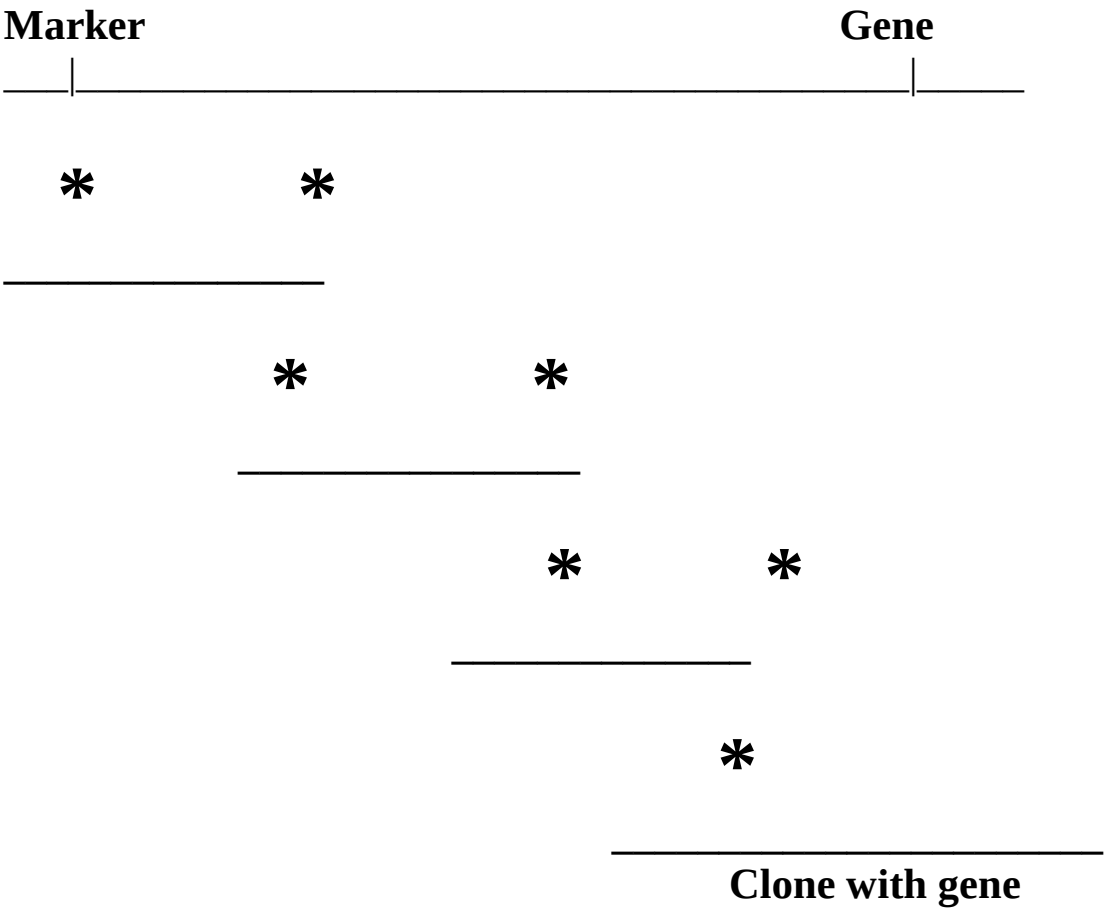
Map-Based Cloning of Qualitative Plant Genes

Map-based cloning

- using the genetic relationship between a gene and a marker as the basis for beginning a search for a gene

Chromosome walking

- moving toward a gene by using a probe for a marker near a gene to select a genomic clone near the gene and moving toward the gene by repeatedly selecting for overlapping clones until you have a clone that contains the gene
- original approach used to find a large insert clone that contains your gene of interest; replaced by high density mapping



Chromosome landing

- using high density mapping in the region of a gene to find markers that cosegregate with the gene of interest
- this marker used to select a genomic clone that contains the gene

```
                    Marker  Gene
_____________________|__|_____________________

                      *

                  _____________
                  Clone with gene
```

Steps Involved in Map-based Cloning

Traditional Approach – linkage base approach

- **Identify a cosegregating marker in a "large" mapping population**
  - Perform high density molecular mapping

