every change in surface climate at our site has an accompanying change in deep circulation (and vice versa)—emphasizing even more strongly the uniqueness of the coeval extremes in climate and deep circulation that follow the Lake Agassiz flood outburst.

Wider climatic implications. Despite the complex nature of the climate record, the fact that large-scale changes in deep-ocean circulation are accomplished just as quickly in the natural world as is predicted in computer simulations [e.g., (*12*)] confirms that they were sufficiently rapid to be a plausible mechanism for driving the similarly abrupt changes seen in paleoclimate records. This observation helps to define the minimum time scale for altering ocean circulation and demonstrates that it is rapid enough to be relevant for human societies. The fact that these past ocean circulation changes are associated with the largest climate variations in the past underscores the importance of understanding the minimum freshwater forcing capable of affecting such large circulation changes—particularly given the concerns about the impact of future warming on the Greenland ice sheet.

References and Notes

1. E. R. Thomas *et al.*, *Quat. Sci. Rev.* **26**, 70 (2007).
2. S. J. Johnsen *et al.*, *Nature* **359**, 203 (1992).
3. W. Dansgaard *et al.*, *Nature* **364**, 218 (1993).
4. R. B. Alley *et al.*, *Geology* **25**, 483 (1997).
5. R. B. Alley, A. M. Ágústdóttir, *Quat. Sci. Rev.* **24**, 1123 (2005).
6. E. J. Rohling, H. Pälike, *Nature* **434**, 975 (2005).
7. D. C. Barber *et al.*, *Nature* **400**, 344 (1999).
8. D. W. Leverington, J. D. Mann, J. T. Teller, *Quat. Res.* **57**, 244 (2002).
9. J. T. Teller, D. W. Leverington, J. D. Mann, *Quat. Sci. Rev.* **21**, 879 (2002).
10. A. N. LeGrande *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 837 (2006).
11. E. Bauer, A. Ganopolski, *Paleoceanography* **19**, PA3014 (2004).
12. A. P. Wiersma, H. Renssen, *Quat. Sci. Rev.* **25**, 63 (2006).
13. I. R. Hall, G. G. Bianchi, J. R. Evans, *Quat. Sci. Rev.* **23**, 1529 (2004).
14. D. W. Oppo, J. McManus, J. D. Cullen, *Nature* **422**, 277 (2003).
15. L. D. Keigwin, J. P. Sachs, Y. Rosenthal, E. A. Boyle, *Paleoceanography* **20**, PA2003 (2005).
16. C. R. W. Ellison, M. R. Chapman, I. R. Hall, *Science* **312**, 1929 (2006).
17. B. Hansen, S. Østerhus, *Prog. Oceanogr.* **45**, 109 (2000).
18. B. Hansen, W. R. Turrell, S. Østerhus, *Nature* **411**, 927 (2001).
19. A. Biastoch, R. H. Kase, D. B. Stammer, *J. Phys. Oceanogr.* **33**, 2307 (2003).
20. C. N. Wold, *Paleoceanography* **9**, 917 (1994).
21. M. S. McCartney, *Prog. Oceanogr.* **29**, 283 (1992).
22. C. Hillaire-Marcel, A. de Vernal, G. Bilodeau, G. Wu, *Can. J. Earth Sci.* **31**, 63 (1994).
23. A. P. Wiersma, H. Renssen, H. Gosse, T. Fichefet, *Clim. Dyn.* **27**, 831 (2006).
24. C. Kissel, *C.R. Acad. Sci. Paris* **337**, 908 (2005).
25. J. S. Stoner, J. E. T. Channell, C. Hillaire-Marcel, *Paleoceanography* **11**, 309 (1996).
26. C. Kissel *et al.*, *Earth Planet. Sci. Lett.* **171**, 489 (1999).
27. C. Hillaire-Marcel, A. de Vernal, D. J. W. Piper, *Geophys. Res. Lett.* **34**, L15601 (2007).
28. C. Hillaire-Marcel, A. de Vernal, G. Bilodeau, A. J. Weaver, *Nature* **410**, 1073 (2001).
29. W. B. Curry, D. W. Oppo, *Paleoceanography* **20**, PA1017 (2005).
30. U. S. Ninnemann, C. D. Charles, *Earth Planet. Sci. Lett.* **201**, 383 (2002).
31. M. Latif *et al.*, *J. Clim.* **19**, 4631 (2006).
32. D. Klitgaard-Kristensen, H. P. Sejrup, H. Haflidason, S. Johnsen, M. Spurk, *J. Quat. Sci.* **13**, 165 (1998).
33. H. Renssen, H. Goosse, T. Fichefet, J.-M. Campin, *Geophys. Res. Lett.* **28**, 1567 (2001).
34. R. E. Came, D. W. Oppo, J. F. McManus, *Geology* **35**, 315 (2007).
35. W. J. Schmitz, M. S. McCartney, *Reviews of Geophysics* **31**, 29 (1993).
36. R. R. Dickson, J. Brown, *J. Geophys. Res.* **99**, 12319 (1994).
37. S. O. Rasmussen, B. M. Vinther, H. B. Clausen, K. K. Andersen, *Quat. Sci. Rev.* **26**, 1907 (2007).
38. S. O. Rasmussen *et al.*, *J. Geophys. Res.* **111**, D06102 (2006).
39. B. M. Vinther *et al.*, *J. Geophys. Res.* **111**, D13102 (2006).

40. We thank the Centre for Ice and Climate (especially B. Vinther, S. Johnsen, and S. Rasmussen) at the Niels Bohr Institute at University of Copenhagen, for supplying the unpublished GRIP and NGRIP data. We are also grateful for technical assistance from D.-I. Blindheim, E. Bjørseth, E. V. Galaasen, O. Hansen, and I. V. Johansen from Bjerknes Centre for Climate Research (BCCR)/University of Bergen (UoB) and C. Wandres and F. Dewilde from Laboratoire des Sciences du Climat et de l'Environnement (LSCE), Commisariat à l'Energie Atomique (CEA), Centre National de la Recherche Scientifique (CNRS), Université Versailles-Saint-Quentin (UVSQ). This work was supported by the BCCR, the UoB, and the Norwegian Research Council, and at LSCE by the CEA and CNRS through the Institut des Sciences de l'Univers, Programme National d'Etude du Climat (INSU-PNEDC-Impair) project. Additional funding was provided by the PACLIVA EU EVK2-2002-00143 and the Agence National de la Recherche, Integration des Contraintes Paléoclimatiques: Réduire les Incertitudes sur l'Evolution du Climat des Périodes Chaudes (ANR-PICC) projects. The GIFA ^{14}C dates were obtained by AMS Artemis from LMC14, National Service of INSU. We are grateful to M. Paterne (LSCE) for her help with the GIFA ^{14}C data. The ship and scientific technology for the P.I.C.A.S.S.O. cruise were provided by the Institut Polaire Paul Emile Victor (IPEV) within the framework of the International Marine Global Changes (IMAGES) program. H.F.K. and U.S.N. analyzed the planktonic and benthic stable isotope data in this study and prepared samples for the KIA AMS ^{14}C dates. E.C. was responsible for the paleoceanographic team in LSCE. C.K. and C.L. are responsible for the magnetic and $CaCO_3$ analyses reported here and C.L. was chief scientist on P.I.C.A.S.S.O. T.O.R. is responsible for the x-ray fluorescence scans.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1148924/DC1
Materials and Methods
Figs. S1 to S4
Table S1
References

8 August 2007; accepted 8 November 2007
Published online 6 December 2007;
10.1126/science.1148924
Include this information when citing this paper.

