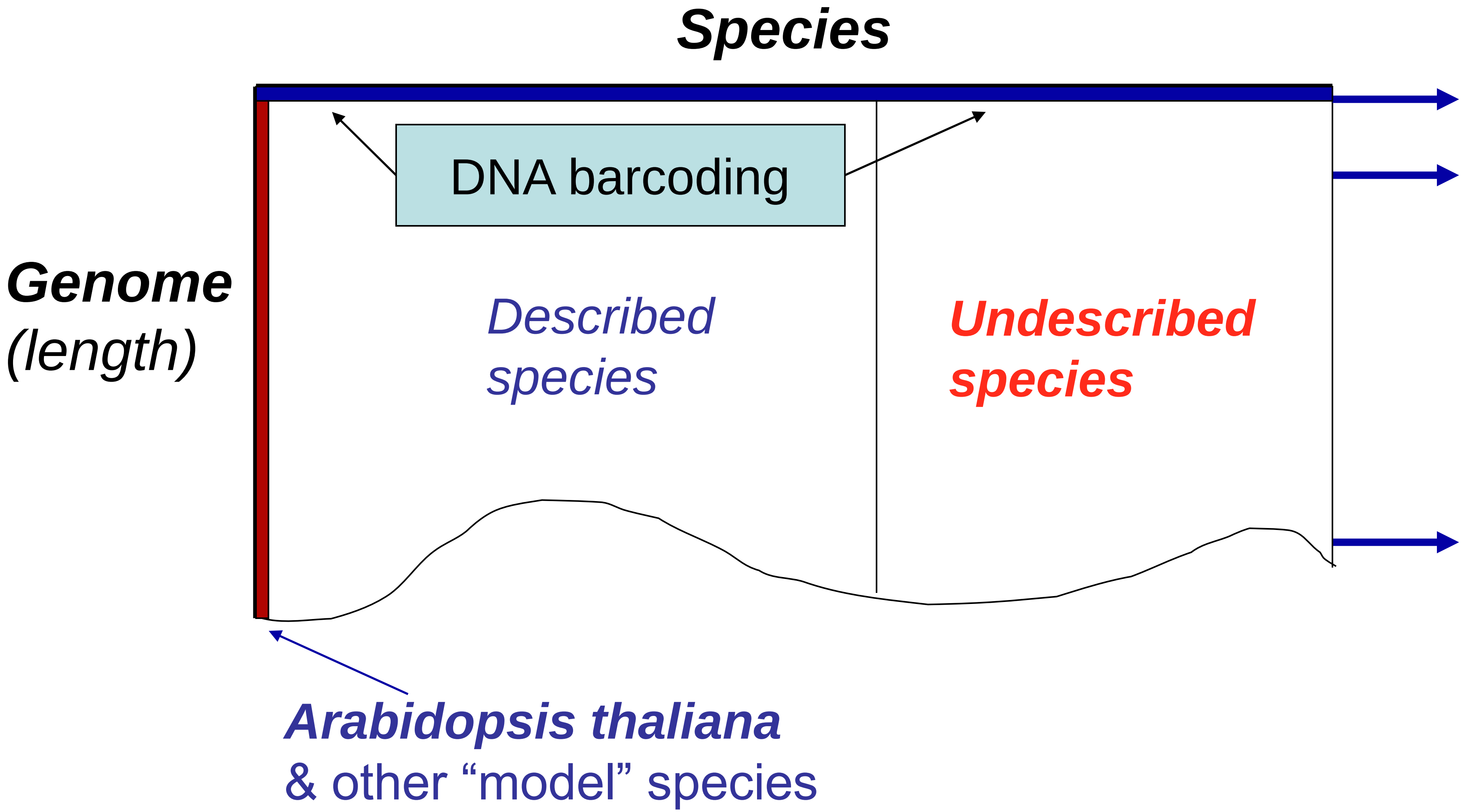
How problematic is 30%+ error (impression)

-- Can think of as question of **resolution**:

* Lower resolution 'images' can work well

-- Depends on the application: e.g. forensics vs. bioinventory

DNA Barcoding as **“horizontal” genomics**



Overview

Choosing a multi-locus barcoding system

Are plants harder to barcode than animals?

-- CBOL Plant Working Group (PNAS 2009)

-- Fazekas & al., 2008 (*PLoS ONE* 2008)

-- Fazekas & al. 2009 (*Mol. Ecol. Resources 2009)*

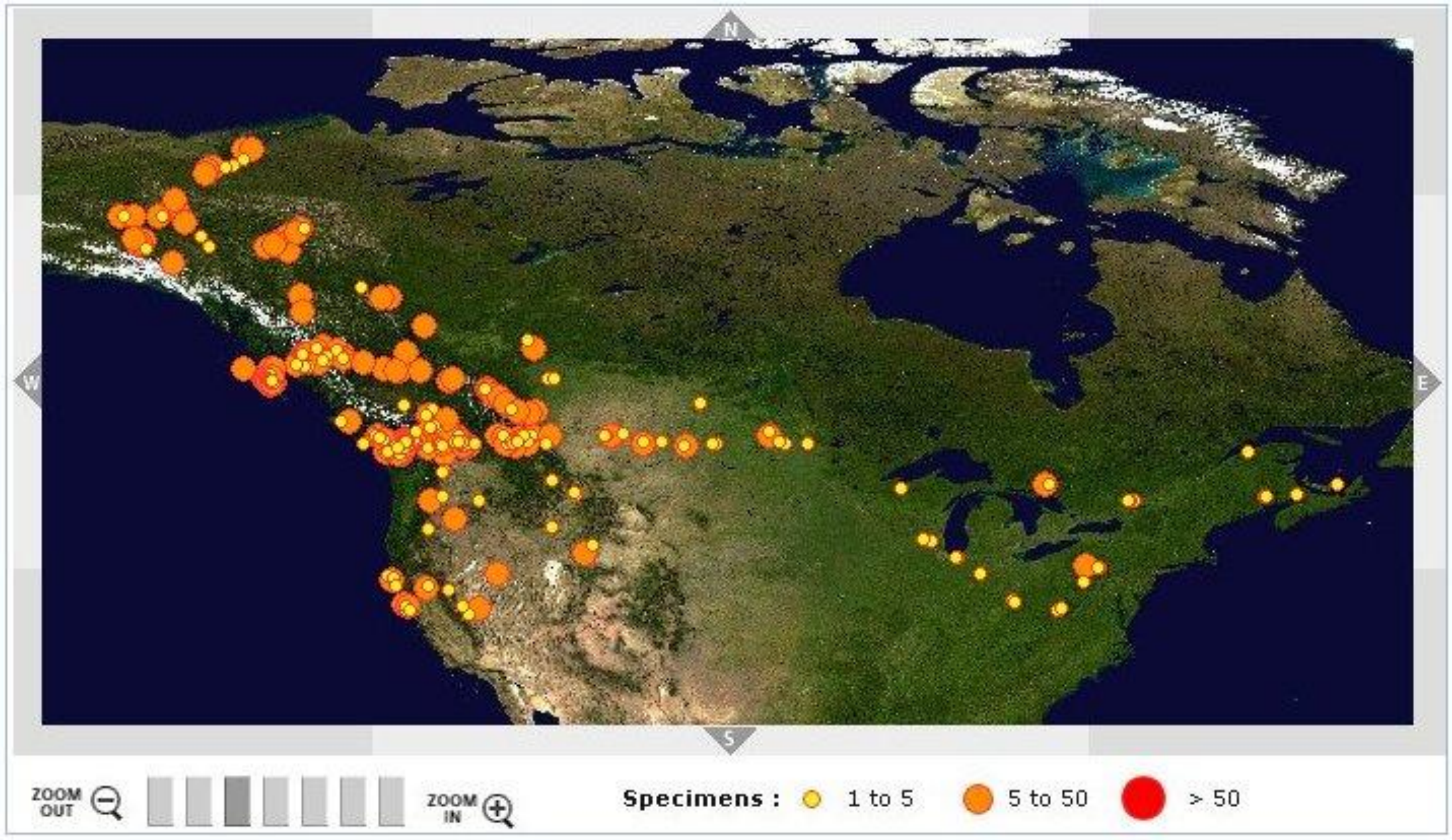
Plant DNA barcoding studies

-- **Poaceae & *Salix*** of BC, Canada/ **GrassBoL campaign**

-- Eco-applications: **Below-ground ecology**

Samples collected for grasses –

BOLD specimen map



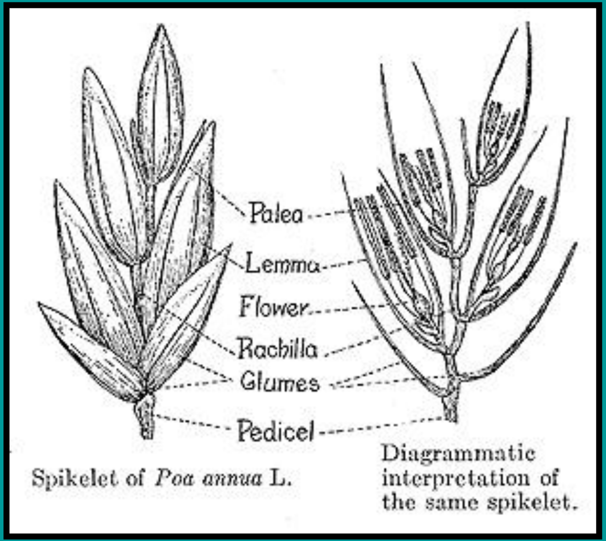
Economically important:

Wheat, rice & maize provide > 50% of human calories

Wheat (*Triticum*)

Rice (*Oryza*)

