Evidence that rice, and other cereals, are ancient aneuploids

Klaas Vandepoele, Cedric Simillion and Yves Van de Peer

Bioinformatics and Evolutionary Genomics Division, Department of Plant Systems Biology, Flanders Interuniversity Institute for Biotechnology (VIB),
Ghent University, K.L. Ledeganckstraat 35, B-9000 Gent, Belgium. E-mail: Klaas.Vandepoele@gengenp.rug.ac.be

Introduction

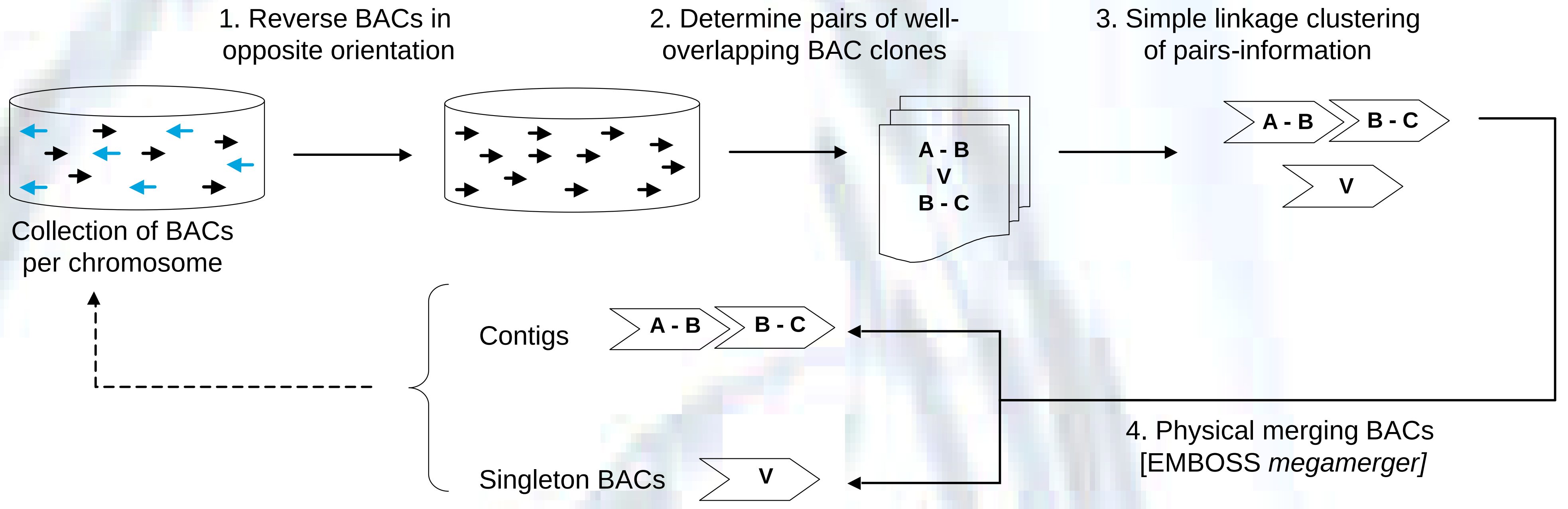
Genome sequencing projects show that entire genome duplications have probably occurred in the evolutionary past of vertebrates, fungi and plants. Large-scale duplication events have been considered important for the evolution of many organisms because they provide a way to considerably increase the genetic material on which evolution can work. Since duplicated genes are redundant, one of the copies is, at least in theory, freed from functional constraint, and can therefore evolve a new function. By applying novel techniques to detect heavily degenerated block duplications in *Arabidopsis thaliana*, we have shown that the genome of this dicotyledonous model plant has been reshaped by not one, but most probably three polyploidization events. Apart from *Arabidopsis thaliana*, rice (*Oryza sativa*) is currently the only plant species for which sequences of the nuclear genome have been published. In strong contrast with *Arabidopsis*, where initial sequencing of the genome sequence already revealed numerous duplicated segments, no clear evidence for large-scale gene or complete genome duplications in rice has been detected so far.