- **Find a large insert clone (BAC or YAC) to which the marker probe hybridizes**
  - Chromosome walking
  - Chromosome landing

- **Determine that the gene is on the clone**
  - Find an appropriate recombinant genotype

- **Identify a candidate gene on the clone**
  - Search for a cDNA clone

- **Perform genetic complementation to rescue the wild-type phenotype**
  - Transform a plant without the gene and look for phenotypic rescue

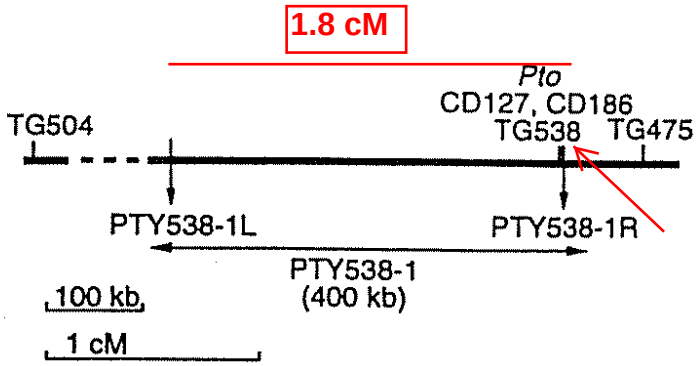
- **Sequence the gene and determine if the function is known**
  - Determine the molecular sequence and compare it against a sequence database

Example of Map-based Cloning

Cloning the tomato *Pto* gene

- First example of map-based cloning in plants
- Provides resistance against bacterial speck disease
- Disease caused by *Pseudomonas syringa* pv. tomato
- Pathogen expresses the avirulence gene avrPto
- Gene-for-gene interaction between *Pto* and *avrPto*
- Martin et al. [Science (1993) 262:1432]

1. A genetic population of 251 F2 plants was screened with DNA probes, and the locus TG538 cosegregated with *Pto*.



From: Martin et al. 1993. Science 262:1432

2. A YAC library was screened with the TG538 probe, and the clone PTY538-1 was identified.

Chromosome landing step

3. Primers that marked the end of the 400 kb clone were created, and the population was screened.

- PCR marker PTY538-1L was 1.8 cM from *Pto*
- PCR marker PTY538-1R cosegregated with *Pto*
- PCR marker PTY538-1R may be to the left of *Pto*; need to find a recombinant between this marker and TG538

Is the Pto gene located on the clone?

4. 1300 plants from F2, F3 and cultivars were screened

- One plant with the *Pto* allele at the TG538 and the *pto* PTY538-1R allele was found. Thus, *Pto* must be located on the YAC clone PTY538-1.

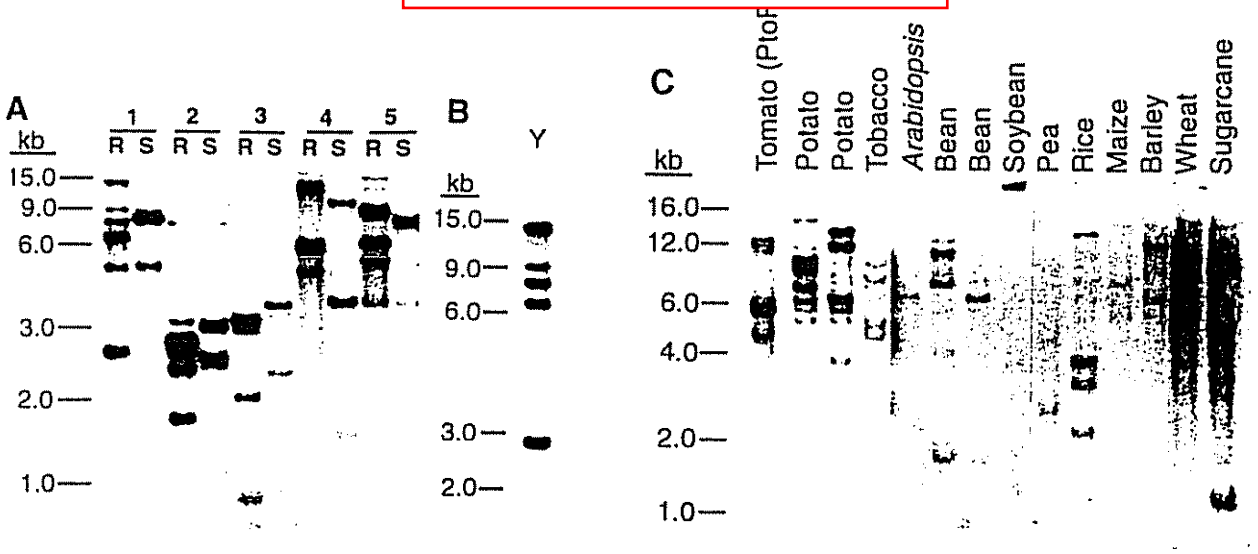
Recombination data needed to prove your markers on each side of the gene and the clone contains the gene.

