D' (*12*), ranging between values of 0 (a fully random association) to 1 (a complete association between allele pairs). Because the underlying model assumes random mating within the population from which the sample was taken, it is only appropriate to calculate this statistic for the Leeds sample. D' was high (>0.5) for all markers across the 1.4-Mb sequence examined, defining a region of approximately 400 kb where $D' > 0.9$, with a maximum D' at locus *c* (Fig. 2). Interpolation of D' values at loci *b*, *c*, and *d* suggests that the genetic polymorphism controlling the morph phenotype is located within 100 kb on either side of locus *c*. Further consideration of the "missing" (repulsion-phase) haplotypes separately for *carbonaria* and *typica* (Table 2) emphasizes the deficit of repulsion-phase *carbonaria* haplotypes across all six loci. In contrast, *typica* haplotypes show only weak deficits of *carbonaria*-type marker alleles, which is consistent with the view that these alleles were segregating in the population before the genesis of *carbonaria*. At loci *a* and *b*, the deficit for *typica* is greater because the *carbonaria*-type alleles (C and A, respectively) were rare in the ancestral population. These two loci suggest that *carbonaria*-to-*typica* haplotype introgression has been weak.

Bombyx mori chromosome 17 and its orthologs in other lepidopterans are rich in major color-patterning genes, such as *black moth* and *wild wing spot* (*13*). These *Bombyx mori* genes do not map closely to the *carbonaria* locus (fig. S5). However, *Bicyclus anynana* LG17 contains two pigment-patterning mutants, *067* and *Bigeye*, that both affect eyespot size, with *Bigeye* predicted to reside within the *carbonaria* region (*14*). The *Bigeye* and *carbonaria* phenotypes are clearly very different, but they share a large increase in the proportion of melanized scales. The *carbonaria* core region also overlaps the mimetic patterning locus in four *Heliconius* species, collectively referred to as the *Yb-P-Yb/Sb-Cr* locus (*15*–*17*). The *B. betularia* genes identified in this region so far correspond entirely with those described for the *Yb-P-Yb/Sb-Cr* region. A major feature distinguishing *Heliconius* forms is the amount and distribution of black, as with the various *B. betularia* morphs (*4*). This unlikely coincidence suggests that the control of melanin pattern formation in these deeply diverged lepidopterans may have a common genetic basis, the functional units of which have yet to be identified.

The rapid spread of an initially unique haplotype, driven by strong positive selection, is expected to generate the profile of linkage disequilibrium we have observed (*18*), establishing that UK industrial melanism in the peppered moth was seeded by a single recent mutation that spread to most parts of mainland Britain and also colonized the Isle of Man (fig. S4). Paradoxically, although the *carbonaria* morph is now strongly disadvantageous and consequently rare in the United Kingdom, the rapidity of its decline (*19*) has minimized the eroding effect of *typica* introgression on the molecular footprint of strongly positive selection created during its ascendency.

References and Notes

1. M. E. N. Majerus, *Melanism: Evolution in Action* (Oxford Univ. Press, Oxford, 1998).
2. L. M. Cook, *Q. Rev. Biol.* **78**, 399 (2003).
3. I. J. Saccheri, F. Rousset, P. C. Watts, P. M. Brakefield, L. M. Cook, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 16212 (2008).
4. L. M. Cook, J. Muggleton, *Entomol. Gaz.* **54**, 211 (2003).
5. D. R. Lees, in *Genetic Consequences of Man Made Change*, J. A. Bishop, L. M. Cook, Eds. (Academic Press, London, 1981), pp. 129–176.
6. D. L. Stern, *Evolution, Development and the Predictable Genome* (Roberts and Company Publishers, Greenwood Village, CO, 2010).
7. M. E. N. Majerus, N. I. Mundy, *Trends Genet.* **19**, 585 (2003).
8. J. R. True, *Trends Ecol. Evol.* **18**, 640 (2003).
9. P. J. Wittkopp, S. B. Carroll, A. Kopp, *Trends Genet.* **19**, 495 (2003).
10. A. E. van't Hof, I. J. Saccheri, *PLoS ONE* **5**, e10889 (2010).
11. Materials and additional results are available as supporting material on *Science* Online.
12. P. Hedrick, S. Kumar, *Eur. J. Hum. Genet.* **9**, 969 (2001).
13. M. R. Goldsmith, A. S. Wilkins, *Molecular Model Systems in the Lepidoptera* (Cambridge Univ. Press, Cambridge, 1995).
14. P. Beldade, S. V. Saenko, N. Pul, A. D. Long, *PLoS Genet.* **5**, e1000366 (2009).
15. R. Papa, A. Martin, R. D. Reed, *Curr. Opin. Genet. Dev.* **18**, 559 (2008).
16. M. Joron *et al.*, *PLoS Biol.* **4**, e303 (2006).
17. M. R. Kronforst, D. D. Kapan, L. E. Gilbert, *Genetics* **174**, 535 (2006).
18. P. Pfaffelhuber, A. Lehnert, W. Stephan, *Genetics* **179**, 527 (2008).
19. L. M. Cook, S. L. Sutton, T. J. Crawford, *J. Hered.* **96**, 522 (2005).

