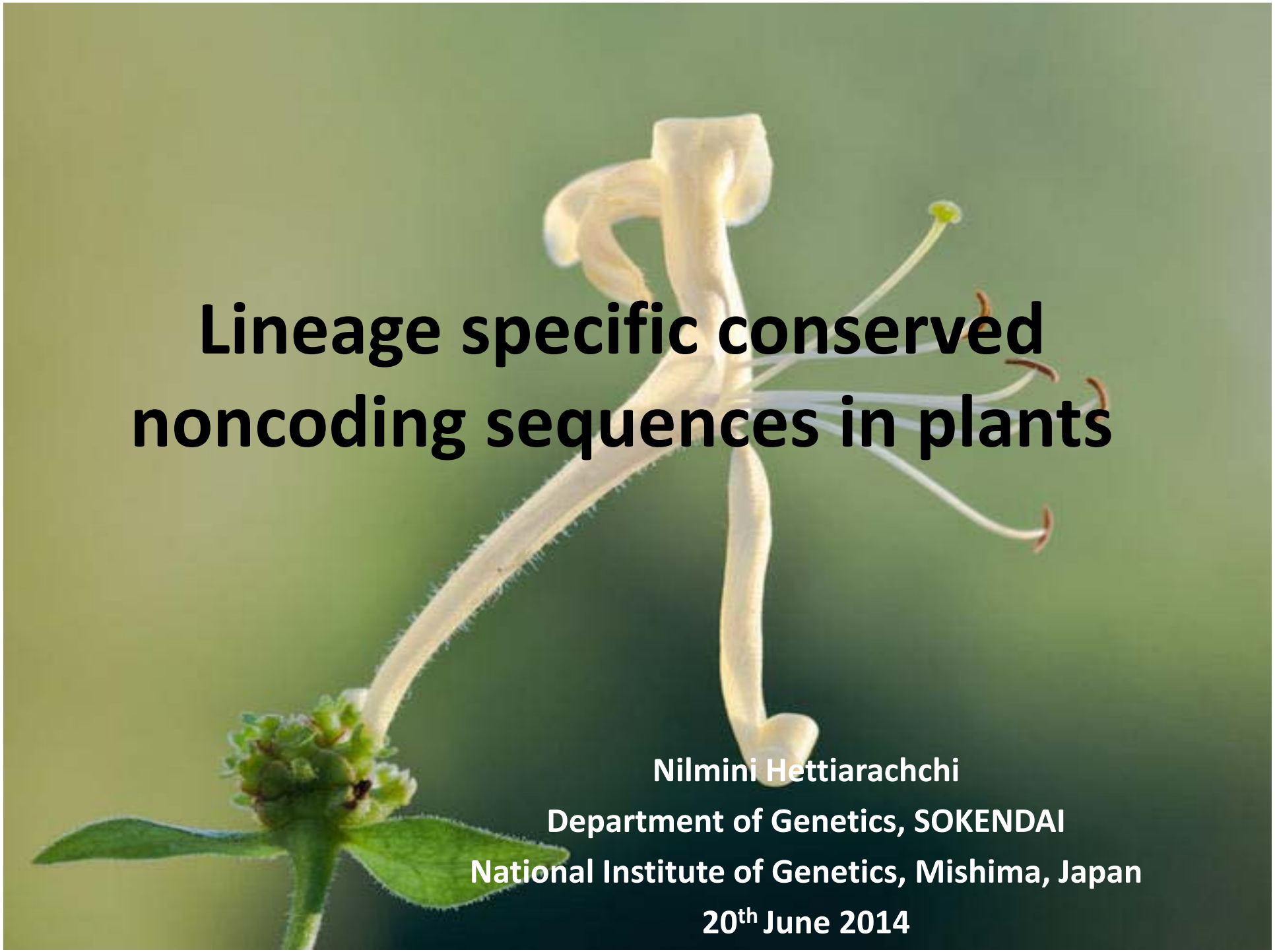
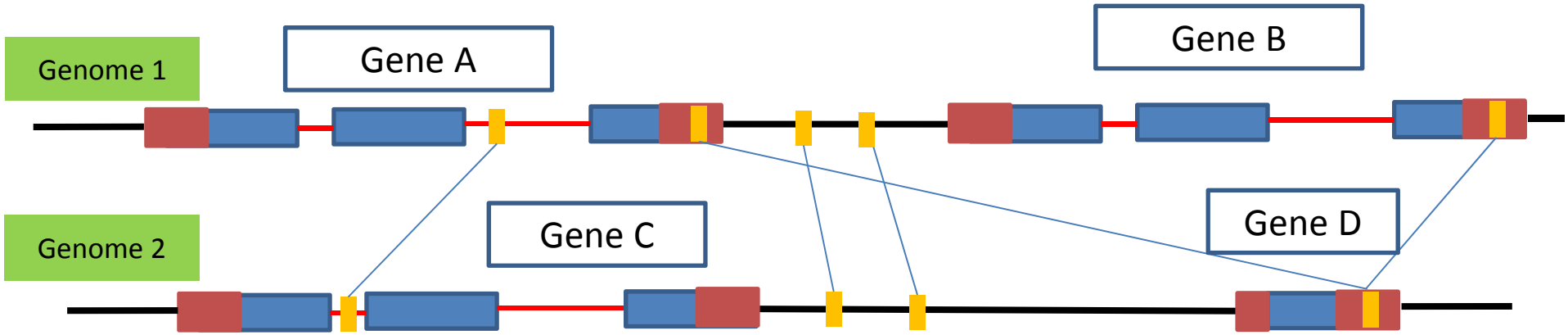