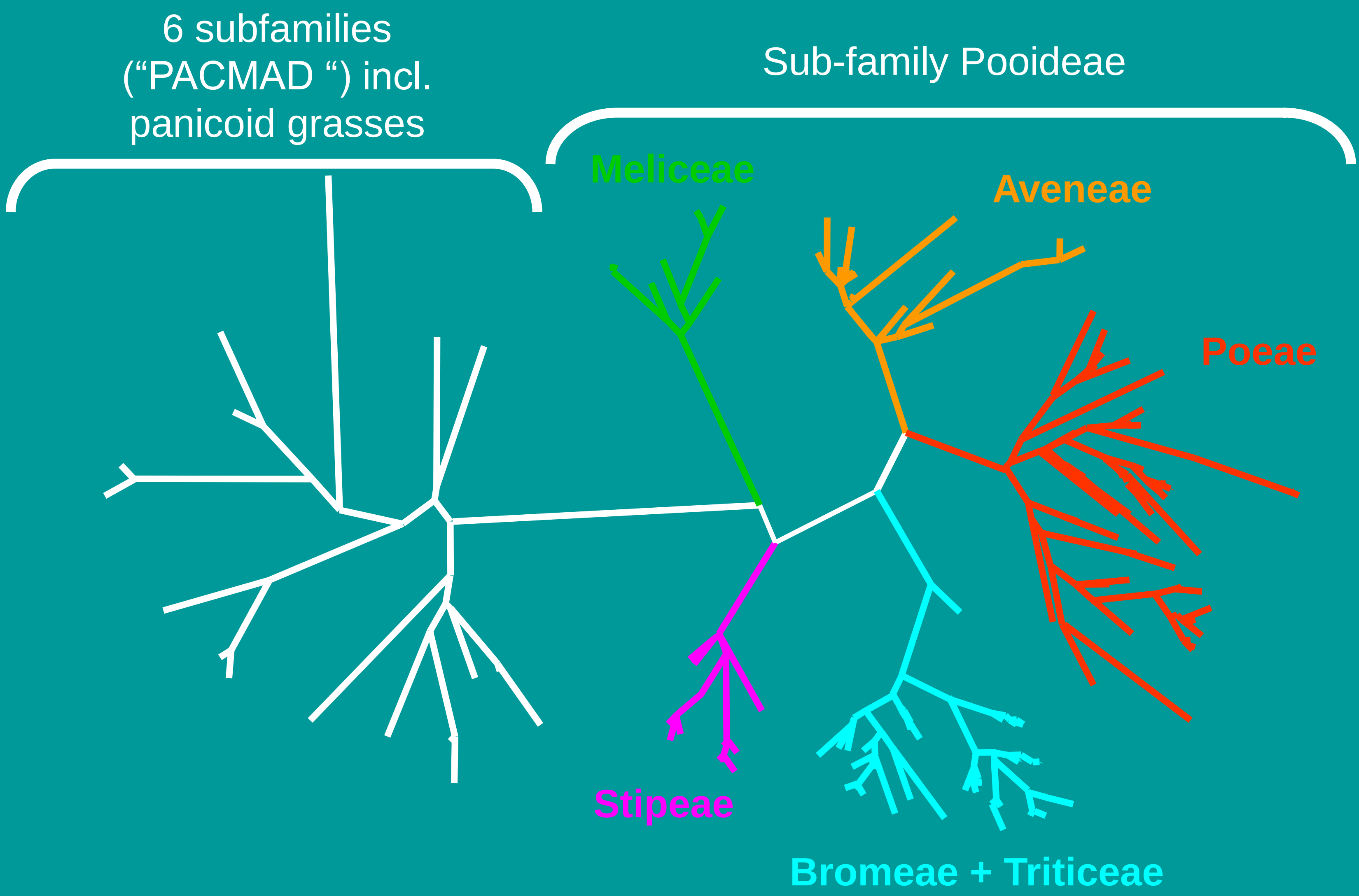
Maize (*Zea*)



Identification is often problematic –

Highly reduced reproductive morphology

Our grass sampling concords with grass family systematics



Grasses in British Columbia

Difficult groups: ***Poa*** **(29 species)** ***Bromus*** **(21 species)**
Festuca **(15 species)**

Agrostis **(13 species)** ***Elymus*** **(11 species)** ***Glyceria*** **(10 species)**
Calamagrostis **(8 species)**

Sequence success:

	Grasses	Willows
*rbc*L	550 – 94%	541 – 98%
*mat*K	550 – 89%	541 – ~90%
*trn*H-*psb*A	419 – 95%	200 – 99%

(Willow herbarium material from 1940s – 2000s – oldest samples: 1946 & 1947)

For comparison: Fazekas et al. (2008)
***Multiple land plants -- rbc*L - 100%, *mat*K – 87.6%, *trn*H-*psb*A – 99.2%**

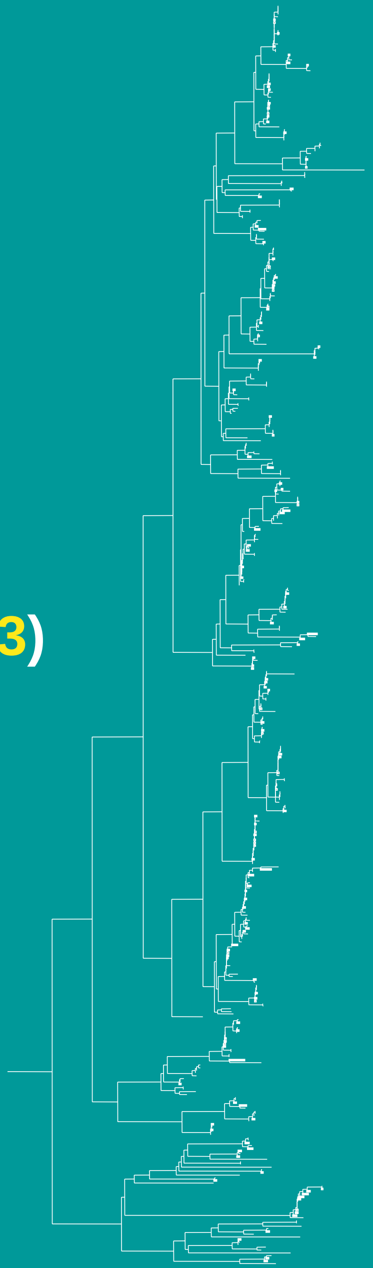
Grasses in British Columbia

*rbc*L + *mat*K (+ *trn*H-*psb*A)

190 species (native & introduced)
- incl. 70% of British Columbian species

68% spp. represented by 2-16 samples (mean 3)

73 genera (native & introduced)