Acknowledgments: This work was funded by Natural Environment Research Council grant NE/C003101/1 to I.J.S. Cytogenetic experiments were financed by Grant Agency of the Academy of Sciences of the Czech Republic grant IAA600960925, Entomology Institute project Z50070508, and Grant Agency of the Czech Republic 521/08/H042 grants to F.M. and M.D. Specimens for the UK-wide survey were made available by several museums and private collectors, listed in table S4. J. Delf produced fig. S3. GenBank accession numbers are listed in table S1.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1203043/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 to S4
References

19 January 2011; accepted 30 March 2011
Published online 14 April 2011;
10.1126/science.1203043

The Selaginella Genome Identifies Genetic Changes Associated with the Evolution of Vascular Plants

Jo Ann Banks,[1]* Tomoaki Nishiyama,[2,3] Mitsuyasu Hasebe,[3,4,5] John L. Bowman,[6,7] Michael Gribskov,[8] Claude dePamphilis,[9,10,11] Victor A. Albert,[12] Naoki Aono,[4] Tsuyoshi Aoyama,[4,5] Barbara A. Ambrose,[13] Neil W. Ashton,[14] Michael J. Axtell,[9] Elizabeth Barker,[14] Michael S. Barker,[15] Jeffrey L. Bennetzen,[16] Nicholas D. Bonawitz,[17] Clint Chapple,[17] Chaoyang Cheng,[3] Luiz Gustavo Guedes Correa,[18] Michael Dacre,[19] Jeremy DeBarry,[16] Ingo Dreyer,[20] Marek Elias,[21,22] Eric M. Engstrom,[23] Mark Estelle,[24] Liang Feng,[25] Cédric Finet,[26] Sandra K. Floyd,[6] Wolf B. Frommer,[27] Tomomichi Fujita,[28] Lydia Gramzow,[29] Michael Gutensohn,[30,31] Jesper Harholt,[32] Mitsuru Hattori,[33,34] Alexander Heyl,[35] Tadayoshi Hirai,[3,36] Yuji Hiwatashi,[4,5] Masaki Ishikawa,[3] Mineko Iwata,[3] Kenneth G. Karol,[13] Barbara Koehler,[18] Uener Kolukisaoglu,[37,38] Minoru Kubo,[3] Tetsuya Kurata,[3,39] Sylvie Lalonde,[27] Kejie Li,[8] Ying Li,[8,40] Amy Litt,[13] Eric Lyons,[41] Gerard Manning,[19] Takeshi Maruyama,[42] Todd P. Michael,[43,44] Koji Mikami,[45] Saori Miyazaki,[4,46] Shin-ichi Morinaga,[4,47] Takashi Murata,[4,5] Bernd Mueller-Roeber,[48] David R. Nelson,[49] Mari Obara,[3,50] Yasuko Oguri,[3] Richard G. Olmstead,[51] Naoko Onodera,[3,52] Bent Larsen Petersen,[32] Birgit Pils,[53,54] Michael Prigge,[24] Stefan A. Rensing,[55,56,57] Diego Mauricio Riaño-Pachón,[58,59] Alison W. Roberts,[60] Yoshikatsu Sato,[3] Henrik Vibe Scheller,[41,61] Burkhard Schulz,[30] Christian Schulz,[62] Eugene V. Shakirov,[63] Nakako Shibagaki,[64] Naoki Shinohara,[3,65] Dorothy E. Shippen,[63] Iben Sørensen,[32,66] Ryo Sotooka,[65] Nagisa Sugimoto,[3] Mamoru Sugita,[33] Naomi Sumikawa,[4] Milos Tanurdzic,[67] Günter Theißen,[29] Peter Ulvskov,[32] Sachiko Wakazuki,[3] Jing-Ke Weng,[17,68] William W.G.T. Willats,[32] Daniel Wipf,[69] Paul G. Wolf,[70] Lixing Yang,[16] Andreas D. Zimmer,[55,71] Qihui Zhu,[16] Therese Mitros,[72] Uffe Hellsten,[73] Dominique Loqué,[61] Robert Otillar,[73] Asaf Salamov,[73] Jeremy Schmutz,[73] Harris Shapiro,[73] Erika Lindquist,[73] Susan Lucas,[73] Daniel Rokhsar,[72,73] Igor V. Grigoriev[73]

Vascular plants appeared ~410 million years ago, then diverged into several lineages of which only two survive: the euphyllophytes (ferns and seed plants) and the lycophytes. We report here the genome sequence of the lycophyte *Selaginella moellendorffii* (Selaginella), the first nonseed vascular plant genome reported. By comparing gene content in evolutionarily diverse taxa, we found that the transition from a gametophyte- to a sporophyte-dominated life cycle required far fewer new genes than the transition from a nonseed vascular to a flowering plant, whereas secondary metabolic genes expanded extensively and in parallel in the lycophyte and angiosperm lineages. *Selaginella* differs in posttranscriptional gene regulation, including small RNA regulation of repetitive elements, an absence of the trans-acting small interfering RNA pathway, and extensive RNA editing of organellar genes.