Lineage specific conserved noncoding sequences in plants

Nilmini Hettiarachchi
Department of Genetics, SOKENDAI
National Institute of Genetics, Mishima, Japan

20th June 2014

Conserved Noncoding Sequences (CNSs)



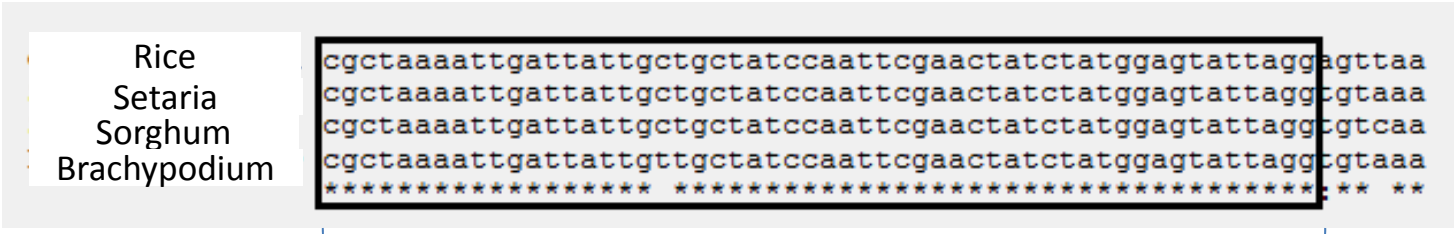
Protein coding region

UTR

Intergenic

Intronic

CNSs



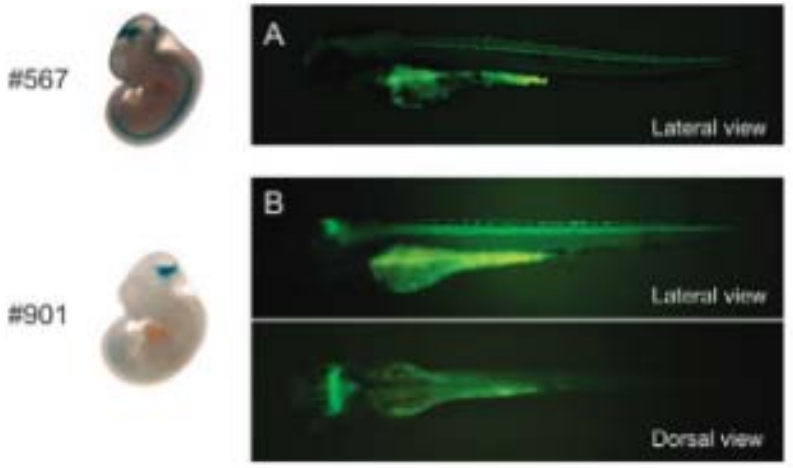
Significance of CNS Studies in Animals

- Functional constraint ➡ **Regulatory function**

- CNSs found near genes involved in regulation of **transcription and development** (Sandelin et al. 2004; Shin et al. 2005; Venkatesh et al. 2006)

