

Science

 **AAAS**

**Rates of Molecular Evolution Are Linked to Life
History in Flowering Plants**

Stephen A. Smith, *et al.*

Science **322**, 86 (2008);

DOI: 10.1126/science.1163197

***The following resources related to this article are available online at
www.sciencemag.org (this information is current as of October 2, 2008):***

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/322/5898/86>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/322/5898/86/DC1>

This article **cites 34 articles**, 12 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/322/5898/86#otherarticles>

This article appears in the following **subject collections**:

Evolution

<http://www.sciencemag.org/cgi/collection/evolution>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

Rates of Molecular Evolution Are Linked to Life History in Flowering Plants

Stephen A. Smith* and Michael J. Donoghue

Variable rates of molecular evolution have been documented across the tree of life, but the cause of this observed variation within and among clades remains uncertain. In plants, it has been suggested that life history traits are correlated with the rate of molecular evolution, but previous studies have yielded conflicting results. Exceptionally large phylogenies of five major angiosperm clades demonstrate that rates of molecular evolution are consistently low in trees and shrubs, with relatively long generation times, as compared with related herbaceous plants, which generally have shorter generation times. Herbs show much higher rates of molecular change but also much higher variance in rates. Correlates of life history attributes have long been of interest to biologists, and our results demonstrate how changes in the rate of molecular evolution that are linked to life history traits can affect measurements of the tempo of evolution as well as our ability to identify and conserve biodiversity.

Variation in the rate of molecular evolution has been attributed to a number of factors, including differences in body size, metabolic rate, DNA repair, and generation time (e.g., 1–4). In plants, differences in rates of molecular evolution have been noted between annuals and perennials (5) and between woody and herbaceous species (6, 7). These differences have been presumed to reflect differences in generation time (the time from seed germination to the production of fruits/seeds). However, in plants the relationship between life history and the average length of time before a nucleotide is copied either through replication or repair [nucleotide generation time (1)] is complicated by the fact that somatic mutations can accumulate during growth and can be transmitted through gametes (8, 9). Variation in breeding system and/or seed-banking by annual plants (9) may also affect the ability to detect a correlation between molecular rate and generation time.

Previous studies have been inconclusive with respect to the extent and the correlates of rate heterogeneity in plants (5, 7, 9, 10). Studies focused on individual smaller clades, or on single gene regions, have yielded results of uncertain generality (7), whereas broader phylogenetic studies have suffered from limited taxon sampling and, hence, comparisons among very distant relatives (11). Some tests have failed to account for phylogenetic relatedness (9).

We assembled molecular sequence data for five major branches within the flowering plants: three clades of asterids (Apiales, Dipsacales, and Primulales), one clade of rosids (Moraceae/Urticaceae), and one of monocotyledons (Commelinidae). We used group-to-group profile alignments (12) that take advantage of previously recognized clades within the groups analyzed (13) and yield denser data matrices (containing less missing data) than

those produced using other strategies (14). Specifically, we identified alignable clusters of homologous gene regions, which were then concatenated with profile alignment (13). To minimize missing data, only phylogenetically informative clusters (with at least four taxa) were used. The gene regions varied among the five matrices but in each case included markers from the chloroplast, nuclear, and mitochondrial genomes (figs. S1 and S2 and table S2). The average gene region in our analyses contained 305 species; the smallest contained 10 species. This process resulted in an Apiales matrix of 1593 species by 9522 sites (>15 megabases); for Dipsacales, it was 366 by 11374 (>4 megabases); for Primulales, 529 by 11505 (>6 megabases); for Moraceae/Urticaceae, 457 by 7820 (>3.5 megabases); and for Commelinidae, 4657 by 22391 sites (>104 megabases).

Phylogenetic trees (Fig. 1) were inferred under maximum-likelihood (ML) with RAxML (vers.7.0.0) (15), with gene regions treated as separate partitions (13). We conducted 100 rapid bootstrap analyses, using every 10th bootstrap tree as a starting tree for a full ML search, and chose the tree with the highest likelihood score; owing to the size of the Commelinidae matrix, only a single ML search was conducted. For all clades but Commelinidae, we used nonparametric rate smoothing (16) to set branch lengths proportional to time; we used the PATHd8 method (17) for the exceptionally large commelinid analysis. Published studies were used to calibrate each phylogeny, using multiple calibration points to limit the impact of clade-specific rate heterogeneity (13, 18–21). For Apiales and Primulales, we separately calibrated the major subclades identified in previous analyses, which also accommodated the fact that our analyses included some taxa not represented in previous studies.