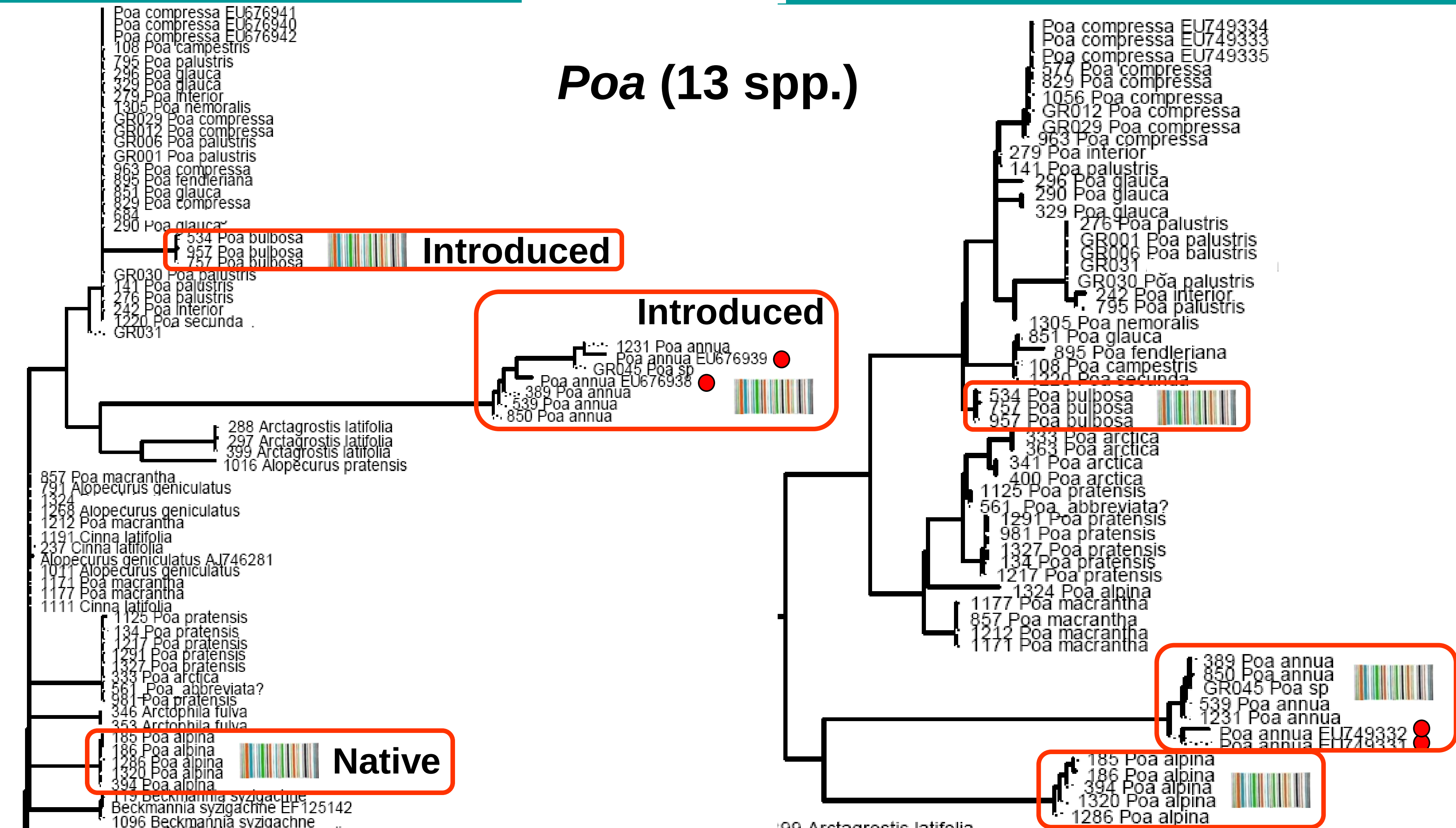


*rbc*L

● = GenBank

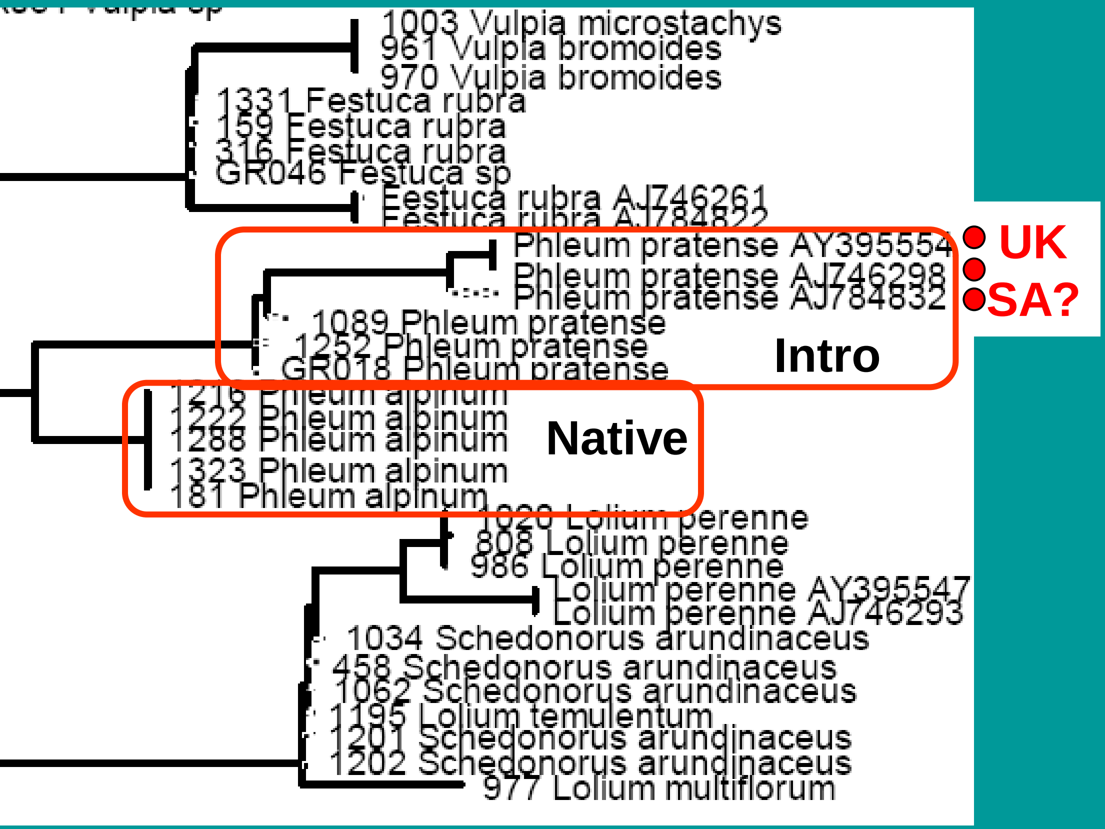
*mat*K

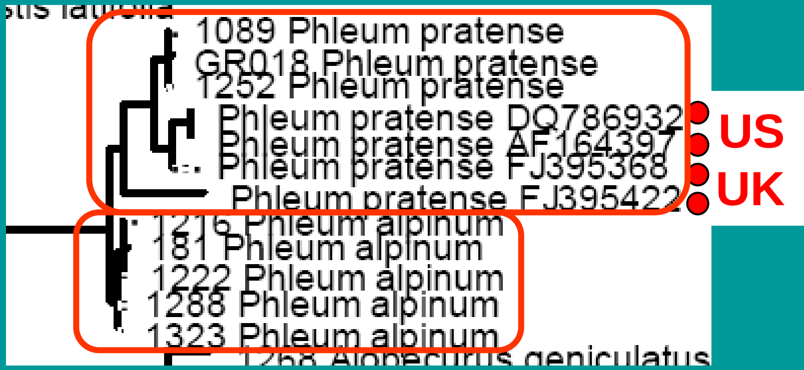
***Poa* (13 spp.)**



*rbc*L

*mat*K





● = GenBank

(not due to sequence length difference)

Species assignment in Grasses

NJ distance (K2P) bootstrap ≥70%

	Species monophyletic	**Genera monophyletic ***
*rbc*L+*mat*K	of 128 – 40% (1/2 100% supp)	of 27 – 70% (all ≥80% supp)
+*trn*H-*psb*A	of 128 – 40% (5 diff. spp.)	–

* Non-monophyletic grass genera:
Ammophila (2 spp.), ***Calamagrostis*** (9 spp.), ***Cynosurus*** (2 spp.), ***Elymus*** (9 spp.), ***Festuca*** (9 spp.), ***Leymus*** (5 spp.), ***Poa*** (13 spp.)

GrassB of L

arcoding ife

An international initiative to barcode the grasses & grass-like plants

Possible GrassBoL objectives

- Bring together grass researchers from diverse fields (taxonomists, ecologists, agronomists, molecular biologists) & coordinate DNA barcoding efforts
- Identify funding opportunities
- Coordinate protocols & primers for core & supplementary barcoding loci in Poales
- Develop barcoding applications

GrassBoL Organizers

Andy Lowe

Hugh Cross

Adelaide University &
State Herbarium of South Australia

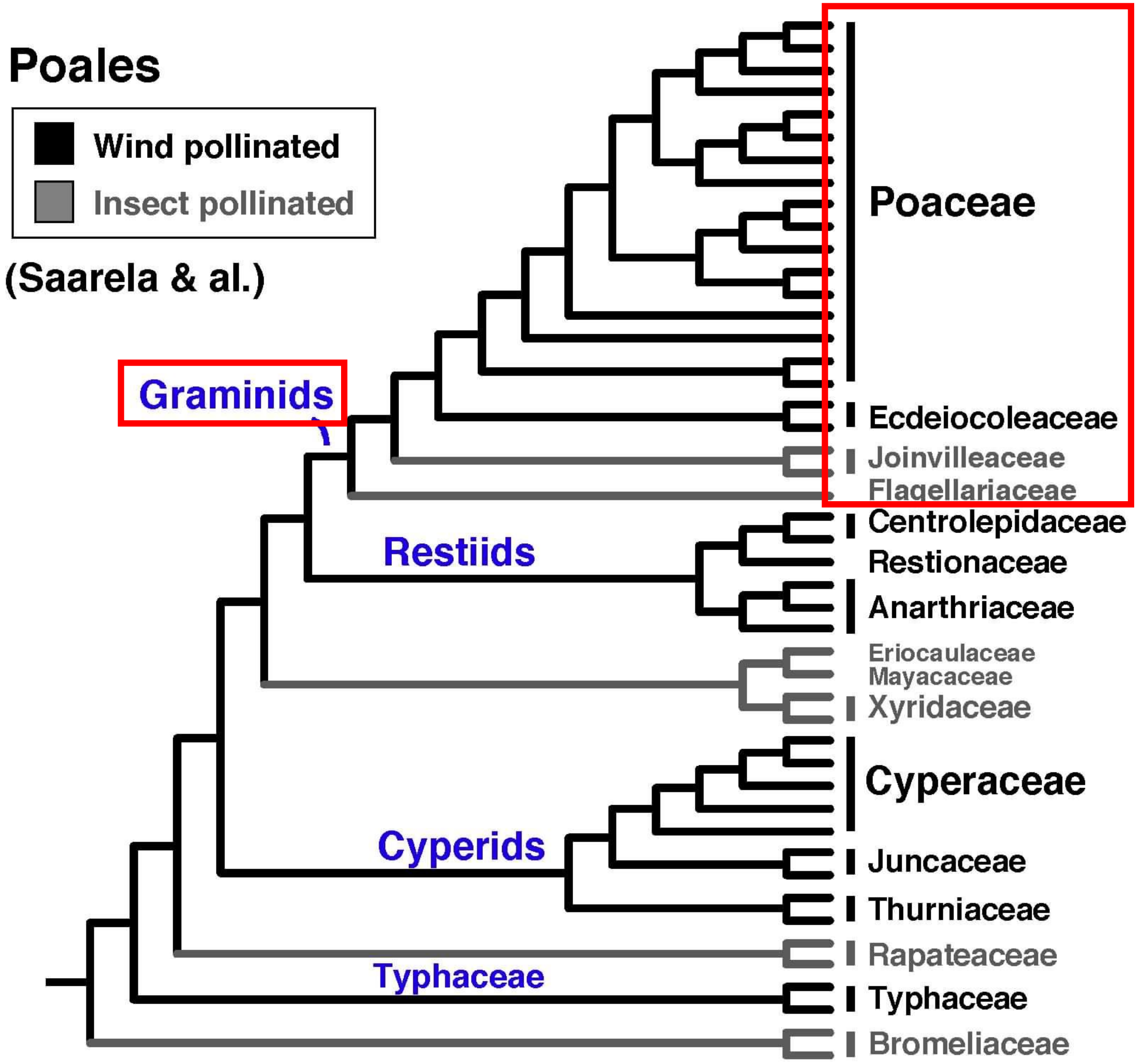
Sean Graham

University of British Columbia

GrassB$^{o}_{arcoding}$ ^{f} L$_{ife}$

- Grasses & rels important **economically/ecologically**
- **‘Difficult’ to ID/ key**
- Feasible to DNA barcode --
  - **Good universality** for core & supplementary loci
  - Plastid genome has **elevated rate**
  - Phylogenetically **discrete** (1 major clade, **Poales**)
  - Substantial **systematic expertise** & interest
  - Substantial **herbarium resources**
  - Tractable to get **DNA from herb. specimens**
- Substantial **genomic resources** to develop nuclear/ next-gen markers