Selaginella moellendorffii, like all lycophytes, has features typical of vascular plants, including a dominant and complex sporophyte generation (Fig. 1, A and B) having vascular tissues with lignified cell types. Lycophytes also share traits with nonseed plants, most notably

the release of haploid spores (Fig. 1C) from the sporophyte and a gametophyte generation that develops independently of the sporophyte. Because the lycophytes are an ancient lineage that diverged shortly after land plants evolved vascular tissues (Fig. 2A) (*1*), we sequenced the Selaginella genome to provide a resource for identifying genes that may have been important in the early evolution of developmental and metabolic processes specific to vascular plants.

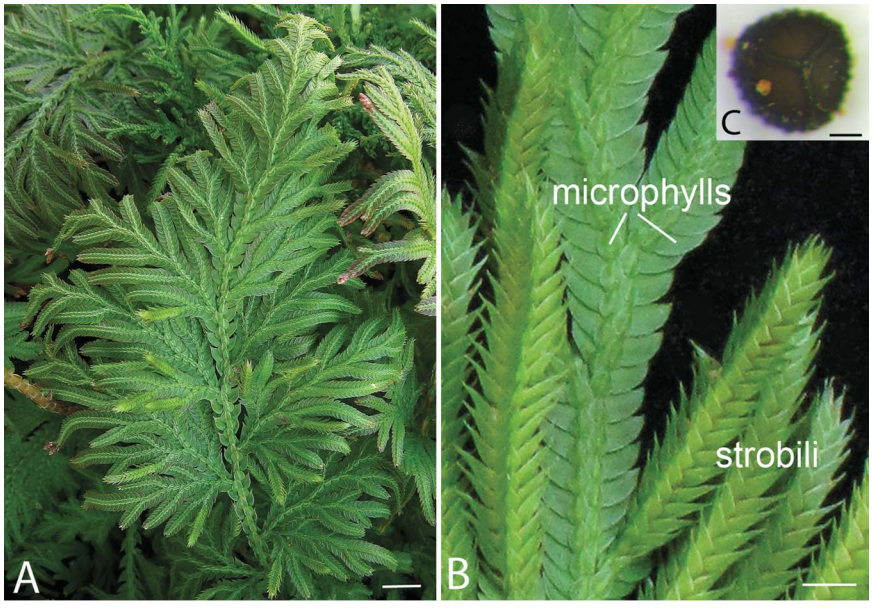

Fig. 1. Selaginella morphology. (**A**) The diploid sporophyte body. Bar, 10 mm. (**B**) A shoot with two ranks of microphylls ("leaves") and strobili. Each microphyll of a strobilus has either a mega- or a microsporangium where mega- or microspores are produced. Bar, 2 mm. (**C**) An orange microspore on top of a dark megaspore. These single-celled haploid spores represent the beginning of the independent haploid gametophyte generation. The microgametophye produces motile sperm and the megagametophyte eggs. Bar, 0.1 mm.

The Selaginella genome was sequenced by whole-genome shotgun sequencing (*2*). The assembled genome size (212.6 Mbp) is twice that determined by flow cytometry (*3*), indicating that the assembled genome includes two haplotypes of ~106 Mbp that are 98.5% identical at the nucleotide level. A deduced haplotype has 22,285 predicted protein-coding genes, of which 37% are supported by expressed sequence tag sequences, and 58 microRNA (miRNA) loci (*2*, *4*). The Selaginella genome lacks evidence of an ancient whole-genome duplication or polyploidy (*2*), unlike all other sequenced land-plant genomes (*5*–*7*). Gene density in Selaginella and *Arabidopsis*, which has a slightly larger genome size, is very similar (*2*), and both genomes have gene-poor regions rich in transposable elements (TEs) and other repetitive sequences (*2*). Although fewer genes and smaller introns (*2*) contribute to a genome size smaller than *Arabidopsis*, this is offset by a greater proportion of TEs in Selaginella (37.5% versus 15% in *Arabidopsis*) (*2*). Long terminal repeat retrotransposons are the most abundant TEs, occupying one-third of the Selaginella genome (*2*).