- Many CNSs function as **cis-regulatory elements or enhancers** – associated with tissue specific expression during development (Woolfe et al. 2005; Pennachio et al. 2006)

- Ancient vertebrate CNEs overlap with **functionally verified human enhancers** – direct tissue specific expression during embryonic development (Lee et al. 2011)



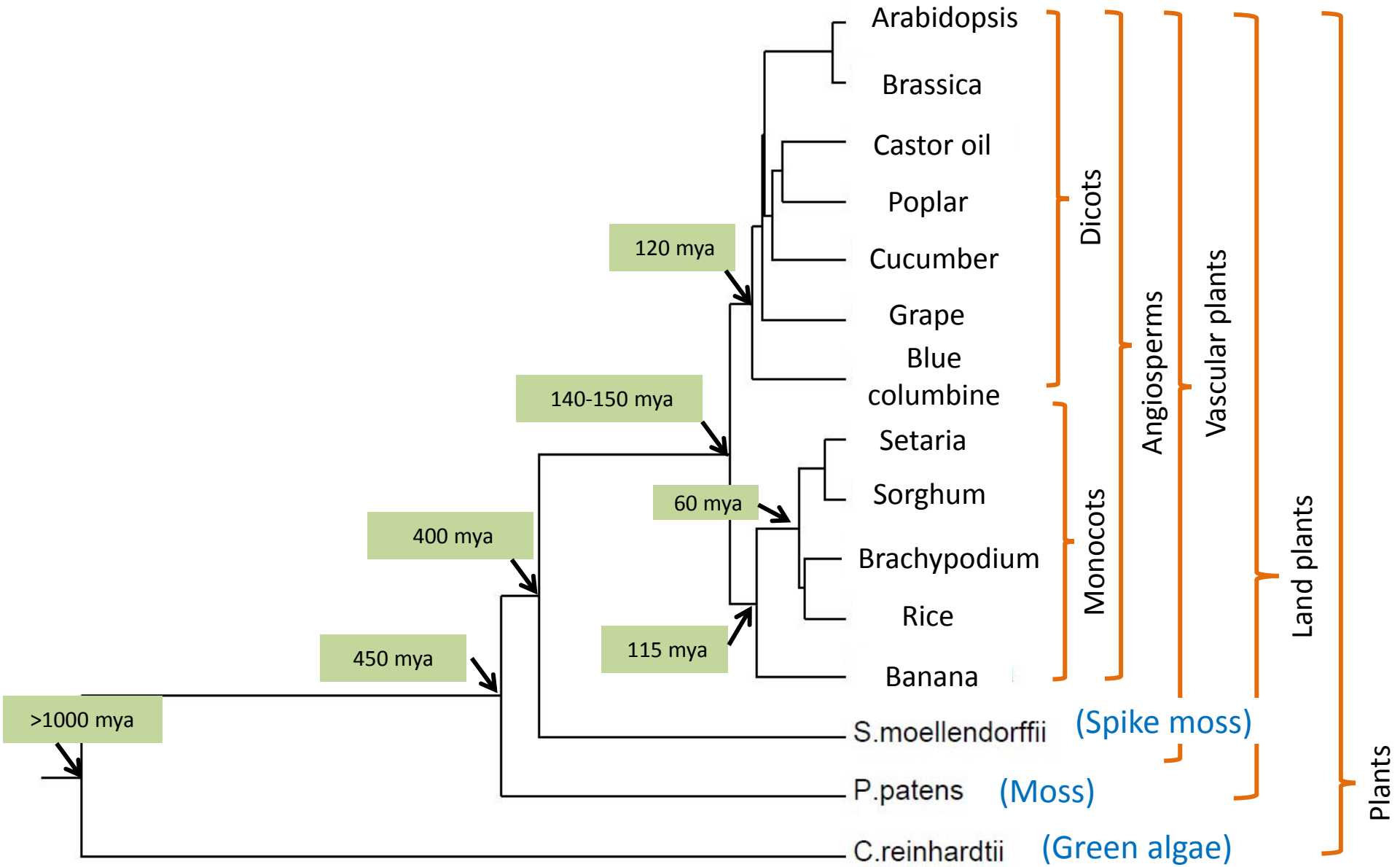
(Lee et al. 2011)

