

## COMPARATIVE GENOME MAPPING

Reference: Keller and Feuillet, 2000; Schmidt, 2002.

### Some terms:

- Orthologous loci: loci in different species originating from the same ancestral locus  
paralogous loci: loci in different (or the same) species that arose due to a duplication of an ancestral locus  
synteny: a similar linear order of markers in two different species (or on two different linkage groups, as in soybean or *Brassica*)  
microsynteny: synteny at the sequence level, sub megabase pairs levels.

### *What do you gain from comparative mapping?*

1. Genome evolution: Allow us to understand the evolution of species (Schmidt, 2002). Maps constructed in one species can be compared by means of common markers (or common single gene traits) with closely related species. These comparative maps can be used to study genome evolution—how the genome has been rearranged through time—and to make inferences about gene organization, repeated sequences, etc.
2. Development of linkage maps: One can use anchor markers, identified from the mapping of several related species, in constructing a molecular linkage map for a new species.
3. Gene cloning: Specially among grass species, to which most crop species belong, one can use information from a simple species such as rice to clone genes from complex species like barley (e.g. Kilian et al., 1997). Positional cloning of *Lr10* leaf rust disease resistance locus in bread wheat by chromosomal walking in *Triticum monococcum*, 1/3 rd genome size of bread wheat (*T. aestivum*) (Stein et al. 2000). Map based cloning may be easier in some species than others—e.g., rice (with a small genome) vs. wheat (with a massive genome).

### Review of a few examples:

- I. The *Solanaceae*: Initial comparative mapping work was done with the *Solanaceae* (tomato, potato, pepper, eggplant, jimsonweed (*Datura*), tobacco) by Steve Tanksley and colleagues at Cornell.
  - A. Tomato and pepper (Tanksley et al., 1988)
    1. Basic chromosome number is  $x = 12$  for both species.
    2. Most genes are conserved, but some are duplicated in pepper.
    3. Linkage order is mixed up—i.e., many rearrangements since their common ancestor.
    4. Pepper has 4x DNA content of tomato—probably not due to gene duplication (due to non-coding, repetitive DNA—retrotransposons?).

- B. Tomato and potato (Bonierbale et al., 1988; Tanksley et al., 1992)
1. Much more closely related than tomato-pepper,  $x=12$  for both.
  2. Virtually identical maps with five paracentric inversions accounting for all differences.
  3. Potato has  $\frac{1}{2}$  as much recombination as tomato, despite similar physical lengths (1200 vs 600 cM).
  4. More closely related species (tomato-potato) have fewer rearrangements than more distantly related taxa (tomato-pepper), as expected.

## II. The *Gramineae* (grasses)

- A. Rice and maize: (Ahn and Tanksley, 1993)
1. Diverged 50 Myr ago.
  2. Rice:  $x=12$ ; maize:  $x=10$ ; maize 6x nuclear DNA.
  3. Single copy gene of rice always duplicated in maize and 72% of the duplication still exist in maize genome.
  4. Loss of 28% of duplicated copy of maize genes could have resulted from deletions or loss of entire chromosomes or chromosomal segments. Results indicate larger deletion encompassing several loci – from translocation or inversions.
  5. Pairs of homologous chromosomes in maize: ch 2 and 10 are similar and coliner to rice ch 4. Ch 3 and 8 have the same gene content but shuffling of gene orders.
  6. 6-fold more DNA but did not increase in recombination in the conserved region as compared to that in rice. Same recombination rate. Recombination in genic regions??
  7. But maize map length (1723 cM) ,  $< 2x$  the rice map (1055 cM) - a bit less than 2x (polyploidy) due to loss of dup. regions?
- B. Wheat, rye, barley Mike Gale's group in England (Devos and Gale, 1997)–see Fig. 1
1. Wheat A, B, and D genomes are homologous.
  2. Wheat shows synteny with barley and rye.
  3. Regions of no synteny are “fast-evolving” and non-coding–i.e. more synteny seen with cDNA than with random genomic probes.
- C. Grasses as a single genetic system
1. The results in the early synteny papers led to Bennetzen & Freeling's (1993) suggestion of using grasses as a single genetic system.
    - a. 10-11 thousand species in the *Gramineae*
    - b. Many of the species are cross compatible (allopolyploids, etc.).
  2. Mapping experiment have shown:
    - a. Large degree of collinearity among diverse grasses; within 5-10 cM regions, good conservation across grass species were found.

- b. Most genes are homologous across species--i.e. all grasses have essentially the same genes.
  - c. Fine structure mapping may show insertions of repeated sequences, etc.
  - d. All this led to the development of the circular grass genome figure--“Grass genomes line up and form a circle”– (Moore et al., 1995). See also Devos and Gale (1997), Figs. 4 and 6.
3. Comparative QTL mapping: (Paterson et al., 1995)  
Locations of some major QTL (genes) controlling domestication traits are similar across several major grass crop species.

### III. The *Cruciferaeae*

- A. Within *Brassica* (Lagercrantz and Lydiate, 1996)
1. *B. nigra* (B genome;  $x = 8$ ), *B. oleracea* (C;  $x = 9$ ), and *B. rapa* (A;  $x = 10$ )
  2. The B genome is triplicated--others are too; therefore, all derived from hexaploid ancestor (?), i.e., genome content is similar among the three species.
  3. Rearrangements among the A, B, and C genomes--fusion/fission of chromosomes to change numbers, but some conservation of linkage blocks remain.
  4. Do rearrangements occur at specific points in the genome? That way, the overall shuffling could still permit localized sets of loci to remain colinear.
- B. *Brassica-Arabidopsis* (Lagercrantz, 1998)
1. *Arabidopsis* represents one copy of the ancestral *Brassica* genome, i.e., three copies of *Arabidopsis* genes in *Brassica* (see Figure 2).
  2. Conserved segments between species ~8 cM (see Figure 1).
  3. ~90 rearrangements since divergence much higher than other species .
  4. Why so much rearrangement? Figure 4 and Table 3 show the putative number of rearrangements that are required to bring the genomes into rough equivalence. The high number for the crucifers may be due to the recent polyploidization of *Brassica*, which induces many rearrangements.
- C. *Arabidopsis*--soybean (Grant et al., 2000)
1. Compared putative protein sequences of genes coded by soybean RFLP and flanking regions of SSR markers with *Arabidopsis* databases.
  2. *Arabidopsis* appears to be duplicated.
  3. Strong synteny observed between soybean and *Arabidopsis* with few interchanges explaining most of the map differences (only considered three soybean linkage groups).
  4. This synteny may not be identified using either RFLP cross-hybridization or direct nucleotide comparisons because too many alterations may have obscured the functional relationships

IV. Bottom line on comparative mapping:

- A. Some conservation do exist among all plants
- B. Synteny increases as the divergence time between species decreases.
- C. Possible that small linkage blocks are well conserved, but often shuffled.
- D. questions:
  1. How small are the conserved areas over large evolutionary time?
  2. Do species without particular traits still have the genes—e.g., resistances to diseases—so that finding a resistance to some disease of wheat, for example, could be possible in rice, which doesn't express the disease symptoms?
  3. Why in some families we see a higher level conservations than those in others?
  4. Where do we see the real applicability?