IWGSC WEBINAR

Finding a name for your new baby wheat gene

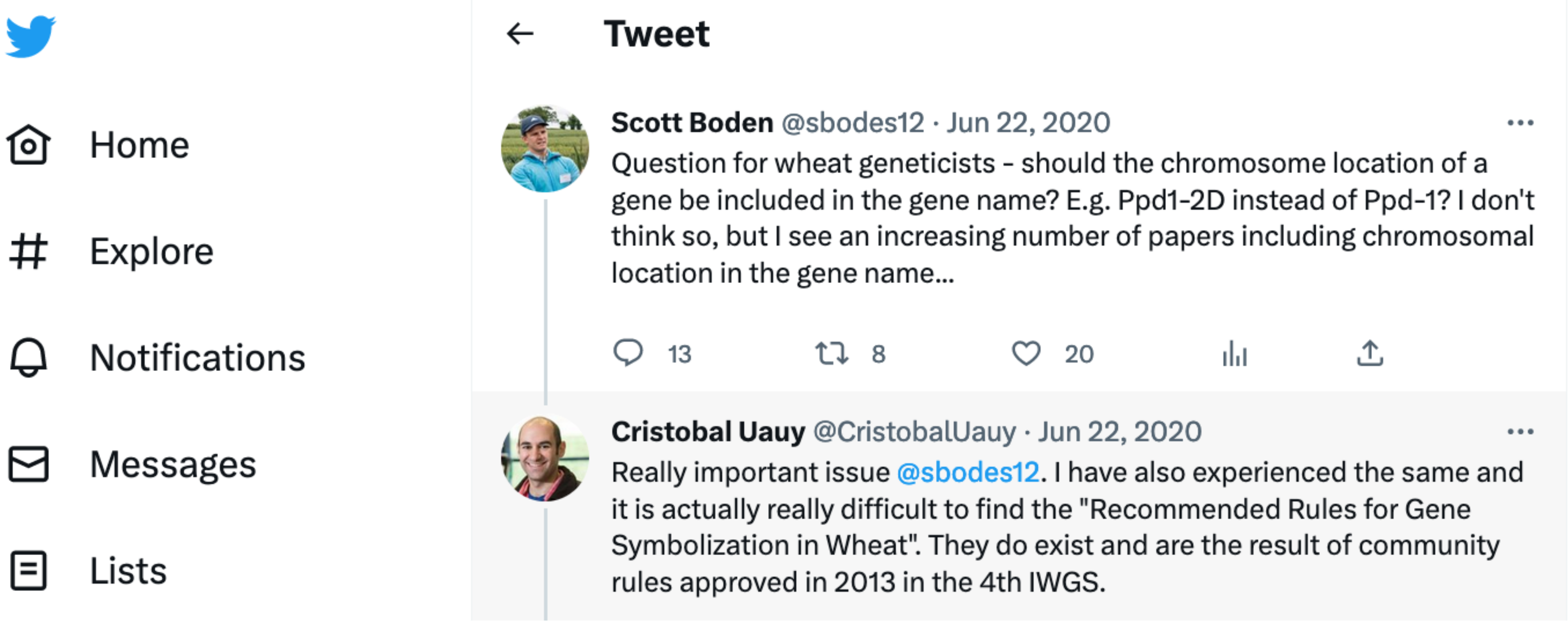
Wednesday 7 June
9:00am EDT | 10:30 pm ACST

Scott Boden
Group Leader and ARC Future Fellow
University of Adelaide
Australia

An international multi-discipline initiative

S. A. Boden[1] · R. A. McIntosh[2] · C. Uauy[3] · S. G. Krattinger[4,28] · J. Dubcovsky[5,28] · W. J. Rogers[6,28] · X. C. Xia[7] · E. D. Badaeva[8] · A. R. Bentley[9,28] · G. Brown-Guedira[10,28] · M. Caccamo[11,28] · L. Cattivelli[12,28] · P. Chhuneja[13] · J. Cockram[11,28] · B. Contreras-Moreira[14] · S. Dreisigacker[9,28] · D. Edwards[15,28] · F. G. González[16,28] · C. Guzmán[17,28] · T. M. Ikeda[18,28] · I. Karsai[19,28] · S. Nasuda[20] · C. Pozniak[21,28] · R. Prins[22,23] · T. Z. Sen[24,28] · P. Silva[25] · H. Simkova[26] · Y. Zhang[27] · the Wheat Initiative[28]

Why update the guidelines?



Scott Boden @sbodes12 · Jun 22, 2020

Question for wheat geneticists - should the chromosome location of a gene be included in the gene name? E.g. Ppd1-2D instead of Ppd-1? I don't think so, but I see an increasing number of papers including chromosomal location in the gene name...

Cristobal Uauy @CristobalUauy · Jun 22, 2020

Really important issue @sbodes12. I have also experienced the same and it is actually really difficult to find the "Recommended Rules for Gene Symbolization in Wheat". They do exist and are the result of community rules approved in 2013 in the 4th IWGS.

An example: one gene, many names

VRN-A1

MADS14

?