The *Physcomitrella* Genome Reveals Evolutionary Insights into the Conquest of Land by Plants

Stefan A. Rensing,[1] Daniel Lang,[1] Andreas D. Zimmer,[1] Astrid Terry,[2] Asaf Salamov,[3] Harris Shapiro,[3] Tomoaki Nishiyama,[4] Pierre-François Perroud,[5] Erika A. Lindquist,[3] Yasuko Kamisugi,[6] Takako Tanahashi,[7,8] Keiko Sakakibara,[9] Tomomichi Fujita,[10] Kazuko Oishi,[11] Tadasu Shin-I,[11] Yoko Kuroki,[12] Atsushi Toyoda,[12] Yutaka Suzuki,[13] Shin-ichi Hashimoto,[14] Kazuo Yamaguchi,[4,15] Sumio Sugano,[13] Yuji Kohara,[11,16] Asao Fujiyama,[12,17,18] Aldwin Anterola,[19] Setsuyuki Aoki,[20] Neil Ashton,[21] W. Brad Barbazuk,[22] Elizabeth Barker,[21] Jeffrey L. Bennetzen,[23] Robert Blankenship,[5] Sung Hyun Cho,[5] Susan K. Dutcher,[24] Mark Estelle,[25] Jeffrey A. Fawcett,[26] Heidrun Gundlach,[27] Kousuke Hanada,[28,29] Alexander Heyl,[30] Karen A. Hicks,[31,32] Jon Hughes,[33] Martin Lohr,[34] Klaus Mayer,[27] Alexander Melkozernov,[35] Takashi Murata,[7,8] David R. Nelson,[36] Birgit Pils,[37] Michael Prigge,[25] Bernd Reiss,[31] Tanya Renner,[38] Stephane Rombauts,[26] Paul J. Rushton,[39] Anton Sanderfoot,[40] Gabriele Schween,[1] Shin-Han Shiu,[28] Kurt Stueber,[31] Frederica L. Theodoulou,[41] Hank Tu,[3] Yves Van de Peer,[26] Paul J. Verrier,[42] Elizabeth Waters,[37] Andrew Wood,[19] Lixing Yang,[23] David Cove,[5,6] Andrew C. Cuming,[6] Mitsuyasu Hasebe,[7,8,43] Susan Lucas,[2] Brent D. Mishler,[44] Ralf Reski,[1] Igor V. Grigoriev,[3] Ralph S. Quatrano,[5*] Jeffrey L. Boore[3,44,45]

We report the draft genome sequence of the model moss *Physcomitrella patens* and compare its features with those of flowering plants, from which it is separated by more than 400 million years, and unicellular aquatic algae. This comparison reveals genomic changes concomitant with the evolutionary movement to land, including a general increase in gene family complexity; loss of genes associated with aquatic environments (e.g., flagellar arms); acquisition of genes for tolerating terrestrial stresses (e.g., variation in temperature and water availability); and the development of the auxin and abscisic acid signaling pathways for coordinating multicellular growth and dehydration response. The *Physcomitrella* genome provides a resource for phylogenetic inferences about gene function and for experimental analysis of plant processes through this plant's unique facility for reverse genetics.

Here, we report the draft genome sequence of the moss *Physcomitrella patens*, the first bryophyte genome to be sequenced. The embryophytes (land plants) began to diverge about 450 million years ago (Ma). Bryophytes, comprising hornworts, mosses, and liverworts, are remnants of early diverging lineages of embryophytes and thus occupy an ideal phylogenetic position for reconstructing ancient evolutionary changes and illuminating one of the most important events in earth history—the conquest of land by plants (Fig. 1). The terrestrial environment

involves variations in water availability and temperature, as well as increased exposure to radiation. Adaptation entailed dramatic changes in body plan (*1*) and modifications to cellular, physiological, and regulatory processes. Primary adaptations included enhanced osmoregulation and osmoprotection, desiccation and freezing tolerance, heat resistance, synthesis and accumulation of protective "sunscreens," and enhanced DNA repair mechanisms. Fossil evidence suggests that early land plants were structurally similar to extant bryophytes (*2*); they probably had a dominant haploid phase and were dependent on water for sexual reproduction, having motile male gametes.

The genome sequence of *P. patens* allows us to reconstruct the events of genome evolution that occurred in the colonization of land, through comparisons with the genome sequences of several angiosperms (*Arabidopsis thaliana*, *Oryza sativa*, and *Populus trichocarpa*), as well as aquatic unicellular green algae (*Ostreococcus tauri*, *Ostreococcus lucimarinus*, and *Chlamydomonas reinhardtii*).

Features of the Whole Genome

General genome properties. The draft genome sequence of *P. patens* ssp. *patens* (strain Gransden 2004) was determined by whole-genome shotgun sequencing, assembling into 480 mega–base pairs of scaffold sequence with a depth of ~8.6× (*3*, *4*); expressed sequence tag (EST) coverage of the assembly is over 98%. The sequence contains 35,938 predicted and annotated *P. patens* gene models (tables S1 to S5). Most predicted genes are supported by multiple types of evidence (table S4), and 84% of the predicted proteins appear complete. About 20% of the analyzed genes show alternative splicing (table S6), a frequency similar to that of *A. thaliana* and *O. sativa* (*5*).

Repetitive sequences and transposons. An ab initio approach detected 14,366 repetitive elements comprising 1381 families [average member number 10, length 1292 bp (table S7)]. The largest repetitive sequence is from the "AT-rich, low complexity" class (23% of the repetitive fraction), and 15 families account for over 84% of the repetitive fraction (table S8).

Long terminal repeat retrotransposons (LTR-Rs) are generally the most abundant class of transposable elements, contributing substantially to flowering plant genome size (*6*). Of the 4795 full-length LTR-Rs in *P. patens*, 46% are gypsy-like and 2% are copia-like. *P. patens* contains about three times as many full-length LTR-Rs as *A. thaliana*, but about one-third as many as *O. sativa*. The density among the three genomes is lowest in *P. patens* (fig. S1). Although about half of the *P. patens* genome consists of 157,127 LTR-Rs, only 3% exist as intact full-length elements. The remainder is made of diverged and partial remnants, often fragmented by mutual insertions (fig. S2). Nested regions are common, with 14% of LTR-Rs inserted into another LTR-R (table S9). The genome also contains 895 solo LTR-Rs, probably as a result of unequal crossing-over or DNA repair. Periodic retrotransposition activity peaks are discernible over the past 10 million years (My) (Fig. 2). Only one full-length element is inserted within a gene, which suggests strong selection against transposon insertion into genes ($P < 0.001$).