5. A leaf cDNA library was screened with PTY538-1 DNA
   - 200 clones were found and 30 analyzed
   - 50 plants with recombination in the *Pto* region was screened
   - cDNA clone CD127 cosegregated with *Pto*.

Finding cDNA clones (=expressed genes) located on the clone!!!

6. Southern hybridization with CD127 detected many polymorphic fragments
   - **PANEL** A: The clone could be detecting exons or a multigene family
   - 14 other cross hybridizing clones were found that all represented six different classes of related genes
   - **PANEL B**: Analyses of PTY538-1 detected all (but one) of the genomic fragments that were detected with Southern hybridization.
   - It was concluded that PTY538-1 contained a multigene family.
   - **PANEL C:** Clone hybridized to multiple fragments in multiple species; clone detects a multigene family in other species.

Clone recognizes a multigene family in tomato



7. Transformation vectors were created with CD127 (1.2 kb insert; vector pPTC5) and CD186 (2.4 kb insert; vector pPTC8).

- Two pPTC8 resistant plants were found.
- No pPTC5 resistant plants were found.

Select the appropriate clone to work with

8. One resistant plant (genotype=*Pto*/*Pto*) was crossed to a susceptible cultivar (genotype=*pto*/*pto*).

- 9 plants contained the CD186 sequence and were resistant to the bacterial speck pathogen
- 13 without CD186 were susceptible to bacterial speck.

Proof that the candidate gene is located on that piece of DNA

9. The clone CD186 was sequenced, and a 321 amino acid opening reading frame was found.

- The sequence of the open reading frame was similar to a serine-threonine protein kinase.

Modern Alternative – Using reference sequence as a guide

Perform traditional mapping to develop tightly linked marker

- Same approach used linkage analysis

- **Use the marker to search the reference genome sequence**
  - Locate the marker to a specific region of the genome

- **Develop markers from the sequence that you think might flank the candidate gene**
  - Use your SNP or SSR resources to idenetify/develop new markers near your gene

- **Confirm the markers flank the gene**
  - Develop a large F_2 population
  - Screen the F_2 population to discover markers that flank your phenotype

- **Select the candidate gene**
  - Evaluate the putative function of the genes in the candidate interval and select one that matches your functional expection
  - Requires understand of the function, biology, and molecular expression pattern of the expected candidate

- **Perform complementation test**
  - Same as for traditional map-based cloning procedure

- **Discover putative causative mutation in the gene**
  - Sequence multiple genotypes to determine the sequence variation associated with the different alleles

Cloned Quantitative Trait Loci

Species	Trait	QTL	Gene	Allelic Variant	Reference
Tomato	Soluble Solids (sugars)	*Brix9-2-5*	Apoplastic invertase	Intron repeat	PNAS 97:4718
Tomato	Fruit weight	*fw2.2*	RAS protein	5' region	Science 289:85
Rice	Heading date	*Hdl*	Zinc finger	Exon and intron differences	Plant Cell 12:2473
Rice	Heading date	*Hd6*	Protein kinase	Truncate protein	PNAS 98:7922
Arabidopsis	Flowering time	*ED1*	Blue light photorecptor cryptochrome-2	Single AA change	Nature Genetics 29:435

