ELSEVIER

Re-evaluating the relevance of ancestral shared synteny as a tool for crop improvement

Michel Delseny

In addition to the *Arabidopsis* and rice genomic sequences, numerous expressed sequence tags (ESTs) and sequenced tag sites are now available for many species. These tools have made it possible to re-evaluate the extent of synteny and collinearity not only between *Arabidopsis* and related crops or between rice and other cereals but also between *Arabidopsis* and rice, between *Arabidopsis* and other dicots, and between cereals other than rice. Major progress in describing synteny relies on statistical tests. Overall, the data point to the occurrence of ancestral genome fragments in which a framework of common markers can be recognised. Micro-synteny studies reveal numerous rearrangements, which are likely to complicate map-based cloning strategies that use information from a model genome.

Addresses
Laboratoire Génome et Développement des Plantes, UMR 5096, CNRS–IRD–UP, University of Perpignan, 66860 Perpignan, France
e-mail: delseny@univ-perp.fr

Current Opinion in Plant Biology 2004, **7**:126–131

This review comes from a themed issue on
Genome studies and molecular genetics
Edited by Joseph R Ecker and Doug Cook

1369-5266/$ – see front matter

DOI 10.1016/j.pbi.2004.01.005

Abbreviations
BAC bacterial artificial chromosome
EST expressed sequence tag

Introduction

One of the major arguments in favour of sequencing the *Arabidopsis* and rice genomes was that they were considered to be 'relatively simple' compared to the genomes of other cultivated crops. They should serve as reference models to provide information on the more complex genomes of major crops. Particularly appealing was the preliminary observation that the rice genome is essentially collinear with that of other cereals [1]. As a result, a concentric circle model for aligning cereal genomes was deduced [2].

The conservation of synteny between the *Arabidopsis* and *Brassica* genomes became rapidly accepted, but synteny was predicted to be much more limited when genomic comparisons were extended to botanical families beyond the Crucifereae [3]. Subsequently, further work has been carried out at both the macro-scale and the level of sequence, although only very limited regions have been tested at the level of sequence. These studies broadly confirmed previous results but also revealed numerous rearrangements and breakages of synteny, indicating that the use of the original model [2,4] to facilitate the cloning of important genes from complex genomes might not be as easy as anticipated [5•,6•]. Since the time these reviews were written, much new information and many new tools have become available, allowing a more precise and general reassessment of collinearity between plant genomes and its use for crop improvement. Among the major new resources are the almost complete sequences of the *Arabidopsis* and rice genomes [7–10], corresponding websites, an enormous number of plant expressed sequence tags (ESTs) in the dbEST database [11•], large collections of *Arabidopsis* and rice full-length cDNAs [12], and databases for annotating and comparing cereal genomes [13–15]. Surprisingly, the sequencing of the *Arabidopsis* genome revealed extensive segmental duplications that presumably corresponding to several ancient polyploidy events, the most recent being dated somewhere between 30 and 50 million years ago [16–18]. EST sequencing also revealed that most plant genes belong to multigene families. These data have now to be taken into account when attempting to compare genomes. In this review, we describe the improved strategies for comparative mapping at both macro-scale and micro-scale levels. We summarise what has been learned about the most extensively studied species and discuss the chance of isolating important genes using comparative mapping strategies.

Technological difficulties in comparative mapping and their solutions

The basis of comparative mapping is the use of homologous probes that cross-hybridise between the two species being compared. This unavoidably restricts the number of loci that can be analysed. A second difficulty comes from the fact that many genes are members of gene families. Accordingly, it is often difficult to determine if a gene mapped in the second species is orthologous or paralogous to that in the first species. In high-throughput comparative mapping, probe selection has to be optimised. Within the European Union programme EuDicotMap, an attempt was made to first select as probes *Arabidopsis* ESTs that have a high nucleotide-similarity score with rice ESTs (BLASTN e value of $<10^{-20}$). It was assumed that such probes would cross-hybridise with genomes that are more closely related than *Arabidopsis* and rice [19••]. However, many probes did not