Helitrons (rolling-circle transposons) are an ancient class of transposons present in animals, fungi, and plants (*7*). Different from all eukaryotic genomes sequenced so far, the *P. patens* genome contains only a single Helitron family (table S10) with 19 members. High sequence similarity (96%) suggests that they have been active within the past 3 My. Presumably, multiple Helitron families evolved in all plant lineages, including *P. patens*, but we predict that a rapid process of DNA removal has excised all members that have not been active recently, a process that has been demonstrated in other plant genomes (*6*).

Gene and genome duplications. Gene and genome duplications are major driving forces of gene diversification and evolution (*8*). In *P. patens*, the K_s distribution plot (i.e., the frequency classes of synonymous substitutions) among paralogs shows a clear peak at around 0.5 to 0.9 (fig. S3), which suggests that a large-scale duplication, possibly involving the whole genome, has occurred. The presence of this peak confirms EST-based data (*9*). Additional evidence for a large-scale duplication comes from the identification of 77 nonoverlapping duplicated segments containing at least five paralogous gene pairs. All duplicated segments have an average K_s of 0.5 to 0.7.

Tandemly arrayed genes (TAGs) can contribute substantially to genome size. However, only ~1% of the protein-encoding genes in *P. patens* occur in tandem array, in contrast to *A. thaliana* (~16%), *O. sativa* (~14%), and *P. trichocarpa* (11%) (*10*–*12*). The majority of *P. patens* TAG clusters are made up of two genes that are not separated by an intervening gene (fig. S4). Compared with non-TAG genes, genes in TAGs are significantly shorter ($P < 0.001$) in terms of gene, coding sequence (CDS), and intron length, whereas their G/C content is significantly higher (table S11). Functional analysis of TAGs compared with paralogous non-TAG clusters reveals that photosynthesis proteins, particularly antenna proteins, are significantly ($P < 0.05$) enriched among the TAGs [section 3.6, St 58 A/B (*13*)]. Other enriched categories are glyoxylate and dicarboxylate metabolism, carbon fixation, and ribosome assembly (fig. S5). Apparently, *P. patens* has increased the genetic playground for photosynthesis and related carbon-based metabolism in its recent past.

Comparison of the K_s of *P. patens* TAGs with paralogs that were established during the large-scale genome duplication (K_s ~0.5 to 0.9) sug-

[1]Plant Biotechnology, Faculty of Biology, University of Freiburg, Schaenzlestrasse 1, D-79104 Freiburg, Germany. [2]U.S. Department of Energy (DOE) Joint Genome Institute and Lawrence Livermore National Laboratory, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA. [3]DOE Joint Genome Institute and Lawrence Berkeley National Laboratory, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA. [4]Advanced Science Research Center, Kanazawa University, 13-1 Takara-machi Kanazawa 920-0934, Japan. [5]Department of Biology, Washington University, 1 Brookings Drive, St. Louis, MO 63130–4899, USA. [6]Centre for Plant Sciences, University of Leeds, Leeds LS2 9JT, UK. [7]National Institute for Basic Biology, Okazaki 444-8585, Japan. [8]Department of Basic Biology, School of Life Science, The Graduate University for Advanced Studies, Okazaki 444-8585, Japan. [9]School of Biological Sciences, Monash University, Clayton Campus, Melbourne, VIC 3800, Australia. [10]Department of Biological Sciences, Faculty of Science, Hokkaido University, Sapporo 060-0810, Japan. [11]Genome Biology Laboratory, Center for Genetic Resource Information, National Institute of Genetics, Mishima 411-8540, Japan. [12]RIKEN Genomic Sciences Center, Kanagawa 230-0045, Japan. [13]Laboratory of Functional Genomics, Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo 108-8639, Japan. [14]Department of Molecular Preventive Medicine, School of Medicine, The University of Tokyo, Tokyo 113-8654, Japan. [15]Division of Life Science, Graduate School of Natural Science and Technology, Kanazawa University, Kanazawa 920-1192, Japan. [16]Department of Genetics, School of Life Science, The Graduate University for Advanced Studies, Mishima 411-8540, Japan. [17]National Institute of Informatics, Tokyo 101-8403, Japan. [18]Department of Informatics, School of Multidisciplinary Sciences, The Graduate University for Advanced Studies, Tokyo 101-8403, Japan. [19]Department of Plant Biology, Southern Illinois University, Carbondale, IL 62901–6509, USA. [20]Life-Science Informatics Unit, Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan. [21]University of Regina, 3737 Wascana Parkway, Regina, SK S4S 0A2, Canada. [22]Donald Danforth Plant Science Center, 975 North Warson Road, St. Louis, MO 63132, USA. [23]Department of Genetics, Davison Life Sciences Complex, University of Georgia, Athens, GA 30602–7223, USA. [24]Department of Genetics, Washington University, 660 South Euclid Avenue, St. Louis, MO 63108, USA. [25]Department of Biology, Indiana University, 1001 East Third Street, Bloomington, IN 47405–3700, USA. [26]VIB Department of Plant Systems Biology, Ghent University, Technologie Park 927, 9052 Ghent, Belgium. [27]MIPS/IBI Institute for Bioinformatics, GSF Research Center for Environment and Health, Ingolstaedter Landstrasse 1, D-85764 Neuherberg, Germany. [28]Department of Plant Biology, 166 Plant Biology Building, Michigan State University, East Lansing, MI 48824–1312, USA. [29]RIKEN Plant Science Center, 1-7-22 Suehiro, Tsurumi, Yokohama, Kanagawa 230-0045, Japan. [30]Free University, Institute for Biology, Applied Genetics Neubau, Albrecht-Thaer-Weg 6, D-14195 Berlin, Germany. [31]Max-Planck Institute of Plant Breeding Research, Carl-von-Linne-Weg 10, D-50829 Cologne, Germany. [32]Biology Department, Kenyon College, Gambier, OH 43022, USA. [33]Pflanzenphysiologie, Justus Liebig University, Senckenbergstrasse 3, D-35390 Giessen, Germany. [34]Institute of General Botany, Johannes Gutenberg-University, D-55099 Mainz, Germany. [35]Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287–1604, USA. [36]University of Tennessee-Memphis, 101 Molecular Science Building, 858 Madison Avenue, Memphis, TN 38163, USA. [37]Department of Bioinformatics, Biozentrum, Am Hubland, Würzburg University, D-97074 Würzburg, Germany. [38]Biology Department, San Diego State University, North Life Sciences Room 102, 5500 Campanile Drive, San Diego, CA 92182–4614, USA. [39]Department of Biology, Gilmer Hall, 485 McCormick Road, University of Virginia, Charlottesville, VA 22903, USA. [40]Department of Plant Biology, University of Minnesota, 250 Biological Science Center, 1445 Gortner Avenue, St. Paul, MN 55108, USA. [41]Biological Chemistry Department, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK. [42]Biomathematics and Bioinformatics Department, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK. [43]ERATO, Japan Science and Technology Agency, Okazaki 444-8585, Japan. [44]Department of Integrative Biology, 3060 Valley Life Sciences Building, University of California, Berkeley, CA 94720, USA. [45]Genome Project Solutions, 1024 Promenade Street, Hercules, CA 94547, USA.

