The Origins of Genomic Duplications in *Arabidopsis*

Todd J. Vision,[1]* Daniel G. Brown,[2] Steven D. Tanksley[3]

Large segmental duplications cover much of the *Arabidopsis thaliana* genome. Little is known about their origins. We show that they are primarily due to at least four different large-scale duplication events that occurred 100 to 200 million years ago, a formative period in the diversification of the angiosperms. A better understanding of the complex structural history of angiosperm genomes is necessary to make full use of *Arabidopsis* as a genetic model for other plant species.

A. thaliana has one of the smallest angiosperm genomes (*1*). It is a well-behaved diploid with only five haploid chromosomes (*2*). Despite this, much of the genome is internally duplicated (*3*–*14*). It has been hypothesized that the duplicated blocks originated in a single polyploidy event and have since been scrambled by chromosomal rearrangements (*5*). This hypothesis predicts that each region of the *Arabidopsis* genome should be present in exactly two copies. Recent comparative mapping results suggest that some regions are present in three or more copies (*6*, *8*), but it is not clear how prevalent such regions are. Here, we use the nearly complete genome sequence of *Arabidopsis* to study the evolutionary origins of duplicated blocks on a genome-wide scale. Sequence and map data used in our analyses, along with more detailed results, are available on the Internet (*15*).

Duplicated blocks were identified by the presence of neighboring genes with high sequence similarity to neighboring genes elsewhere in the genome. We considered only protein-coding genes because little conservation exists between noncoding duplicated regions in *Arabidopsis* (*5*, *16*). We used BLAST to identify genes with high sequence similarity (*17*, *18*). Our data set contained 20,269 composite open reading frames (cORFs), of which 2796 represented tandem arrays of related genes or the same gene present on overlapping clones (*19*). After removing low-quality matches (*20*), there were matches between 18,569 pairs of cORFs; 64% of the cORFs had at least one match. To identify duplicated blocks, we considered the proximity and transcriptional orientation of matches in both segments (*21*). We allowed singleton (nonmatching) genes within duplicated blocks because gene loss is known to follow duplication (*8*, *22*) and because some genes may be transposed from their original positions. We also did not prohibit inversions within duplicated blocks because small-scale inversions are not uncommon in eukaryotic genomes (*8*, *23*). To ensure that spurious duplicated blocks would not be identified, we chose a conservative set of parameter values based on the outcome of randomization tests (*24*).

We identified 103 duplicated blocks containing seven or more matching cORFs (Fig. 1). Candidates with fewer than seven genes are much more abundant in real than randomized data, suggesting that many smaller blocks may be present. There are duplications between all chromosomes except chromosome 2 with itself. Over 81% of cORFs fall within the bounds of at least one block. However, only 28% of these are actually present in duplicate. Interestingly, pericentromeric genes account for much of the genome that is not covered by blocks. Nearly 25% of all cORFs fall within two or more blocks, and one region, near ATAP22 on chromosome 4, falls within five blocks. Such extensive overlap among blocks provides prima facie evidence for multiple duplication events.

The number of independent duplication events can be inferred from patterns of sequence divergence between duplicated genes. A single polyploidization event will produce a unimodal distribution of divergence estimates with homogeneity among blocks. Many small independent events can also result in unimodality but with heterogeneity among blocks. A limited number of asynchronous, independent duplication events will produce a multimodal distribution.

The median estimated amino acid divergence (d_A) was between 0.325 and 0.725 amino acid substitutions per site (Δaa/site) for all but nine blocks (Fig. 2) (*25*). Excluding these nine, there is still significant among-block heterogeneity in d_A (*26*). Thus, we reject the hypothesis they belong to a single age class. The best-fit mixture of normal distributions to these 94 medians is trimodal (*27*). Each block can be assigned to one of the three age classes (labeled C, D, and E in progressing order of age) with greater than 50% posterior probability. The remaining blocks can be assigned, ad hoc, to two younger age classes (A and B) and one older age class (F). Some spatial overlap remains between duplicated blocks within age classes (Fig. 1), but this may be an artifact of erroneous age class assignments and spatially overextended blocks.