respond as expected [6•,19••]. EST resources and other sequence-tagged sites that were developed for each crop were extensively used. Databases were searched for the best homologue to the *Arabidopsis* or the rice gene, which was subsequently mapped on the appropriate population. A threshold e value of BLASTN ($e < 10^{-10}$) or BLASTX ($e < 10^{-15}$) was finally chosen, above which the homologous ESTs were discarded [19••,20••]. In addition, when multiple hits are observed, it is necessary to determine the 3′ end of the mRNA in order to identify the best homologue, as this region is usually the least conserved between members of multigene families. Such a strategy, with a similar $e < 10^{-15}$ or more stringent threshold, was used independently to compare the *Arabidopsis* and *Medicago truncatula* genomes [21]. As a result of the EuDicotMap project, approximately 700 *Arabidopsis* genes have been used as anchor markers and 220 putative homologous loci have been mapped in sugar beet [19••], 227 in *Prunus* [19••], 322 in potato [19••,20••] and 350 in sunflower [19••]. In addition, 212 new loci have been mapped in *Brassica oleracea* [22]. Wheat [23] and maize maps [8,24••] have been similarly compared with the rice genomic sequence. Partial comparisons were made between rice and sorghum chromosomes [25••,26], and an extensive comparison was carried out between maize and sorghum maps [27].

When ESTs or homologous genomic sequences have been mapped to *Arabidopsis*, rice and target genomes, the observed patterns are often rather complex. Therefore, it is essential to evaluate objectively whether the presence of a few genes in the same order on two chromosomal segments in two species occurs by chance or is truly significant. This is not a trivial question and several recent papers have addressed this problem. Whatever the statistics used, the basic assumption is that orthologous loci are compared. Each method essentially uses simulations to calculate the probability that *n* markers in the same order within a given physical or genetic map interval in two chromosomal segments occur by chance. Several computer programmes, such as LineUp [28••] or ADHoRE [29••], have recently been developed to align genomes and to measure synteny [30]. Similar methods, used to compare the potato and *Arabidopsis* genomes, calculated that 6.32 syntenic blocks of three markers would occur by chance within a 20 cM potato map interval and a 1 Mbp *Arabidopsis* interval. This value decreases to 0.41 for four markers and is null for five markers within the same interval [20••]. Such simulations have to be run for each comparison and several studies, in which one of these softwares has been used, now give an evaluation of the quality of the syntenic blocks [19••,20••,31•]. Direct comparisons between rice and *Arabidopsis* at the sequence level have used the same approach. From 10 000 permutations, just 42.4, 1.77, 0.076 and 0.002 syntenic regions in two bacterial artificial chromosomes (BACs) that share 3, 4, 5 or 6 coding sequences in common, respectively, are likely to occur by chance. It was therefore considered that sequences that share four coding sequences in common had met an acceptable and significant probability threshold and should be considered to be syntenic [32••].

Alignment between genomes is limited by several factors that relate to the evolution of genomes. Plant genomes have undergone numerous polyploidy events [17]. Segmental duplications have also been observed in rice [24••,32••,33]; in this case, it was concluded that rice, as well as other cereals, might be an ancient aneuploid [33]. Duplication problems can be overcome by a phylogenetic analysis of the markers [17,34••]. In addition, large genomes have been regularly invaded by retroelements that are responsible for large size differences. This expansion phenomenon is counteracted by deletion of these elements through illegitimate recombination. All these events post-date the divergence between species and contribute to the reshuffling of the original ancestral genomes [5•,35,36].

A final difficulty in evaluating synteny between genomes is that the data are very incomplete and partial. Furthermore, the data are sometimes biased because probes that give simple hybridisation patterns are selected. At the sequence level, only *Arabidopsis* and rice can be compared extensively [8,32••]. At present, comparisons with other species rely on very limited regions, usually no more than a few BAC clones [5•]. At the macro-scale, the resolution of the genetic maps is in the order of one or a few centimorgans, which means than no more than one gene out of several hundred can presently be matched with its homologue in the compared species.