VRN-5A or
VRN1-5A

AP1-A1/
AP1-5A

Why update the guidelines?

- Based on the **wheat gene catalogue**, curated by Bob McIntosh, forms the basis of the updated version.
- Hard to find and lacked information: genome sequence data, technical advances (e.g., CRISPR, markers for SNPs), and the development of new resources (e.g., TILLING populations).
- Expansion of wheat research community: more genes being discovered, more researchers investigating wheat (knowledge transfer from other crops), potential to network with other crops and plants.

What do we hope comes from this update?

- Improved clarity on how to name genes in wheat – an agreed model for all to follow.
- Reduced complexity in the literature, and better flow of knowledge within the wheat community and among crops/model plants.
- A system that can adapt to advances in genome sequencing and technical capabilities.
- Provide a reference for the academic and breeding community.
- Potential to transfer to other cereal crops; e.g., barley, oats, rye.

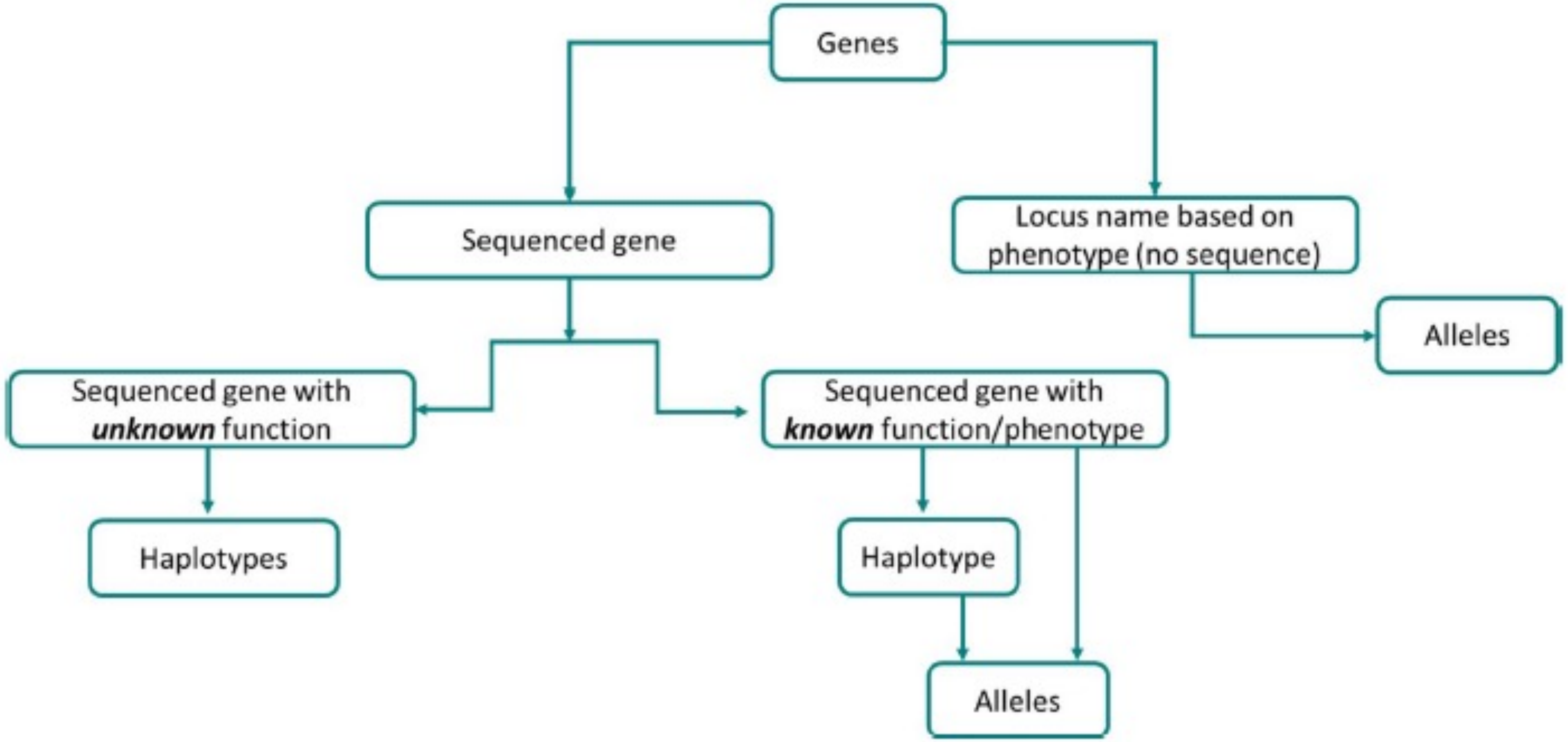
Why should you follow these guidelines?

- Improved clarity – easy to name your new “baby” gene and put it in the context of other wheat and plant research.
- Reduced complexity in the literature – your papers will have bigger and broader impact and get cited more often.
- Easier for the academic and breeding communities to communicate.
- Potential to transfer your knowledge to other cereal crops – wheat could come first for once!