Ancestral states of the life history trait “trees/shrubs” versus “herbs” (a proxy for generation time) (6, 7, 22) were inferred with ML methods (Fig. 1) (13); palms (Arecaceae, Commelinidae), which do not produce true wood (secondary xylem), were scored as trees/shrubs. For each branch on

each phylogeny, we calculated the number of substitutions per nucleotide site per million years using branch lengths estimated from the dated molecular trees. Branch calculations were binned on the basis of inferred life history to produce box plots for each clade (Fig. 1). Outliers (values >1.5 times beyond the first and third quartiles) were excluded as artifacts of divergence-time estimation (e.g., those with zero or near-zero branch lengths). Within each major clade, we noted that trees/shrubs were consistently evolving more slowly than related herbaceous plants. Median rates of nucleotide divergence were 2.7 to 10 times as high in herbs as in trees/shrubs; herbs also showed higher ranges and variances (Fig. 1). None of the tree/shrub lineages examined here showed high rates of molecular evolution, but some herbaceous lineages were inferred to have low evolutionary rates, in the range characteristic of trees/shrubs. This asymmetry in variance may reflect the fact that, although most trees/shrubs are not able to reproduce within the first few years (23, 24), as most herbs can, some herbs take as long as trees to flower. Consistent with the view that generation time influences the rate of molecular evolution within the Commelinidae (Fig. 1), the longer-lived bromeliads [which take up to 18 years to reproduce (25)] have remarkably short branches, with even fewer substitutions per site per million years than palms (0.00059 and 0.0014, respectively). Other factors, such as population size, breeding system, and seed-banking, may also relate to the observed asymmetry; for example, the rate of fixation of mutations by selection increases in large populations. Although we do not dismiss these variables in explaining the observed variance, they are less clearly correlated with the life history distinction than is generation time [e.g., (26)].

To explore whether the difference in rates of molecular evolution has remained constant over time, we compared substitutions per site per million years through 10-million-year segments for each dated phylogeny (Fig. 2) (13). We found that the trend in rate heterogeneity holds through time, with some noteworthy exceptions in the earliest time periods. For example, woody Dipsacales are estimated to have a high rate of evolution before the herbaceous habit is inferred to have evolved in this lineage (Fig. 2B). Fossil data might help to distinguish whether these results are best explained by incorrect reconstructions (i.e., perhaps the first Dipsacales were herbaceous), by faster evolution of woody lineages during earlier times (e.g., due to warmer climate in the early Tertiary), or by the extinction of early woody lineages.

Because these comparisons do not directly take into account phylogenetic relationships or examine the effects of evolutionary change from one life history state to the other, we calculated branch length contrasts (27) around each inferred evolutionary shift in life history (Fig. 3) (13). Specifically, we calculated the average accumulation of molecular changes from each branch tip to the shared ancestor of a tree/shrub clade and

Department of Ecology and Evolutionary Biology, 21 Sachem Street, Post Office Box 208105, Yale University, New Haven, CT 06520–8105, USA.

*To whom correspondence should be addressed. E-mail: stephen.smith@yale.edu

compared this to the average accumulation in its herbaceous sister clade. We started from the most nested clades and worked toward the root, excising any nested contrasts from the more inclusive calculations to avoid measuring any node more than once. We omitted contrasts containing only one tree/shrub or one herb branch to less-

en the impact of incorrectly estimating singleton branches (branch lengths were averaged in clades with two or more species).

Of the 13 contrasts identified using these criteria (Table 1 and Fig. 3), 12 showed a slower rate of molecular evolution in trees/shrubs than in herbs (sign test, $P = 0.00342$). On average, herbs evolve

2.5 times as fast as trees/shrubs. A maximum rate difference of 4.75 times was found between *Dorstenia* (Moraceae) and its tree/shrub sister clade. The single exception occurred within *Sambucus* (Dipsacales), where the tree/shrub species showed a slightly higher rate than the herbs (0.0075 and 0.0061, respectively). This case involved the smallest

Fig. 1. Phylogenies of five angiosperm clades with branch lengths proportional to substitutions per site. Branch colors represent inferred life history states (brown for trees/shrubs; green for herbs). Box plots show substitutions per site per million years for the inferred life history categories; centerline represents the median, hinges mark the first and third quartiles, whiskers extend to the lowest and highest non-outlier. Outliers (not shown) have values >1.5 times beyond the first or third quartiles.