Extent of synteny between *Arabidopsis*, rice and other crops: evidence for ancestral genome fragments

Despite the limitations in comparing genome organisation, evidence for ancestral chromosomal segments is accumulating from both classical genetic mapping and comparative sequencing of homologous regions. The *Arabidopsis* genome has served as a basis for comparison with *Brassica* and, more recently, with crops outside of the crucifer family. Although initial data suggested excellent collinearity between *Arabidopsis* and *Brassica* chromosome segments, present data indicate that the situation is more complex. Because most segments are already duplicated in *Arabidopsis* and because *Brassica* species are cryptic polyploids, there are multiple fragments in *Brassica* that correspond to an *Arabidopsis* region and *vice versa* [22,37,38]. In many cases, the orthology relationship can be clarified by polygenetic analysis. In addition, rearrangements are often detected at the sequence level [6•]. A more rigorous analysis, using genomic probes and statistical treatment, identified 38 significant *Arabidopsis* regions that are collinear with more than 28% of the

B. oleracea genetic map. These regions correspond to 3.3 markers on average in common between 2.1 Mbp in *Arabidopsis* and 2.5 cM on the *B. oleracea* genetic map [31•].

The next most extensively studied pairs are *Arabidopsis*–potato [19••,20••], *Arabidopsis*–sugar beet [19••], *Arabidopsis*–sunflower [19••], *Arabidopsis–Prunus* [19••] and *Arabidopsis*–cotton [34••]. In the two former studies, statistical analysis of the significance of the collinear segments was carried out and the duplicated nature of the *Arabidopsis* genome was taken into account. As a result, 57 *Arabidopsis* segments, which were organised in 14 chromosomal regions, were recognised as having at least one homologous collinear segment in the genome of one of the four analysed species. *Arabidopsis* and sugar beet share 27 collinear blocks, whereas *Arabidopsis* shares 24, 37 and 49 collinear blocks with sunflower, *Prunus* and potato, respectively. These segments span between 16% and 33% of the *Arabidopsis* genome. A more detailed study on the potato–*Arabidopsis* comparison, which included blocks with lower significance, discovered 90 segments covering 41% of the potato genetic map and 50% of the *Arabidopsis* physical map [20••]. From the comparison between *Arabidopsis* and cotton, which used a phylogenetic approach, an ancestral gene order could be inferred for *Arabidopsis* that revealed more synteny with other dicots [34••].

A remarkable point emerges from the comparisons described above [19••]. Out of the 57 *Arabidopsis* blocks that have a homologous collinear segment in the genomes of potato, sugar beet, *Prunus* or cotton, 16 are shared by all four other species and 32 are shared by two or three species. This indicates that these regions correspond to an ancestral genome, and that they should contain several essential genes that have led to their being roughly conserved over probably more than 70–100 million years. These regions should be priority targets for genomic sequencing in these four crops, which will determine the extent to which the gene order is conserved and what percentage of genes are found in similar positions. The observation of a frame of markers that are common to several species within certain *Arabidopsis* chromosomal regions strongly argues for a common ancestral origin of these chromosome segments. Incidentally, these studies also revealed that many more duplicated segments exist in the genomes of these crops than had been anticipated initially [19••,34••]. Several studies reporting micro-synteny between *Arabidopsis* and tomato have been discussed in a recent review [6•].

Attempts to compare the *Arabidopsis* genomes with those of legumes, specifically pea and *Medicago sativa*, in the EuDicotMap project failed to detect any significant collinear region, and an identical conclusion was reached in a recent comparison of *Arabidopsis* and *Medicago truncatula* [21]. It is quite possible, however, that not enough markers have been mapped to detect significant groups of markers because micro-synteny is evident at the sequence level in some regions of *M. truncatula*. Micro-synteny between *Arabidopsis* and *M. truncatula* seems to be more extensive than that between *Arabidopsis* and rice, which is expected as crucifers and legumes diverged more recently than crucifers and cereals; 17 out of 40 sequenced loci showed a significant level of micro-synteny between *Arabidopsis* and *M. truncatula* [21]. Two homologous regions in *M. sativa* and *Lotus japonicus* also show micro-synteny with *Arabidopsis* [39,40].