Disclaimer…

- These guidelines are not meant to be too prescriptive, aim to provide a common reference point for the wheat community.
- We understand there will be exceptions to the rules, and we are happy for the guidelines to adapt to new genetic and genomic advances.
- You won't be policed… we hope to make the guidelines more visible, especially for publication in scientific journals.
- We hope they will promote discussion, and we encourage you to share with colleagues, especially those who are new to wheat research.

Overview



Underlying gene is identified, or gene investigated based on similarity with other species
e.g., *DELLA-D1*, *AP1-A1*, *ZIP4-B2*

Morphological or phenotypic features
e.g., *Rht-1*, *Vrn-1*, *Ph1*

Gene named based on phenotype

- E.g., *VERNALISATION* (*VRN*) – principal characteristic of the trait.
- Acronym can contain 2-4 letters, preferably 3, only classical examples can have single letters (e.g., *Q*).
- Locus name written in uppercase italics.
- *VRN-1* – the first locus identified to control vernalization. Successive loci provided number in ascending order *VRN-2*, *VRN-3*, etc, with number separated from name by hyphen.
- Numbers based on order of discovery, and if more than one is reported in a single publication, order is based on chromosomal location.
- For homeologs of the same set (e.g., *VRN-1*), genes are named *VRN-A1*, *VRN-B1*, *VRN-D1* – not *VRN-1A* etc.
- No reference to chromosome group – brings confusion.

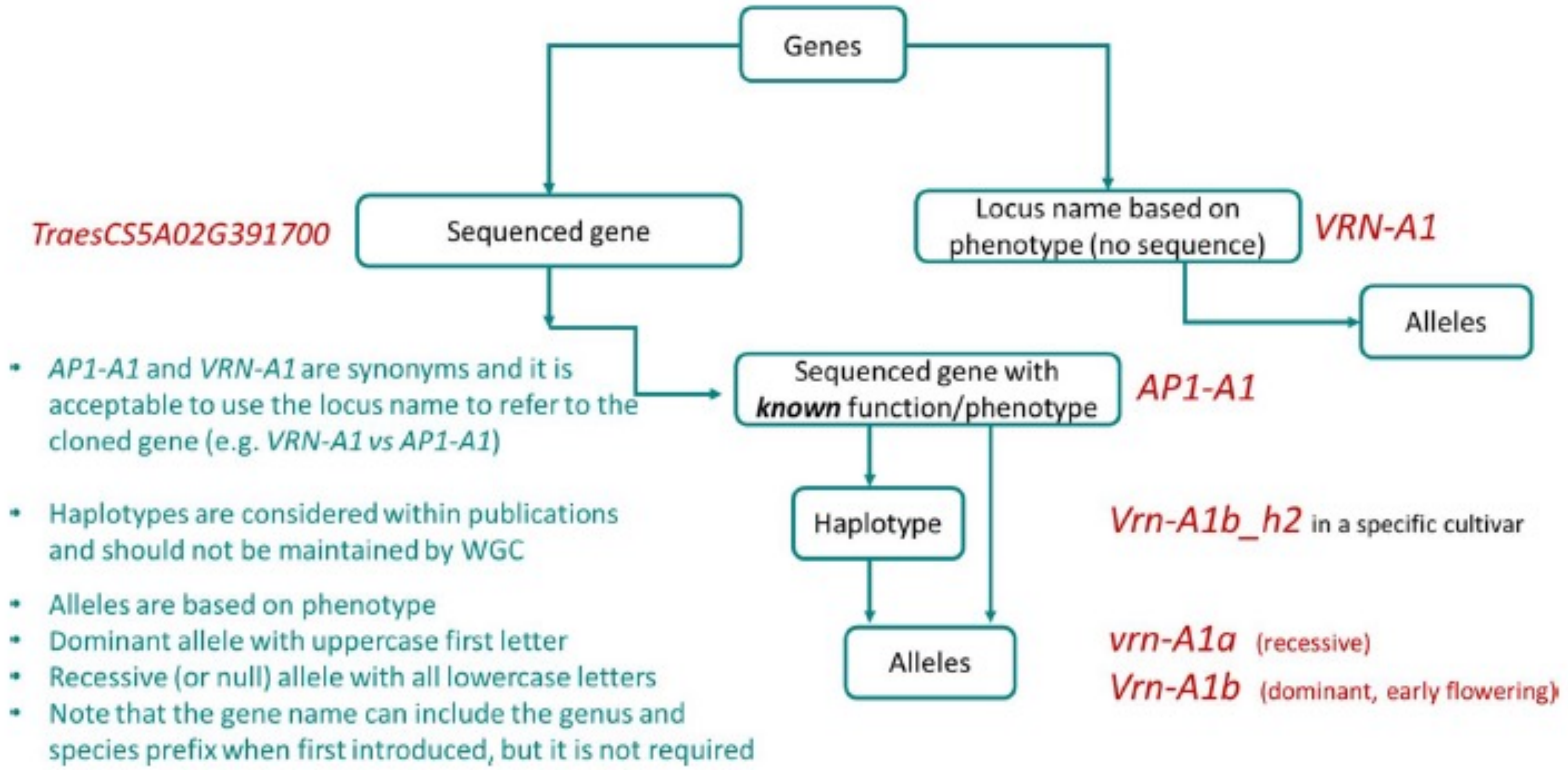
Alleles for genes identified by phenotype

- Dominant allele begins with capital letter, recessive alleles all lower case. E.g., *Vrn-1* is the dominant allele, *vrn-1* is the recessive allele.
- Independent alleles that confer different effects are designated with lower case letters after the number. E.g., *Vrn-A1a*, *Vrn-A1b*.
- The plus sign should NOT be used (e.g., for WT). Heterozygous genotypes designated by both alleles being written with a slash separating each allele. E.g., *Vrn-A1/vrn-A1*.