The blocks in age class C collectively bound 48% of the cORFs in the genome. Adding the number of duplicated pairs of cORFs to singleton cORFs, we estimate that more than 9000 cORFs were duplicated at this time (*28*). This is far larger than chromosome 1, the largest of the present complement, which we estimate to contain fewer than 6000 genes (*29*). Thus, it is likely that age class C represents either a whole-genome polyploidy event or the near-simultaneous duplication of multiple chromosomes.

The progressively older blocks in age classes D, E, and F bound 39, 11, and 3% of the cORFs in the genome, respectively. The true extent of these more ancient duplications is likely to be much greater because gene loss after each duplication event tends to obscure older blocks. Thus, these older age classes may also represent very large-scale duplication events.

We estimated the absolute ages of the duplicated blocks by assuming that the average extent of amino acid substitution (d_A) is linearly related to time (*30*). The average d_A values for age classes B through F (Table 1) yield age estimates of approximately 50, 100, 140, 170, and 200 million years ago (Mya), respectively. Thus, age classes C through F appear to date from the Mesozoic Era (65 to 245 Mya). Age class E is marginally older than the reported age of divergence between rosid and asterid eudicots, 112 to 156 Mya, whereas age class F is within the estimated time window for the divergence of monocots and dicots, 180 to 220 Mya (*31*–*33*). Thus, the older duplicated blocks reported here are likely to be a common feature of diverse groups of angiosperms. Regions contained within blocks 45, 48, 85, 88, and 100 were recently found to share common ancestry with a 105-kilobase genomic sequence from tomato, an asterid (*8*). Phylogenetic analysis in that study suggested that the rosid-asterid divergence occurred before the events leading to age classes C and E (duplicated blocks 45 and 85, respectively) and was nearly contemporaneous with age class E. Thus, the divergence and duplication dates are consistent. The two blocks assigned to age class A are likely to be artifacts of erroroneous genome assembly. In both cases, the copies are nearly identical even at the nucleotide level and are restricted to individual large-insert clones (*34*). Thus, the youngest duplicated

[1]USDA-ARS Center for Agricultural Bioinformatics, 604 Rhodes Hall, Cornell University, Ithaca, NY 14853, USA. [2]Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA. [3]Departments of Plant Breeding and Plant Biology, 252 Emerson Hall, Cornell University, Ithaca, NY 14853, USA.

