

## Introduction

Genomic divergence due to tandem duplications, block duplications, inversions and gene loss is extensive in plant genomes, as well as the number of chromosomal rearrangements. Nevertheless, it is expected that different plant species still exhibit a certain degree of conservation regarding gene content and gene order. Therefore, comparative genomics should enable us to transfer information from a reference model system (with a small genome size) to other species (with larger genome sizes).

We developed a tool, called ADHoRe, that automatically detects genomic regions with statistically significant conserved gene content and order. This tool can be used for analyses within one genome (to look for paralogous regions with duplicated genes) or for comparisons between genomes of different organisms (to look for colinearity). The algorithm was especially designed to cope with micro-rearrangements (events frequently observed within different plant lineages), since these greatly complicate the search for colinear regions. This study is performed since the analysis of gene organization at the micro-level and the evolution of duplicated genes provides a basis for the understanding of plant genome evolution.

## Methods & dataset

We have used the ADHoRe (Automatic Detection of Homologous Regions) algorithm to detect the degree of gene conservation between Arabidopsis<sup>1</sup> and Rice, two distantly related species which diverged from one another around 200 MYA. Genomic data from 7 Rice chromosomes (in total 1.274 BACs containing around 33.000 genes using the RiceGAAS gene annotation system<sup>2</sup>) were used in the comparison with Arabidopsis. In order to detect large segments of conserved gene content, clusters of overlapping Rice BACs were generated, where possible.

Table 1. Overview of the Rice dataset used in this study.

Chromosome	Percent sequenced*	Total dataset				Overlapping BACs		
		MB	# BAC	# annotated genes	Gene Density (gene/kB)	# BACs	MB	Genes
1	100.0	50.68	370	10,300	4.92	266	34.97	6,237
2	44.7	18.40	154	3,692	4.98	2	0.30	38
4	92.4	18.90	143	3,766	5.02	75	10.76	2,064
6	63.0	21.57	159	4,410	4.89	6	0.94	163
7	75.5	20.33	164	4,149	4.90	7	0.86	168
8	46.2	16.85	139	3,398	4.96	24	3.55	615
10	95.6	19.28	145	3,806	5.07	73	10.83	1,892
Total	38.6	166.01	1274	33,521	4.95	453	62.21	11,177

\* TIGR status January 14th

## Results

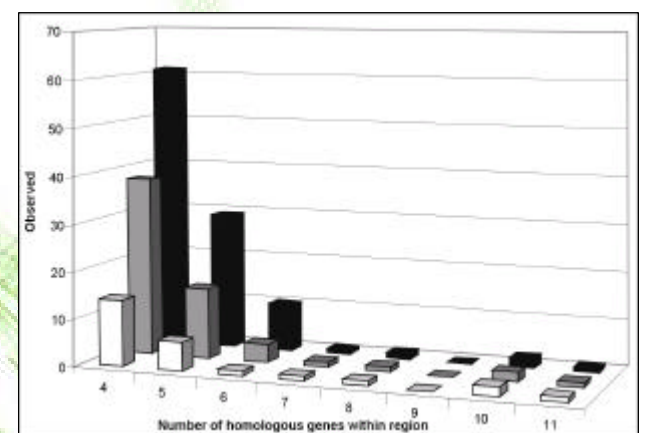


Figure 1. Distribution of the number of conserved genes within colinear regions between Arabidopsis and rice. Black, grey and white histograms represent the 0%, 95% and 99% significance level, respectively.

Within the collection of clustered Rice BACs, several large genomic segments could be detected showing microcolinearity with Arabidopsis. The largest regions are spanning multiple Rice BACs (250 - 300 kb) and contain about 10 homologous genes with conserved gene order in Arabidopsis. These

genes located within a syntenic region are extensively interspersed by non-homologous genes in both plant species. In the largest syntenic regions (containing 7 or more conserved genes), each conserved Arabidopsis gene is interspersed with on average 3 non-homologous genes. This demonstrates the high level of micro-rearrangements occurred since the monocot / dicot divergence.

For individual, non-overlapping Rice BACs, still significant stretches of gene colinearity with Arabidopsis can be found (with a maximum of 5 homologous Rice genes showing conserved gene order (see Figure 3)). In total 70 individual Rice BACs were identified having 4 or more genes colinear in Arabidopsis. With more assembled Rice sequence data becoming available, these fragments potentially can generate longer segments of colinearity. Therefore, the level of large Rice segments showing gene colinearity with Arabidopsis could be much higher than shown in this analysis.

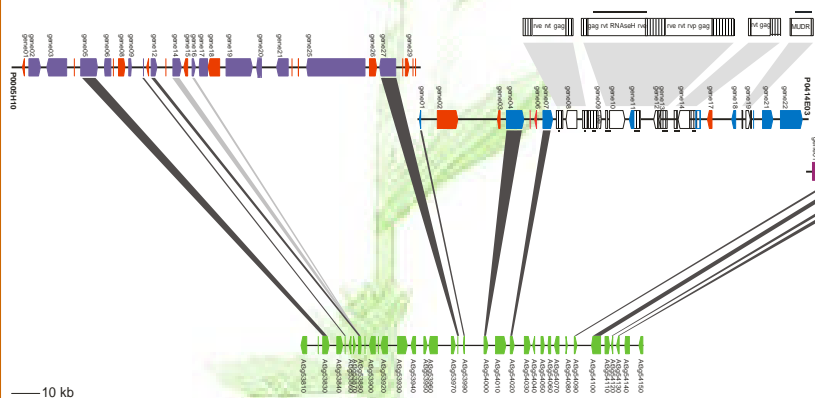
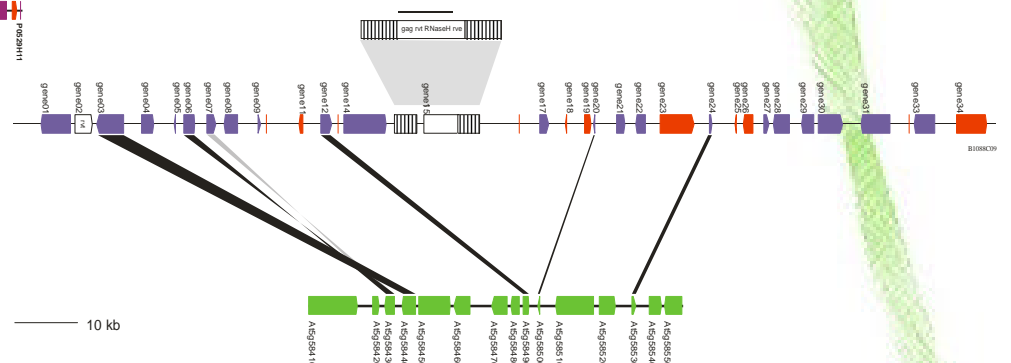


Figure 2. Microsynteny between overlapping Rice BACs and Arabidopsis chromosome 3.

Figure 3. Microsynteny between an individual Rice BAC and Arabidopsis chromosome 5.



## Conclusions

In this study, we could identify some large genome segments showing microcolinearity between Rice and Arabidopsis. This demonstrates that using improved strategies wide-spread genetic colinearity between monocots and dicots can be found.