Other notes

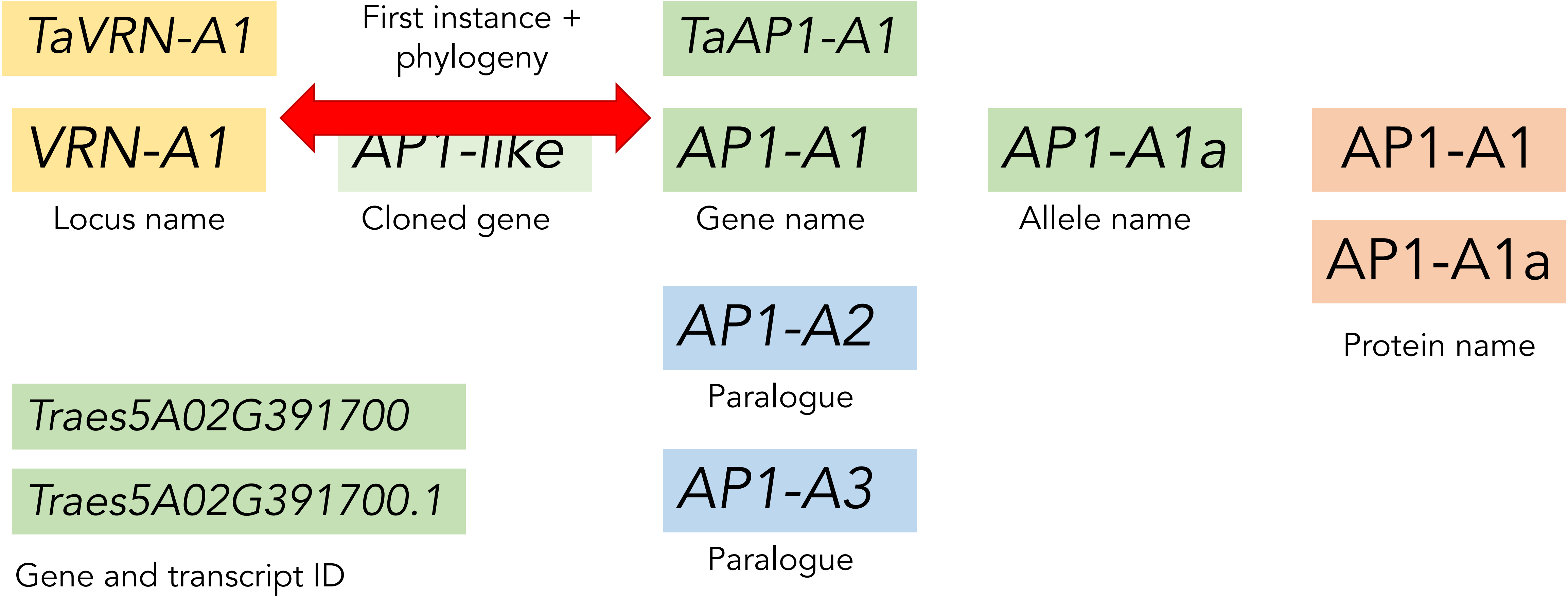
- Inhibitors, suppressors and enhancers – add a prefix of *I*, *Su*, and *En*, or *i*, *su*, and *en* for recessive elements, followed by the symbol of the allele affected. E.g., *Su1-Ph1.*
- Chromosomal aberrations: Deficiency, Duplication, Inversion, Translocation, Transpositions to be identified by *Df*, *Dp*, *Inv*, *T*, *Tp*. E.g., HP(Tp6D) for Hairy Peduncle gene on 6D transposed from 4B. See Koebner and Miller (1986).
- Extra chromosomal units: Mitochondrial, plastid, and chloroplast units should be prefixed with an underlined defining symbol (<u>*Mt*</u>, <u>*Pt*</u>, <u>*Cp*</u>) that precedes the gene name.