Brix9-2-5

Soluble solids in tomato

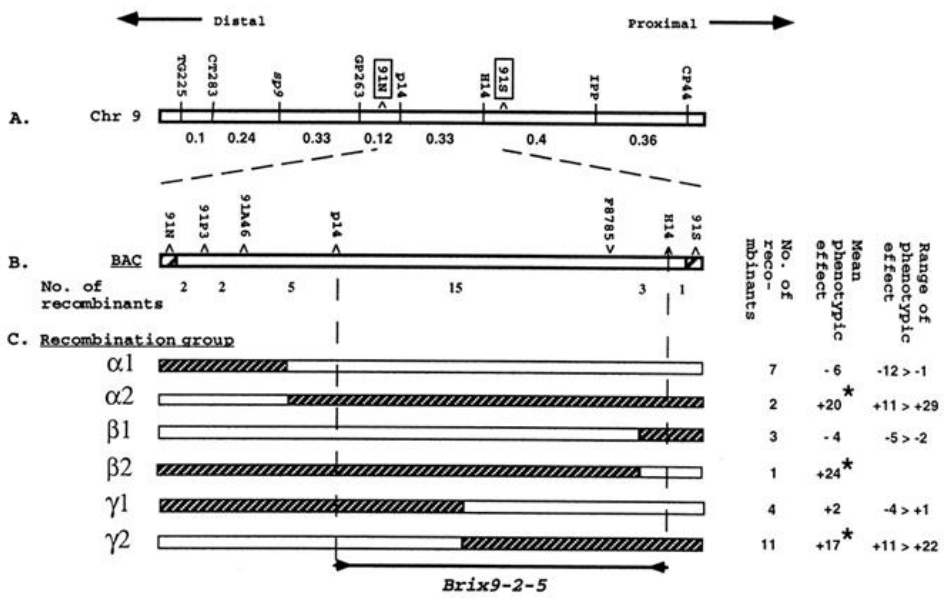
- A major breeding objective is to increase solids
- Total soluble solids consists of:
  - Sugars
  - Acids
- Improves taste and processing quality
  - Wild lines values can reach 15%, three times that of the cultivated varieties

Genetics of soluble solid

- The green-fruited species, *Lycopersicon pennellii*, contains high values
- 50 introgression lines developed, each with a single RFLP fragment from *L. pennellii*
- QTL mapping
  - Placed *Brix9-2-5* on a 9 eM fragment of chromosome 9
- Allelic variation
  - In F1 between cultivated and introgressed line the wild allele of *Brix9-2-5:*
    - Glucose increased 28%
    - Fructose increased 18%
    - Acid content did not change
  - Change is reflected in the above ground tissue

Map-based Cloning of *Brix9-2-5*

- **High-resolution mapping**
  - 7000 F2 progeny of the NIL evaluated with flanking markers previously discovered
    - 145 recombinants
    - Distance between markers was 1 cM
    - 28 recombinants localized between the ends of a BAC clone
  - 28 recombinants analyzed further
    - Brix measured
    - Recombinant subclasses defined by markers
    - *Brix9-2-5* localized to a 18 kb span
  - Analysis of 18 kb fragment
    - Specific primers developed
    - 28 recombinants again analyzed
    - One primer pair
    - Defined a 1 kb fragment
    - Fragment exhibited complete cosegregation with *Brix9-205*
    - SNPs discovered with 1 kb fragment
    - 13 families recombinant with the 1 kb fragment
    - Recombinant families analyzed
    - *Brix9-2-5* placed within a 484-bp interval



- **Molecular analysis**
  - 484-bp interval contains an apoplastic invertase
  - Gene is *Lin5*
    - Member of a small multigene family
    - Gene contains six exon
    - *Brix9-2-5located* to exon 3/intron 3/5' fragment of exon 4
  - Apoplastic invertases cleave sucrose into glucose and fructose
    - Regulate source-sink transitions
  - Allelic variability
    - Three AA differences between *L. esculentum* and *L. pennellii*
    - AA differences are not major
    - Intron differences noted that
    - These may be related to regulation differences between the two species