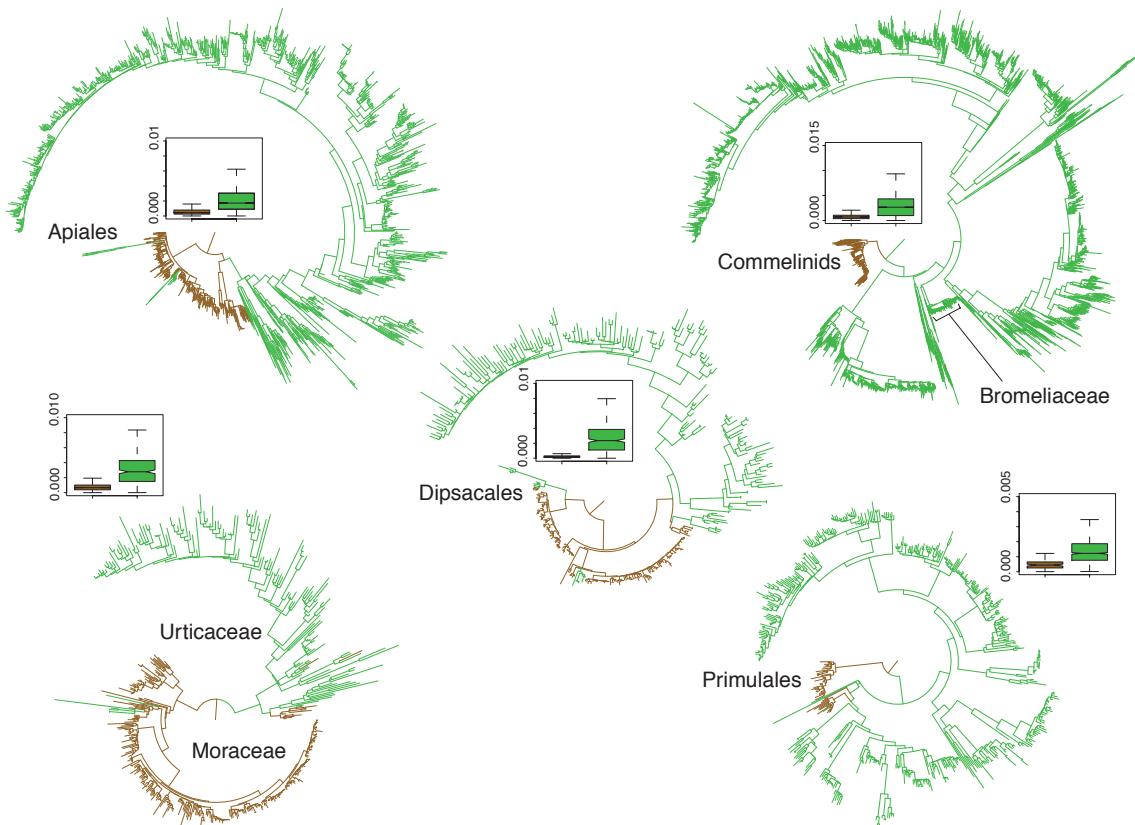
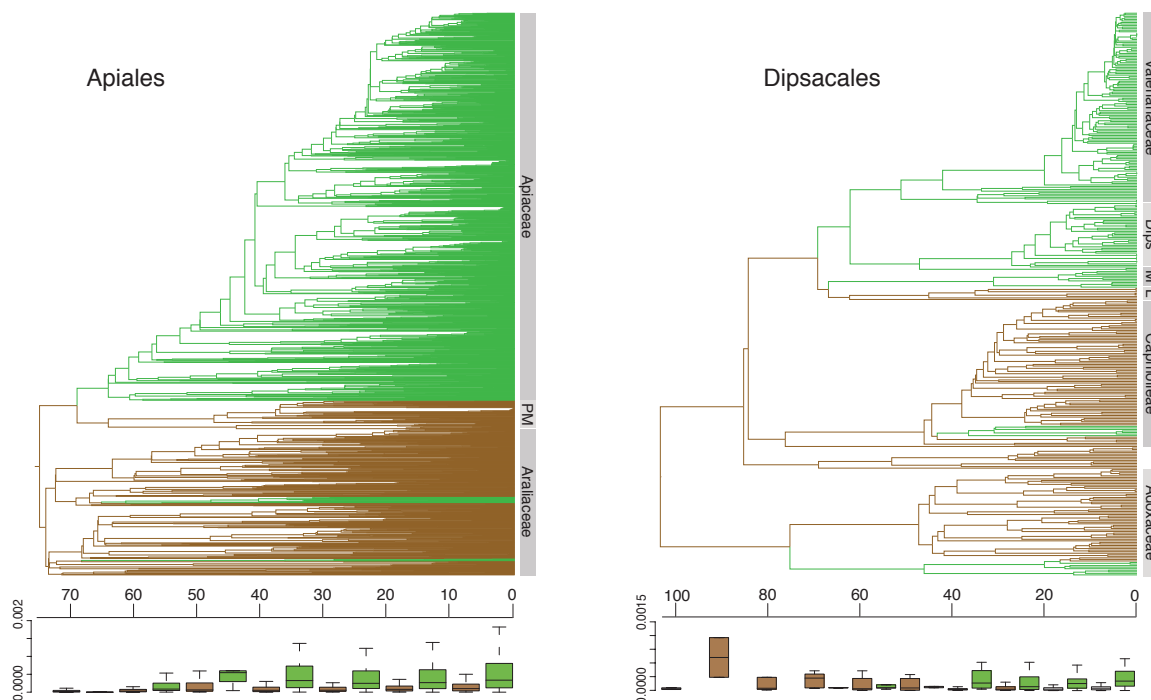