Underlying gene is identified, or gene investigated based on similarity with other species



- *AP1-A1* and *VRN-A1* are synonyms and it is acceptable to use the locus name to refer to the cloned gene (e.g. *VRN-A1 vs AP1-A1*)
- Haplotypes are considered within publications and should not be maintained by WGC
- Alleles are based on phenotype
- Dominant allele with uppercase first letter
- Recessive (or null) allele with all lowercase letters
- Note that the gene name can include the genus and species prefix when first introduced, but it is not required

Gene nomenclature for cloned gene



TaVRN-A1

VRN-A1

Locus name

First instance + phylogeny

AP1-like

Cloned gene

TaAP1-A1

AP1-A1

Gene name

AP1-A2

Paralogue

AP1-A3

Paralogue

AP1-A1a

Allele name

AP1-A1

AP1-A1a

Protein name

Traes5A02G391700

Traes5A02G391700.1

Gene and transcript ID

Genes identified based on homology to other plants

An example with single homologue in wheat

CCA1

Arabidopsis gene

First introduced

TaCCA1

CCA1

Wheat gene(s)

CCA1

Wheat gene

CCA-A1 *CCA-B1* *CCA-D1*

Three wheat homeologs

NB: This gene has been described as CCA1-7A, etc.

Genes identified based on homology to other plants

An example with multiple homologues in wheat

SEP1

Arabidopsis gene

First introduced

TaSEP1-1

SEP1

Wheat gene(s)

SEP1-1

SEP1-2

SEP1-3

...

SEP1-6

Wheat gene(s)

SEP1-A1 *SEP1-B1* *SEP1-D1*

SEP1-A2 *SEP1-B2* *SEP1-D2*

Three wheat homeologs

Priority of names

1. Has my gene already been named in wheat? Is it a well-chosen and explained name?

SEP1-6 = *AGAMOUS LIKE GENE1 (AGLG1)* = *MADS34* = *PANICLE PHYTOMER2 (PAP2)*

This gene has been reported about in multiple publications, with only 1-2 linking the studies with each other – lost opportunity for transfer of knowledge.

---

2. My gene has not been named in wheat, but named homologues exist in other species – what should I choose?

TB-1	= *BRANCHED1 (BRC1)*	= *FINECULM1 (FC1)*	= *TEOSINTE BRANCHED1 (Tb1)*
	Arabidopsis	Rice	Maize

We chose TB1 because the phenotypes aligned strongly with maize, the allele was complementary, and we wanted to connect with maize literature.

Priority of names

- Please include the full gene identifier. E.g., *TraesCSXX02G XXXXXX*. If the gene or sequence you have discovered is not in Chinese Spring, then please provide a cultivar specific gene identifier: e.g., *TraesMAC02GXXXXXX*.

- If the sequence of the gene is altogether novel, please deposit the sequence into the International Sequence Database Collaboration (www.insdc.org) and include accession number in the publication.

Haplotypes

- DNA sequences that include regions up- and down-stream of a gene that are of unspecified length.
- Length of the sequence limits should be reported in a publication.
- Designation: gene name_*hX*, *h* means 'haplotype', and *X* is a sequential number.
- Can be assigned to gene with or without phenotype: e.g., *TaBRI-A1_h1*, *TaBRI-A1_h2* for *BRASSINOSTEROID-INSENSITIVE 1*, and *Rht-A1a_h1* and *Rht-A1a_h2* for *Rht-A1a*.
- Important for communication between academic and breeding communities, will NOT be maintained by WGC.

Alleles

- Alleles are based on phenotype – can be "plant-based phenotype" or "molecular phenotype".
- NOT synonymous mutations used solely for SNP calling. Exception is mutations in miRNA binding sites.
- Please report clearly the basis of the phenotype and allele.
- Natural alleles: different existing alleles designated by lowercase italic letter. E.g., *Rht-B1a*, *Rht-B1b*.
- Chinese Spring is the preferred reference for '*a*' allele.
- Dominant, semi-dominant or co-dominant alleles should be designated with capital letter (e.g., *Rht-B1b*), recessive and null alleles with lowercase (*rht-B1a*).

Summary: Underlying gene is identified, or gene investigated based on similarity with other species

