

Progress in orthologous candidate gene diversity assessment



Dominique This, Agropolis Montpellier (France)
and all partners of the ADOOC project



Why study the diversity of candidate genes within / across GCP species?

- Apply results coming from model crops to useful GCP material
 - Rice, barley, sorghum, chickpea, bean, potato, cassava
- Evolutionary /diversity pattern of a gene can be helpful in functional prediction:
 - Rates of non-synonymous vs synomymous changes
 - Signs of selection compared / neutral markers
 - Orthology / paralogy relationships across species
- Allelic diversity of candidate genes facilitates association mapping with drought tolerance traits and finding the best alleles. It requires:
 - A relevant population structure
 - Comparison with genome-wide neutral markers

But: candidate genes are most often gene families

- From small to large families

ERECTA: 1 to 3 (ER and ER-L genes)

ASR: 0 to 7

MIP (aquaporins): 38 in *Arabidopsis*

- Complexity of gene family evolution

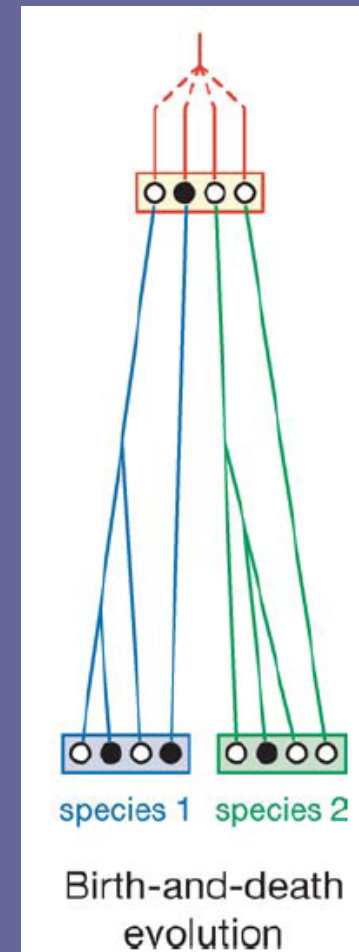
- Redundancy / subfunctionalisation / neofunctionalisation / pseudogene formation

Exple: FEH / cw invertase

Exple: CBF1_3 Vs CBF4

- Consider the whole gene families when possible

- DREB2A: find orthologs



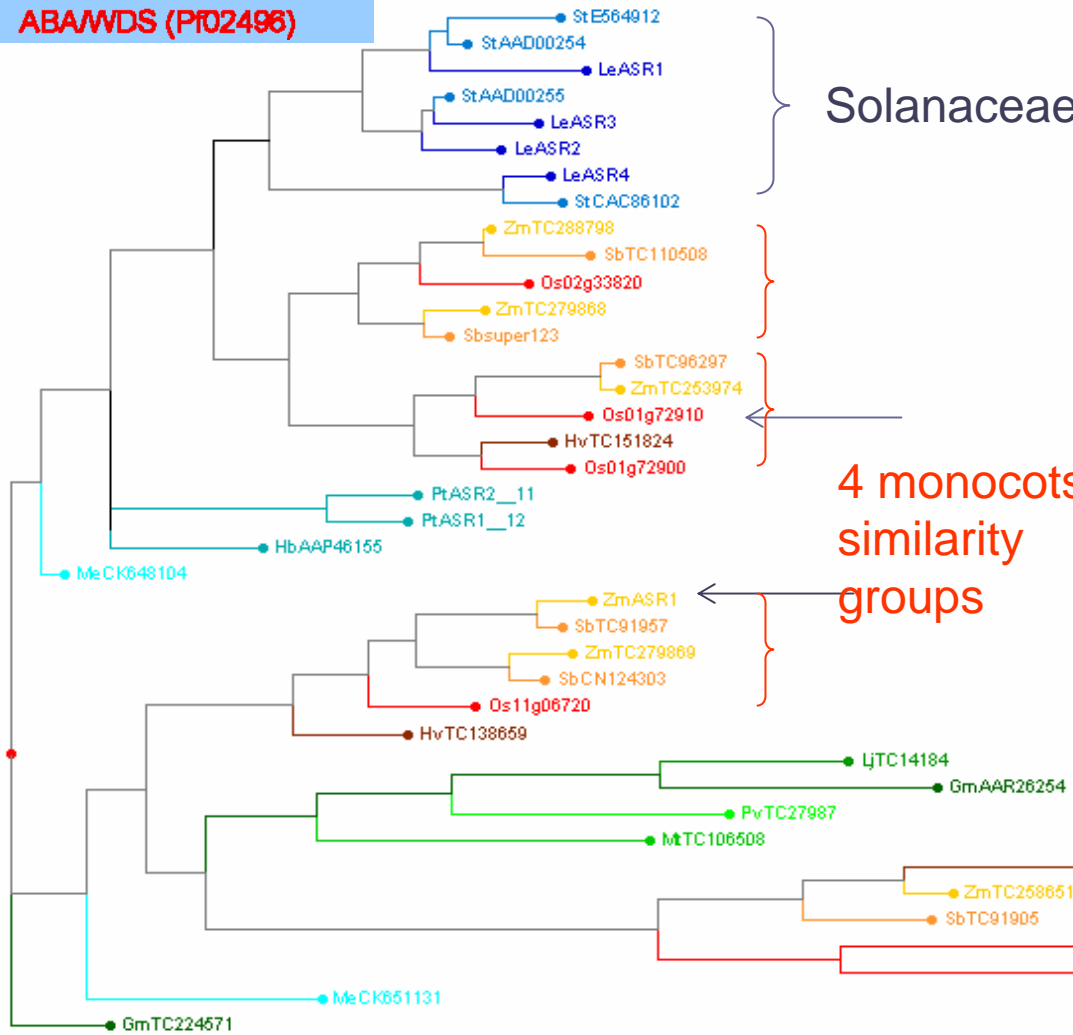
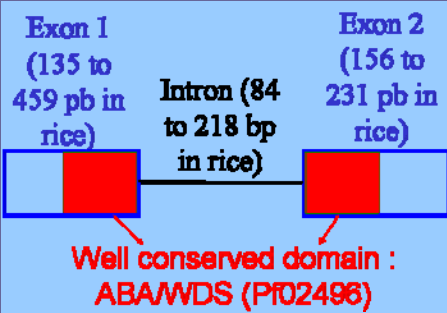
Candidate gene families