Relationships among grasses (Poaceae) & relatives (Poales)



I. Graminids

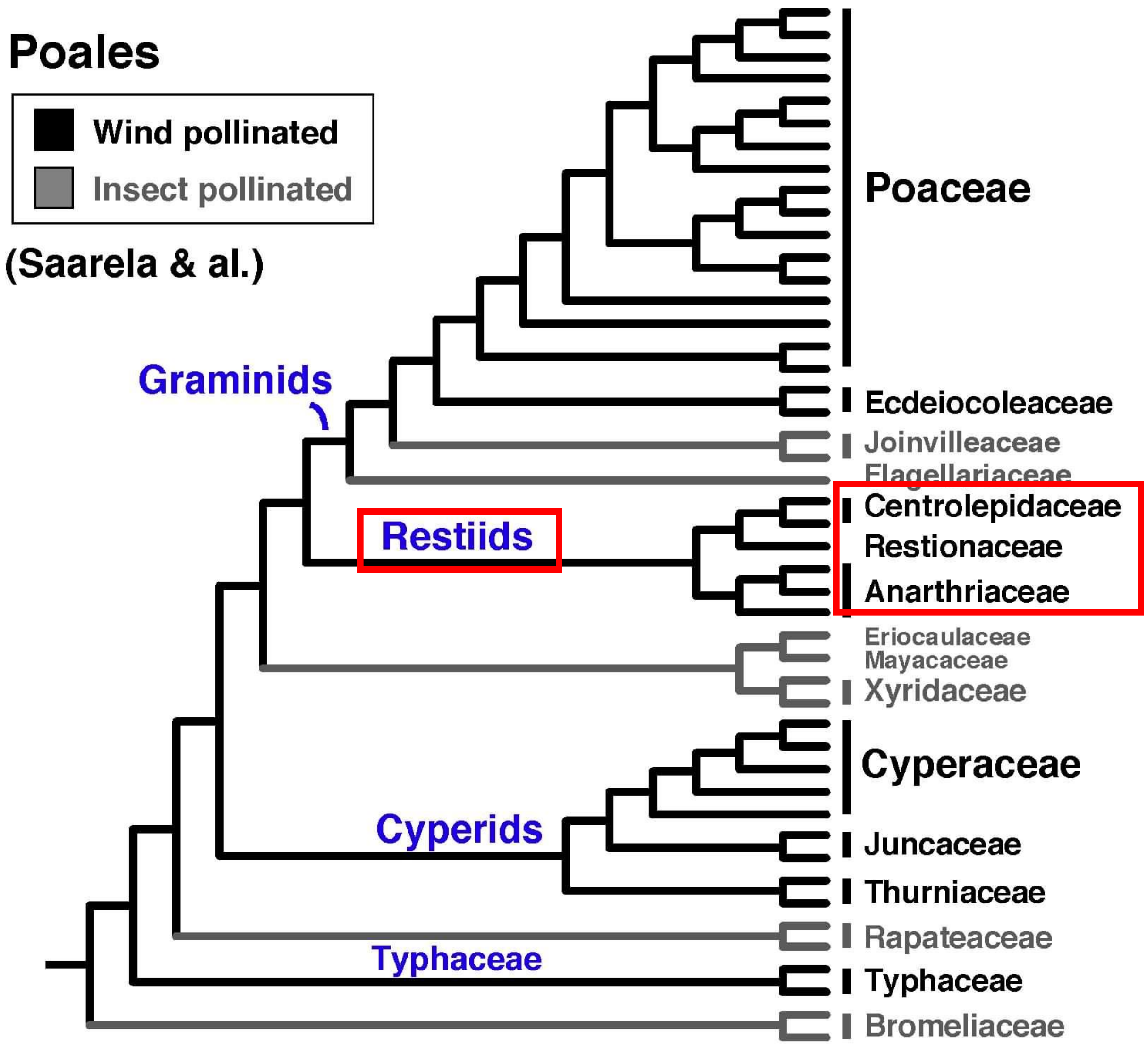
Poaceae: 11K spp., 600-900 genera, 12 subfams. **Cosmopolitan**

Ecdeiocoleaceae: 3 spp., **W Aust**

Joinvilleaceae: 2 spp., **SE Asia**

Flagellariaceae: 4 spp., **S Hemisph**

Relationships among grasses (Poaceae) & relatives (Poales)

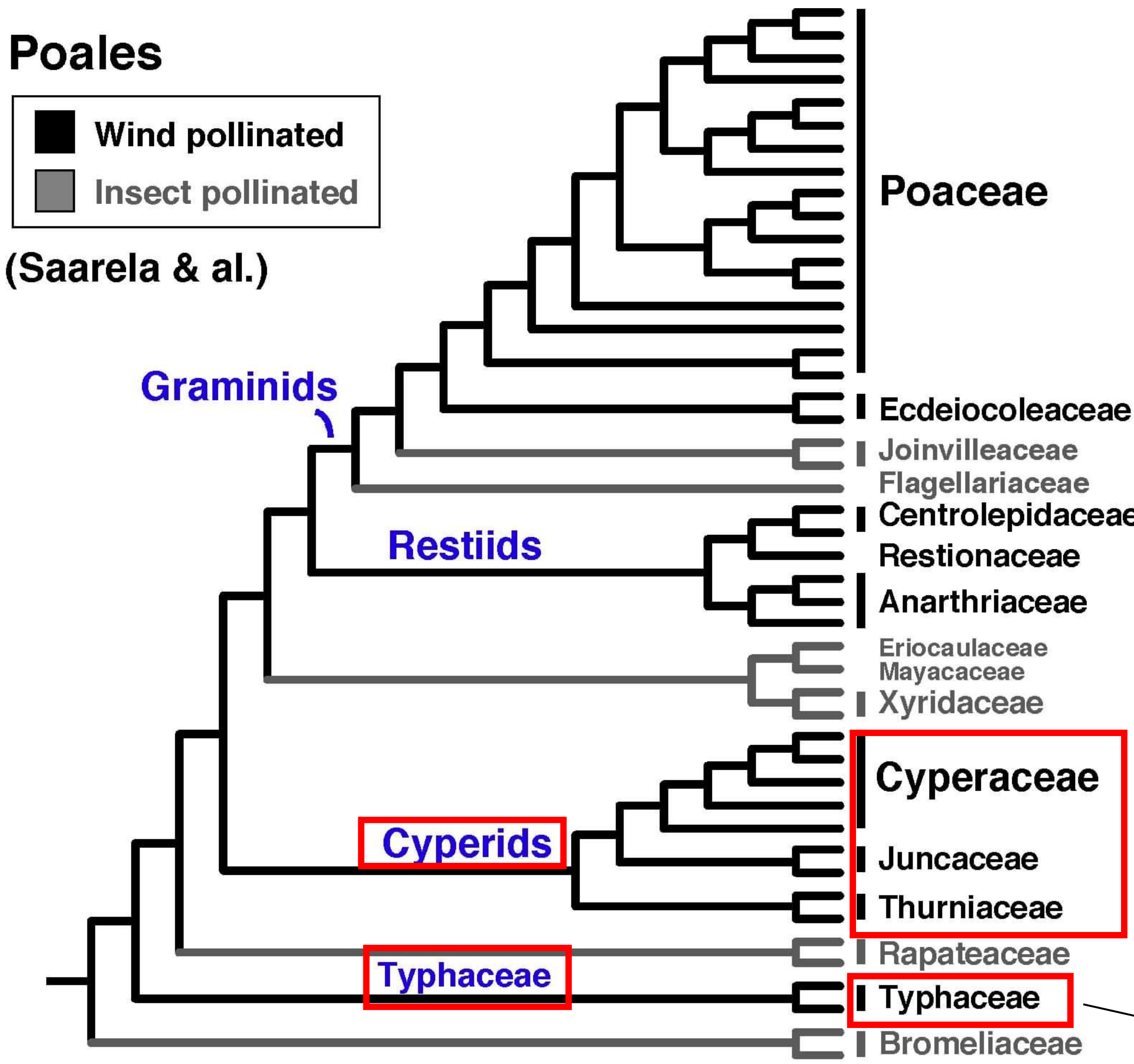


II. Restiids: Southern rushes

Hypodiscus (Restionaceae)

Restionaceae: ~500 spp., 55 gen.
S Africa, Australia, etc
Centrolepidaceae: 35 spp., **S. Hem.**
Anarthriaceae: 10 spp., **W Australia**

Relationships among grasses (Poaceae) & relatives (Poales)



III. Cyperids: sedges & rushes

Cyperaceae: ~5000 spp, 120 gen. **Cosmopolitan**

Juncaceae: ~440 spp., **Cosmopolitan**

Thurniaceae: 4 spp., **S Am, S Africa**

IV. Cattails & bur-reeds

Typhaceae: ~30 spp., 2 gen. **Cosmopolitan**

GrassBoL: Promoting links with active taxonomy

- DNA from **herbarium specimens**
- **Acrive taxonomic expertise**
- **Critical funding components** incl. collecting, **expert ID/training** & digitization

Overview

Choosing a multi-locus barcoding system

Are plants harder to barcode than animals?