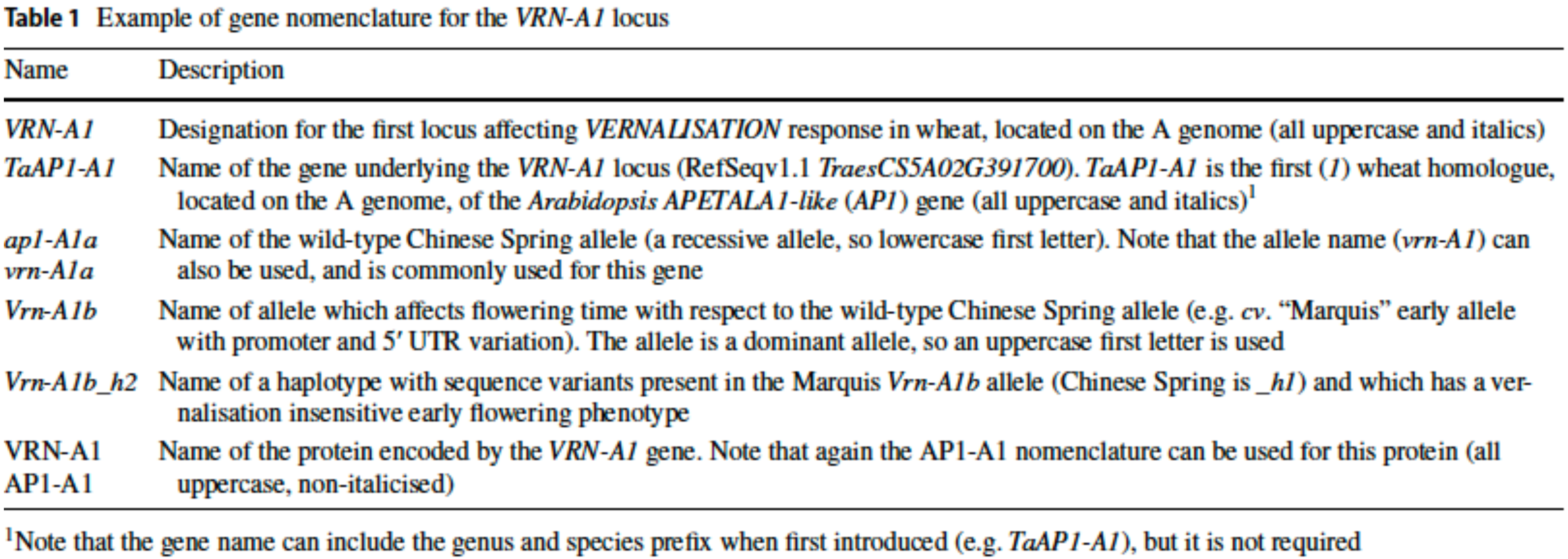
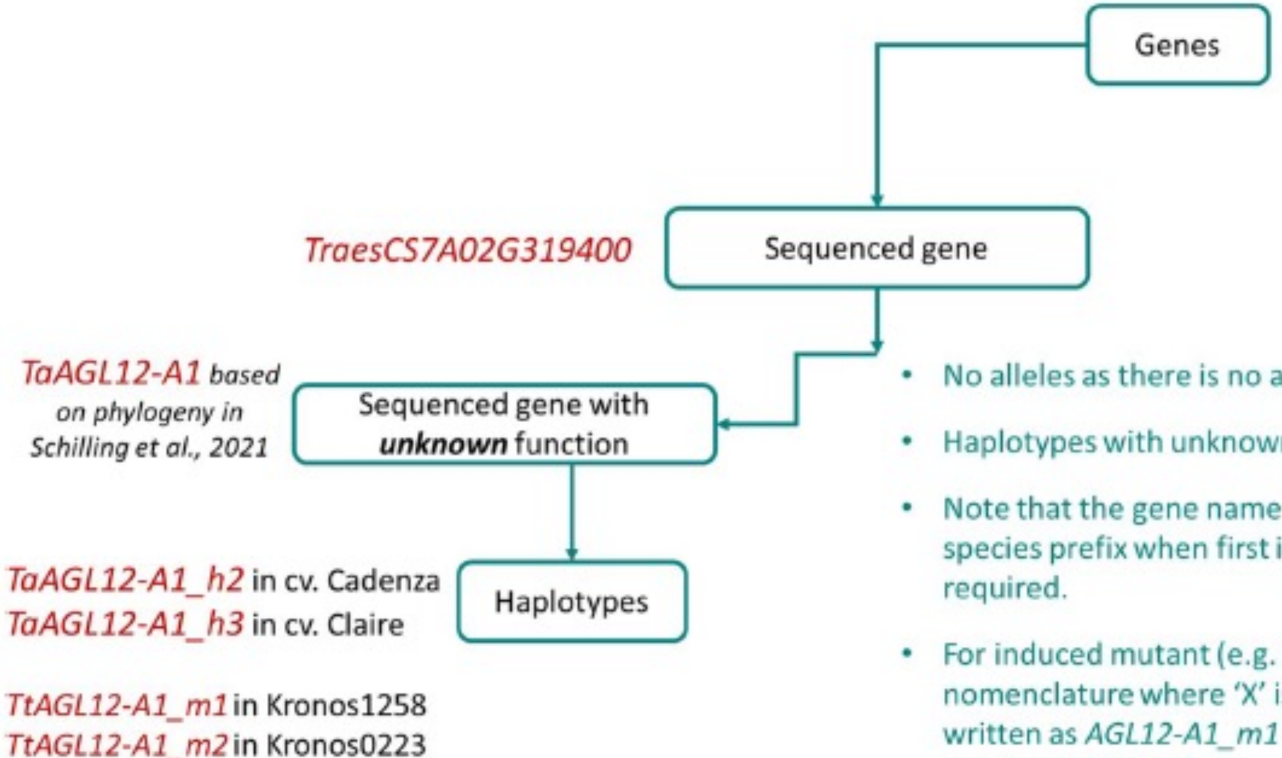