Objectives of the Study

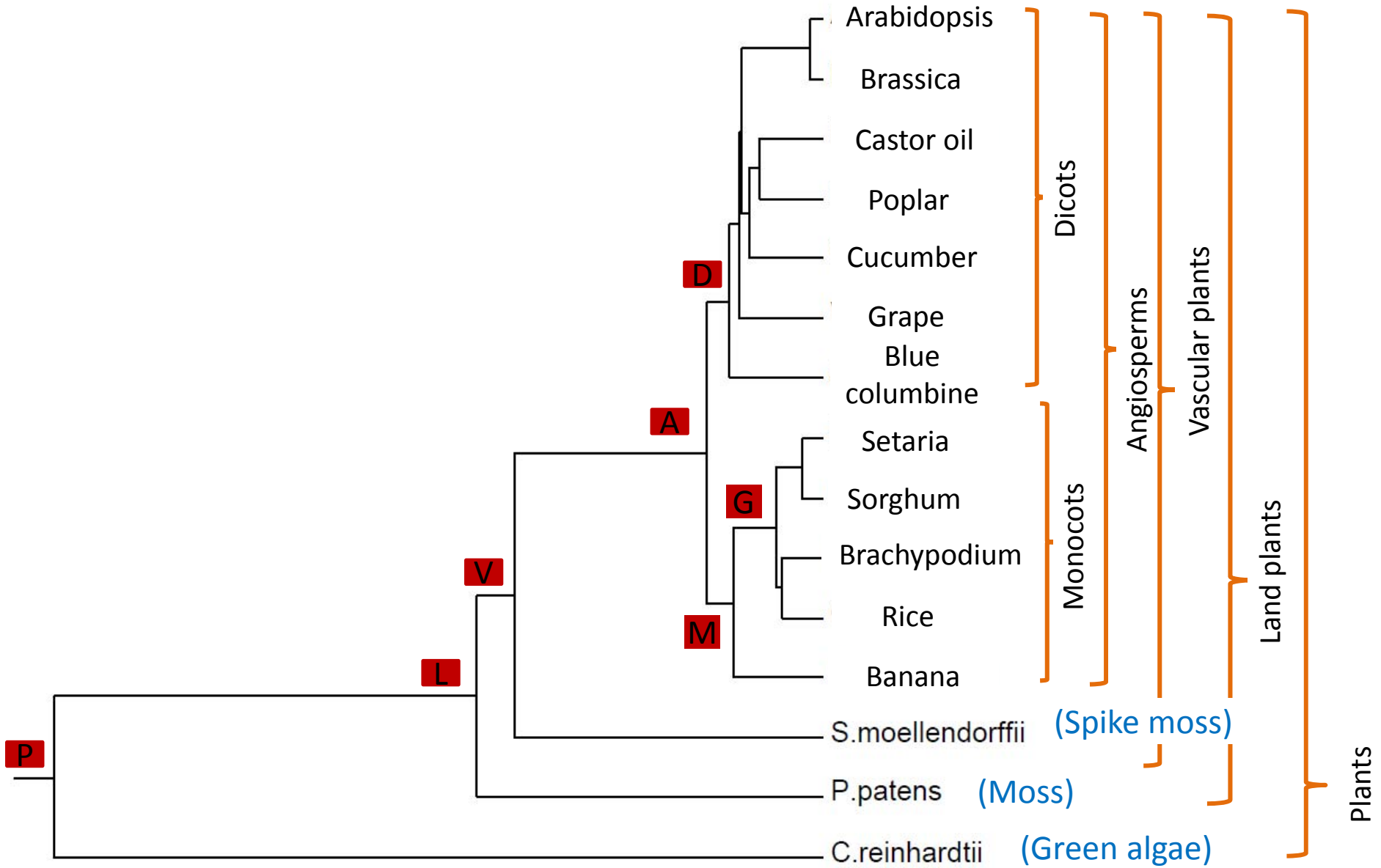
- Identify CNSs specific to a lineage of organisms
- Features of CNSs in lineages
- Identify how CNSs differ between groups
- Find likely target genes for the CNSs
- Functional classification of predicted target genes