*To whom correspondence should be addressed. E-mail: tv23@cornell.edu

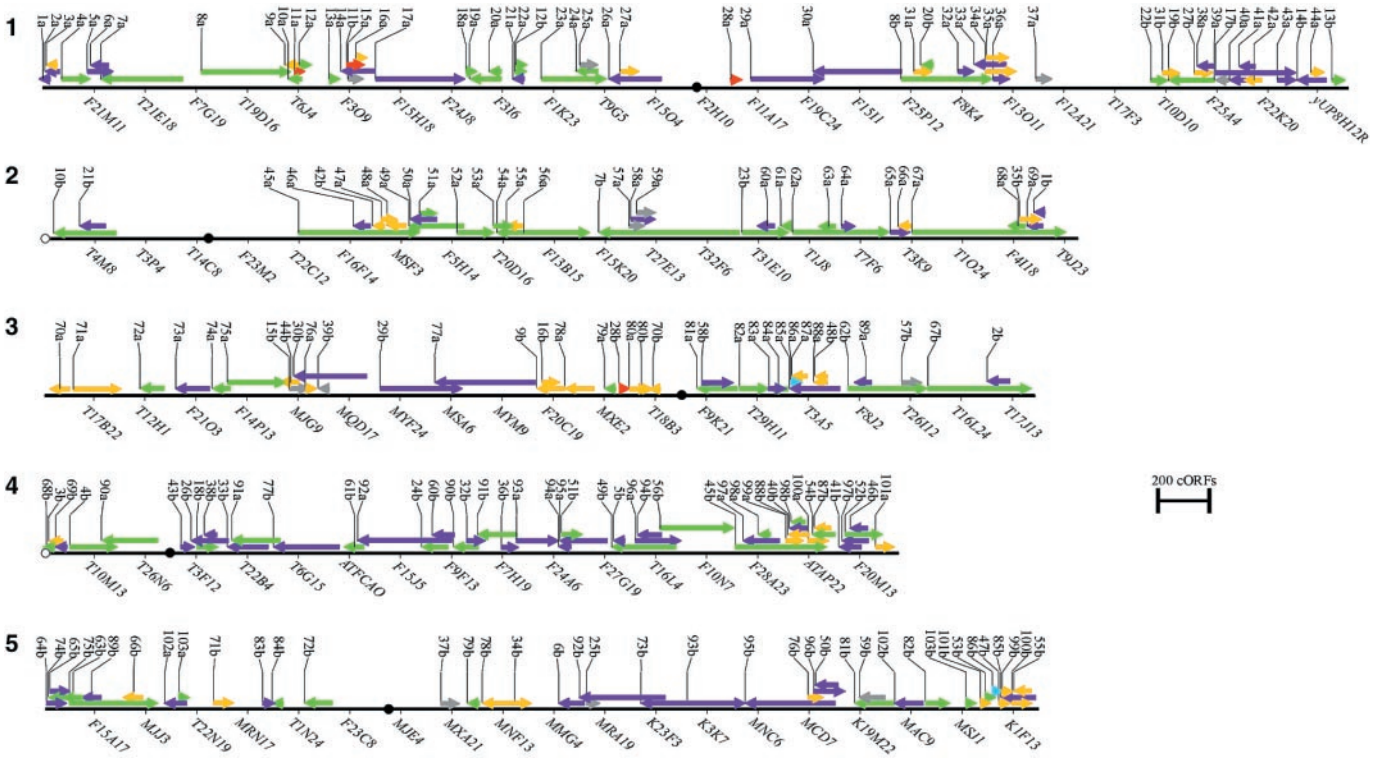

Fig. 1. Genomic map of duplicated blocks in *Arabidopsis*. The two copies of each putative duplicated block (e.g., 1a and 1b) are shown. Color denotes age class (red, A; blue, B; green, C; purple, D; orange, E; gray, F). Centromeres (*9*, *10*, *42*) are shown with black circles, and ribosomal DNA with white circles. Direction of arrowhead indicates the predominant relative orientation of duplicated cORFs within each block (right, direct; left, inverted). Landmarks are given at 200 cORF intervals.

block appears to be the sole member of age class B.

If the *Arabidopsis* genome has experienced multiple large-scale duplications, then the present complement of five chromosomes suggests a history of chromosome fusions. In fact, three such fusions have occurred since *A. thaliana* diverged from its closest extant relatives (*35*). Subchromosomal rearrangements, such as inversions and translocations, are expected to cause the average size of duplicated blocks to decrease with age, as is observed for blocks C through F (Table 1). Inversions can, in some cases, be inferred from our data set by the orientation of neighboring blocks or by gene order and orientation within blocks. Reciprocal translocations would be expected to conserve the orientation of blocks relative to the centromere (*22*) but, contrary to an earlier report (*5*), we do not see an excess of blocks with identical orientations in any age class.

We have estimated the number of deleted genes in each duplicated block by counting the number of singleton cORFs (*36*). The proportion of deleted genes increases with the inferred age of the duplication event (Table 1). A small number of blocks deviate significantly from a 1:1 distribution of singletons between the two copies, suggesting that the loss of duplicate genes between segments may sometimes be biased, as previously observed (*5*).