Attempting to detect blocks of markers that are common to the *Arabidopsis* and rice maps also initially failed completely. When the sequences were compared directly, however, several collinear regions were identified. In a comparison of 190 Mbp of rice genomic sequence, 60 significant syntenic regions were detected, each of which spanned between 4 and 22 orthologues that are common to rice and *Arabidopsis* [32••]. A higher figure of 137 *Arabidopsis*–rice syntenic groups was reported after completion of the draft genome [8]. Several segments that are collinear in *Arabidopsis* and rice have been used to reveal hidden duplications in both species [33]. Traces of a common origin can therefore be recognised in *Arabidopsis* and rice, although the two species diverged more than 200 million years ago and the ancestral genome has been almost completely scrambled and reshuffled.

Comparable studies, attempting to use the rice genome sequence as the reference, have been made in cereals. These studies have met two additional difficulties. The first is that the rice genome sequence is not as completely assembled and annotated as that of *Arabidopsis*. The second is that many markers in maize and wheat are not precisely mapped but are assigned to bins. The first alignment of the rice sequence with 610 maize homologous cDNA markers, using a low stringency, revealed a complex pattern [8]. A more detailed study [24••], involving more than 2600 mapped sequenced markers among which only 656 putative orthologous genes could be identified, gave a similar result, pointing to many more rearrangements than had been anticipated from the concentric circles model [2,4]. The wheat genetic map was also recently compared to the rice sequence [23]. This work also points to numerous rearrangements between the two genomes, with a high frequency of breakdowns in collinearity.

Extensive comparisons have also been made between sorghum and rice [25••,26]. Indeed, a sorghum physical map exists, and an interesting new approach has been developed to align this map with the rice map [25••]. Sorghum BAC clones were selected from the minimum tiling path of chromosome 3. Unique partial sequences were obtained from each BAC clone and could be directly compared with the rice sequence. This approach revealed

excellent conservation between the overall structure and gene order of sorghum chromosome 3 and rice chromosome 1 but also indicated several rearrangements: 50 of the 118 BACs examined did not show any sequence similarity and five BACs showed better collinearity with rice BACs located elsewhere in the rice genome. 63 BACs are collinear with rice BACs from chromosome 1, but four of them do not respect the rice minimum tiling path and therefore correspond to different positions in the two homologous chromosomes. Together, these studies point to a general conservation of large syntenic blocks within cereals, but with many more rearrangements and synteny breakdowns than anticipated.

This trend is even more obvious when synteny is analysed at the sequence level [41,42,43••,44••]. Sequencing of the domestication locus *Q* in *Triticum monococcum* revealed excellent collinearity with the bread wheat genetic map [41]. Following sequencing the leaf-rust-resistance locus *Rph7* from barley, it was observed that this locus is flanked by two *HGA* genes. The orthologous locus in rice chromosome 1 consists of five *HGA* genes. In barley, only four of the five *HGA* genes are present, one is duplicated as a pseudogene and six additional genes have been inserted in between the *HGA* genes. These six genes have homologues on eight different rice chromosomes [42]. The most striking rearrangement was revealed by the comparison of 100 kb around the *Bronze* locus of two maize lines. Not only does the retrotransposon distribution differ between the two lines but the genes themselves could also be different [43••]. Comparison of the low molecular weight glutenin locus between *T. monococcum* and *Triticum durum* also revealed dramatic rearrangements: more than 90% of the sequence diverged because of retroelement insertions and because different genes are present at this locus [44••]. Therefore collinearity can be lost very rapidly within two genomes from the same species.

