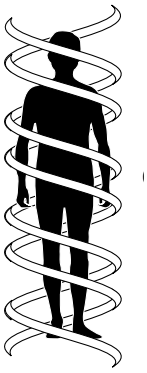
7 Lopez-Bocio, J. *et al.* (2000) Enhanced phosphorus uptake in transgenic tobacco plants that overproduce citrate. *Nat. Biotechnol.* 18, 450–453
8 Ayling, S.M. *et al.* (2000) Transmembrane electric potential difference of germ tubes of arbuscular mycorrhizal fungi responds to external stimuli. *New Phytol.* 147, 631–639
9 Schweiger, P.F. and Jakobsen, I. (1999) The role of mycorrhizas in plant P nutrition: fungal uptake kinetics and genotype variation. In *Plant Nutrition – Molecular Biology and Genetics* (Gissel-Nielsen, G. and Jensen, A., eds), pp. 277–289, Kluwer Academic Publishers
10 Harrison, M.J. and van Buuren, M.L. (1995) A phosphate transporter from the mycorrhizal fungus *Glomus versiforme*. *Nature* 378, 626–629
11 Rosewarne, G.M. *et al.* (1999). A *Lycopersicon esculentum* phosphate transporter (*LePT1*) involved in phosphorus uptake from a vesicular–arbuscular mycorrhizal fungus. *New Phytol.* 144, 507–516
12 Gianinazzi-Pearson, V. *et al.* (2000) Differential activation of $H^+$-ATPase genes by an arbuscular mycorrhizal fungus in root cells of transgenic tobacco. *Planta* 211, 609–613

Comment

# Do plants have more genes than humans?



The most surprising outcome of sequencing the human genome is the small number of predicted genes. Both the International Human Genome Sequencing Consortium and Celera Genomics came to similar conclusions, with estimations of 31 000–32 000 genes[1,2]. However, how much of the actual genome has been sequenced remains speculation because the sequence is in draft form and is not completely contiguous. Therefore, the total number of genes could be higher. The small number of predicted genes was surprising given the large collection of human cDNAs. This discrepancy could be because of post-transcriptional, rather than transcriptional, control of gene function, which can be accomplished by alternative splicing. Indeed, many of the sequenced human genes have alternative splice products. In addition, several other processes (e.g. signal transduction) proceed via further protein modifications, such as glycosylation. Therefore, the number of human protein products could far exceed the number of genes. Interestingly, although it is only one-thirtieth the size of the human genome, the predicted number of genes in *Arabidopsis thaliana* (25 500) is in the same league as the predicted number of human genes[3]. Because both genomes show evidence of genome-wide segmental duplications, this is unlikely to explain the difference in genome size. Alternatively, because the majority of the human genome appears to have expanded intergenic regions, with retroelements as the predominant species, likewise the size variation among plant genomes [some of which are even larger than the human genome (e.g. barley, wheat)] could be because of the insertion of transposable elements into intergenic regions. Although this is a possibility, it might be too simplistic a view, discounting a fundamental difference between plants and animals in the evolution of their gene regulatory mechanisms.

The proposal that rice should be the next plant genome to be sequenced has attracted a lot of attention because it is the staple food for half the population of the world and it has a relatively small genome (four times that of *Arabidopsis* but a one-fourtieth the size of the wheat genome). Therefore, it is not surprising that a draft sequence of the rice genome has been announced twice within the past 12 months by private companies, first Monsanto and now Syngenta[4,5]. However, because of the repetitive DNA in the rice genome, a draft sequence has many gaps, and it might be that sequences are not always anchored to their chromosomal positions. Interestingly, Monsanto recognized this problem and has made its draft available to the public consortium that is seeking to obtain a complete sequence of the rice genome. It is important to close the gaps and determine the exact position of genes in the rice genome[6] because, in spite of the difference in genome size, most of the major crops (including wheat and maize) seem to have a similar genetic layout to rice. Accordingly, the information obtained from rice could be applied to other crops.