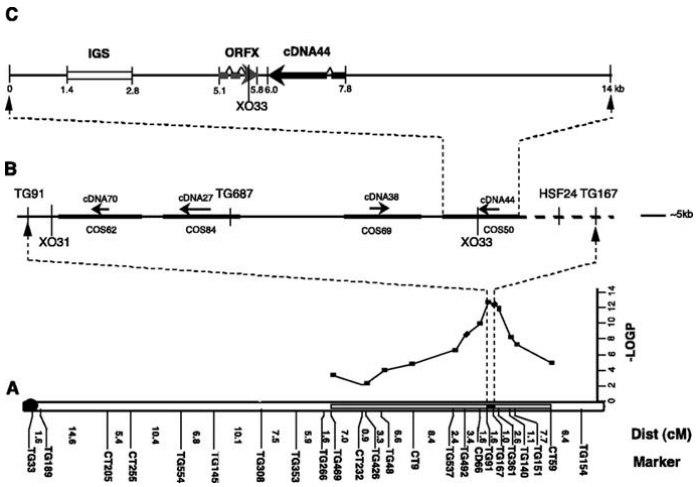
fw2.2

Fruit size in tomato

- A major breeding objective is to increase fruit size
- Genetically mapped in tomato
- *QTL fw2.2*
    - Increases fruit weight by 30%
    - A major domestication trait
- All wild species are small fruited

Map-based Cloning of *fw2.2*

- YAC cloning containing *fw2.2* isolated
- *L. pennellii* (small fruited) eDNA screened
    - Four cDNAs selected
    - High resolution map derived
    - 3472 F2 from a cross between two NILs to develop a map
    - Four cosmid clones obtained
    - Complementation tests performed
    - Mogeor (fresh market) and TA496 (processing) transformed
    - Homozygous lines obtained (because *fw2.2* is partially dominant)
    - Only cos50 progeny showed a statistically significant decrease in size



- cos50 genes
    - Two ORFs on the cosmid
    - One ORF corresponds to eDNA used to isolate cos50
    - Second ORF (*ORFX*) is a 663-nt gene
    - Mapping places *fw2.2* closest to second ORF
        - ORF is expressed at low levels for standard Northern analysis
        - RT-PCR shows the ORF expressed at low levels in pre-anthesis carpels
        - Is a correlated effect on carpel size seen?
            - Carpels, sepals and styles of large-fruited genotypes heavier
            - Large-fruited carpels contained more cells

The ORFX gene

- Three exons, two introns
- Encodes a 163 AA protein, about 22 kd
- Only high homology to plant genes
- Part of a multigene family
- A homolog of human RAS oncogene
- By analogy, ORFX could be involved in control of cell division

Allelic variation

- Compared *L. pennellii* and *L. esculentum* alleles
    - 55nt of 5’ end; 95 nto of 3’ end
    - 42 nt differences in coding region
        - 35 fell with the two introns
        - 4 silent mutations
        - 3 substations
            - Assigned to the 5' region of the gene
            - Allelic effects suggested to reside in 5' region
            - Differences could affect gene expression

Hd1

Heading Date in Rice

- An important agronomic trait
- Rice is photoperiod sensitive (PS)
  - A short-day plant
  - Several PS genes have been mapped
  - One PS gene has been cloned, *Se5*
- Major QTLs have been mapped
  - Mapped as single factors using advanced backcross populations
  - Located on end of chromosome 6
  - *Sel* (historical locus) and *Hdl* (recent QTL) map to the same locus
  - Are these the same locus?
  - Use map-based cloning