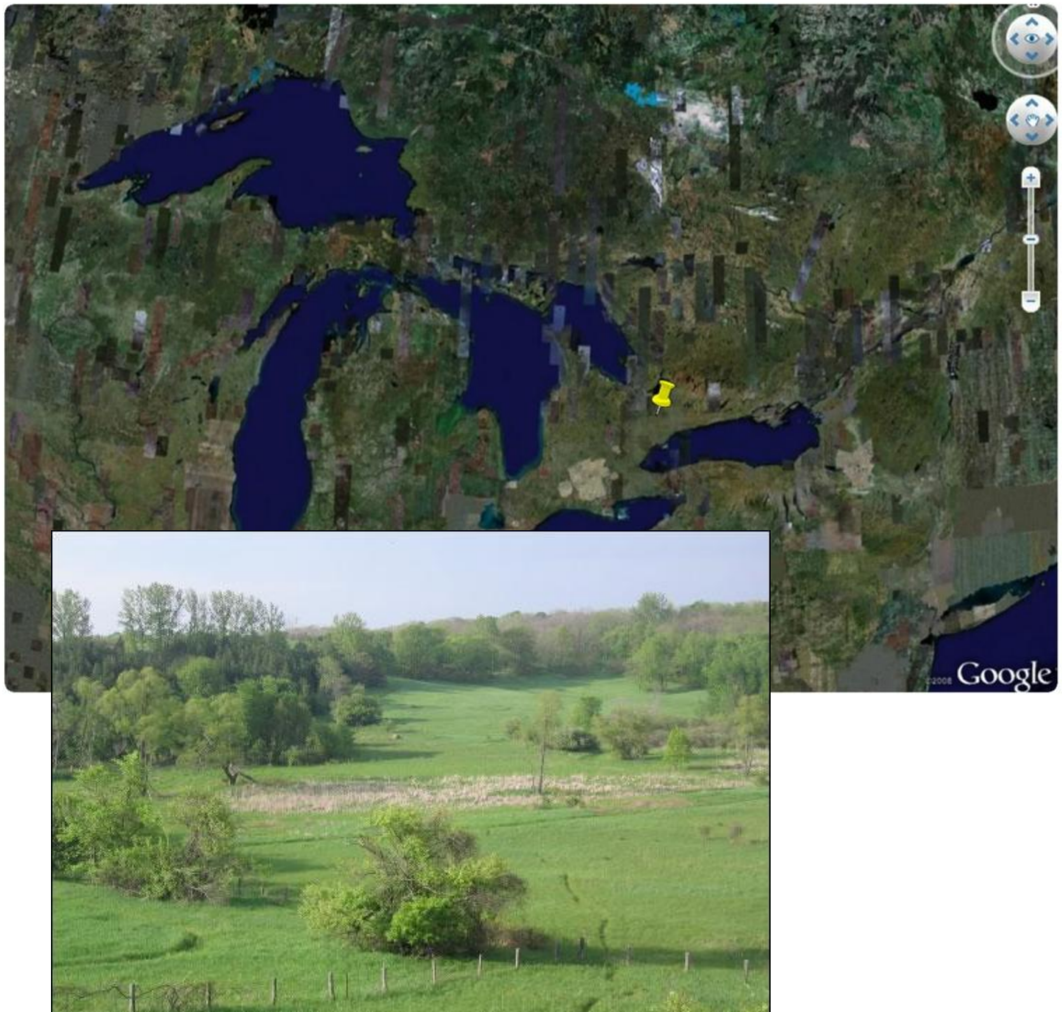
Plant DNA barcoding studies

-- **Poaceae & *Salix*** of BC, Canada; GrassBoL campaign

-- Eco-applications: **Below-ground ecology**

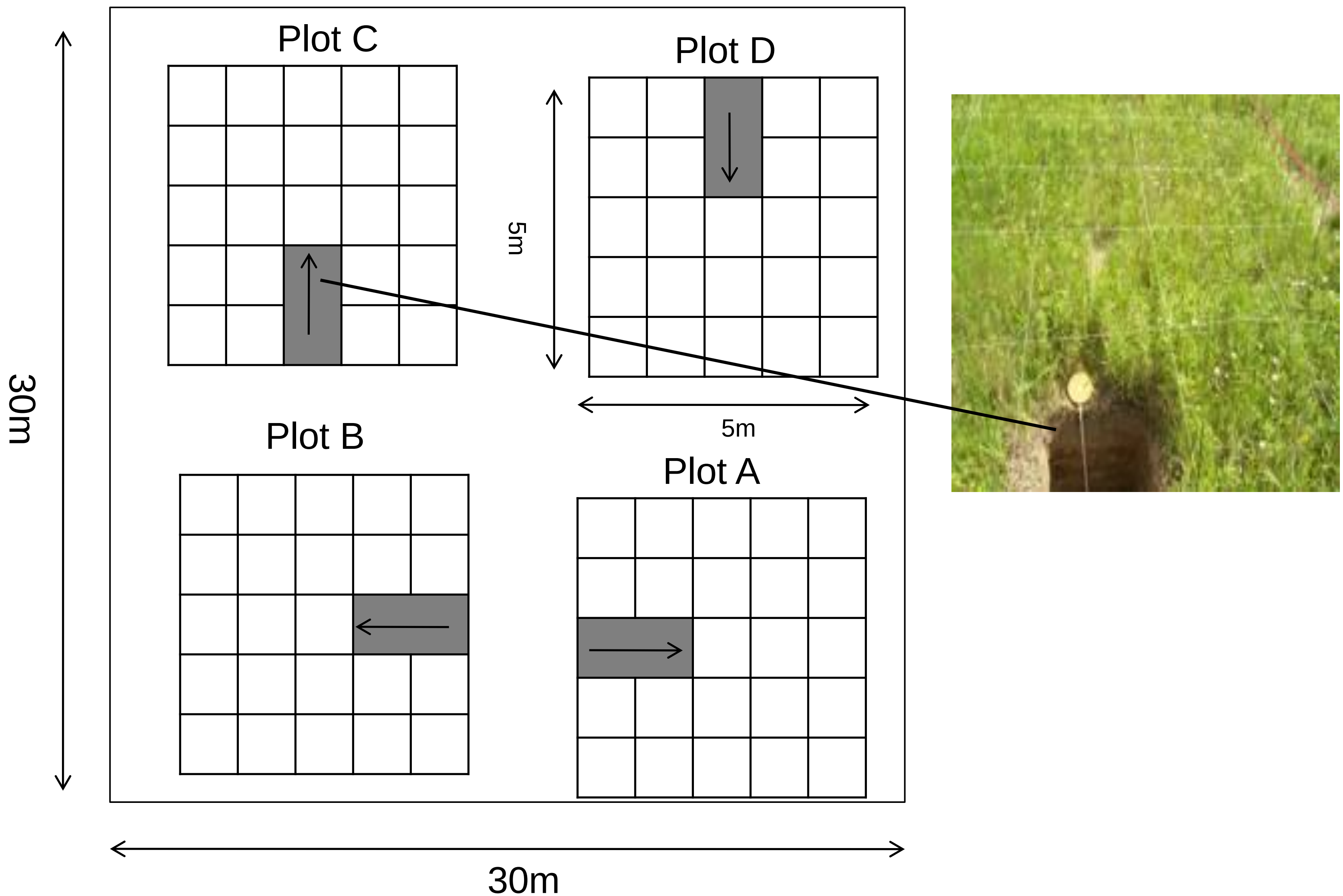
UNDERGROUND ECOLOGY -- An “eco-application” study (Koffler Scientific Reserve, Ontario, Canada)

STUDY LOCATION: Koffler Scientific Reserve

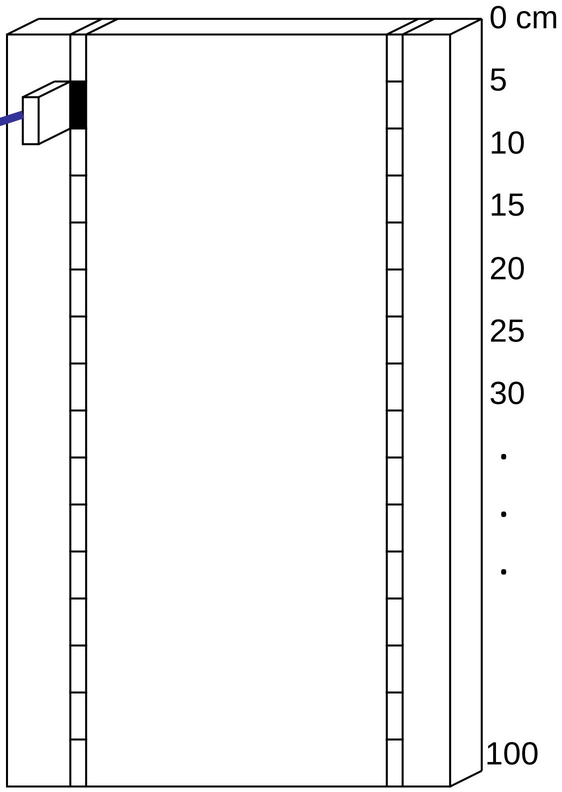


- Extensive and mature examples of several upland forest types and seepage swamp communities (Oak Ridges Moraine)
- **625 plant taxa**
- 157 bird species
- 30 mammal species