- **Extract information from databases**
 - ERECTA, ASR, DREB, VIN, SuSy, SPS (molecular signatures)
- **Get new sequences or complete partial ones**
 - Consensus primer design, PCR and band picking / cloning, sequencing
- **Build phylogenetic trees**
 - Try to identify orthologous sequences
 - Reconciliation of species tree with gene tree (DREB2A)

Assessment of candidate genes diversity in some GCP crops

- Design of specific primers for each target gene
- Initial check on 7 to 8 accessions
- Allelic sequencing on the ADOC sets (~300 accessions, including some related wild species)

ASR phylogeny



taxon	species	Sequences identified
POACEAE	<i>Oryza sativa</i>	6
	<i>Sorghum Bicolor</i>	6
	<i>Hordeum vulgare</i>	3
	<i>Zea mays</i>	6
BRASSICACEAE	<i>Arabidopsis thaliana</i>	0
FABACEAE	<i>Phaseolus vulgare</i>	1+1
	<i>Cicer arietinum</i>	0
	<i>Glycine max</i>	2
	<i>Medicago truncatula</i>	1
	<i>Lotus japonica</i>	1
SOLANACEAE	<i>Solanum tuberosum</i>	3
	<i>Lycopersicon esculentum</i>	4
EUPHORBIACEAE	<i>Manihot esculenta</i>	2
SALICACEAE	<i>Hevea brasiliensis</i>	1
	<i>Populus trichocarpa</i>	3