Large insert clone discovery

- Markers discover *Hd1* is located on PAC P0038C5

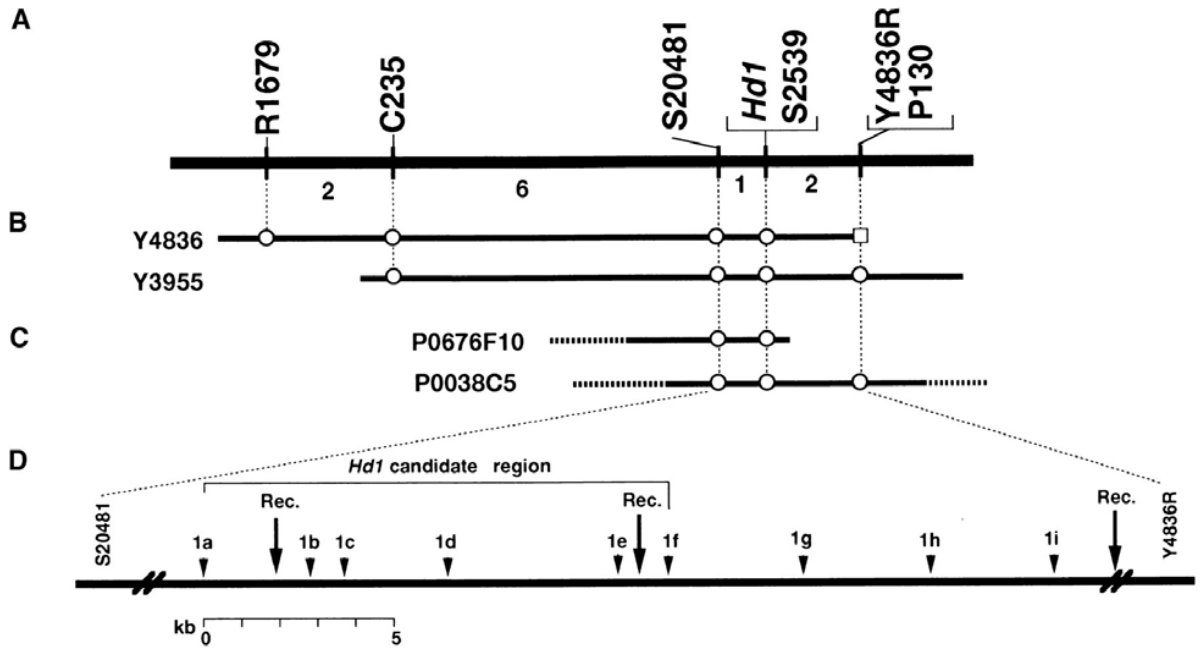


Figure 1. A Fine-Scale, High-Resolution Genetic and Physical Map of the *Hd1* Region on Chromosome 6.

(A) Genetic linkage map showing the relative position of *Hd1* with RFLP markers on chromosome 6. Numbers under the horizontal line are numbers of plants with a recombinant chromosome in the adjacent marker intervals.

(B) and **(C)** Yeast artificial chromosome **(B)** and PAC clones **(C)** spanning the *Hd1* region. A circle indicates the existence of a sequence corresponding to the RFLP markers. Entire insert sequencing was performed on PAC clone P0038C5.

(D) Detailed genetic and physical map showing the relative positions of the candidate regions of *Hd1* and CAPS markers developed based on sequence data. Rec., approximate positions of recombination events that occurred near *Hd1*.

Map-based Cloning of *Hd1*

- *Hd1* mapped between markers S20481 and P130
  - >9000 BC3F3 plants analyzed
  - Only 1 (S20481) and 2 recombinants (P120) discovered
- YACs containing the markers isolated
  - Additional markers discovered
  - P1-dervied artificial clones selected
  - By mapping, clone P0038C5 contains *Hdl*
- P0038C5 shotgun sequenced
  - 26 kb region defined as candidate
  - CAPS markers developed from this region
  - Markers delimited gene to 12 kb region
- Two ORFs discovered in 12 kb region
- Peroxidase gene (not ruled out)
- Ortholog to *Arabidopsis CONSTANS* (CO)
  - Studied further because of its similar role to *Hdl*
  - Work based on Nipponbare allele
  - Kasalath studied and compared to Nipponbare allele
    - 1 two-base substitution
    - 33 bp deletion in first exon
    - 2 bp deletion in second exon

Studied *Sel* from Ginbouzu and its mutants using *Hd1* primers

- 1 nt substition and 36 bp insertion in exon one of Ginbouzu
- Ginbouzu and Nipponbare function equivalenty in PS response
- 43 bp deletion in first exon in mutant HS66
- 433 bp insertion in first intron in mutant HS110
- *Sel* and *Hd1*appear to be allelic