ROOT BARCODING: Sampling Design



Vertical sampling: Section of soil profile



40 soil cubes/plot (each cube is 5 cm^3 in size) -- 160 cubes total

Collection & analysis of root fragments

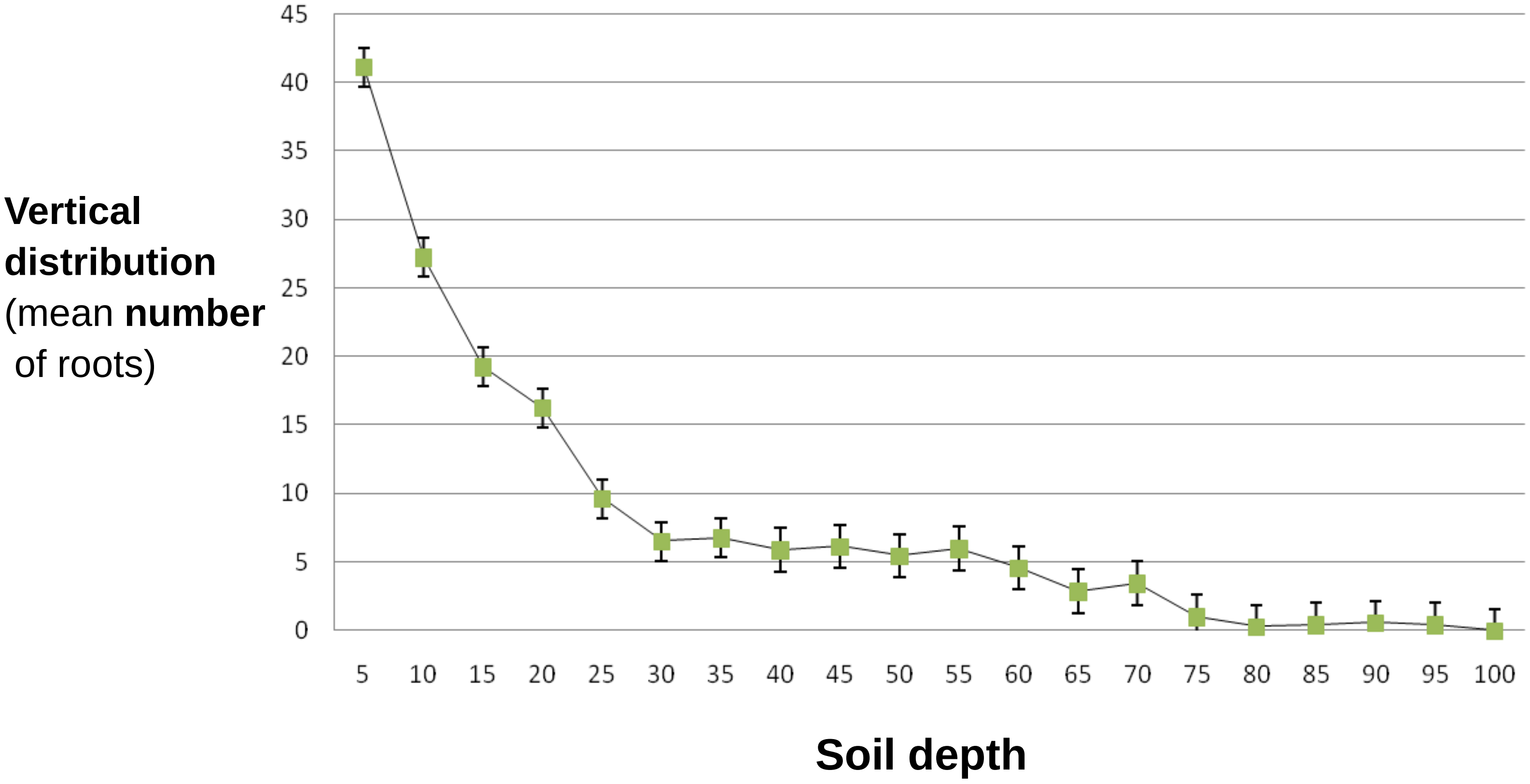
Total: ~3800

Randomly sampled: ~1503

Barcode library of KSR flora (450 spp.) used as reference database for root ID

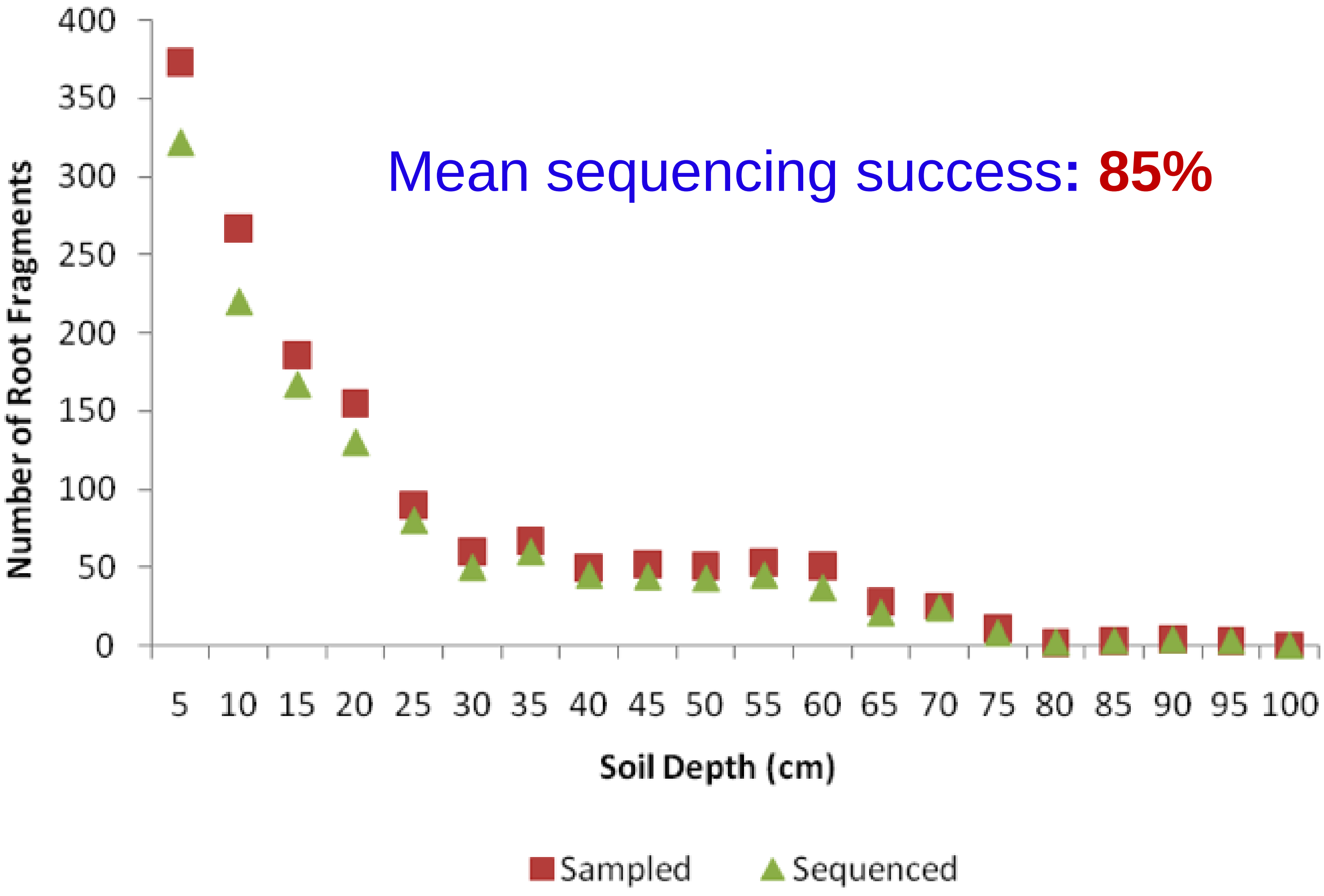
Barcode region used: *rbcL*

Abundance of root fragments

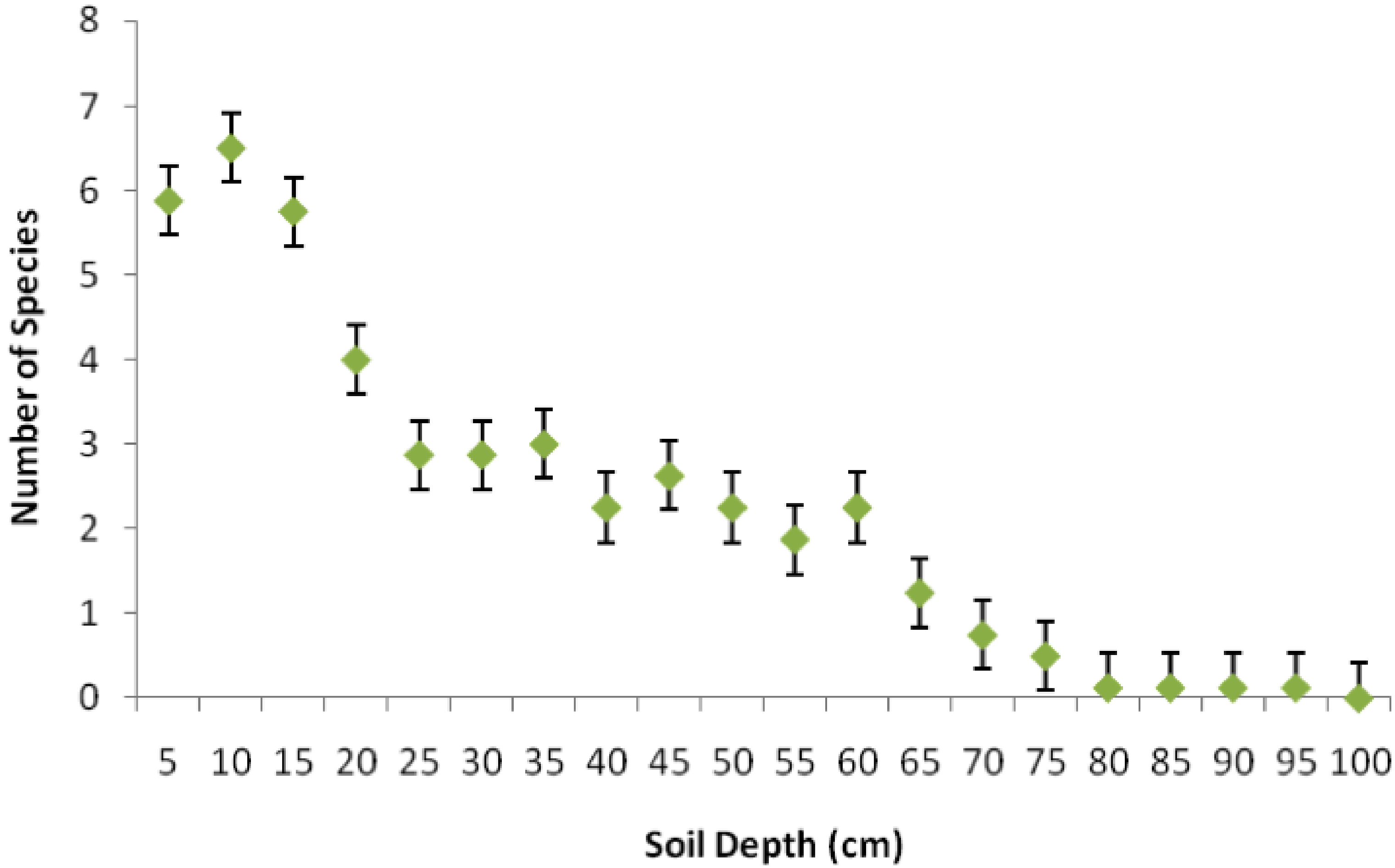


(* No significant effect of **plot** or **column** on root abundance)