*To whom correspondence should be addressed: E-mail: rsq@wustl.edu

gests that most TAGs were established recently (K_s < 0.1). It is noteworthy that *P. patens* TAG partners tend to be located on opposite strands (64.4%, with 36.4% in head-to-head orientation and 28.0% in tail-to-tail orientation), whereas there is a tendency (68 to 88%) for TAGs to be located on the same DNA strand in *A. thaliana*, *O. sativa* (*11*), *Caenorhabditis elegans*, *Homo sapiens*, *Mus musculus*, and *Rattus norvegicus* (*11*, *14*). Significantly fewer substitutions (P < 0.001) are observed within them (average K_s = 0.59) than in those that are located on the same strand (K_s = 1.25). Homologous recombination between TAGs on the same strand may have resulted in loss of such TAGs, whereas gene conversion associated with homologous recombination of TAGs on opposite strands may have resulted in reduction of sequence divergence (K_s) between those. These differences in TAG organization might be connected to the exceptional reliance on sequence similarity for DNA repair observed in *P. patens* (*15*, *16*). Alternatively, the generation and exclusion rate of TAGs on the opposite strand might have been higher than for TAGs on the same strand in the ancestor of *P. patens*.

Gene and domain family expansion patterns. Eukaryotic gene family sizes differ mainly because of different rates of gene duplication and retention, and gene content differences may reflect species-specific adaptations. Overall, lineage-specific gains among domain families occurred at a lower rate (by a factor of about three) in the *P. patens* compared with the *A. thaliana* lineage (Fig. 3A). Similarly, in comparisons with the *O. sativa* and *P. trichocarpa* lineages, gene gain rates in the *P. patens* lineage are substantially lower. Among gene families shared by both *P. patens* and *A. thaliana*, there are consistently fewer families with relatively large gains (≥6) in the *P. patens* lineage (Fig. 3B), which indicates that the gain rate differences noted in Fig. 3A are mainly due to higher retention rates of large families in the *A. thaliana* lineage. In addition, many *P. patens* gene families with higher-than-average gain rates in general also have elevated rates of gene loss (Fig. 3C).

Highly expanded gene families in the *P. patens* lineage are not necessarily highly expanded in the *A. thaliana* lineage ($r^2 = 0.33$, $P < 2 \times 10^{-16}$). Only 36 families with significantly higher-than-average gains are common to both the *P. patens* and *A. thaliana* lineages, whereas 43 are significantly expanded only in *P. patens* (Fig. 3A). Examples of parallel expansion include genes encoding protein kinases and leucine-rich–repeat proteins, as well as Apetala 2 (AP2) and Myb transcription factors. Transcription factor duplicates are retained in the *P. patens* lineage with a rate lower than those in the flowering plant lineage, yet higher than in algae (*17*); for example, the MADS-box and WRKY transcription factor families are intermediate in size compared with flowering plants and algae (table S12 and S13).

Families that significantly expanded only in the *P. patens* lineage include histidine kinases and response regulators. Both families are parts of two-component signaling networks important in plants, fungi, and bacteria. These two families are much larger in *P. patens* than those found in sequenced angiosperm genomes; their increased size suggests a more elaborate use of two-component systems in *P. patens*.

The *P. patens* genome contains genes for each of the core groups of small guanosine triphosphatases (G proteins) (fig. S6, A and B), consistent with increased complexity of vesicle trafficking machinery, not present in green algae, which suggests that such complexity was already present in

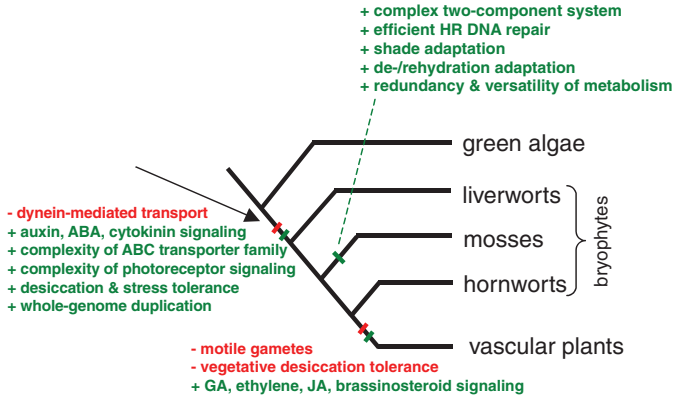

Fig. 1. Land plant evolution. Bryophytes comprise three separate lineages which, together with the vascular plants (including the flowering plants), make up the embryophytes (land plants) (*38*). These four lineages, remnants of the initial radiation of land plants in the Silurian, began to diverge from each other about 450 Ma.

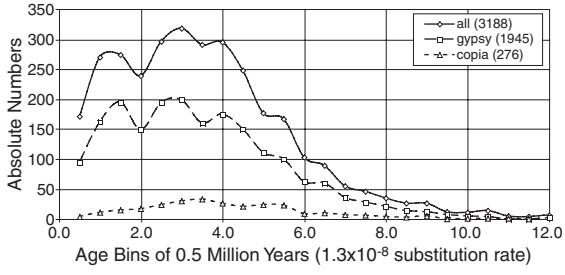

Fig. 2. Periodic cycles of LTR retrotransposon activity. *P. patens* underwent periodic cycles of LTR-R amplifications. The most recent activity peaks at an estimated 1 to 1.5 Ma, preceded by invasion events around 3, 4, and 5.5 Ma. Gypsy-like elements are younger (average 3.2, median 3.0) than copia-like elements (average 3.9, median 3.6), coinciding with an increased full-length copy number by a factor of seven. The gradual decrease between 5 to 12 Ma probably reflects element deterioration leading to loss of ability to detect these elements. Numbers found of each element are shown in parentheses.