Genomes Used in the Analysis

- **Whole genome information – 15 species**



Lineage specific CNSs

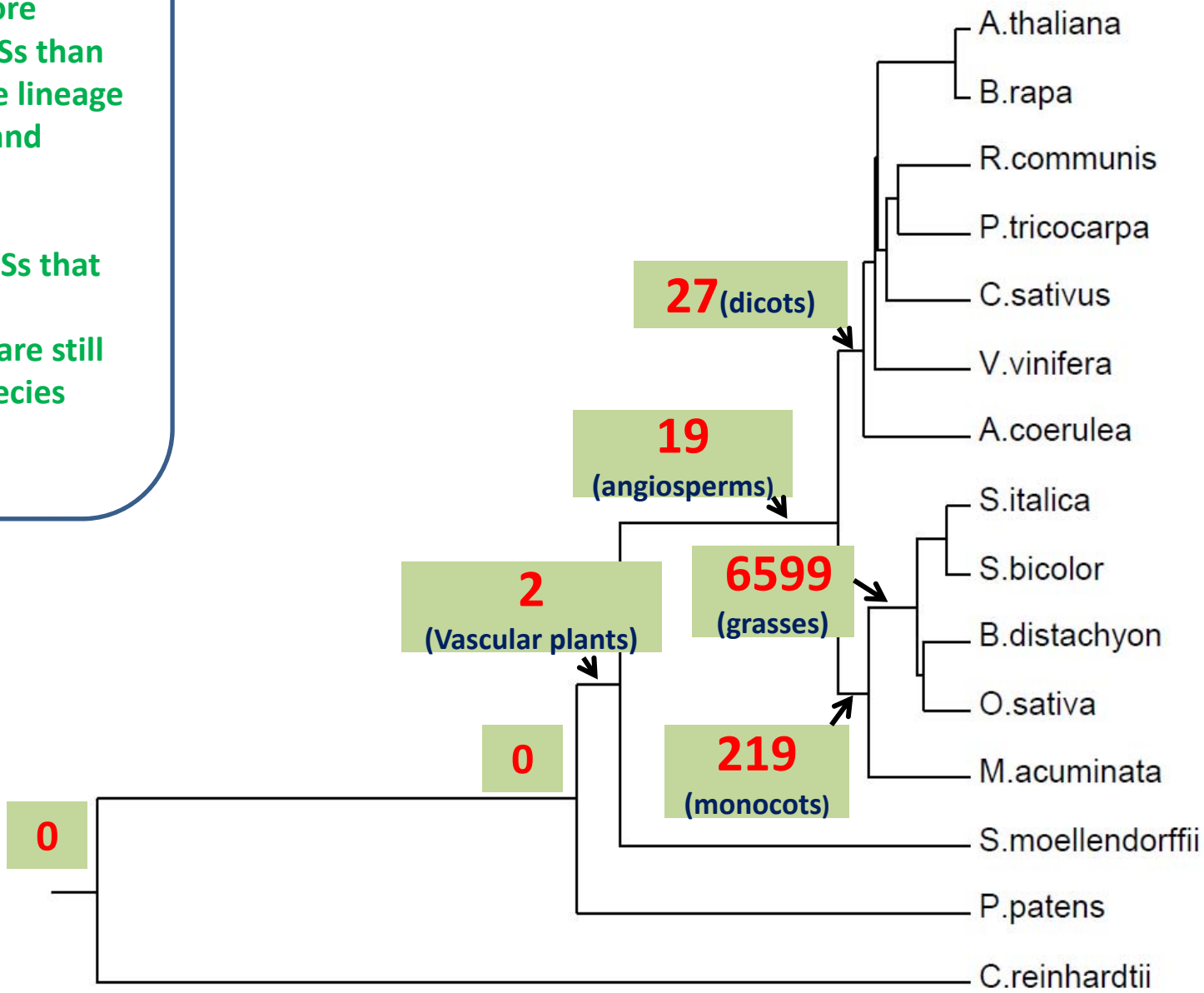


- Many grass specific CNSs – originated in the grass common ancestor
- Monocots have more lineage specific CNSs than dicots – may define lineage specific functions and morphology
- Only few of the CNSs that arose in land plant common ancestor are still kept in all plant species

Results

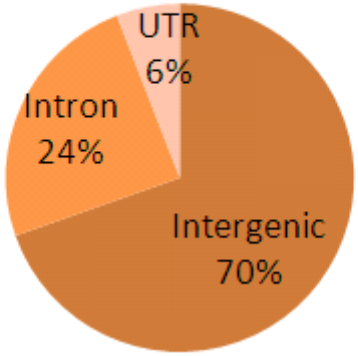
Lineage specific CNSs

Numbers of Lineage specific CNSs in red color

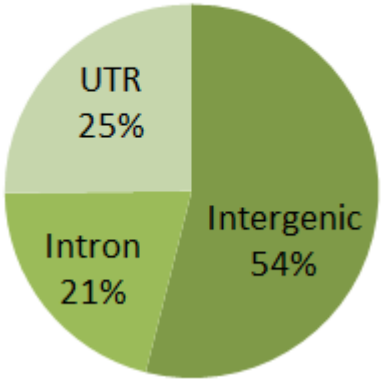


Genomic locations of the lineage specific CNSs

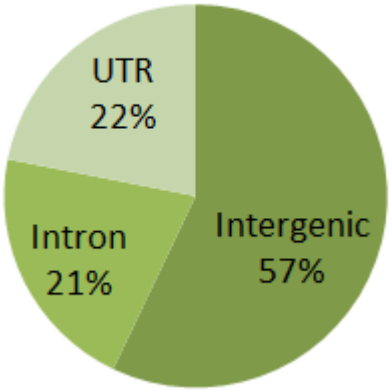
Noncoding genome composition
Oryza sativa japonica



Grass specific



Monocot specific



Functional classification of likely target genes

Grass

Functional group	P-value
Functions related to nucleus	0.0E0
Regulation of transcription	0.0E0
DNA-binding	9.6E-309
Transcription	4.5E-278
Transcription regulator activity	5.7E-272
Transcription factor activity	8.6E-269
Regulation of RNA metabolic processes	4.6E-171
Zinc-finger related	3.4E-106
Activator	1.3E-86

Monocots

Functional group	P-value
Transcription factor activity	2.4E-51
Transcription regulator activity	5.8E-48
Regulation of transcription	1.3E-46
Functions related to nucleus	5.8E-45
DNA binding	1.2E-43
Sequence specific DNA binding	1.8E-17
Zinc-finger related	3.0E-12
Homeodomain related	3.6E-8
Basic-leucine zipper transcription factor	1.3E-7

Underrepresented GO terms

- **Enzymatic activity**
- **Plant defense**

- The conservation level of the target genes
  - Eudicot
  - Monocot
  - Grass [statistically significant (t-test) compared to random samples]

Methylation level of CNSs

- DNA methylation is involved in epigenetic regulation (affects gene expression, genomic imprinting, transposon silencing, timing of replication etc…)