So, is the difference in DNA content between different plant genomes simply a question of leaps and bounds of retrotranspositions that have occurred during speciation? At first glance, this appears to be a good hypothesis. In the few cases of sequence comparisons between chromosomal regions of rice and wheat or rice and corn, it appears that intergenic regions have expanded by insertions of retrotransposons, and that larger genomes contain more junk DNA between genes[7]. Interestingly, repetitive DNA between genomes of sorghum and maize do not cross-hybridize, although they have some common retrotransposon families. This indicates that retrotransposition has occurred since speciation[8].

However, the expansion of genome sizes has also occurred during speciation. For example, maize resulted from an allotetraploidization event some 11.4 million years ago from two closely related progenitors[9]. The result is a genome that nearly doubled in gene content. Today, mutations based on single genes can still be detected in maize. Differentiation of the gene pairs following the allotetraploidization event can explain this disomic nature of the current maize genome. We know from several of these gene pairs that their regulation, but not their function, has changed. For example, *R1* and *B1* are two orthologous genes encoding a transcription factor that activates an anthocyanin pigment-synthesis pathway. However, they condition pigment synthesis in different parts of the plant, indicating that, although their functions are conserved, their developmental regulation has changed. Therefore, polyploidization or gene amplification might be a process by which plants can accelerate morphological changes, permitting a more diverse control of gene expression and a faster response to the environment. Therefore, a fundamental difference between plants and humans might be that although the gene content of plants can be much higher, the number of

 1360-1385/01/$ – see front matter  PII: S1360-1385(01)01925-2

protein products is much lower compared with humans. An example of this is somatic recombination; the germline and the vegetative tissue in plants are contiguous, whereas in animals the germline is separated early on. Therefore, somatic rearrangement of genes does not influence inheritance in animals, but it can affect the inheritance of recombined genes in plants. Hence, many functions in animals, such as immunity and compatibility, are assembled by programmed somatic recombination, which requires fewer genes to start with.

However, the larger number of genes does not necessarily explain the difference in genome size. When rice, sorghum and maize are compared in orthologous regions (two genes in two species derived from the same gene through speciation), distances between genes differ as would be expected for their genome size. In other words, maize has insertions that push genes further apart than in sorghum, and rice has even shorter intergenic distances. Interestingly, it has been determined that many retrotranspositions have occurred since the progenitor of maize and sorghum diverged[9]. Accordingly, retrotransposition has played an important role in the evolution of chromosomes and the organization of their gene content. It is easy to see how they have helped to scramble pieces of chromosomes to prevent pairing of orthologous sequences. Chromosome breakage and fusion, observed by Barbara McClintock[10], might have been an important mechanism that accelerated diploidization. However, even 100 000 genes in a plant genome would not account for a significant portion of the size increase observed among the larger plant genomes, because polyploidization would also add all the pre-existing retroelements at the same time. Finally, natural variation of regulatory elements in plant genes could be a treasure box of new paradigms in gene expression. Comparative genomics in plants should have a great future.

**Joachim Messing**

Waksman Institute of Microbiology, Rutgers, The State University of New Jersey, 190 Frelinghuysen Road, Piscataway, NJ 08854-8020, USA.
e-mail: messing@mbcl.rutgers.edu

**References**

1 International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
2 Venter, J.C. *et al.* (2001) The sequence of the human genome. *Science* 291, 1304–1351
3 *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815
4 Davenport, R.J. (2001) Syngenta finishes, consortium goes on. *Science* 291, 807
5 Dickson, D. and Cyranoski, D. (2001) Commercial sector scores success with whole rice genome. *Nature* 409, 551
6 Messing, J. and Llaca, V. (1998) Importance of anchor genomes for any plant genome project. *Proc. Natl. Acad. Sci. U. S. A.* 95, 2017–2020
7 SanMiguel, P. *et al.* (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274, 765–768
8 Takahashi, C. *et al.* (1999) Genomic relationships between maize and its wild relatives. *Genome* 42, 1201–1207
9 Gaut, B.S. *et al.* (2000) Maize as a model for the evolution of plant nuclear genomes. *Proc. Natl. Acad. Sci. U. S. A.* 97, 7008–7015
10 McClintock, B. (1984) The significance of