To what extent can collinearity be used to isolate important genes?

The usefulness of the collinearity between the genomes of model plants and important crops can be assessed by the number of failures or successes in its exploitation. In most recently reported cases, collinearity has been useful in providing additional markers with which to saturate fine genetical and physical maps, particularly those for *Brassica*, tomato and cereals [5•,6•].

The *Lr21* leaf-rust-resistance gene of bread wheat was successfully isolated using a strategy of shuttle-mapping between diploid wheat as a model and bread wheat [45•]. Most of the time, however, there are breakages in microsynteny that prevent the straightforward identification of a candidate gene for a given trait. This was the case when attempts were made to isolate the leaf-rust-resistance gene *Rph7* [42] or the photoperiod response gene *Phd-H1* [46] from barley. A similar story was reported for the *Rfo* restorer genes isolated from radish: markers flanking these genes in radish are collinear with the *Arabidopsis* sequence, but the gene itself is not present in *Arabidopsis* although many homologues are present elsewhere in the *Arabidopsis* genome [47,48]. An important gene in the establishment of nodule symbiosis was cloned in both *Medicago* and *Lotus* by classical map-based cloning. When the BACs containing these loci were compared with the *Arabidopsis* genome sequence, micro-collinearity was obvious but again the key gene was absent from the *Arabidopsis* genome [39,40].

Conclusions

Recent studies have given stronger support to the concept of an ancestral genome for plant species. Improved methods have been developed for the establishment of statistically significant collinearity between genome segments. At the same time, it has become increasingly obvious that plant genomes belonging to different botanical families have been extensively reshaped during millions of years of evolution, allowing each species to adapt to its ecological niche. Even within a family, many rearrangements can be detected now that we have enough sequenced-based markers and as genomic sequences can now be compared directly. The present evidence from cereals points to extremely fast evolution of the intergenic regions, with consequences for gene conservation. The use of a shuttle-mapping strategy has to be evaluated on a case-by-case basis, and even then, the numerous pitfalls of this approach must be kept in mind. The present information, from both successes and failures, strongly suggests that the development of efficient tools for isolating genes of agronomic importance within each important family should continue to be a priority, and that restricting ourselves to use the two present model species, *Arabidopsis* and rice, would be unwise.

Acknowledgements

The author acknowledges support of the European Union EuDicotMap programme (Bio4-CT 97-2170), of the Montpellier Languedoc-Roussillon Genopole and of the French Plant Genomic Genoplante programme, without which this work would not have been possible. The contribution of numerous colleagues in discussions and comments is gratefully acknowledged; Guillaume Blanc, Richard Cooke, Benoît Piegu, Jérôme Salse and Steve Barnes have been particularly helpful.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Ahn S, Tanksley SD: **Comparative linkage maps of the rice and maize genomes**. *Proc Natl Acad Sci USA* 1993, **90**:7980-7984.

2. Moore G, Devos KM, Wang Z, Gale M: **Cereal genome evolution. Grasses, line up and form a circle**. *Curr Biol* 1995, **5**:737-739.

3. Paterson AH, Lan TH, Reischmann KP, Chang C, Lin YR, Lin SC, Burow MD, Kowalski SP, Katsar CS, DelMonte TA *et al.*: **Toward a unified genetic map of higher plants transcending the monocot–dicot divergence**. *Nat Genet* 1996, **14**:380-382.

4. Devos KM, Gale MD: **Genome relationships: the grass model in current research**. *Plant Cell* 2000, **12**:637-646.

5. • Bennetzen JL, Ma J: **The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis**. *Curr Opin Plant Biol* 2003, **6**:128-133.
The most recent review of collinearity between rice and other cereals at the sequence level.

6. • Barnes S: **Comparing *Arabidopsis* to other flowering plants**. *Curr Opin Plant Biol* 2002, **5**:128-133.
A recent comparison of the *Arabidopsis* genome with dicot crop genomes.