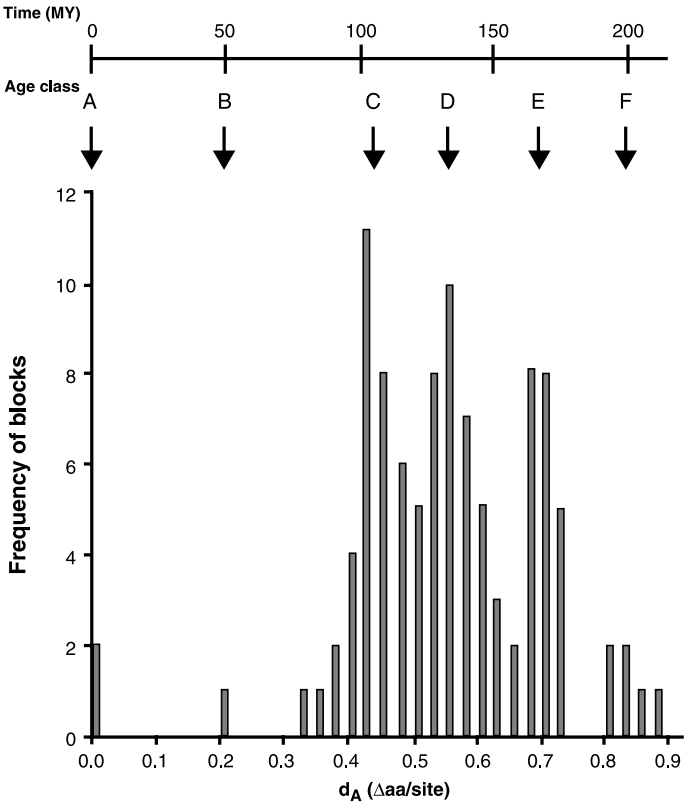

Fig. 2. Mutimodal distribution of block ages. The distribution of median d_A values among blocks. Age classes are marked with arrows. An approximate time scale is given.

The 103 duplicated blocks account for only 15% of the matches in the data set. Some proportion of the remaining matches may lie within undetected duplicated blocks or may have been transposed from their original position. Still, the remaining matches are not

Table 1. Features of the five age classes of duplicated blocks. d_A is the minimum change in amino acids per between dispersed duplicated cORFs, averaged among all cORFs. Retained duplicates, ratio of presently duplicated to inferred ancestral cORFs. Block size, mean number of cORFs (including singletons) per copy.

Age class	No. of blocks	d_A	Retained duplicates	Block size	Estimated age (Mya)
A*	2	0†	0.90	12.5	0
B	1	0.21†	0.38	14.5	50
C	35	0.45‡	0.15	149.5	100
D	36	0.57‡	0.13	128.0	140
E	23	0.71‡	0.11	60.0	170
F	6	0.84†	0.09	57.8	200

*Probable artifact of genome assembly. †Median among all duplicated genes. ‡Mean of best-fit normal distribution to block medians.

randomly dispersed in the genome, suggesting the presence of a separate gene duplication process. Matches not in blocks are 20% more likely to occur on the same chromosome than if the distribution were proportional to size. Of those matches on the same chromosome, the average distance is 86% of that expected between two random points. Similar findings have been reported for *Caenorhabditis elegans* (*37*), a genome that lacks large-scale duplications.

Many insertion mutants in *Arabidopsis* have no obvious phenotypic effect (*38*). This may be due, in part, to redundant functions among duplicated genes. One example appears to be the *shatterproof* genes SHP1 and SHP2, MADS-box regulatory factors that must be simultaneously removed before fruit nondehiscence is observed (*39*). SHP1 and SHP2, on chromosomes 3 and 2, are within duplicated block 67 in age class C.

Our analysis implies that regions of the *Arabidopsis* genome homoeologous to genomes that diverged from *Arabidopsis* before ~100 Mya will be small, generally less than ~10 centimorgans in size. This, coupled with massive gene loss in *Arabidopsis*, has likely been responsible for the difficulty in identifying regions homoeologous between *Arabidopsis* and rice (*40*). Knowledge of the duplication history of *Arabidopsis* should facilitate such mapping efforts. For example, it has recently been proposed that segments of *Arabidopsis* chromosome 4 and rice chromosome 2 are homoeologous (*41*). The *Arabidopsis* segment is within duplicated block 92 (age class D), implying that the rice segment is also homoeologous to a part of *Arabidopsis* chromosome 5.