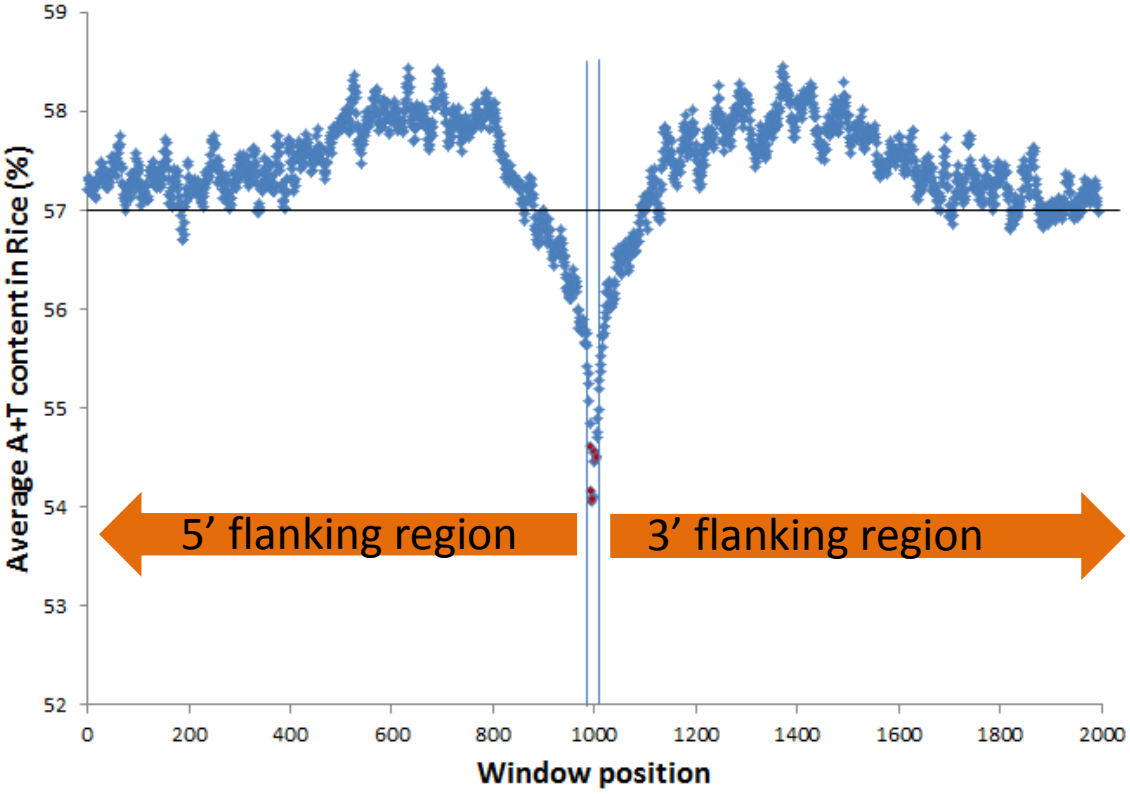
- Used the single base pair resolution DNA methylation map (Cokus et al. 2008)