Plant TEs and *MIRNA* loci are important sources of small RNAs (sRNAs) that function

[1]Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN 47907, USA. [2]Advanced Science Research Center, Kanazawa University, Kanazawa 920-0934, Japan. [3]ERATO, Japan Science and Technology Agency, Okazaki 444-8585, Japan. [4]National Institute for Basic Biology, Okazaki 444-8585, Japan. [5]Department of Basic Biology, School of Life Science, The Graduate University for Advanced Studies, Okazaki 444-8585, Japan. [6]School of Biological Sciences, Monash University, Clayton Campus, Melbourne, Victoria 3800, Australia. [7]Section Plant Biology, University of California, Davis, CA 95616, USA. [8]Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA. [9]Department of Biology and Huck Institutes of Life Sciences, Pennsylvania State University, University Park, PA 16802, USA. [10]Graduate Program in Plant Biology, Pennsylvania State University, University Park, PA 16802, USA. [11]Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park, PA 16802, USA. [12]Department of Biological Sciences, University at Buffalo, Buffalo, NY 14260, USA. [13]The New York Botanical Garden, Bronx, NY 10458, USA. [14]Department of Biology, University of Regina, 3737 Wascana Parkway, Regina, SK S4S 0A2, Canada. [15]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA. [16]Department of Genetics, University of Georgia, Athens, GA 30602, USA. [17]Department of Biochemistry, Purdue University, West Lafayette, IN 47907, USA. [18]Department of Molecular Biology, University of Potsdam, Potsdam-Golm 14476, Germany. [19]Razavi Newman Center for Bioinformatics, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA. [20]Heisenberg-Group Biophysics and Molecular Plant Biology, University of Potsdam, Potsdam-Golm 14476, Germany. [21]Charles University in Prague, Faculty of Science, Department of Botany, 128 01 Prague 2, Benatska 2, Czech Republic. [22]University in Prague, Faculty of Science, Chittussiho 10, 710 00 Ostrava, Czech Republic. [23]Department of Biology, The College of William and Mary, Williamsburg, VA 23187, USA. [24]Section of Cell and Developmental Biology, University of California, San Diego, La Jolla, CA 92093, USA. [25]Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA. [26]Laboratoire de Reproduction et Développement des Plantes, Ecole Normale Supérieure de Lyon, Lyon F-69364, France. [27]Department of Plant Biology, Carnegie Institution for Science, Stanford, CA 94305, USA. [28]Faculty of Science, Hokkaido University, Sapporo 060-0810, Japan. [29]Department of Genetics, Friedrich Schiller University Jena, D-07743 Jena, Germany. [30]Department of Horticulture and Landscape Architecture, Purdue University, W. Lafayette, IN 47907, USA. [31]Institute of Biology, Martin Luther University Halle-Wittenberg, Halle/Saale 06120, Germany. [32]Department of Plant Biology and Biotechnology, University of Copenhagen, Frederiksberg C 1871, Denmark. [33]Center for Gene Research, Nagoya University, Nagoya 464-8602, Japan. [34]Department of Chemistry, School of Science, The University of Tokyo 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. [35]Institute for Biology–Applied Genetics, Dahlem Centre of Plant Sciences, Freie Universität Berlin, 14195 Berlin, Germany. [36]Graduate School of Life and Environmental Sciences, Gene Research Center, University of Tsukuba, Tsukuba 305-8572, Japan. [37]Center for Life Science Automation, University of Rostock, D-18119 Rostock, Germany. [38]Studiengangskoordination Nano-Science, Universität Tübingen, Tübingen 72076, Germany. [39]Plant Global Education Project, Graduate School of Biological Sciences, Nara Institute of Science and Technology, Nara 630-0192, Japan. [40]Division of Biomedical Statistics and Informatics, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA. [41]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA. [42]Graduate School of Life Science, Hokkaido University, Sapporo 060-0810, Japan. [43]Department of Plant Biology and Pathology, Rutgers, The State University of New Jersey and Waksman Institute of Microbiology, 190 Frelinghuysen Road, Piscataway, NJ 08854–8020, USA. [44]The Genome Analysis Center, Monsanto, St. Louis, MO 63167, USA. [45]Faculty of Fisheries Sciences, Hokkaido University, Hakodate 041-8611, Japan. [46]National Institute of Genetics, Mishima 411-8540, Japan. [47]Graduate School of Arts and Sciences, The University of Tokyo, Tokyo 153-8902, Japan. [48]Bernd Max-Planck Institute of Molecular Plant Physiology, Potsdam-Golm 14476, Germany. [49]Department of Microbiology, Immunology and Biochemistry, University of Tennessee, Memphis, TN 38163, USA. [50]Japan Science and Technology Agency, Sapporo 060-0819, Japan. [51]Department of Biology, University of Washington, Seattle, WA 98195–5325, USA. [52]Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada. [53]Wellcome Trust Centre for Human Genetics, Oxford, OX3 7BN, UK. [54]Sias AG, CH-8634 Hombrechtikon, Switzerland. [55]Faculty of Biology, University of Freiburg, Freiburg 79104, Germany. [56]BIOSS Centre for Biological Signaling Studies, University of Freiburg, Freiburg 79104, Germany. [57]Freiburg Initiative for Systems Biology (FRISYS), University of Freiburg, Freiburg 79104, Germany. [58]GabiPD team, Bioinformatics Group, Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm 14476, Germany. [59]Departamento de Ciencias Biológicas, Universidad de los Andes, Bogotá D.C., Colombia. [60]Department of Biological Sciences, University of Rhode Island, Kingston, RI 02881, USA. [61]Joint BioEnergy Institute, Feedstocks Division, Emeryville, CA 94608, USA. [62]Department of Evolution and Biodiversity of Plants, Ruhr-University Bochum, Bochum 44780, Germany. [63]Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843, USA. [64]Graduate School of Engineering, Osaka University, Osaka 565-0871, Japan. [65]School of Science, Hokkaido University, Sapporo 060-0810, Japan. [66]Department of Plant Biology, Cornell University, Ithaca, NY 14850, USA. [67]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA. [68]The Jack H. Skirball Center for Chemical Biology and Proteomics, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA. [69]UMR INRA 1088/CNRS 5184, Université Bourgogne Plant-Microbe-Environment, Dijon 21065, France. [70]Department of Biology, Utah State University, Logan, UT 84322-5305, USA. [71]Plant Biotechnology, University of Freiburg, Freiburg 79104, Germany. [72]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA. [73]U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA.