Table 1 Example of gene nomenclature for the *VRN-A1* locus

Name	Description
VRN-A1	Designation for the first locus affecting *VERNALISATION* response in wheat, located on the A genome (all uppercase and italics)
TaAP1-A1	Name of the gene underlying the *VRN-A1* locus (RefSeqv1.1 *TraesCS5A02G391700*). *TaAP1-A1* is the first (*1*) wheat homologue, located on the A genome, of the *Arabidopsis APETALA1-like* (*AP1*) gene (all uppercase and italics)[1]
ap1-A1a *vrn-A1a*	Name of the wild-type Chinese Spring allele (a recessive allele, so lowercase first letter). Note that the allele name (*vrn-A1*) can also be used, and is commonly used for this gene
Vrn-A1b	Name of allele which affects flowering time with respect to the wild-type Chinese Spring allele (e.g. *cv*. "Marquis" early allele with promoter and 5′ UTR variation). The allele is a dominant allele, so an uppercase first letter is used
Vrn-A1b_h2	Name of a haplotype with sequence variants present in the Marquis *Vrn-A1b* allele (Chinese Spring is *_h1*) and which has a vernalisation insensitive early flowering phenotype
VRN-A1 AP1-A1	Name of the protein encoded by the *VRN-A1* gene. Note that again the AP1-A1 nomenclature can be used for this protein (all uppercase, non-italicised)

[1] Note that the gene name can include the genus and species prefix when first introduced (e.g. *TaAP1-A1*), but it is not required

Induced mutants and gene edited lines



Genes

TraesCS7A02G319400 Sequenced gene

TaAGL12-A1 based on phylogeny in Schilling et al., 2021

Sequenced gene with ***unknown*** function

Haplotypes

TaAGL12-A1_h2 in cv. Cadenza
TaAGL12-A1_h3 in cv. Claire

TtAGL12-A1_m1 in Kronos1258
TtAGL12-A1_m2 in Kronos0223

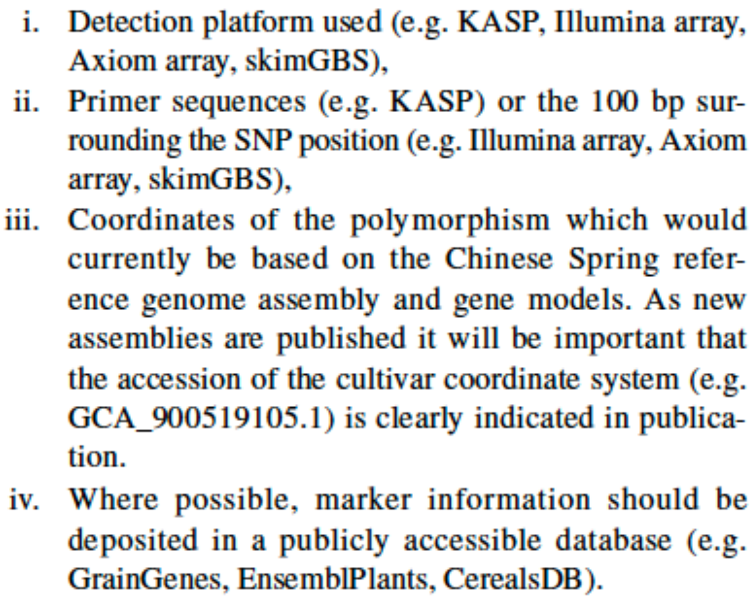
- No alleles as there is no associated phenotype
- Haplotypes with unknown effect are designated *_hX*
- Note that the gene name can include the genus and species prefix when first introduced, but it is not required.
- For induced mutant (e.g. Kronos EMS) use the '*_mX*' nomenclature where 'X' is a number. This could also be written as *AGL12-A1_m1* if species (*T. turgidum)* was previously defined.

Induced mutants and gene edited lines

- Same designation for induced mutants and gene edited lines.
- Designation: gene name*_mX, m* means ‘mutant’, and *X* is a sequential number. E.g., *AGL12-A1_m1*, *AGL12-A1_m2*.
- Additional mutants for same gene generated in later studies should be labelled sequentially on from first study, so long as the new mutations cause a unique effect.
- When reporting a mutant from a TILLING population, the identifier (e.g., CAD1715) should be provided.
- Details of the mutation should be provided when mutation is first introduced. E.g., V63M or null, W91*.
- If the mutation becomes more widely used in crosses or breeding, then an allele designation is warranted.

SNPs and derived markers

General comments



i. Detection platform used (e.g. KASP, Illumina array, Axiom array, skimGBS),
ii. Primer sequences (e.g. KASP) or the 100 bp surrounding the SNP position (e.g. Illumina array, Axiom array, skimGBS),
iii. Coordinates of the polymorphism which would currently be based on the Chinese Spring reference genome assembly and gene models. As new assemblies are published it will be important that the accession of the cultivar coordinate system (e.g. GCA_900519105.1) is clearly indicated in publication.
iv. Where possible, marker information should be deposited in a publicly accessible database (e.g. GrainGenes, EnsemblPlants, CerealsDB).