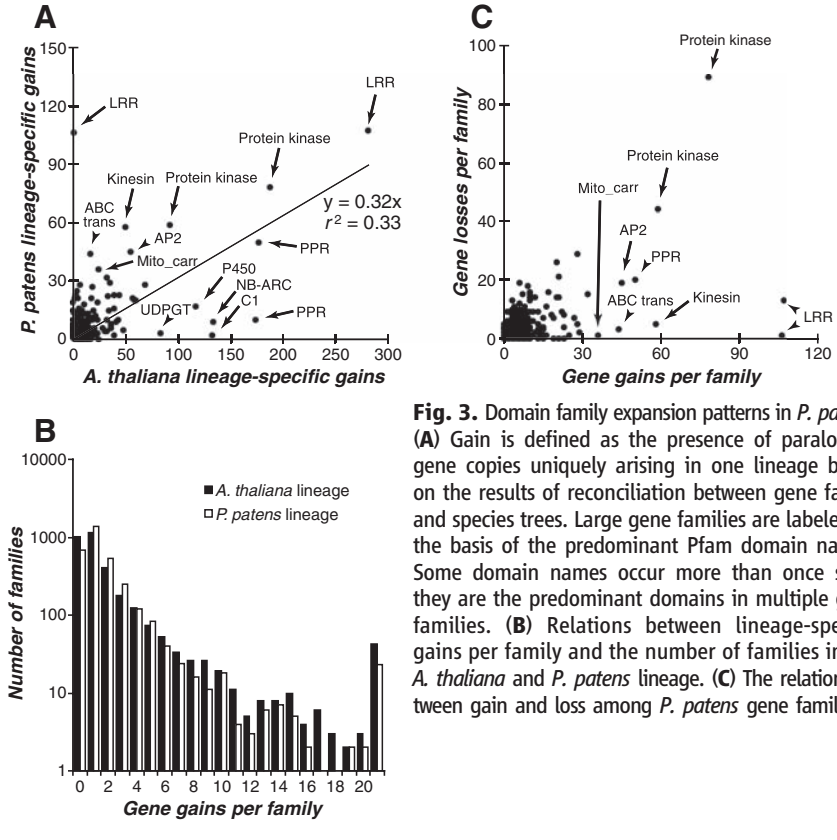

Fig. 3. Domain family expansion patterns in *P. patens*. (**A**) Gain is defined as the presence of paralogous gene copies uniquely arising in one lineage based on the results of reconciliation between gene family and species trees. Large gene families are labeled on the basis of the predominant Pfam domain names. Some domain names occur more than once since they are the predominant domains in multiple gene families. (**B**) Relations between lineage-specific gains per family and the number of families in the *A. thaliana* and *P. patens* lineage. (**C**) The relation between gain and loss among *P. patens* gene families.

the last common ancestor of land plants. *P. patens* also has a large adenosine triphosphate binding cassette (ABC) superfamily [121 members; (tables S14 and S15); St 29_ABDI/C/F/G, 9, 57, 110 to 113], similar in size to that in *A. thaliana* (130) and *O. sativa* (129), but larger than *O. tauri* (~50) and twice that of humans and *Drosophila melanogaster* (48 and 56, respectively). In flowering plants, most ABC-containing proteins are membrane-bound transporters for lipids, hormones, secondary metabolites, metals, and xenobiotics and control certain ion channels. The sessile habit and metabolic diversity of land plants appears to require a large repertoire of ABC proteins.

Adaptations to the Terrestrial Environment

Desiccation tolerance. Desiccation tolerance (DT) is widespread in reproductive structures of vascular plants, but vegetative DT is rare, except among bryophytes (*18*). Evolution of this trait was important in facilitating the colonization of the land, but was lost subsequently in vascular plants. DT in seeds is dependent on the phytohormone abscisic acid (ABA) to induce expression of seed-specific genes, such as late embryogenesis abundant proteins (LEAs), a group of proteins that accumulate during desiccation. *P. patens* is highly dehydration-tolerant (*19*) and contains orthologs of LEA genes and other genes expressed during the DT response in the poikilohydric moss *Tortula ruralis* (*20*) and in flowering plants (*21*).

ABA signaling also operates in the *P. patens* drought response (*21*). The genome contains putative homologs of the *A. thaliana* ABA receptors, one of which appears to have been specialized for a role in seed development, and the transcription factor ABI5, which implicates it in the regulation of ABA-mediated gene expression. Particularly interesting is ABI3, the seed-specific transcription factor of the B3 family (St 132), which, when mutated, results in the loss of desiccation-tolerance in seeds (*22*). The *P. patens* genome contains four *ABI3*-like genes, one of which (*PpABI3A*) functions to potentiate ABA responses in *P. patens* and partially complements the *A. thaliana abi3-6* mutant (*23*).

Finding these genes in *P. patens* and similar sequences in liverworts (*Riccia fluitans* and *Marchantia polymorpha*) suggests that desiccation tolerance gene networks likely originated in the last common ancestor of extant land plants.

Metabolic pathways. Cytochrome P450 enzymes that incorporate oxygen into small lipophilic compounds are represented by 250 to 350 members in genomes of flowering plants, 71 genes in *P. patens*, and 40 in *C. reinhardtii*. Specific examples of P450s lacking in *P. patens* are related to the absence and regulation of key molecules in flowering plants. One P450 required for the synthesis of gibberellic acid (CYP88) is absent, as is the enzyme needed to make *S*-lignols (CYP84) required for the accumulation of lignin. The CYP86 family includes fatty acid omega-hydroxylases involved in the formation of cutin, which prevents dehydration of plant tissues. The presence of CYP86 in *P. patens*, but not in green algae, suggests that cutin may have evolved in the ancestral land plants as an innovative mechanism to survive a terrestrial habitat.

Most enzymatic steps in carotenoid and chlorophyll biosynthetic pathways are more complex in terms of paralog frequencies in *P. patens* than in *A. thaliana* and *C. reinhardtii* (Fig. 4 and table S16). This is consistent with previous interpretations that the *P. patens* genome encodes seemingly redundant metabolic pathways and contains a network of genes for functions like phototoxic stress tolerance (*9*). Unlike light harvesting complex (LHC) proteins, most genes (79%) of the carotenoid and chlorophyll metabolic pathways are not TAGs and were acquired during the whole-genome duplication, i.e., since the divergence from the lineage leading to flowering plants (*9*).

One striking exception is the genes involved at the branching point of siroheme and heme/chlorophyll formation (Fig. 4 and table S16). UROS and UMT are encoded by single copy genes in *P. patens* (Fig. 4), whereas conserved ancient paralogs encode UROD (St 76). These paralogs had already been acquired before the split of green algae and land plants (St 76) and probably are functionally divergent (*24*, *25*). Note that both the UROD3 (St 76) and CPX2 (St 59) subfamilies are present in algae and *P. patens*, but have been lost in flowering plants.