Initial diversity assessment / ASR

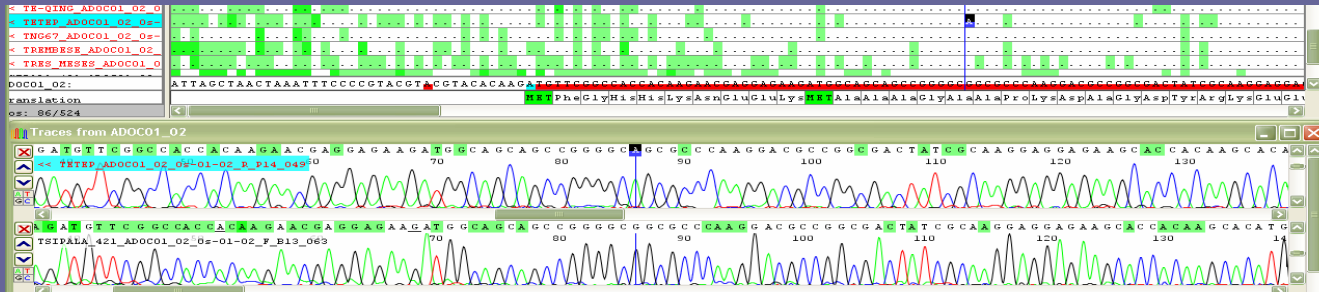
Species	ADOC Name	Identification	Fragment size	No of polymorphic sites	ns /s changes
Rice	ADOC_01_01_Osa	Os01g72900	560	3	0/1
	ADOC_01_02_Osa	Os01g72910	621	6	2/2 + 1aa
	ADOC_01_03_Osa	Os02g33820	1028	11	0/0
	ADOC_01_04_Osa	Os04g34600		na	na
	ADOC_01_05_Osa	Os11g06720	862	3	1/1
	ADOC_01_06_Osa	Os01g73250	722	15	0/0
Barley	ADOC_01_01_Hvu	TC138659	467	4	0/1
	ADOC_01_02_Hvu	TC151824	518	64	9/14
	ADOC_01_03_Hvu	TC140709	865	4	3/0
Sorghum	ADOC_01_01_Sbi	CN124303	1251	8	0/0
	ADOC_01_02_Sbi	TC91905	788	2	1/1
	ADOC_01_03_Sbi	TC96297	797	22	2/2
	ADOC_01_04_Sbi	TC110508	448	4	1/1
	ADOC_01_05_Sbi	TC91957		na	na
	ADOC_01_06_Sbi	super123bis	611	8	0/2
Cassava	ADOC_01_01_Me	CK64810	488	10	3/4
Bean	ADOC_01_01_Pvu	TC27987		na	na
	ADOC_01_02_Pvu	New		na	na
Potato	ADOC_01_01_Stu	U76610	1309	14	1/1
	ADOC_01_02_Stu	U76611	540	24	0/7
	ADOC_01_03_Stu	AJ320154	1307	37	5/7