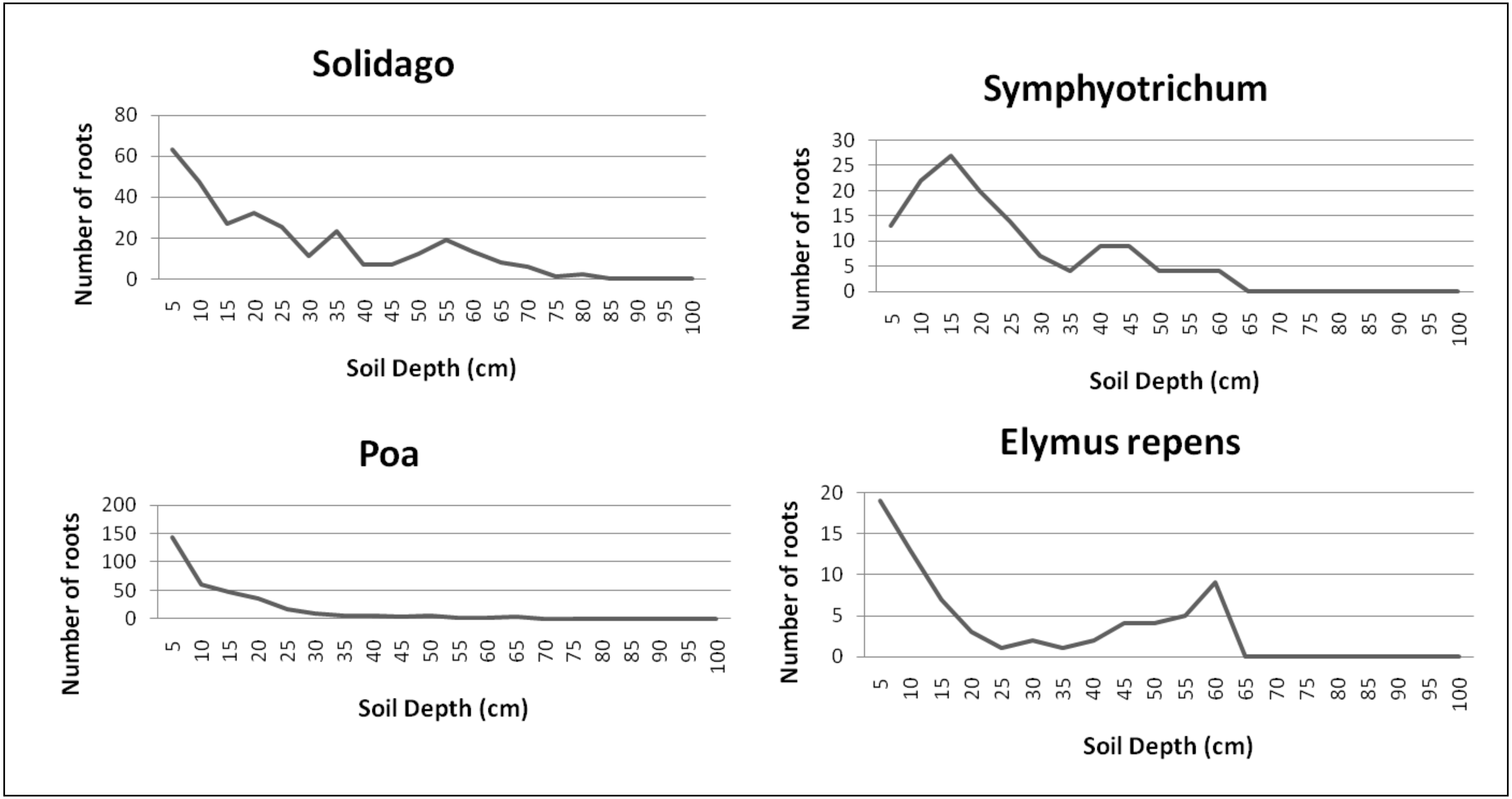
Root barcodes: Sequencing success



Root barcodes: Soil depth and species diversity



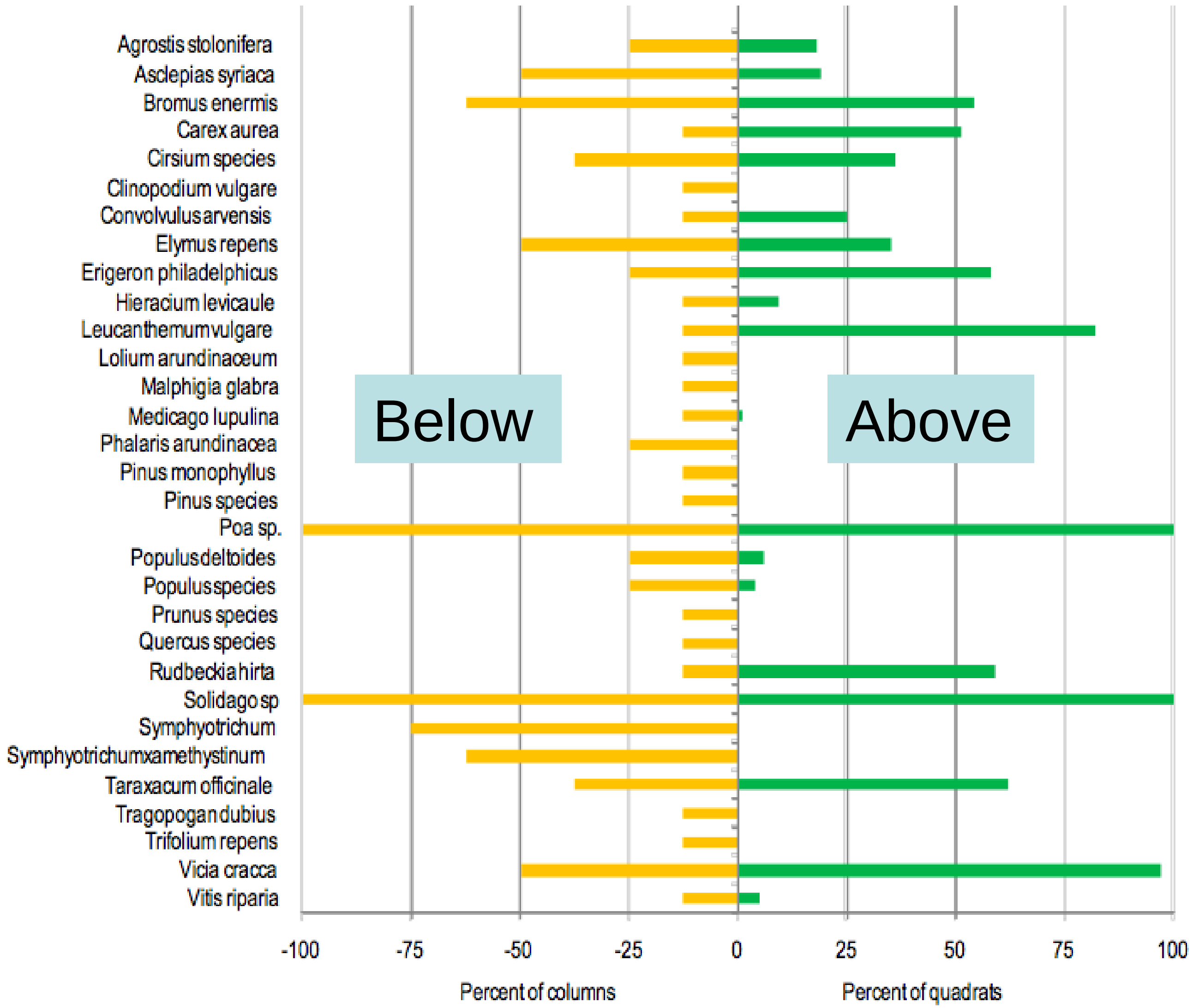
Root Barcodes: Abundance variation across species