Signaling pathways. The phytohormones and light receptors for morphogenesis found in flowering plants are absent in the unicellular algae, but are present in *P. patens*, e.g., genes for all four classes of cytokinin signaling pathways found in flowering plants. These include at least three

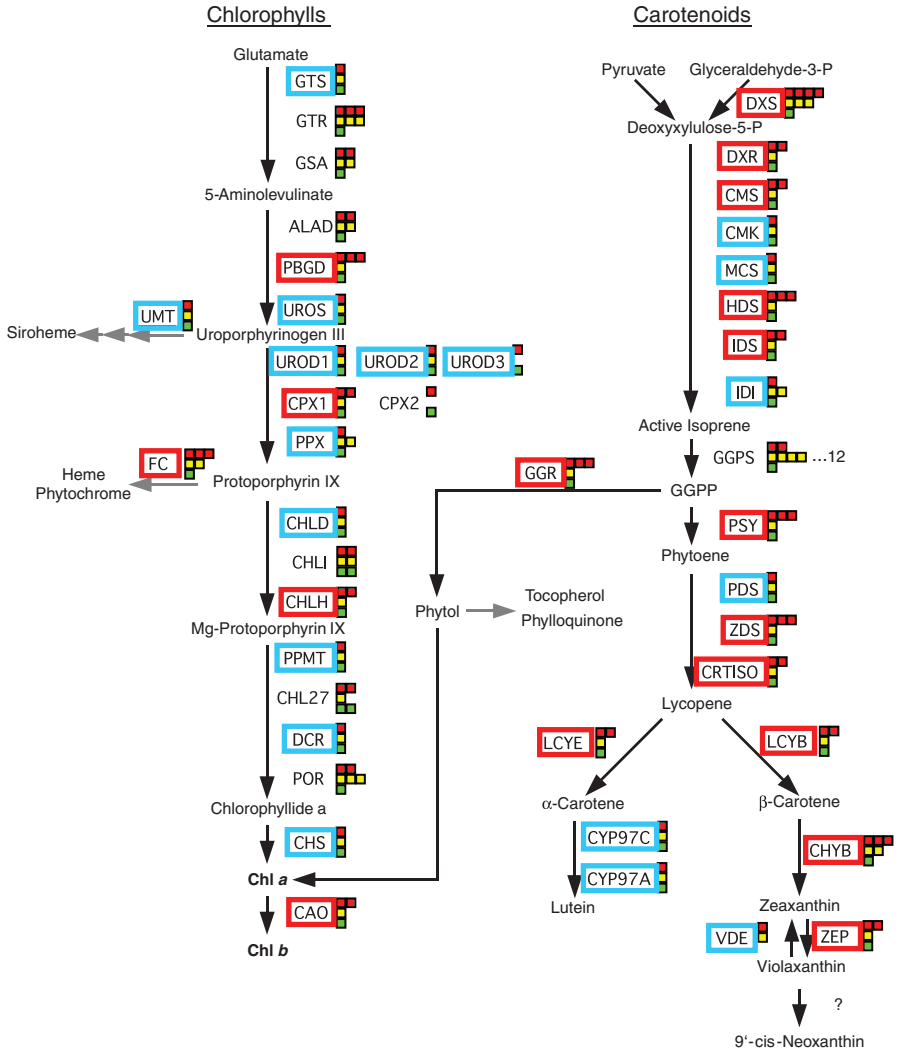

Fig. 4. Paralog frequencies in the biosynthetic pathways of chlorophylls and carotenoids in *P. patens*, *A. thaliana*, and *C. reinhardtii*. Denoted are products that accumulate to significant amounts, major intermediates, and known enzymes of both pathways (for full names of enzymes, see table S16). Major pathways are indicated by black arrows; branch-points leading to the formation of related compounds (italicized) are indicated by gray arrows. For each reaction, colored squares symbolize the number of (iso-) enzymes in *P. patens* (red), *A. thaliana* (yellow), and *C. reinhardtii* (green). Enzymes for which *P. patens* has more paralogs than *A. thaliana* and *C. reinhardtii* are boxed in red, those encoded by unique genes in *P. patens* are boxed in blue.

cytokinin receptors, two of which have been confirmed by EST evidence, which make *P. patens* the earliest diverging species that contains genes for all members of the cytokinin signal transduction pathway known today.

Ten gene families implicated in auxin homeostasis and signaling have been analyzed [(table S17), St 25, 33_A/B, 41, 45, 71, 73_7, 77, 85, 88, 89]. The *C. reinhardtii*, *O. lucimarinus*, and *O. tauri* genomes do not encode these, but the *P. patens* genome encodes members of each family [although, on the basis of phylogenies of the GH3 and ILL proteins, St 71 and 85, *P. patens* might not conjugate IAA to alanine, leucine, aspartic acid, or glutamic acid consistent with empirical data (*26*)]. Angiosperms dedicate a larger proportion of their genomes to auxin signaling; only one (AUX1/LAX; St 41) of the 10 families has as many members as angiosperm genomes. On the basis of analysis of *A. thaliana* and our phylogenetic analyses, the auxin signaling pathway has undergone substantial functional diversification within vascular plants since they diverged from bryophytes.

Although no ethylene responses have been noted in mosses, the *P. patens* genome codes for six putative ETR-like ethylene receptors, at least one of which is known to bind ethylene (*27*). Two putative 1-aminocyclopropane-1-carboxylate (ACC) synthases, catalyzing a critical step in ethylene biosynthesis, were also found. Two transcription factors with strong similarity to the EIN3 ethylene signaling family are also apparent as are six N-RAMP-type (natural resistance–associated macrophage protein) channel proteins, one or more of which might be involved in ethylene signaling, similar to EIN2 in *A. thaliana*.

Protective proteins. Adaptation to land also required the evolution of proteins that protect against stresses such as variation in temperature, light, and water availability. One example of this is the expansion of the heat shock protein 70 (HSP70) family to nine cytosolic members in *P. patens* (St 24), whereas all algal genomes sequenced to date encode one single cytosolic HSP70 (*28*).

The complement of the LHC genes is significantly expanded in *P. patens* when compared with algae and vascular plants [St 58_A (table S18A)]. Although several LHC homologs were already present in the last common ancestor of all land plants, more have been retained after the whole-genome duplication in *P. patens*, and more of these genes are present in TAGs than in *A. thaliana* (table S19). Redundancy and expansion of these abundantly expressed proteins probably contributes to a robustness of the photosynthetic antenna, i.e., the capacity to deal with high light intensities. The photoprotective early light–induced proteins (ELIPs) expanded extensively in *P. patens* [St 58_B (table S18B)]. Numerous ELIP-like proteins with supposedly free radical scavenging activity may reflect adaptation to dehydration and rehydration cycles and associated avoidance of photo-oxidative damage.

DNA repair. DNA damage repair maintains genomic integrity. Double-strand breaks (DSBs) can be repaired by nonhomologous end-joining (NHEJ), but are more precisely repaired using a second copy of the sequence. The introduction of linear DNA into a cell mimics DNA damage, and mosses, uniquely among plants, but like yeast, show a strong preference for the use of a homologous sequence for the incorporation of linear DNA into the genome.