Our understanding of plant evolution and our use of *Arabidopsis* as a genetic model for other plants will clearly depend on a deeper appreciation for the complex duplication history of this small genome.

References and Notes

1. K. Arumuganthan, E. D. Earle, *Plant Mol. Biol. Rep.* **9**, 208, (1991).
2. D. W. Meinke, J. M. Cherry, C. D. Dean, S. Rounsley, M. Korneef, *Science* **282**, 662 (1998).
3. J. M. McGrath, M. M. Jansco, E. Pichersky, *Theor. Appl. Genet.* **86**, 880 (1993).
4. I. Bancroft, *Yeast* **17**, 1 (2000).
5. G. Blanc, A. Barakat, R. Guyot, R. Cooke, M. Delseny, *Plant Cell* **12**, 1093 (2000).
6. D. Grant, P. Cregan, R. C. Shoemaker, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4168 (2000).
7. S. P. Kowalski, T. H. Lan, K. A. Feldmann, A. H. Paterson, *Genetics* **138**, 499 (1994).
8. H.-M. Ku, T. Vision, J. Liu, S. D. Tanksley, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9121 (2000).
9. X. Lin *et al.*, *Nature* **402**, 761 (1999).
10. K. Mayer *et al.*, *Nature* **402**, 769 (1999).
11. A. H. Paterson *et al.*, *Nature Genet.* **14**, 380 (1996).
12. T.-H. Lan *et al.*, *Genome Res.* **10**, 776 (2000).
13. A. H. Paterson *et al.*, *Plant Cell* **12**, 1523 (2000).
14. N. Terryn *et al.*, *FEBS Lett.* **445**, 237 (1999).
15. Supplementary information is available at: www.igd.cornell.edu/~tvision/arab/science_supplement.html
16. We used GENSCAN [C. Burge, S. Karlin, *J. Mol. Biol.* **268**, 78 (1997)] to provide gene models in unannotated clones. Because we only require that most predicted exons overlap true exons in the same translation frame and that predicted gene densities are approximately correct, ab initio predictions are sufficient for our purposes.
17. S. F. Altschul *et al.*, *Nucleic Acids Res.* **25**, 3389 (1997).
18. BLASTP scores were obtained for all pairs of genes. The (i,j) element of matrix **M** was assigned the alignment score for proteins i and j if the score was 100 bits or greater. Row and column indices denote the position of each protein within each chromosome. Chromosome order and orientation are arbitrary.
19. Matching genes within 15 positions of each other were collected into the row and column with the smallest index and were assigned the maximum of the component scores. This was iterated until convergence, thereby combining both tandemly duplicated genes and single-copy genes shared by overlapping clones. A single gene may occur in two positions if the tiling path information is wrong, but extensive clone overlap would be needed to generate a spurious duplicated block and the high sequence similarity would have flagged the error.
20. Only the five highest scores in each row and column were retained, a conservative approach that sacrifices sensitivity for specificity.
21. To identify duplicated blocks, we first calculated a weight for each pair of nonzero elements $M_{i1,j1}$ and $M_{i2,j2}$, for all $j_2 > j_1$ with corresponding transcriptional orientations $T(i_1)$, $T(i_2)$, $T(j_1)$, $T(j_2)$, where $T \in \{-1, 1\}$. The weight was $W = -k + \ell^{(r+c)} - 1_{[T(i_1)\cdot T(j_1)=T(i_2)\cdot T(j_2)=\mathrm{sgn}(i_2-i_1)]}\cdot m$, where k, ℓ, and m are constants; $r = |i_1 - i_2|$ and $c = |j_1 - j_2|$ are the row and column distances; $1_{[T(i_1)\cdot T(j_1)=T(i_2)\cdot T(j_2)=\mathrm{sgn}(i_2-i_1)]}$ is an indicator function equaling 1 if the transcriptional orientations of the two pairs of linked cORFs are equivalent but otherwise equaling 0; and sgn(x) equals −1, 0, or 1 for x greater than, less than, or equal to 0, respectively. When cORFs were composed of multiple genes, transcriptional orientations were taken to be those of the highest scoring gene pair. Weights were assigned to edges of a directed acyclic graph in which nodes were nonzero elements of **M** and edges connected all nodes (i_1, j_1) and (i_2, j_2) for which $j_2 > j_1$. We computed minimum-weight paths between every pair of nodes connected by an edge [T. H. Cormen, C. E. Leiserson, R. L. Rivest, *Introduction to Algorithms* (MIT Press, Cambridge, MA, 1990), pp. 536–538], identified paths with negative weight, combined overlapping paths, combined paths from either side of the diagonal, and accepted the resulting sets of nodes as duplicated blocks. Errors in the order and orientation of genes may hinder our ability to detect duplicated blocks but are unlikely to generate false ones.