7. The *Arabidopsis* Genome Initiative: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana***. *Nature* 2000, **408**:796-815.

8. Goff SA, Rickle D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H *et al.*: **A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*)**. *Science* 2002, **296**:92-100.

9. Yu J, Hu S, Wang J, Wong GKS, Li S, Lin B, Deng Y, Dai L, Zhou Y, Zhang X *et al.*: **A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*)**. *Science* 2002, **296**:79-92.

10. Delseny M: **Towards an accurate sequence of the rice genome**. *Curr Opin Plant Biol* 2003, **6**:101-105.

11. • Rudd S: **Expressed sequence tags: alternative or complement to whole genome sequence**. *Trends Plant Sci* 2003, **8**:321-329.
An excellent review on plant ESTs. The authors discuss the availability, quality and limitations of plant ESTs, as well as their use in describing genomes, in identifying polymorphisms and in comparative genomics.

12. Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yasaki J, Ishikawa M, Yamada H, Oaka H *et al.*: **Collection, mapping and annotation of over 28 000 cDNA clones from *japonica* rice**. *Science* 2003, **301**:376-379.

13. Ware DH, Jaiswal P, Ni J, Yap IV, Pan X, Clark KY, Teytelman L, Schmidt SC, Zhao W, Chang K *et al.*: **Gramene, a tool for grass genomics**. *Plant Physiol* 2002, **130**:1606-1613.

14. Yuan Q, Ouyang S, Lin J, Suh B, Cheung F, Sultana R, Lee D, Quackenbush J, Buell CR: **The TIGR rice genome annotation resource: annotating the rice genome and creating resources for plant biologists**. *Nucleic Acids Res* 2003, **31**:229-233.

15. Dong Q, Roy L, Freeling M, Walbot V, Brendel V: **ZmDB, an integrated database for maize genome research**. *Nucleic Acids Res* 2003, **31**:244-247.

16. Simillion C, Van de Poele K, Van Montagu MCE, Zabeau M, Van de Peer Y: **The hidden duplication past of *Arabidopsis thaliana***. *Proc Natl Acad Sci USA* 2002, **99**:13627-13632.

17. Blanc G, Hokamp K, Wolfe KH: **A recent polyploidy superimposed on older large scale duplications in the *Arabidopsis* genome**. *Genome Res* 2003, **13**:137-144.

18. Ermolaeva MD, Wu M, Eisen JA, Salzberg SL: **The age of the *Arabidopsis thaliana* genome duplication**. *Plant Mol Biol* 2003, **51**:859-866.

19. •• Dominguez I, Graziano E, Gebhardt C, Barakat A, Berry S, Arus P, Delseny M, Barnes S: **Plant genome archaeology: evidence for conserved ancestral chromosome segments in dicotyledonous plant species**. *Plant Biotechnol* 2003, **1**:91-99.
The authors of this paper are the first to attempt to compare several dicotyledonous plant genomes, including the *Arabidopsis* genome as a model. This study revealed collinear chromosome segments that have markers in the same order in several species, suggesting that they are ancestral segments. This paper also includes statistical treatments and describes the evolutionary story of the *Arabidopsis* genome.

20. •• Gebhardt C, Walkemeier B, Henselewski H, Barakat A, Delseny M, Stüber K: **Comparative mapping between potato (*Solanum tuberosum*) and *Arabidopsis thaliana* reveals structurally conserved domains and ancient duplications in the potato genome**. *Plant J* 2003, **34**:529-541.
One of the most detailed comparisons between *Arabidopsis* and a dicot species from outside the *Brassica* tribe. The authors also include a statistical treatment of the data.

21. Zhu H, Kim DJ, Back JM, Choi HK, Ellis LC, Küester H, McCombie WR, Peng HM, Cook DR: **Syntenic relationships between *Medicago truncatula* and *Arabidopsis* reveal extensive divergence of genome organisation**. *Plant Physiol* 2003, **131**:1018-1026.