Fig. 2. Dated phylogenies for Apiales and Dipsacales with substitutions per site per million years plotted for 10-million-year intervals through the life of the clade. Branch colors represent inferred life history states (brown for trees/shrubs; green for herbs). Box plots as in Fig. 1. PM, Pittosporaceae and Myodocarpaceae; Dips, Dipsacaceae; M, Morinaceae; L, Linnaecae.



Downloaded from www.sciencemag.org on October 2, 2008

numbers of species (three shrubby species versus three herbs) and also presented the greatest difficulty in assigning life history states (the herbaceous species are subshrubby and the woody species mature rapidly). As such uncertainties are inherent in large comparative analyses, we explored whether alternative phylogenetic hypotheses (13) affected the results for the smallest clade examined here, the Dipsacales, as well as the effect of scor-

ing all *Sambucus* species as trees/shrubs. These alternatives (Table 1 and Fig. 3) yielded a similarly strong historical correlation ($P = 0.00049$), as did the exclusion of these contrasts altogether ($P = 0.00195$).

On the basis of our trees and broader phylogenetic studies of angiosperms [reviewed in (28); see also (29)], the likely direction of evolution of plant habit was from trees/shrubs to the herbaceous

condition in Apiales, Dipsacales, and Primulales, and with less certainty in Moraceae/Urticaceae. The palms (Arecaceae) within the Commelinidae present the one clear instance in our sample of the evolution of trees/shrubs from herbaceous ancestors (30). From our comparisons and a broader analysis of monocotyledons (11), the shift to the tree/shrub habit in palms was associated with a marked decrease in the rate of molecular evolution (palms evolve 2.7 times as slow as their sister commelinids), as predicted by the hypothesis that generation time drives the rate of molecular evolution.

Differences in rates of evolution associated with generation time may be reflected most clearly in synonymous substitutions within coding sequences (31). We analyzed 1208 commelinid *rbcL* sequences, pruning species lacking an *rbcL* sequence in GenBank from our Commelinidae phylogeny and using RAxML to estimate branch lengths for several partitions of the data (Table 2) (13). As expected, estimated amino acid branch lengths showed the least difference in rate between life history classes (2.1 times as fast in herbs), with first and second nucleotide positions being next smallest (3.2 times as fast). The rate difference in the full Commelinidae data set (all species, all genes) fell between these two values (2.7 times as fast in herbs). The third positions showed the greatest difference in rate (4.98 times as fast in herbs). These findings are similar to those based on a much smaller sample of *rbcL* sequences from grasses and palms (11).

Our findings highlight the need for the methods used to date phylogenies to address the form of clade-dependent heterogeneity documented here. A rate of nucleotide substitution obtained from an herbaceous group cannot be used to calibrate a clade of trees/shrubs, or vice versa, without confounding age estimates. Likewise, relaxed clock methods [e.g., (32)] are likely to estimate that slowly evolving groups are younger, and that rapidly evolving groups are older, than their true ages. It may be possible to avoid mixing clades with very different life histories in designing dating studies. Otherwise, as we have attempted here, the use of multiple calibration points spanning clades that differ in life history may help alleviate this problem. Also, as shown here for Commelinidae, the use of amino acid sequences (or the removal of third sites) may be useful. Bayesian models that do not assume an autocorrelated rate of molecular evolution [e.g., (33)] are promising, but current methods are incapable of analyzing large data sets.