***Hd1* Gene Structure**

- 59% identical to CO
- 79% identical in C-tenninal
- encodes a 395 *AA* protein
- A zinc-finger domain protein
- *Hd1*diverges downstream of the zinc finger

Functional Complementation

- 7.1 kb fragment inserted
- Homozygous or heterozygous transgenics flowered earlier under short days

Expression of *Hd1*

- Gene not detectable using standard Northem analysis
- RT-PCR procedure applied
  - Product observed in Nipponbare, Kasalath'and NIL (*Hdl*)
  - Amount of Nipponbare did not change in LD to SD transition
  - Ginbouzu and its *sel* mutant produced product

Molecular and Genetic Control of Flowering by *Hd1*

- Zinc finger structure suggests *Hd1* acts as a transcriptional regulator
- Positive regulator under SD and negative regulator under SD
- CO only effective under SD
- *Hd1* expression unaffected by photoperiod
- Evolutionary comparisons suggest the same genes control LD and SD flowering

Allelic Difference Detected

- Nucleotide substitutions (lines)
- Deletions (boxes)

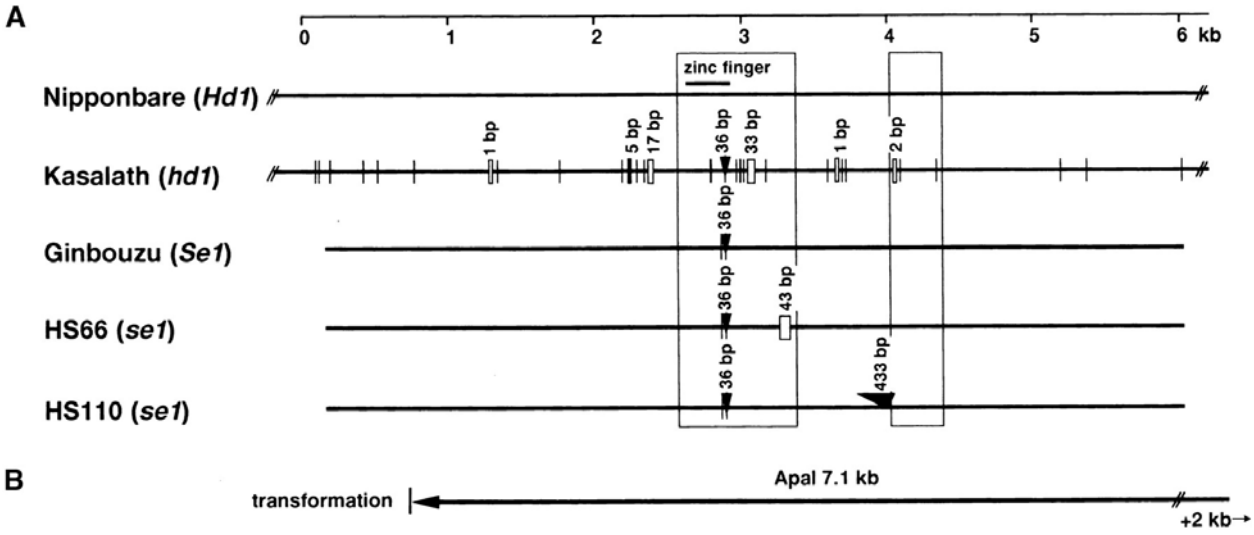


Figure 2. Scheme of the Structural Differences in the Candidate Region of *Hd1* in Nipponbare and Kasalath and the Corresponding Genomic Region of *se1* Mutants HS66 and HS110 and Their Original Variety, Ginbouzu.

(A) Comparison of genomic sequences of Nipponbare and Kasalath *Hd1* alleles. Boxes show the predicted open reading frames based on the Genscan software maize model. Vertical lines without labels represent single-base substitutions between Nipponbare and Kasalath. Small rectangular boxes and arrowheads represent deletions and insertions, respectively.

(B) A 7.1-kb ApaI genomic fragment containing the entire *Hd1* candidate sequence used in the complementation analysis. This fragment does not contain another predicted gene, encoding peroxidase, which was found in the candidate genomic region of *Hd1*