22. Babula D, Kaczmarek M, Barakat A, Delseny M, Quiros CF, Sadowski J: **Chromosomal mapping of *Brassica oleracea* based on ESTs from *Arabidopsis thaliana*: complexity of the comparative map**. *Mol Genet Genomics* 2003, **268**:656-665.

23. Sorrells ME, La Rota M, Bermudez-Kandianis CE, Greene RA, Kentety R, Munkvold JD, Miftahudin, Mahmoud A, Ma XF, Gustafson PJ *et al.*: **Comparative DNA sequence analysis of wheat and rice genomes**. *Genome Res* 2003,**13**:1818-1827.

24. •• Salse J, Piegu B, Cooke R, Delseny M: **New *in silico* insight into the synteny between rice (*Oryza sativa* L.) and maize (*Zea mays* SL.) highlights reshuffling and identifies new duplications in the rice genome**. *Plant J* 2004, in press.
The authors describe the most elaborate maize–rice comparison to date, which was based on the analysis of 2600 mapped maize ESTs. Once again, this work provides evidence of rearrangements and duplications in the maize and rice genomes.

25. •• Klein PE, Klein RR, Vrebalov J, Mullet JE: **Sequence-based alignment of sorghum chromosome 3 and rice chromosome 1 reveals extensive conservation of gene order and one major chromosomal rearrangement**. *Plant J* 2003, **34**:605-621.
This paper describes an original, powerful and cost-effective comparison of two maps, which was based on the partial sequencing of ordered BAC clones.

26. Yu Y, Rambo T, Currie J, Saski C, Kim HR, Collura K, Thompson S, Simons J, Yang TJ, Nah G *et al.*: **In depth view of structure, activity and evolution of rice chromosome 10**. *Science* 2003, **300**:1566-1569.

27. Bowers JE, Abbey C, Anderson S, Chang C, Draye X, Hoppe AA, Jessup R, Lemke C, Lennington J, Li Z *et al.*: **A high density genetic recombination map of sequence-tagged sites for sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses**. *Genetics* 2003, **165**:367-380.

28. •• Hampson S, McLysaght A, Gaut B, Baldi P: **LineUp: statistical detection of chromosomal homology with application to plant comparative genomics**. *Genome Res* 2003, **13**:1-12.
The authors present a key tool for comparing genomes and duplicated segments. LineUp is a refinement of previous methods used to establish the duplicated nature of the maize genome.

29. •• Van de Poele K, Saeys Y, Simillion E, Raes J, Van de Peer Y: **The automatic detection of homologous regions (ADHoRE) and its application to micro-collinearity between *Arabidopsis* and rice**. *Genome Res* 2002, **12**:1792-1801.
The automatic detection of homologous regions (ADHoRE) software searches for micro-collinearity at the sequence level and takes into account different types of rearrangements.

30. Housworth EA, Postlethwait J: **Measures of synteny conservation between species pairs**. *Genetics* 2002, **162**:441-448.

31. • Lukens L, Zou F, Lydiate D, Parkin I, Osborn T: **Comparison of a *Brassica oleracea* genetic map with the genome of *Arabidopsis thaliana***. *Genetics* 2003, **164**:359-372.
This comparison of the genetic maps of *B. oleracea* and *A. thaliana* used genomic sequences rather than cDNA. The authors discuss statistical problems and improve previous software that was originated in Gaut's laboratory.

32. •• Salse J, Piegu B, Cooke R, Delseny M: **Synteny between *Arabidopsis thaliana* and rice at the genome level: a tool to identify conservation in the ongoing rice genome sequencing project**. *Nucleic Acids Res* 2002, **30**:2316-2328.
A statistically validated estimation of conserved synteny between rice and *Arabidopsis* at the sequence level.

33. Van de Poele K, Simillion C, Van de Peer Y: **Evidence that rice and other cereals are ancient aneuploids**. *Plant Cell* 2003, **15**:2192-2202.