*To whom correspondence should be addressed. E-mail: banksj@purdue.edu

to epigenetically regulate TE and gene activity (*8*). Several observations suggest that some aspects of epigenetic or posttranscriptional gene regulation in Selaginella are unique among plants. For one, the proportion of sRNAs 23 to 24 nucleotides (nt) in length is extraordinarily small in the Selaginella sRNA population (*2*) compared to angiosperms (*9*). Nearly three-quarters of the Selaginella sRNAs (*4*) map to *MIRNA* loci and are predominantly 21 nt in length (*2*). In angiosperms, 24-nt siRNAs, which are generated primarily from TEs, function to silence TE activity through the RNA-dependent DNA methylation pathway (*10*–*12*) and accumulate massively in specific cells of the female gametophyte (*13*). Because the Selaginella sRNA population was generated from sporophytic tissues, the 24-nt siRNA pathway may only be deployed during gametophyte development in Selaginella. A second distinction is the absence of *DCL4*, *RDR6*, and *MIR390* loci in Selaginella, which are required for the biogenesis of trans-acting siRNAs (tasiRNAs) in angiosperms (*2*). Their absence suggests that tasiRNA-regulated processes in angiosperms, including leaf polarity (*14*) and developmental phase changes in the sporophyte (*15*, *16*), are regulated differently in Selaginella, and possibly reflects the independent origins of foliar organs in the lycophyte and angiosperm lineages (*17*, *18*). Finally, the Selaginella plastome sequence reveals an extraordinarily large number of RNA-edited sites (*2*), as do other lycophyte organellar genomes (*19*, *20*). This coincides with an exceptionally large number of PPR genes in Selaginella (>800) (*2*), some of which guide RNA editing events in angiosperms (*21*).

Because Selaginella is a member of a vascular plant lineage that is sister to the euphyllophytes, we used comparative and phylogenetic approaches to identify gene origins and expansions coinciding with evolutionary innovations and losses in land plants. To identify such genes without regard to function, we compared the proteomes of the green alga Chlamydomonas, the moss Physcomitrella, Selaginella, and 15 angiosperm species; identified gene families that are related by homology by hierarchical clustering (*2*); and then mapped them onto a phylogenetic tree (Fig. 2B). The 3814 families with gene members present in all plant lineages define the minimum set of genes that were likely to be present in the common ancestor of all green plants and their descendants and include genes essential for plant function. The transition from single-celled green algae to multicellular land plant approximately doubled the gene number with the acquisition of 3006 new genes. The transition from nonvascular to vascular plant is associated with a gain of far fewer new genes (516) than the transition from a basal vascular plant to a basal euphyllophyte whose descendants include the angiosperms (1350). These numbers show that the evolution of traits specific to euphyllophytes or angiosperms required the evolution of about three times more new genes than the transition from a plant having a dominant gametophyte and simple, leafless, and nonvascularized sporophyte (typified by modern bryophytes) to a plant with a dominant, vascularized, and branched sporophyte with leaves.