For known protein coding genes

Gene name

GW2-B1_C2504T

Change and position relative to ATG

GW2-B1_S25Y

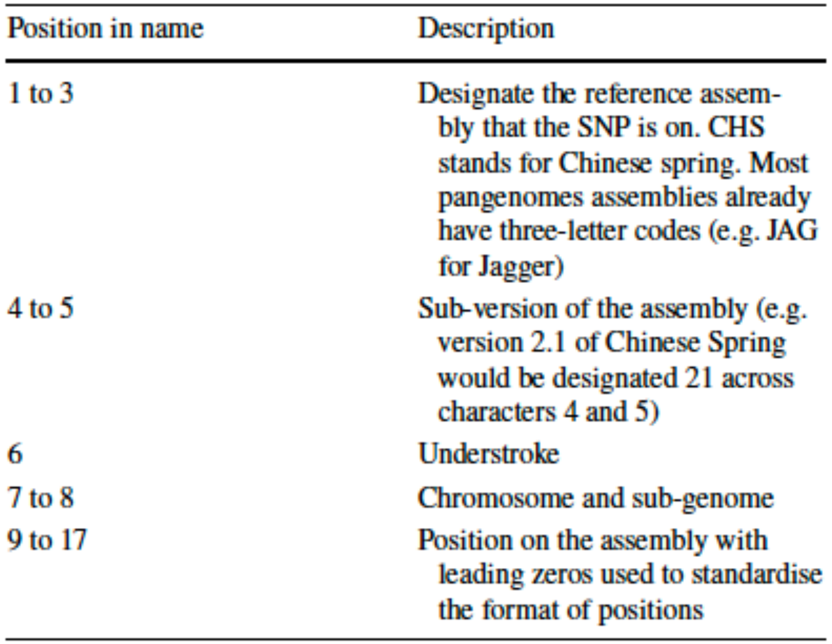
Or Met

GW2-B1_53_72del

Position of deletion

For known anonymous DNA sequences

CHI21_6A001234567



Position in name	Description
1 to 3	Designate the reference assembly that the SNP is on. CHS stands for Chinese spring. Most pangenomes assemblies already have three-letter codes (e.g. JAG for Jagger)
4 to 5	Sub-version of the assembly (e.g. version 2.1 of Chinese Spring would be designated 21 across characters 4 and 5)
6	Understroke
7 to 8	Chromosome and sub-genome
9 to 17	Position on the assembly with leading zeros used to standardise the format of positions

Other genetic topics

- **Details are also provided in the guidelines for:**
  - Pseudogenes
  - Gene complexes (relevant especially for grain quality traits)
  - QTL
  - Historical markers
  - Grain quality genes
  - Pathogen or pest resistance genes

Germplasm

- Germplasm associated with a formally named gene is clearly defined and referenced to an internationally available germplasm collection.
- The collection *must* allow worldwide access within framework for FAO of the United Nations International Treaty on Plant Genetic Resources for Food and Agriculture.
- When specific reference germplasm cannot be assured, proposers are encouraged to use a temporary name.

Closing thoughts

- These guidelines are intended to be for the benefit of the wheat research community – so please use them!
- Not too prescriptive, but a common reference point. Accommodate historical conventions while adapting to new advances.
- Support their use with colleagues, particularly when reviewing publications.
- Could be adopted by other cereal research communities?
- Should benefit academics and breeders.
- Go forth and discover!

Acknowledgements

- Scott Boden (Uni of Adelaide, ARC)
- Cristobal Uauy (JIC, BBSRC)
- Bob McIntosh (Uni of Sydney)
- Simon Krattinger (KAUST)
- Jorge Dubcovsky (UC Davis, USDA)
- W. John Rogers (UNCPBA, CONICET)
- Xianchun Xia (Institute of Crop Science, CAAS)
- Katerina Badaeva (NI Vavilov Institute, Russian State Fund)
- Alison Bentley, Susanne Dreisigacker (CIMMYT, Accelerating Genetic Gains in Maize and Wheat)
- Gina Brown-Guedira (NCSU, USDA-ARS)
- Mario Caccamo, James Cockram (NIAB, BBSRC)
- Luigi Cattivelli (CREA, PRIMA)
- Parveen Chhuneja (Punjab Agricultural Uni, Government of India)
- Bruno Contreras-Moreira (EEAD-CSIC, PRIMA)
- Dave Edwards (Uni of Western Australia, ARC)
- Fernanda Gonzalez (INTA)
- Carlos Guzman (Uni de Cordoba, Ramon y Cajal Program)
- Tatsuya Ikeda (Western Region Agricultural Research Center, NARO)
- Ildiko Karsai (Centre of Agricultural Research, Innovation and Tech Ministry of Hungary)
- Shuhei Nasuda (Kyoto University, NBRP-Wheat)
- Curtis Pozniak (Uni of Saskatchewan, Genome Canada)
- Renee Prins (Cengen, Stellenbosch Uni, S.A. Winter Cereal Industry Trust)
- Taner Sen (Albany, USDA-ARS)
- Paula Silva (INIA, ANII)
- Hana Simkova (Institute of Exp. Bot., European Regional Development Fund)
- Yijing Zhang (Fudan University)
- **The Wheat Initiative**

Article can be found at:

Updated guidelines for gene nomenclature in wheat
Theoretical and Applied Genetics (2023) 136:72
doi: 10.1007/s00122-023-04253-w

scott.boden@adelaide.edu.au
@sbodes12

*****Fully-funded PhD position available in Boden Group*****
Please contact me for further information – come discover a gene!