34. •• Bowers JE, Chapman BA, Rong J, Paterson AH: **Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events**. *Nature* 2003, **422**:433-438.

An excellent illustration of the use of phylogenetic analysis to identify orthologues and to reconstruct an ancestral genome.

35. Devos KM, Brown JKM, Bennetzen JL: **Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis***. *Genome Res* 2002, **12**:1075-1079.

36. Li W, Gill BS: **The colinearity of the *Sh2/A1* orthologous region in rice, sorghum and maize is interrupted and accompanied by genome expansion in the Triticeae**. *Genetics* 2002, **160**:1153-1162.

37. Li G, Gao M, Yang B, Quiros CF: **Gene for gene alignment between *Brassica* and *Arabidopsis* genomes by direct transcriptome mapping**. *Theor Appl Genet* 2003, **107**:168-180.

38. Parkin IA, Lydiate DJ, Trick M: **Assessing the level of collinearity between *Arabidopsis thaliana* and *Brassica napus* for *A. thaliana* chromosome 5**. *Genome* 2002, **45**:356-366.

39. Endre G, Kereszt A, Kevei Z, Milhacea S, Kalo P, Kiss GB: **A receptor kinase gene regulating symbiotic nodule development**. *Nature* 2002, **417**:962-966.

40. Stracke S, Kistner C, Yoshida S, Mulder L, Sato S, Kaneko T, Tabata S, Sandal N, Stougoard J, Szczyglowski K *et al.*: **A plant receptor-like kinase required for both bacterial and fungal symbiosis**. *Nature* 2002, **417**:959-962.

41. Faris JD, Fellers JP, Brooks SA, Gill BS: **A bacterial artificial chromosome contig spanning the major domestication locus *Q* in wheat and identification of a candidate gene**. *Genetics* 2003, **164**:311-321.

42. Brunner S, Keller B, Feuillet C: **A large rearrangement involving genes and low copy DNA interrupts the micro-collinearity between rice and barley at the *Rph7* locus**. *Genetics* 2003, **164**:673-683.

43. Fu H, Dooner HK: **Intraspecific violation of genetic colinearity**
•• **and its implications in maize**. *Proc Natl Acad Sci USA* 2002, **99**:9573-9578.
A fascinating paper showing that a 100 kbp region around the major *Bronze* locus has undergone important rearrangements that resulted in the violation of collinearity within the same species.

44. Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu ZD,
•• Dubcowski J, Keller B: **Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and Am genomes of wheat**. *Plant Cell* 2003, **15**:1186-1197.
The authors present a detailed sequence analysis of an orthologous locus in two closely related species. This work reveals the extremely fast evolution of the wheat genome during the past 5 million years.

45. Huang L, Brooks SH, Li W, Fellers JP, Trick HN, Gill BS: **Map based**
• **cloning of leaf rust resistance gene *Lr21* from the large and polyploid genome in bread wheat**. *Genetics* 2003, **164**:655-664.
The authors successfully illustrate the shuttle-mapping strategy using a diploid and a hexaploid wheat.

46. Dunford RP, Yano M, Kurata N, Sasaki T, Huestis G, Rocheford T, Laurie DA: **Comparative mapping of the barley *Phd-H1* photoperiod response gene region, which lies close to a junction between two rice linkage segments**. *Genetics* 2002, **161**:825-834.

47. Brown GG, Formanova N, Jin H, Wargachuk R, Dondy C, Patil P, Laforest M, Zhang J, Cheung WY, Landry BS: **The radish *Rfo* restorer gene of Ogura cytoplasmic mole sterility encodes a protein with multiple pentatricopeptide repeat**. *Plant J* 2003, **35**:262-272.

48. Desloire S, Gherbi H, Laloui W, Marhadour S, Clouet V, Catholico L, Falentin C, Giancola S, Renard M, Budar F *et al.*: **Identification of the fertility restoration locus *Rfo* in radish as a member of the pentatricopeptide repeat family**. *EMBO Rep* 2003, **4**:588-594.