22. K. H. Wolfe, D. Shields, *Nature* **387**, 708 (1997).
23. A. McLysaght, C. Seoighe, K. H. Wolfe, in *Comparative Genomics*, D. Sankoff, J. H. Nadeau, Eds. (Kluwer, New York, 2000), pp. 47–58.
24. Random matrices were derived from **M** by permutation of its rows. Using parameter values of $k = 5$, $l = 1.14$, and $m = 25$, only one block defined by six or more pairs was identified in 1000 permutations of the chromosome 2 versus 4 submatrix compared with 34 blocks of five pairs. Thus, blocks of seven pairs are unlikely to arise by chance, though real duplicated blocks may be overextended or erroneously merged.
25. Amino acid alignments were obtained using CLUSTALW version 1.7 [J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* **22**, 4673 (1994)]. Estimates of d_A were obtained using PAML [Z. Yang, *Phylogenetic Analysis by Maximum Likelihood (PAML)*, Version 3.0. (University College, London, 2000)] with the JTT substitution matrix [D. W. Jones, W. R. Taylor, J. M. Thornton, *CABIOS* **8**, 275 (1992)]. The smallest estimate of d_A was used for matches between multiple genes. The median is more robust to outliers than the mean, though it will still be affected by the absence of highly diverged homologous genes from the sample.
26. Homogeneity in d_A among blocks was rejected by a single-classification analysis of variance ($P < 0.0001$).
27. Mixture models of normal distributions, with parameters for means, variances, and mixing proportions, were fit using an expectation-maximization algorithm. Models were compared by likelihood ratio tests [M. Lynch, B. Walsh, *Genetics and Analysis of Quantitative Traits*. (Sinauer, Sunderland, MA, 1997), pp. 359–364]. The medians of samples drawn from a population of any distribution approach a normal distribution as the sample size increases. [W. Feller, *An Introduction to Probability Theory and its Applications* (John Wiley & Sons, New York, ed. 2, 1957), pp. 238–241]. The approximation should be adequate for samples of this size (average = 29), though it is not strictly valid because there are differing numbers of matches in each block. The log likelihoods of the one, two, and three distribution models are −5.5, −128.5, and −183.6, respectively, and each differs from the next by three degrees of freedom.
28. Because numerous small blocks were not counted and up to 20% of the genome sequence has not been analyzed, this is likely to be an underestimate.
29. This estimate is the product of the average gene density of chromosomes 2 and 4, ~210 genes/megabase, and the estimated length of chromosome 1, which is 27.9 megabases [T. Mozo *et al.*, *Nature Genet.* **22**, 271 (1999)].
30. A rate of $9 \times 10^{-10} \pm 9 \times 10^{-10}$ nonsynonymous base substitutions·site^{-1}·lineage^{-1}·year^{-1} has been estimated for nuclear genes in the grasses [B. S. Gaut, *Evol. Biol.* **30**, 93 (1998); S. V. Muse, *Plant Mol. Biol.* **42**, 25 (2000)], though this must be treated with caution due to many sources of uncertainty. Because 75% of all possible sense nucleotide substitutions in the genetic code are nonsynonymous and there are three positions in each codon, we assume that there are 2.25 nonsynonymous sites in each codon in converting from amino acid substitutions to nonsynonymous base substitutions. Accounting for patterns of codon usage in *Arabidopsis* (GenBank Release 119.0), one obtains a nearly identical conversion factor (2.30 nonsyn-