Maize ASR1 orthologs more conserved



One branch of the tree more polymorph

Allelic sequencing / ASR

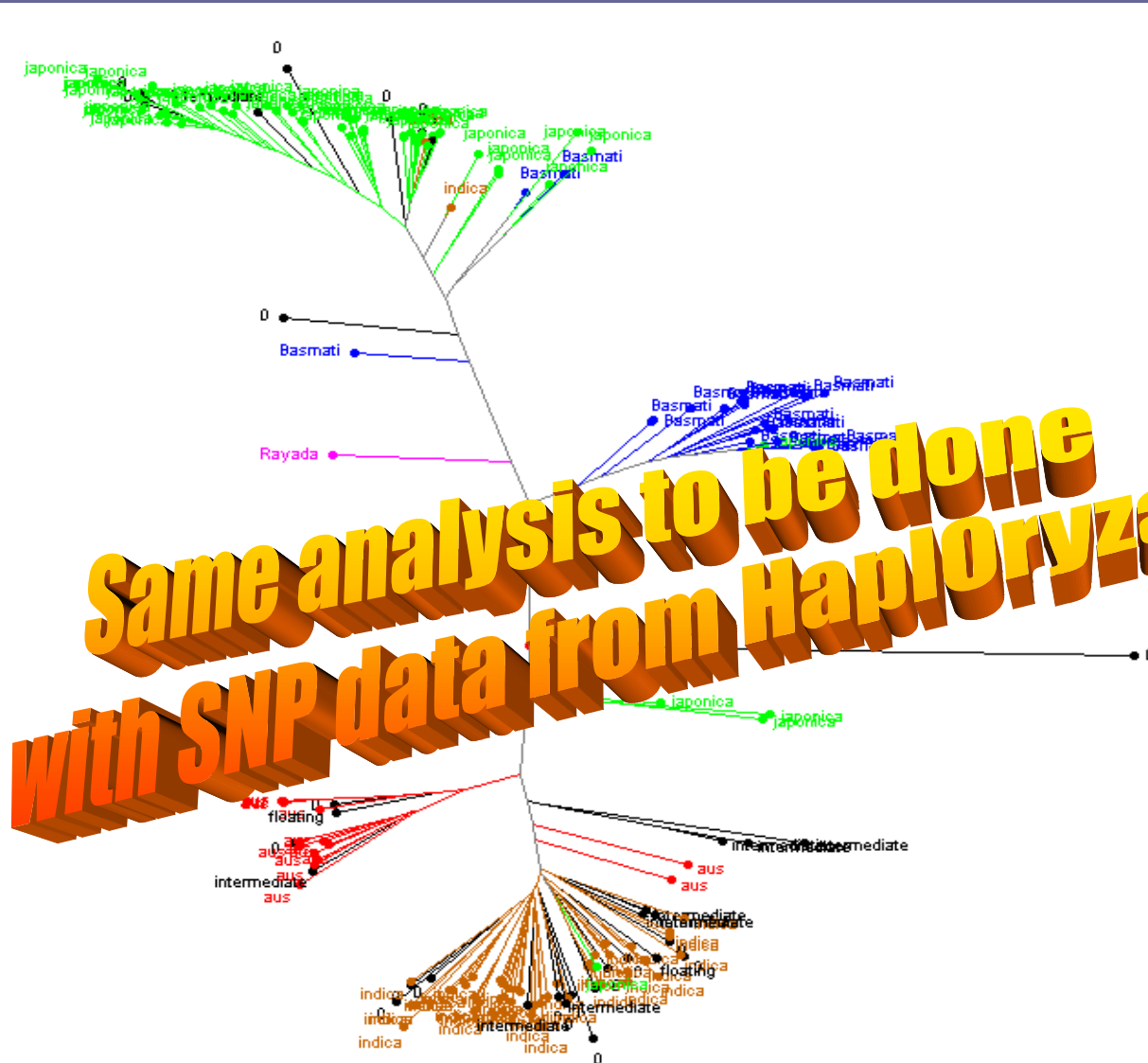


	Nb sequences	length	Nb polymorphic sites	Nb Haplotypes
ADOC01_02_Os	254	524	6	7
ADOC01_03_Os	145 / no MITE	755	13	2
	129 / MITE (1)	322	1	2
	129 / MITE (2)	204	11	8
ADOC01_05_Os	247	573	4	5
ADOC01_01_Hwu	212	373	4 (1 SSR in intron)	3
ADOC01_02_Hwu	194	419	62 (1 SSR in intron)	23
	(1): not considering the MITE insertion. (2) MITE only			

Comparison SSR vs ASR / rice

Different colors represent different isoenzymatic groups

Different colors represent different Groups of haplotypes



175 samples, 1536 SNPs (HaplOryza), simple matching, neighbor joining

A few highlights on other candidate gene families

- **ERECTA and ERECTA-like genes**

- Both groups clearly identified
- From 16 to 86 polymorphic sites
- More haplotypes in barley (many singletons)

- **SuSy and SPS genes**

- Similarity of *SuSy* and *SPS* genes in rice and chickpea

- **ERECTA-like domain of Os01g69030 confirmed**

- **ERECTA-like domain**

- Very few polymorphism
- A promoter sequence included for chickpea

- **Vacuolar invertases**

- No polymorphism in rice, but variation in barley

See poster 1.11 for more information

Link with other projects

- **Genotyping projects (SP1)**
 - SSR or SNP markers (GCP composite, HapOryza)
 - Comparative mapping projects?
- **Mutant / physiological analysis (SP2)**
 - « rice mutants » (Andy Pereira)
 - Research works undertaken by gene specialists
- **Large scale phenotyping projects**
 - Drought tolerance traits?
- **Candidate gene / Association mapping projects in other crops**
 - maize cf M. Warburton
- **SP4 projects**
 - Stress gene catalog / orthology finding cf S. Wanchana

Product delivery and impact on users

- **A list of consensus primers**
 - Identify candidate gene families in other crop species
- **A list of specific primers**
 - Extend the diversity analysis in your favourite germplasm
 - Map our candidate genes
- **Allelic sequence diversity for some genes**
 - Use SNP (SSR) information for association mapping
 - Define specific haplotypes for functional analysis

Special thanks

- Romain Philippe / Brigitte Courtois / Claire Billot (CIRAD)
- Roland Schafleitner (CIP)
- Spurthi Nayak / Rajeev Varshney (ICRISAT)
- Pierre Mournet / Dominique Brunel (CNG)
- Millicent Sanciangco / Ken Mc Nally (IRRI)

The whole ADOC community for their input

The organizers for the invitation

And the audience for listening...