Results

1. The genome sequence of *Oryza sativa* - assembly

- **Input:** 2,987 IRGSP genomic BAC sequences (average size 140kb)
- **Assembly:** Two rounds of ASGAR

- **Output:** 1,025 genomic scaffolds
  (498 contigs and 527 singletons)
  Total size scaffolds: 330.47 Mb
  (average size of 322 kb/scaffold)
- **Annotation:** RiceGAAS proteins mapped on scaffolds


1. Reverse BACs in opposite orientation
2. Determine pairs of well-overlapping BAC clones
3. Simple linkage clustering of pairs-information
Collection of BACs per chromosome
A - B
V
B - C
A - B
B - C
V
Contigs
A - B
B - C
Singleton BACs
V
4. Physical merging BACs [EMBOSS *megamerger]*

ASGAR: Automatic Sequence-to-Genome Assembly Routine

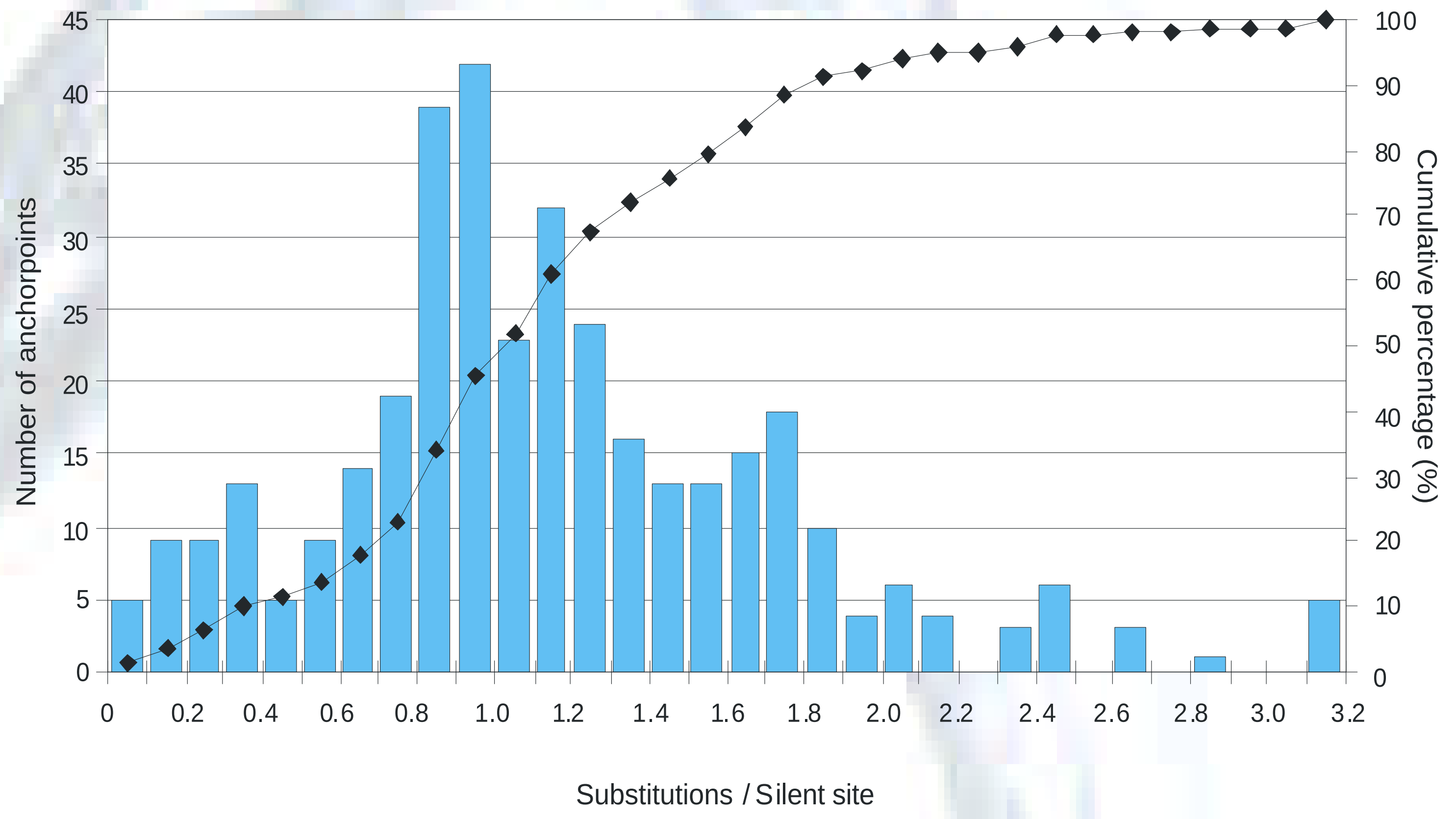
2. Detection and absolute dating of duplicated segments