onymous sites per codon) [D. A. Benson *et al*., *Nucleic Acids Res.* **28**, 15 (2000)].
31. *Arabidopsis* is a member of the rosid lineage of dicot angiosperms [P. S. Soltis, D. E. Soltis, M. W. Chase, *Nature* **402**, 402 (1999)].
32. K. H. Wolfe, M. Gouy, Y.-W. Yang, P. M. Sharp, W. H. Li, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6201 (1989).
33. Y.-W. Yang, K.-N. Lai, P.-Y. Tai, W.-H. Li, *J. Mol. Evol.* **48**, 597 (1999).
34. The two presumed redundancies involve clones T6J4/F13B4 and F21N10/K17E7.
35. M. Koch, J. Bishop, T. Mitchell-Olds, *Plant Biol.* **1**, 529 (1999).
36. This may be inflated by undetected homologies, overextended blocks, and transposition of genes from their original positions.
37. C. Semple, K. H. Wolfe, *J. Mol. Evol.* **48**, 555 (1999).
38. C. Somerville, personal communication.
39. S. J Liljegren *et al., Nature* **404**, 766 (2000).
40. K. M. Devos, J. Beales, Y. Nagamura, T. Sasaki, *Genome Res.* **9**, 825 (1999).
41. A.-M. van Dodeweerd *et al.*, *Genome* **42**, 887 (1999).
42. G. P. Copenhaver, W. E. Browne, D. Preuss, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 247 (1998).
43. We thank C. Aquadro, J. Doyle, R. Durrett, T. Mitchell-Olds, C. Somerville, M. Yanofsky, and L. Zhang for helpful comments. This research was funded by grants from the National Science Foundation and the Office of Naval Research. T.J.V. is supported in part by the Cornell Theory Center.

23 October 2000; accepted 15 November 2000

Strange Magnetism and the Anapole Structure of the Proton

R. Hasty,[1] A. M. Hawthorne-Allen,[5] T. Averett,[9] D. Barkhuff,[4] D. H. Beck,[1] E. J. Beise,[3]* A. Blake,[2] H. Breuer,[3] R. Carr,[2] S. Covrig,[2] A. Danagoulian,[1] G. Dodson,[4] K. Dow,[4] M. Farkhondeh,[4] B. W. Filippone,[2] J. Gao,[2] M. C. Herda,[3] T. M. Ito,[2] C. E. Jones,[2] W. Korsch,[6] K. Kramer,[9] S. Kowalski,[4] P. Lee,[2] R. D. McKeown,[2] B. Mueller,[7] M. Pitt,[5] J. Ritter,[5] J. Roche,[9] V. Savu,[2] D. T. Spayde,[3] R. Tieulent,[3] E. Tsentalovich,[4] S. P. Wells,[8] B. Yang,[4] T. Zwart[4]

The violation of mirror symmetry in the weak force provides a powerful tool to study the internal structure of the proton. Experimental results have been obtained that address the role of strange quarks in generating nuclear magnetism. The measurement reported here provides an unambiguous constraint on strange quark contributions to the proton's magnetic moment through the electron-proton weak interaction. We also report evidence for the existence of a parity-violating electromagnetic effect known as the anapole moment of the proton. The proton's anapole moment is not yet well understood theoretically, but it could have important implications for precision weak interaction studies in atomic systems such as cesium.