In a second approach, we analyzed the phylogenies of genes known to function in *Arabidopsis* development (*2*). We identified 424 monophyletic groups of developmental genes, each group containing putatively all genes descended from a common land-plant ancestral gene (table S6). Selaginella and Physcomitrella genes are present in 377 (89%) and 356 (84%) of the 424 land-plant orthologous gene groups, respectively, indicating that the common ancestor of land plants had most of the gene families known to direct angiosperm development. Conspicuous expansions of families within different lineages resulted in different numbers of land-plant orthologs in each genome (table S6). The 27 vascular plant-specific orthologous groups likely represent genes associated with developmental innovations of vascular plants. Among them are genes regulating the meristem (*CLV1* and *CLV2*), hormone signaling (*GID1* and *CTR1*), and flowering (*TFL2* and *UFO*). Homologs of genes involved in the specification of xylem (*NST* and *VND*) (*22*) and phloem (*APL*) (*23*) in *Arabidopsis* are present in Physcomitrella and Selaginella, suggesting that the developmental programs for patterning and differentiation of vascular tissues were either present in, or co-opted from, preexisting genetic programs in the ancestral land plant. The 43 groups lacking genes from Physcomitrella and Selaginella (table S6) likely identify genes that were necessary for euphyllophyte or angiosperm developmental innovations. Among this group are genes that regulate light signaling (*FAR1*, *MIF1*, *OBP3*, and *PKS1*), shoot meristem development (*AS2* and *ULT1*), hormone signaling and biosynthesis (*BRI1*, *BSU1*, *ARF16*, *ACS*, and *ACO*), and flowering (*HUA1*, *EMF1*, *FT*, *TFL1*, and *FD*). Altogether, these results suggest that the evolutionary transitions from a nonvascular plant to a vascular angiosperm included the stepwise addition of components of some developmental pathways, especially those regulating meristem and hormone biology, as previously noted for the gibberellin signaling pathway (*24*, *25*).

Genes involved in secondary metabolism were also investigated because plants synthesize numerous secondary metabolites that they use to interact with their environment. Three gene families involved in their biosynthesis, including those encoding cytochrome P450-dependent monooxygenases (P450s), BAHD acyltransferases (BAHDs), and terpene synthases (TSs), were analyzed. The largest of these in Selaginella is the P450 family, accounting for 1% of its predicted proteome (table S7) (*2*). All three families show similar evolutionary trends, with the inferred ancestral vascular plant having a small number of genes that radiated extensively but independently within the lycophyte and angiosperm lineages (figs. S6 to S13). *BAHD* and *TS* genes, which are known to be involved in the biosynthesis of volatile odorants, are apparent only in seed plants (figs. S12 to S13), likely reflecting the coevolution of seed plants with animals that pollinate flowers or disperse seeds. The independent diversification of these gene families plus the large number of Selaginella genes suggest that Selaginella not only has the potential to synthesize a repertoire of secondary metabolites that rivals the angiosperms in complexity, but that many of them are likely to be unique. Some have been shown to be of pharmaceutical value [e.g., (*26*)].

We have used the compact Selaginella genome sequence to uncover genes associated with major evolutionary transitions in land plants. Understanding their functions in Selaginella and other taxa, as well as acquiring the genome sequences of other informative taxa, especially charophytes, ferns, and gymnosperms, will be key to understanding the evolution of plant form and function.

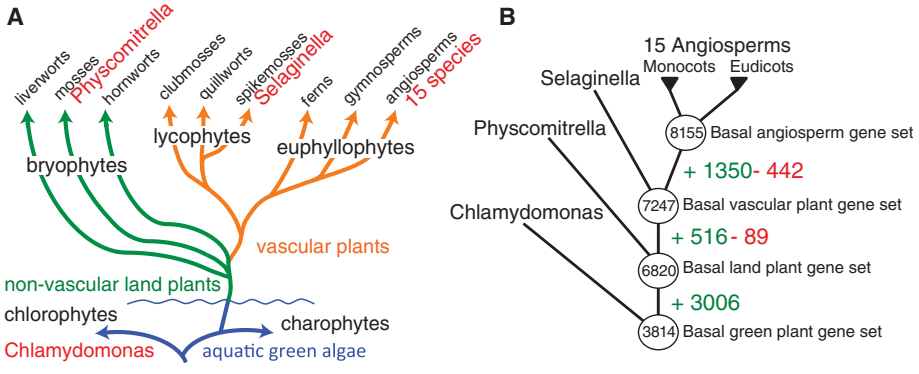

Fig. 2. (**A**) Phylogeny of plants. Taxa in red have sequenced genomes. (**B**) Gene family gains (+) and losses (–) mapped onto the plant phylogenetic tree. The minimum numbers of gene families present in the ancestors of different plant lineages are circled.

References and Notes

1. P. Kenrick, P. R. Crane, *Nature* **389**, 33 (1997).
2. Details are given in the supporting materials on *Science* Online.
3. W. Wang *et al.*, *BMC Plant Biol.* **5**, 10 (2005).