Cell-cycle control is tightly connected to DNA-damage repair (*29*). Proteins known to be involved in these processes in both vertebrates and *A. thaliana* are ATM, ATR, CHK1, CHK2, PARP1, BRCA1, BRCA2, and BARD1. Although *P. patens* encodes the first four of these, there are no homologs found of BRCA1, BRCA2, and BARD1. RAD51 and the RAD51 paralogs (RAD51B, RAD51C, RAD51D, XRCC2, and XRCC3) are important for repair that results in homologous recombination in vertebrates and in *A. thaliana*; *P. patens* encodes all but XRCC3. However, although *A. thaliana* encodes one RAD51, *P. patens* encodes two (*30*). Other genes involved in DSB repair, chromatin remodeling, and processing of recombination intermediates known from *A. thaliana* (INO80, RAD54, MRE11, RAD50, NBS1, RecQ helicases (WRN, BLM, MUS81) are also present in *P. patens*. Additionally, both plant species, but not metazoans, encode SRS2, whereas *P. patens*, like other plants, lacks RAD52. In *A. thaliana* and yeast, the KU70/KU80 complex, DNA ligase IV, and XRCC4 contribute to NHEJ. These genes are also encoded by the *P. patens* genome. In addition, both plant species, but not yeast (*31*), encode the DNA-dependent protein kinase catalytic subunit (DNA-PKcs).

In our phylogenetic analyses, *P. patens* homologs of RAD54B, as well as Centrins and CHD7, cluster with algal and metazoan homologs, whereas flowering plant homologs do not (St 12_2, 28_2, 28_7). Although RAD51 and RAD54 interact in chromatin remodeling in humans (*32*), Centrins are important for genome stability in *C. reinhardtii* (*33*) and in nucleotide excision and DSB repair in *A. thaliana* (*34*). CHD7 is a chromodomain DNA helicase, important for chromatin structure, mutation of which causes developmental aberrations in mammals (*35*).

DNA damage is repaired by multisubunit macromolecular complexes of dynamic composition and conformation (*36*). The special features of the *P. patens* genome (no BRCA1, BRCA2, and BARD1, duplicated RAD51, and phylogenetically conserved RAD54B, Centrins, and CHD7) may well reflect the specific needs of a haploid genome for genome integrity surveillance and account for the efficiency of homology-dependent DSB repair in the *P. patens* genome.

Conclusions for Land Plant Evolution

P. patens occupies a position on the evolutionary tree that, through comparisons with aquatic algae and vascular plants, allows reconstruction of evolutionary changes in genomes that are concomitant to the conquest of land. From this, we conclude that the last common ancestor of all land plants (i) lost genes associated with aquatic environments (e.g., flagellar components for gametic motility); (ii) lost dynein-mediated transport; (iii) gained signaling capacities, such as those for auxin, ABA, cytokinin, and more complex photoreception; (iv) gained tolerance for abiotic stresses, such as drought, radiation, and extremes of temperature; (v) gained more elaborate transport capabilities; and (vi) had an overall increase in gene family complexity. Some of these events may have been enabled by the opportunities for evolutionary novelty created by one or more duplications of the whole genome.

These comparisons also enable reconstruction of the genomic events that occurred after the split of vascular plants and mosses. For example, the former acquired even more elaborate signaling [e.g., through gibberellic acid (GA), jasmonic acid (JA), ethylene, and brassinosteroids], but lost vegetative dehydration tolerance and motile gametes, whereas the latter gained an elaborate use of two-component systems, efficient homology-based DNA repair, and adaptation to shade and de-/rehydration cycles, as well as a redundant and versatile metabolism. The *P. patens* genome sequence provides a resource for the study of both gene function (*37*) and evolutionary reconstruction.

References and Notes

1. S. K. Floyd, J. L. Bowman, *Int. J. Plant Sci.* **168**, 1 (2007).
2. P. R. Kenrick, P. Crane, *Nature* **389**, 33 (1997).
3. Materials and methods are available as supporting material on *Science* Online.
4. Version 1.1 of the *P. patens* genome assembly and annotation can be accessed through the JGI Genome Portal at www.jgi.doe.gov/Physcomitrella; sequences have been deposited at DDBJ/EMBL/GenBank under the project accession ABEU01000000.
5. Y.-Y. Shen *et al.*, *Nature* **443**, 823 (2006).
6. C. Vitte, J. L. Bennetzen, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17638 (2006).
7. V. V. Kapitonov, J. Jurka, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8714 (2001).
8. M. Lynch, J. S. Conery, *Science* **290**, 1151 (2000).
9. S. A. Rensing *et al.*, *BMC Evol. Biol.* **7**, 130 (2007).
10. International Rice Genome Sequencing Project, *Nature* **436**, 793 (2005).
11. C. Rizzon, L. Ponger, B. S. Gaut, *PLoS Comput Biol* **2**, e115 (2006).
12. G. A. Tuskan *et al.*, *Science* **313**, 1596 (2006).
13. St stands for supplementary tree; these can be accessed via www.cosmoss.org/bm/supplementary_trees/Rensing_et_al_2007/.
14. C. Semple, K. H. Wolfe, *J. Mol. Evol.* **48**, 555 (1999).
15. H. Puchta, *J. Exp. Bot.* **56**, 1 (2005).
16. Y. Kamisugi *et al.*, *Nucleic Acids Res.* **34**, 6205 (2006).
17. S. Richardt, D. Lang, W. Frank, R. Reski, S. A. Rensing, *Plant Physiol.* **143**, 1452 (2007).
18. M. Oliver, J. Velten, B. Mishler, *Integr. Comp. Biol.* **45**, 788 (2005).
19. W. Frank, D. Ratnadewi, R. Reski, *Planta* **220**, 384 (2005).
20. L. Saavedra *et al.*, *Plant J.* **45**, 237 (2006).
21. A. C. Cuming, S. H. Cho, Y. Kamisugi, H. Graham, R. S. Quatrano, *New Phytol.* **176**, 275 (2007).
22. J. Giraudat *et al.*, *Plant Cell* **4**, 1251 (1992).
23. H. H. Marella, Y. Sakata, R. S. Quatrano, *Plant J.* **46**, 1032 (2006).
24. G. Hu, N. Yalpani, S. P. Briggs, G. S. Gurmukh, S. Johal, *Plant Cell* **10**, 1095 (1998).
25. H. P. Mock, B. Grimm, *Plant Physiol.* **113**, 1101 (1997).
26. A. E. Sztein, J. D. Cohen, I. G. de la Fuente, T. J. Cooke, *Am. J. Bot.* **86**, 1544 (1999).
27. W. Wang *et al.*, *Plant Cell* **18**, 3429 (2006).

28. W. Wang, B. Vinocur, O. Shoseyov, A. Altman, *Trends Plant Sci.* **9**, 244 (2004).
29. P. Sung, H. Klein, *Nat. Rev. Mol. Cell Biol.* **7**, 739 (2006).
30. U. Markmann-Mulisch *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2959 (2002).
31. J. M. Daley, P. L. Palmbos, D. Wu, T. E. Wilson, *Annu. Rev. Genet.* **39**, 431 (2005).
32. Y. Zhang *et al.*, *Nat. Struct. Mol. Biol.* **14**, 639 (2007).
33. I. Zamora, W. F. Marshall, *BMC Biol.* **3**, 15 (2005).
34. L. Liang, S. Flury, V. Kalck, B. Hohn, J. Molinier, *Plant Mol. Biol.* **61**, 345 (2006).
35. E. A. Hurd *et al.*, *Mammal Genome* **18**, 94 (2007).
36. O. Llorca, *Curr. Opin. Struct. Biol.* **17**, 215 (2007).
37. R. S. Quatrano, S. F. McDaniel, A. Khandelwal, P.-F. Perroud, D. J. Cove, *Curr. Opin. Plant Biol.* **10**, 182 (2007).
38. D. G. Kelch, A. Driskell, B. Mishler, in *Molecular Systematics of Bryophytes*, B. Goffinet, V. Hollowell, R. Magill, Eds. (Missouri Botanical Garden Press, St. Louis, MO, 2004), pp. 3–12.