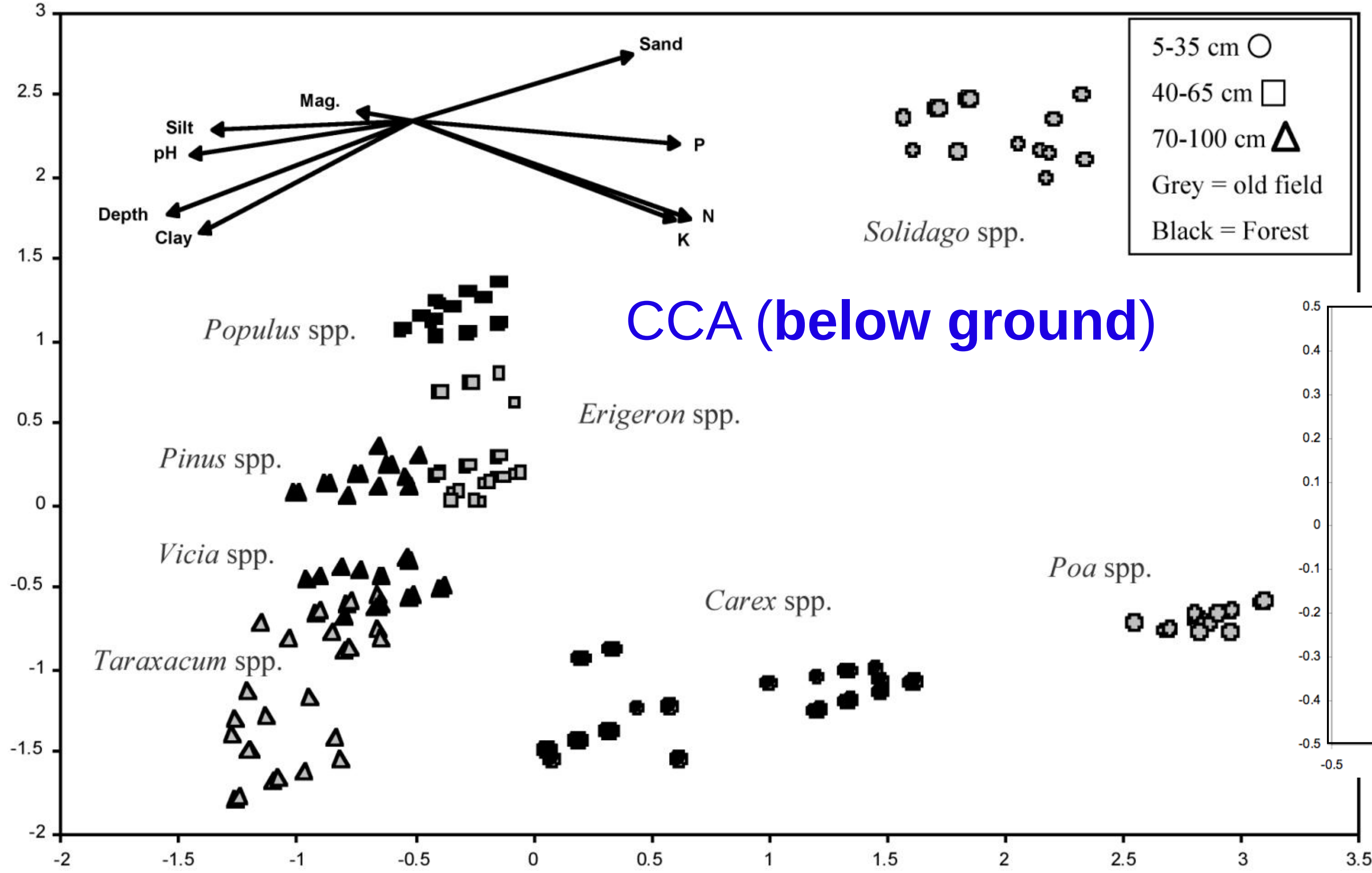
Root Barcodes: Some co-existence patterns observed

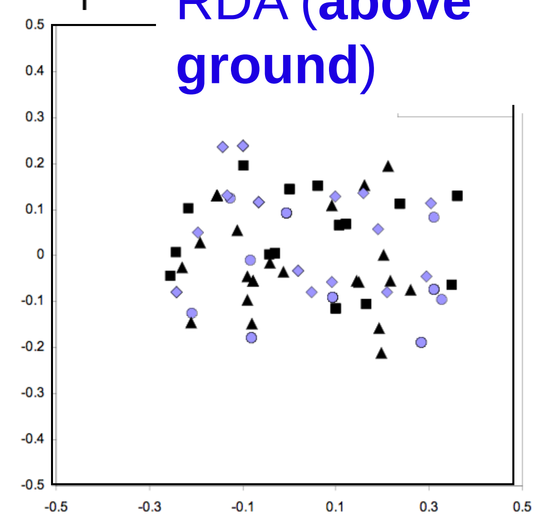
	Poa
Solidago (Asteraceae)	Chisquare : 48.337 **P<0.0002**
Symphyotrichum (Asteraceae)	Chisquare : 13.518 **P<0.0001**
Elymus repens (Poaceae)	Chisquare: 1.700 P<0.1923
Bromus enermis (Poaceae)	Chisquare: 0.585 P<0.4444

Below vs. above-ground abundance poorly correlated



Differences in **below-ground** vs. **above-ground** community Composition & structure





Summary

Plant barcoding system: *rbc*L + *mat*K (++)

Plants **are** harder to barcode, precisely, than animals

-- Causes of greater **gene-tree paraphyly**?

Plant DNA barcoding studies

-- **Regional & monographic campaigns** (eg Grasses of BC)

-- **GrassBoL campaign**

-- Era of eco-applications (e.g. Below-ground ecology)

Summary

- Barcoding can successfully be used for identifying roots
- Depth has a significant effect on root distribution and species number
- Only in few spp. there is correspondence between above and below ground abundance
- Species from two different families co-existed more often than expected
- Below ground community structure

Taxa identified below ground

Total: 28 taxa

Asclepediaceae:
Asclepias syriaca

Asteraceae:
Erigeron philadelphicus
Euthamia graminifolia
Leucanthemum vulgare
Rudbeckia hirta
Symphyo. x amythestinum
Taraxacum officinale
Tragopogon dubius

Brassicaceae:
Alliaria petiolata

Convolvulaceae:
Convolvulus arvensis

Lamiaceae:
Clinopodium vulgare

Fabaceae:
Medicago lupilina
Trifolium aureum
Vicia cracca

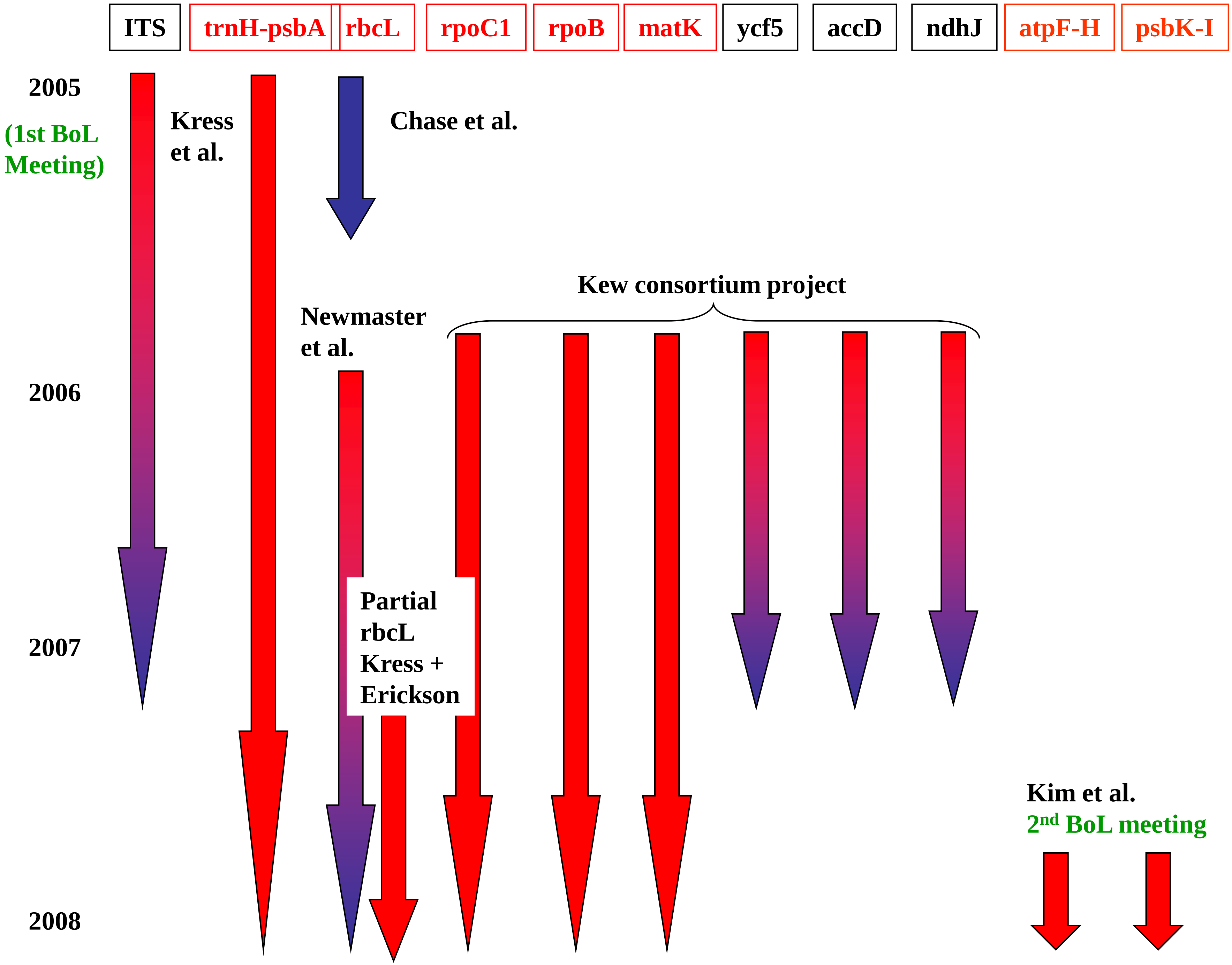
Cyperaceae:
Carex aurea

Salicaceae:
Populus deltoides

Melanthiaceae:
Trillium erectum

Poaceae:
Elymus repens
Lolium arundinaceum
Phalaris arundinacea
Bromus enermis
Agrostis stolonifera

Acer sp.
Cirsium sp.
Poa sp.
Solidago sp.
Symphyotrichum sp.


ITS
trnH-psbA
rbcL
rpoC1
rpoB
matK
ycf5
accD
ndhJ
atpF-H
psbK-I
2005
(1st BoL Meeting)
Kress et al.
Chase et al.
Kew consortium project
Newmaster et al.
2006
Partial rbcL Kress + Erickson
2007
Kim et al.
2nd BoL meeting
2008