4. M. J. Axtell, J. A. Snyder, D. P. Bartel, *Plant Cell* **19**, 1750 (2007).
5. M. S. Barker, H. Vogel, M. E. Schranz, *Genome Biol. Evol.* **1**, 391 (2009).
6. H. Tang *et al.*, *Science* **320**, 486 (2008).
7. H. Tang *et al.*, *Genome Res.* **18**, 1944 (2008).
8. M. Ghildiyal, P. D. Zamore, *Nat. Rev. Genet.* **10**, 94 (2009).
9. D. Chen *et al.*, *Bioinformatics* **26**, 1391 (2010).
10. K. D. Kasschau *et al.*, *PLoS Biol.* **5**, e57 (2007).
11. K. Nobuta *et al.*, *Nat. Biotechnol.* **25**, 473 (2007).
12. S. H. Cho *et al.*, *PLoS Genet.* **4**, e1000314 (2008).
13. D. Bourc'his, O. Voinnet, *Science* **330**, 617 (2010).
14. D. H. Chitwood *et al.*, *Genes Dev.* **23**, 549 (2009).
15. A. Peragine, M. Yoshikawa, G. Wu, H. L. Albrecht, R. S. Poethig, *Genes Dev.* **18**, 2368 (2004).
16. Z. Xie, E. Allen, A. Wilken, J. C. Carrington, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12984 (2005).
17. S. K. Floyd, J. L. Bowman, *Curr. Biol.* **16**, 1911 (2006).
18. C. J. Harrison *et al.*, *Nature* **434**, 509 (2005).
19. S. Tsuji *et al.*, *J. Plant Res.* **120**, 281 (2007).
20. F. Grewe *et al.*, *Nucleic Acids Res.* **39**, 2890 (2011).
21. V. Knoop, *Cell. Mol. Life Sci.* **68**, 567 (2011).
22. T. Demura, H. Fukuda, *Trends Plant Sci.* **12**, 64 (2007).
23. M. Bonke, S. Thitamadee, A. P. Mähönen, M. T. Hauser, Y. Helariutta, *Nature* **426**, 181 (2003).
24. K. Hirano *et al.*, *Plant Cell* **19**, 3058 (2007).
25. Y. Yasumura, M. Crumpton-Taylor, S. Fuentes, N. P. Harberd, *Curr. Biol.* **17**, 1225 (2007).
26. Y. Cao *et al.*, *Fitoterapia* **81**, 253 (2010).

Acknowledgments: Selaginella sequences were deposited at GenBank with the accession numbers GL377566 to GL378322.1 and HM173080. Genome sequencing and analysis were performed by the U.S. Department of Energy (DOE), Joint Genome Institute, supported by the Office of Science of the U.S. DOE, Contract DE-AC02-05CH11231 (I.V.G., U.H., D.L., E.L., S.L., T.M., R.O., D.R., A.S., J.S., H.V.S.). Support was provided by NSF 0844413 (J.A.B.); Japan Society for the Promotion of Science (M.H., T.N., T.F., K.M., T.M., M.S.); Ministry of Education, Culture, Sports, Science, and Technology, Japan (M.H., T.N., T.F.); NSF 0515435 and Australian Research Council FF0561326 (J.L.B.); NSF 0519970 (M.G.); NSF 0638595 (C.D.); NSF 0922742 (V.A.A.); The Lewis B and Dorothy Cullman Program (B.A.A., A.L.); NSF 1020443 (B.A.A.); Natural Sciences and Engineering Research Council of Canada (NSERC) 2982 (N.W.A.); NIH GM84051 (M.J.A.); NSERC (E.I.B.); NIH T32 GM007757 and NSERC PGS-D (M.S.B.); NSF 0607123 (J.L.B.); Life Sciences Research Foundation (NDB); National Human Genome Research Institute (NHGRI) HG004164 (M.D.); German Science Foundation (Deutsche Forschungsgemeinschaft, DFG) DR 430/4-2 (I.D.); Czech Ministry of Education 21620828 (M.E.); Jeffress Memorial Trust J-938 (E.M.E.); NSF 0744800 (M.E. and M.P.); DOE DE-FG02-04ER15542 (W.B.F. and S.L.); The Danish Council for Independent Research, Technology and Production Sciences 009-066624/274-09-0314 (J.H.); The Villum Kann Rasmussen Foundation (J.H., I.S., B.P., P.U., W.W.); NIH T32-HG00035, NSF 1020660, and NSF 1036466 (K.G.K.); NSF 0735191 (E.L.); NHGRI HG004164 (G.M); U.S. Department of Agriculture (USDA) DE-FG02-08ER64630 (T.P.M.); NSF 0228660 (R.G.O.); The Danish Council for Strategic Research 09-063090 (B.L.P. and P.U.); Marie Curie FP6 RTN ZOONET (B.P.); DFG RE 837/10-2, BMBF FRISYS 0313921 (S.A.R.); Bundesministerium fuer Bildung und Forschung, Germany, GABI-FUTURE grant 0315046 (D.M.R.); USDA NRI 2007-35318-18389 (A.W.R.); NSF 0421604 (C.S.); NIH GM065383 (D.E.S.); Burgundy Regional Council 20100112095254682-1 (D.W.); and DFG RE 837/10-2 (A.D.Z.). K. Wall, D. Hurley, and S. Hentel provided computational assistance.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1203810/DC1
SOM Text
Figs. S1 to S14
Tables S1 to S8
References