39. We thank K. Zhou and S. Pitluck at DOE Joint Genome Institute (JGI) for GenBank submissions and G. Werner and his group at JGI for support of gene annotation tools. Discussions with M. Oliver and K. Fisher on desiccation tolerance are greatly appreciated. N. Lyons efficiently handled many of the administrative tasks throughout the project and the detailed preparation of the final manuscript. Part of this work was funded by German National Science Foundation (DFG) grant RE 837/10 to R.R., by a U.K. Biotechnology and Biological Sciences Research Council (BBSRC) grant (24/P11357) to A.C.C. and by grants to R.S.Q. from the NSF (IBN 0112461 and 0425749-1) and Washington University. This work was also supported by Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Culture, Sports, Science and Technology of Japan (to T.T., T.F., Y. Kuroki, A.T., Y.S., S.H., K.Y., S.S., Y. Kohara, A.F., T.M., T.N., and M.H). This work was performed under the auspices of the U.S. Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under contract no. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract no. DE-AC02-05CH11231, and Los Alamos National Laboratory under contract no. W-7405-ENG-36.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1150646/DC1
Materials and Methods
SOM Text
Figs. S1 to S8
Tables S1 to S23
References

18 September 2007; accepted 21 November 2007
Published online 13 December 2007;
10.1126/science.1150646
Include this information when citing this paper.

Hidden Degrees of Freedom in Aperiodic Materials

Bertrand Toudic,[1,2]* Pilar Garcia,[1,2] Christophe Odin,[1,2] Philippe Rabiller,[1,2] Claude Ecolivet,[1,2] Eric Collet,[1,2] Philippe Bourges,[3] Garry J. McIntyre,[4] Mark D. Hollingsworth,[5] Tomasz Breczewski[6]

Numerous crystalline materials, including those of bioorganic origin, comprise incommensurate sublattices whose mutual arrangement is described in a superspace framework exceeding three dimensions. We report direct observation by neutron diffraction of superspace symmetry breaking in a solid-solid phase transition of an incommensurate host-guest system: the channel inclusion compound of nonadecane/urea. Strikingly, this phase transition generates a unit cell doubling that concerns only the modulation of one substructure by the other—an internal variable available only in superspace. This unanticipated pathway for degrees of freedom to rearrange leads to a second phase transition, which again is controlled by the higher dimensionality of superspace. These results reveal nature's capacity to explore the increased number of phases allowed in aperiodic crystals.

Aperiodicity (*1*, *2*) plays a central role in the structure and physical properties of materials as diverse as quasi-crystals (*3*), superconductors (*4*), and zeolites (*5*). In a number of solid-state supramolecular assemblies—a class of materials ubiquitous both in nature and in the laboratory—modern diffraction methods have revealed incommensurate relations between the basic structures of the noncovalently assembled constituents. In host-guest systems, for example, aperiodicity is an essential element that controls phenomena as diverse as crystal growth (*6*), polar ordering of guests (*7*), and absorption and molecular transport (*8*). Aperiodic materials are described in the framework of a higher-dimensional space called superspace, which decomposes into two orthogonal subspaces: the usual three-dimensional physical space and an internal one (*9*, *10*). The increase in dimensionality theoretically allows many more phases (*11*, *12*); however, the role of aperiodicity in the structural changes of supramolecular assemblies has been largely ignored. In the transformations of such materials, aperiodicity resulting from different interacting length scales should play a major role, coupling structure and dynamics, in particular through the mechanism of free sliding (*13*, *14*). Here, we highlight the richness of transformations associated with the new degrees of freedom hidden in superspace, taking as an example a prototypical host-guest system of nonadecane/urea.

Supramolecular chemistry and crystal engineering enable the formation of a diverse array of aperiodic host-guest architectures in which guest molecules are confined with their own periodicity to nanochannels. Although such host structures may be held together by simple van der Waals forces (*8*), covalent bonds (*5*, *15*), or hydrogen bonds, as in the diverse structures formed by certain dipeptides (*16*), a common feature is that the host is more rigid than the guest component. This is typically the case for urea inclusion compounds, in which urea molecules are connected by hydrogen bonds to form helical ribbons, which are woven together to form a honeycomb array of linear, nonintersecting, hexagonal tunnels that contain straight-chain hydrocarbons and analogs (Fig. 1) (*17*). Guests such as nonadecane pack end to end within van der Waals contact of each other and undergo substantial torsions, librations, and translations, as well as diffusive motions or jumps about the sixfold axis of the host (*18*). For this host-guest system, the ratio of repeat lengths along the channel axis of the urea host ($\mathbf{c}_{host}$) and nonadecane guest ($\mathbf{c}_{guest}$) is not a rational number, so the crystal is said to be incommensurate.

Although the locations ($\mathbf{G}_{hkl}$) of diffraction peaks of typical three-dimensional crystals can be defined with three indices (*h*, *k*, and *l*), an additional parameter or parameters must be used to describe the diffraction patterns of aperiodic materials such as incommensurate channel inclusion compounds. In the linear tunnels of urea, $\mathbf{c}_{host}$ and $\mathbf{c}_{guest}$ are parallel, so one extra parameter is required to describe the system. Thus, for this nanotubular intergrowth structure, with its single incommensurate direction (**c**), a four-dimensional superspace description gives the positions of all of the Bragg peaks (*8*, *9*):

$$\mathbf{G}_{hklm} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*_{host} + m\mathbf{c}^*_{guest} \qquad (1)$$

where $\mathbf{a}^*$, $\mathbf{b}^*$, $\mathbf{c}^*_{host}$, and $\mathbf{c}^*_{guest}$ are the conventional reciprocal unit vectors, which differ

[1]Université de Rennes 1, Institut de Physique de Rennes (IPR), 35042 Rennes Cedex, France. [2]CNRS, UMR 6251, IPR, 263 Avenue du Général Leclerc, 35042 Rennes Cedex, France. [3]Laboratoire Léon Brillouin, CEA-CNRS, CE-Saclay, 91191 Gif-sur-Yvette, France. [4]Institut Laue-Langevin, 38042 Grenoble Cedex 9, France. [5]Department of Chemistry, Kansas State University, Manhattan, KS 66506, USA. [6]Facultad de Ciencias, Universidad del Pais Vasco, Apdo 644, Bilbao, Spain.

*To whom correspondence should be addressed. E-mail: bertrand.toudic@univ-rennes1.fr