

COMPARATIVE GENOME MAPPING

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. Map based cloning may be easier in some species than others—e.g., rice (with a small genome) vs. wheat (with a massive genome).

- **Colinearity:** Conservation of the gene order within a chromosomal segment between different species.
- **Microlinearity:** Conservation of the order of genes in DNA fragments that are bigger than 50 kb. Deletion, inversions and duplications are detected at this level.
- **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.

Examples of comparative mapping.

Solanaceae:

Initial comparative mapping work was done with the *Solanaceae* by Steve Tanksley and colleagues at Cornell.

A. Tomato and pepper (Tanksley et al., 1988)

1. Basic chromosome number $x = 12$ 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 four 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.

C. Tomato and Eggplant (Doganlar et al. 2002)

- a. Much more closely related than tomato–pepper, $x=12$ for both.
- b. Molecular genetic linkage map based on tomato cDNA, genomic DNA, and EST markers was constructed for eggplant, *Solanum melongena*.
- c. The map consists of 12 linkage groups, spans 1480 cM, and contains 233 markers.
- d. Comparison of the eggplant and tomato maps revealed conservation of large tracts of colinear markers, a common feature of genome evolution in the *Solanaceae* and other plant families.
- e. Overall, eggplant and tomato were differentiated by 28 rearrangements, which could be explained by 23 paracentric inversions and five translocations during evolution from the species' last common ancestor.
- f. No pericentric inversions were detected. Thus, it appears that paracentric inversion has been the primary mechanism for chromosome evolution in the *Solanaceae*.
- g. On the basis of the number of chromosomal disruptions and an approximate divergence time for Solanum, ≈ 0.19 rearrangements per chromosome per million years occurred during the evolution of eggplant and tomato from their last ancestor.
- h. This result suggests that genomes in Solanaceae, or at least in Solanum, are evolving at a moderate pace compared to other plant species.

Gramineae

A. Rice and maize:(Ahn and Tanksley, 1993)

- a. Diverged 50 Myr ago. Rice: $x=12$; maize: $x=10$; maize 6x more nuclear DNA.
- b. Single copy gene of rice always duplicated in maize and 72% of the duplication still exist in maize genome.
- c. Loss of 28% of duplicated copy of maize genes could have resulted from deletions or loss of entire chromosomes or chromosomal segments.
- d. 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.
- e. 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**?
- f. 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. Grasses as a single genetic system

1. ***Grasses as a single genetic system*** [Bennetzen & Freeling's (1993)].
 - a. 10-11 thousand species in the Gramineae.
 - b. Many of the species are cross compatible (allopolyploids, etc.).
2. ***Mapping experiment have shown that:***
 - a. Large degree of collinearity among diverse grasses; within 5-10 cM regions, - good conservation across grass species.
 - b. Most genes are homologous across species--i.e. all grasses have essentially the same genes.

- c. 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.

C. Wheat, rye, barley: (Devos and Gale, 1997)

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.

Cruciferaeae

A. 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. *B. nigra* (B genome; $x = 8$), *B. oleracea* (C; $x = 9$), and *B. rapa* (A; $x = 10$). The B genome is triplicated—others are too; therefore, all derived from hexaploid ancestor (?).
3. Conserved segments between species ~8 cM (see Figure 1).
4. ~90 rearrangements since divergence much higher than other species.
5. 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.

Gresshoff Group (2003) showed microlinearity in a region that contains a gene involved in nodulation.

What do you gain from comparative mapping?

1. Maps constructed in one species can be compared by means of common markers (or common single gene traits) with closely related species.
2. 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.
3. Map based cloning may be easier in some species than others—e.g., rice (with a small genome) vs. wheat (with a massive genome).

IV. Summary on comparative mapping:

- A. Some conservations 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.

Big questions are:

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?