In 1933, the German physicist Otto Stern discovered that the magnetism of the proton was anomalously large, a factor of three larger than expected from the basic theory of quantum mechanics. This experiment turned out to be the first glimpse of the internal structure of the constituents of the atomic nucleus and a tantalizing hint at the existence of quarks. Widespread applications of the proton's magnetic properties, such as the magnetic resonance imaging (MRI) techniques used in biology and medicine, have been developed despite a lack of fundamental understanding of the basic dynamics that generates this magnetism. After the key discovery of internal structure in the proton in a high-energy electron scattering experiment at the Stanford Linear Accelerator Center in the late 1960s (*1*), the theory of Quantum Chromodynamics (QCD), which describes the interaction between quarks and the gluons that bind the quarks into the atomic nuclei observed in the periodic table, was developed. Despite almost 30 years of intense theoretical effort, QCD has been unable to produce numerical predictions for the basic properties of nucleons such as their degree of magnetism.

In 1988, Kaplan and Manohar (*2*) proposed that the study of the weak magnetic force (analogous to the usual magnetic force associated with electromagnetism) would allow a separation of the proton's magnetism into the three contributing flavors of quarks: up, down, and strange. Measuring the contribution from strange quark-antiquark pairs is of special interest because it relates directly to the "sea" of virtual quark-antiquark pairs in the proton, a phenomenon predicted by QCD. In 1989, it was noted that the weak magnetic force could be isolated using its unique property of lack of mirror symmetry, or parity violation (*3*). The basic idea was to study the preference for the proton's interaction with electrons that have spin counterclockwise, over those with clockwise spin, relative to their direction of travel. We report data obtained using this method that, when combined with our previously published results (*4*, *5*), allow the first unambiguous determination of the proton's weak magnetism. We also report a measurement of a parity-violating, time-reversal-even electromagnetic contribution to proton structure, referred to in the literature as the proton's anapole moment.

The weak magnetic moment of the proton, designated as μ_Z, is associated with the interaction due to exchange of the weak neutral Z particle. The weak and electromagnetic forces are related through the mixing angle θ_W, a fundamental parameter of the standard model of electroweak interactions that relates the Z boson to its charged counterparts, the $W^{\pm}$ [we use $\sin^2\theta_W = 0.23117 \pm 0.00016$ (*6*) in our analysis]. As a result, one can write μ_Z as

$$\mu_Z = (\mu_p - \mu_n) - 4\sin^2\theta_W\mu_p - \mu_s \quad (1)$$

where $\mu_{p(n)}$ are the usual magnetic moments of the proton (or neutron) [$2.97\mu_N$ and $-1.91\mu_N$, respectively, where $\mu_N = \frac{e\hbar}{2m_p} = 3.152451238(24) \times 10^{-14}$ MeV-T (*6*)]. The individual contributions of up, down, and strange quark-antiquark pairs to the proton's magnetic moment, μ_u, μ_d, and μ_s, are defined by the relation $\mu_p = \frac{2}{3}\mu_u - \frac{1}{3}\mu_d - \frac{1}{3}\mu_s$. The neutron's magnetic moment is similarly constructed, interchanging only μ_u and μ_d. Thus, a measurement of μ_Z combined with the known μ_p and μ_n, provides the third observable required to uniquely determine the u, d, and s quark pieces of the proton's magnetic moment. The study of parity-violating electron-nucleon (e-N) scattering enables a determination of the weak magnetic form factor $G_M^Z(Q^2)$, and as a result the equivalent strange piece $G_M^s(Q^2)$, the momentum-dependent counterpart of μ_Z, where Q^2 is the relativistic four-momentum transferred to the proton by the electron. The Q^2 dependence of these quantities is sensitive

[1]Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [2]Kellogg Radiation Laboratory, California Institute of Technology, Pasadena, CA 91125, USA. [3]Department of Physics, University of Maryland, College Park, MD 20742, USA. [4]Bates Linear Accelerator Center, Laboratory for Nuclear Science and Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [5]Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061–0435, USA. [6]Department of Physics and Astronomy, University of Kentucky, Lexington, KY 40506, USA. [7]Physics Division, Argonne National Laboratory, Argonne, IL 60439, USA. [8]Department of Physics, Louisiana Tech University, Ruston, LA 71272, USA. [9]Department of Physics, College of William and Mary, Williamsburg, VA 23187, USA.

*To whom correspondence should be addressed. E-mail: beise@physics.umd.edu