4 February 2011; accepted 8 April 2011
Published online 5 May 2011;
10.1126/science.1203810

Chromatin "Prepattern" and Histone Modifiers in a Fate Choice for Liver and Pancreas

Cheng-Ran Xu,[1] Philip A. Cole,[2] David J. Meyers,[2] Jay Kormish,[1]* Sharon Dent,[3] Kenneth S. Zaret[1]†

Transcriptionally silent genes can be marked by histone modifications and regulatory proteins that indicate the genes' potential to be activated. Such marks have been identified in pluripotent cells, but it is unknown how such marks occur in descendant, multipotent embryonic cells that have restricted cell fate choices. We isolated mouse embryonic endoderm cells and assessed histone modifications at regulatory elements of silent genes that are activated upon liver or pancreas fate choices. We found that the liver and pancreas elements have distinct chromatin patterns. Furthermore, the histone acetyltransferase P300, recruited via bone morphogenetic protein signaling, and the histone methyltransferase Ezh2 have modulatory roles in the fate choice. These studies reveal a functional "prepattern" of chromatin states within multipotent progenitors and potential targets to modulate cell fate induction.

Early pluripotent cells of the mammalian embryo develop into multipotent endoderm, ectoderm, and mesoderm germ layers. In pluripotent cells, silent genes that will be activated later in development often exist with histone modifications and/or bound transcription factors that reflect the chromatin being "poised" for activity (*1*–*3*). It is unclear whether such poised states exist for silent genes in germ layer cells and, if so, whether genes poised for different tissue fates exhibit different chromatin features. Furthermore, it is not known whether enzymes that establish chromatin states can control germ layer fate choices. Embryonic germ layer cells are few in number, they have not been purified, and chromatin analysis on small cell populations is challenging (*4*). Yet germ layer cells represent the first lineage-restricted, multipotent progenitors of the embryo and a paradigm for all subsequent fate decisions.

Ventral foregut endoderm cells undergo a fate choice for liver or ventral pancreas progenitors (*5*, *6*). FoxA1 or FoxA2, GATA4 or GATA6, vHNF1, and Hnf6 (also known as Oc1) are necessary in the endoderm for both liver and ventral pancreas induction (*7*). In the absence of any set of the factors, the earliest liver marker genes *Alb1*, *Afp*, and *Ttr* and the ventral pancreas transcription factor gene *Pdx1* fail to be activated, or expression is delayed, and tissue buds fail to form (*7*). It is not clear how the same group of factors can be necessary for both liver and ventral pancreas and how signaling promotes the different fates. We sought to map chromatin states at silent liver- and pancreas-specific regulatory sequences in endoderm cells, to discover the factors or relevant histone-modifying enzymes, and test the enzymes' functions in the liver-versus-pancreas decision.

We used fluorescence-activated cell sorting (FACS) with the ENDM1 antibody to isolate ventral foregut endoderm cells from embryonic day 8.25 (E8.25) mouse embryos with four to six somite pairs (4-6S) (*8*) (fig. S1), just prior to the induction of hepatic and pancreatic fates (*5*, *9*). We also used the liver-specific antibody Liv2 to isolate nascent hepatoblasts expressing *Alb1*, *Afp*, and *Ttr* from E9.5 embryos (fig. S2) (*10*). Chromatin marks in ENDM1$^+$ and Liv2$^+$ populations were identified with a low–cell number chromatin immunoprecipitation (ChIP) protocol (*4*) for H3K9acK14ac, H3K4me2, H3K4me3, H3K9me3, H3R17me2a, H3K27me3, H3K36me2, H3K36me3, H3K79me2, H4K20me3, H3T3ph, H3S10ph, the histone variant H2A.Z, and the chromatin remodelers Brg1 and SNF2. We assessed the liver-specific promoter and enhancer of *Alb1* (*11*, *12*), the liver-specific promoters of *Afp* and *Ttr* genes (*13*, *14*), and the I, II, III, and IV upstream elements and local promoter of the pancreatic determination gene *Pdx1* (fig. S3). The I, II, and III upstream elements and promoter of *Pdx1* reconstitute pancreas-specific activation (*15*); the IV element may function later (*16*, *17*). All of the target genes are silent in endoderm cells, and only the liver genes become activated in hepatoblasts.

[1]Institute for Regenerative Medicine, Epigenetics Program, Department of Cell and Developmental Biology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA. [2]Department of Pharmacology and Molecular Sciences, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. [3]Department of Molecular Carcinogenesis, University of Texas M. D. Anderson Cancer Center, Smithville, TX 78957, USA.

*Present address: Biochemistry and Molecular Biology, University of Calgary, Calgary, Alberta T2N 4N1, Canada.
†To whom correspondence should be addressed. E-mail: zaret@upenn.edu