Application of the ADHoRe algorithm (G=25 & Q=0.9, P<0.001) yielded 193 statistically significant duplicated segments:

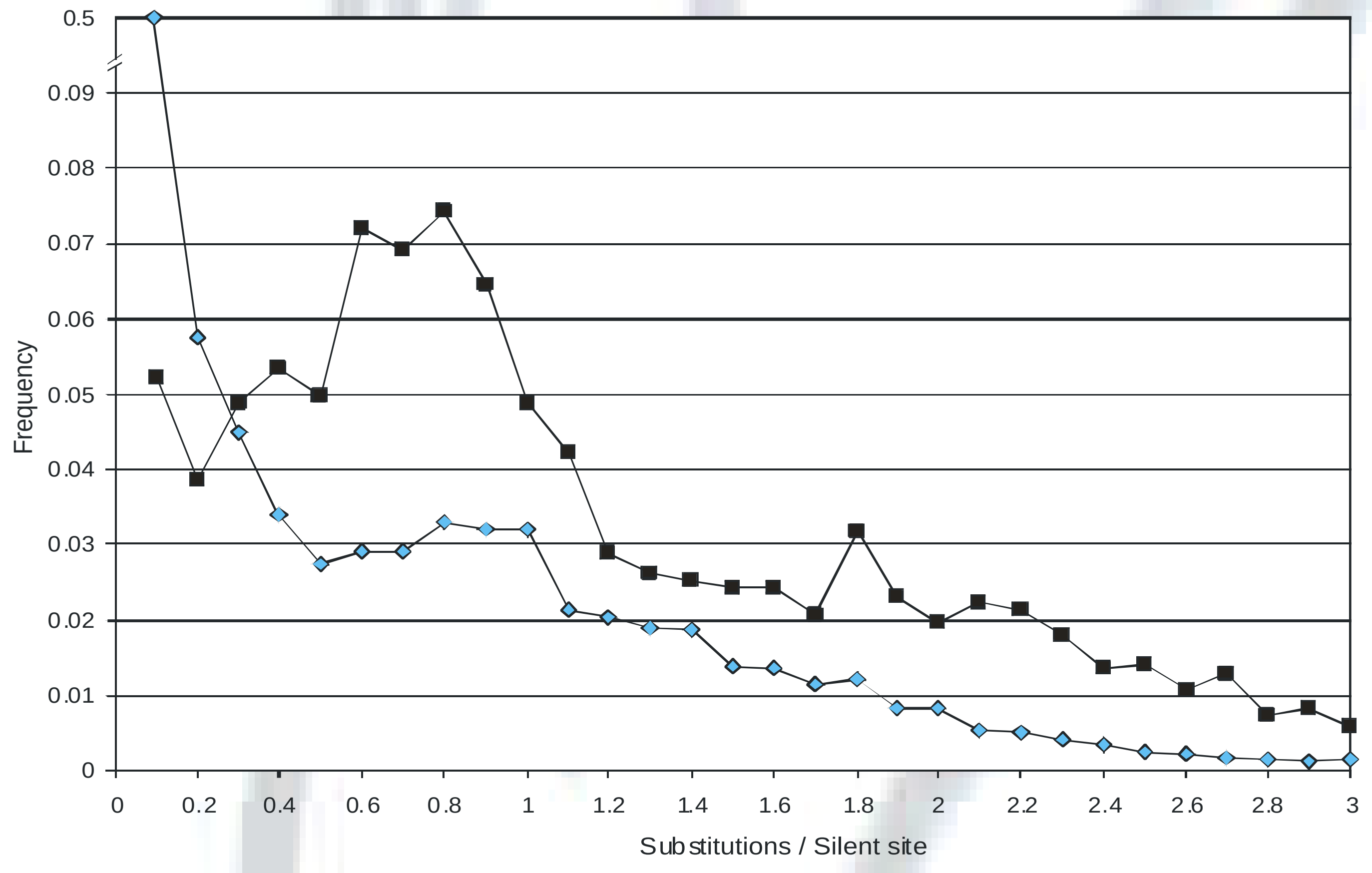
- 150 containing 3 or 4 paralogous gene pairs (so-called anchor points)
- 43 containing five or more duplicates

The complete set of block duplications, omitting tandem duplications, contains 862 anchor points and includes nearly 15% of all rice proteins in our annotated non-redundant genome data set.