- **25** CNSs do not show methylation marks

- **2** CNSs show methylation in CG,CHG or CHH context

- Dicot specific CNSs are **not predominantly modified by DNA methylation**

CNSs are flanked by a sharp drop of A+T content



Average A+T content in the rice genome

A+T content within rice CNSs

Grass/Monocot – specific CNSs

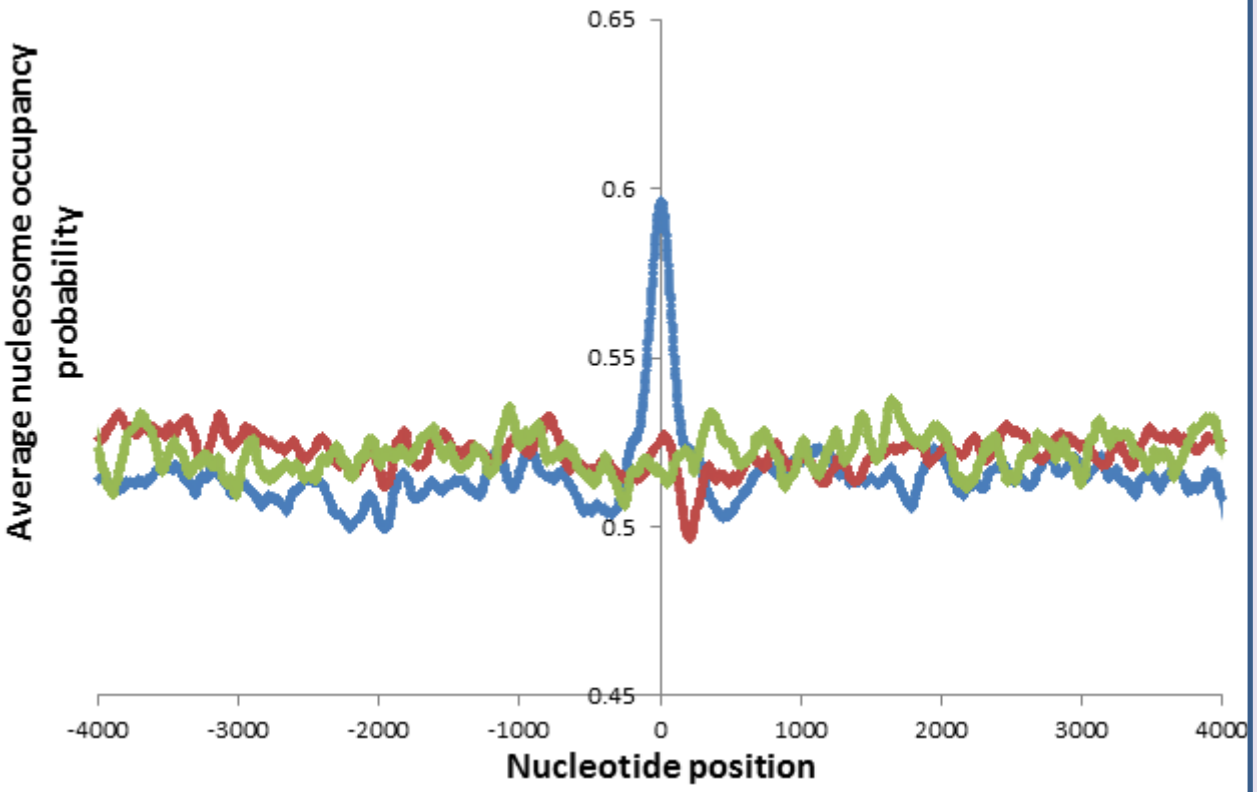
- **Compare base composition at the boundaries (not conserved) of CNSs and within CNSs**
- **1000 base 5’ – 3’ flanking regions**
- **Sharp drop in A+T frequency starting just outside the borders of the CNS**
- **Statistically significant difference between (T-test 95% confidence level – $p < 0.05$) the A+T content in the flanking regions, and the A+T content within CNSs**
- **Dicot specific CNSs (T-test 95% confidence level – $p < 0.05$)**

Explanations for the A+T drop of the flanking regions

- Nucleosome positioning
  - A+T rich sequences ➡ low propensity to form nucleosomes (Jansen and Verstrepen 2011)

- These regions may have a higher recombination rate (Recombination hotspots)
  - Recombination hotspots are associated with increased GC content (Spencer et al. 2006)

Prediction of nucleosome positioning



CNSs and the flanking regions

Random sample with same AT content

Random sample without AT preference

Grass/Monocot – specific CNSs

- Considered sequences of 8000 bases with CNS positioned centrally
- Calculated the average nucleosome occupancy probability for each site
- A clear peak in the predicted nucleosome occupancy probability, which directly coincides with the CNS region
- **These CNS regions tend to have a well-positioned nucleosome in them**

Do these regions have a high recombination rate?

- Determine if CNSs overlap with recombination hotspots?
  - None of the dicot specific CNSs overlap with recombination hotspots (Horton et al.2012)
  - AT drop cannot be due to recombination hotspots

Lineage specific genes and lineage specific CNSs

	Lineage specific genes	Lineage specific CNSs
Eudicots	2439	27
Grasses	444	6599
Monocots	113	219

- Lineage specific genes and CNSs follow opposite patterns
- Lineage specific genes are predominantly plant defense related

Summary

- Large number of specific CNSs originated in the grass common ancestor
- Stronger constraint on CNSs located in UTR
- CNS - Likely target genes regulating transcription
- Dicot specific CNSs are not predominantly modified by DNA methylation
- CNSs are flanked by a sharp drop of A+T content
- CNSs tend to have a well positioned nucleosome in them
- The found CNSs may imply regulatory function – experimental verification is required to clarify the function

Thank you !