K_s-based dating of the anchor points showed that 47% has a K_s value (number of silent substitutions per silent site) between 0.6 and 1.1, which corresponds with an age of 46 and 85 MY, respectively. The median, a K_s value of 0.87, corresponds with 67 MY.

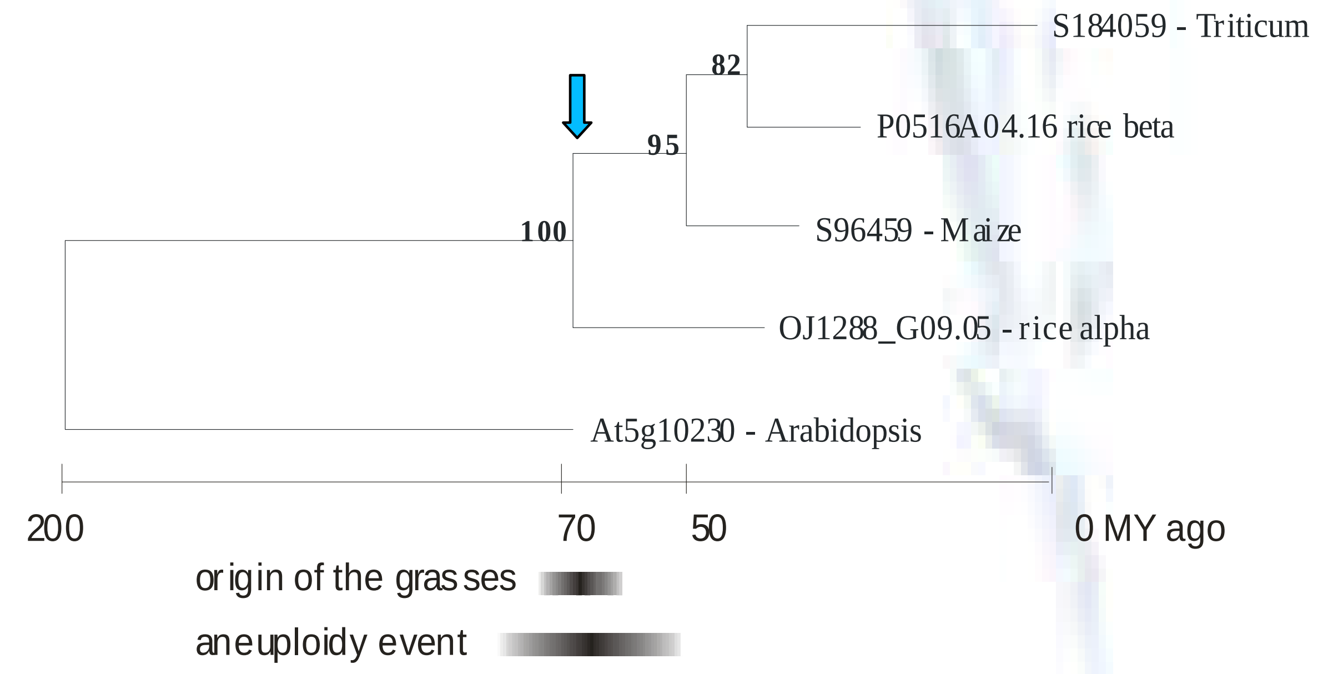


3. The duplication history of *Arabidopsis* and rice



Frequency distribution of pairs of duplicates in Arabidopsis (blue) and rice (purple) as a function of the number of silent substitutions per silent site. All frequencies were corrected for the total number of dated gene duplicates per genome.

4. Relative dating by phylogenetic means



Thirty-two out of 45 tree topologies (*i.e.*, 74%) including two copies of rice (*i.e.*, retained duplicates found in large duplicated segments) and at least one copy of another cereal, show a topology as depicted above in which one rice gene branches off before the divergence of rice and the other cereals. Such tree topologies suggest a duplication prior to the divergence of the cereals rice, barely, wheat, maize, and sorghum, estimated at 50 MY ago, and may have occurred just before or after the origin of the grasses, estimated at 70 MY ago.

Conclusion

Dating of block duplications in the rice genome, their non-uniform distribution over the different rice chromosomes (data not shown), and comparison with the duplication history in *Arabidopsis*, indicates that rice is not an ancient polyploid as previously suggested. Our findings do, however, show that rice, and other cereals, have undergone an aneuploidy event in their evolutionary past, just before or after the origin of the grasses.