We hope that our results will also focus new attention on the extent to which molecular and morphological evolution are coupled [see (34, 35)]. Are rates of morphological evolution also slower in trees/shrubs than in herbs [e.g., (36)]? Until this question is addressed, we urge caution in assuming that morphological change scales with molecular change and in using molecular branch lengths alone to assess “feature diversity” and design conservation strategies [e.g., (37)]. A related issue is the likely success of “barcoding” methods

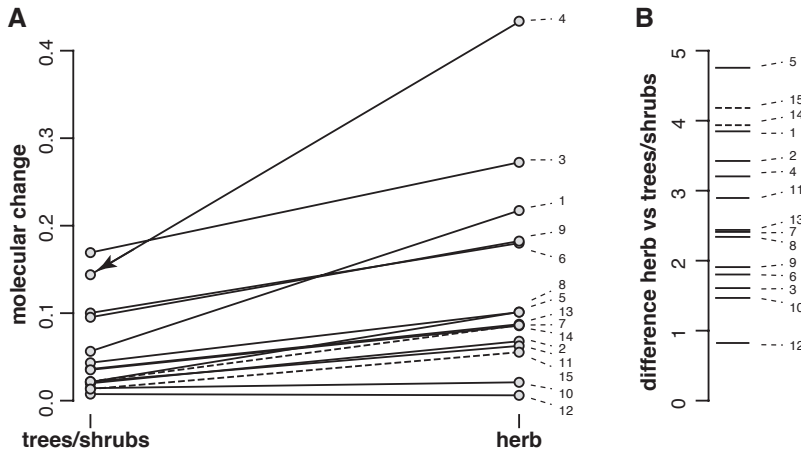


Fig. 3. Branch-length contrasts for trees/shrubs versus herbs. (A) Lines are drawn between the accumulated average molecular branch lengths for each tree/shrub clade and its sister herbaceous clade (numbers correspond to those in Table 1). All evolutionary shifts were inferred to be from trees/shrubs to herbs except for the evolution of palms within monocotyledons (arrowhead in contrast 4). Contrasts 1 to 13 were used in an initial sign test ($P = 0.00342$). Alternative contrasts within the Dipsacales (14 and 15) are marked by dotted lines and were substituted for 11 to 13 in one test ($P = 0.00049$); contrasts 11 to 15 were omitted in a third test ($P = 0.00195$). (B) Magnitude of change between each tree/shrub clade and its herbaceous sister clades; values above 1 show higher rates of molecular evolution in herbs than in trees/shrubs.

Table 1. Branch length contrasts 1 to 13 derive from the trees in Fig. 1 [see (13) for more exact locations of the nodes in question]. Plants in the first taxon in each pair of representative taxa are trees/shrubs; plants in the second are herbs. Within Dipsacales, we explored alternative contrasts, substituting contrasts 14 and 15 for 11 to 13 in one test and omitting contrasts 11 to 15 in another.

Major clade	Representative taxa	Trees/shrubs	Herbs	Difference
Apiales	1 <i>Astrotricha</i> – <i>Hydrocotyle</i>	0.0564	0.2173	3.8538
	2 <i>Aralia</i> – <i>Panax</i>	0.0198	0.0679	3.4224
	3 Pittosporaceae–Apiaceae	0.1692	0.2724	1.6097
Commelinidae	4 Arecaceae–remaining Commelinidae	0.1363	0.4350	3.1915
Moraceae–	5 <i>Brosimum</i> – <i>Dorstenia</i>	0.0213	0.1013	4.7527
Urticaceae	6 Moraceae–Urticaceae	0.1002	0.1800	1.7967
	7 <i>Cecropia</i> / <i>Coussapoa</i> – <i>Boehmeria</i>	0.0361	0.0873	2.4169
Primulales	8 <i>Ardisia</i> –sister Myrsinaceae	0.0433	0.1011	2.3330
	9 Theophrastaceae–Myrsinaceae/Primulaceae	0.0953	0.1824	1.9138
Dipsacales	10 <i>Symphoricarpos</i> – <i>Triosteum</i>	0.0142	0.0210	1.4747
	11 Linnaeaeae–Morinaceae	0.0217	0.0626	2.8912
	12 Woody <i>Sambucus</i> –herbaceous <i>Sambucus</i>	0.0075	0.0061	0.8130
	13 <i>Viburnum</i> – <i>Adoxa</i>	0.0352	0.0856	2.4304
	14 Linnaeaeae–Morinaceae/Dipsacaceae/Valerianaceae	0.0219	0.0863	3.9432
	15 “Woody” <i>Sambucus</i> – <i>Adoxa</i>	0.0133	0.0554	4.1783

Table 2. Branch length contrast estimates for different partitions of *rbcl* sequences from Commelinidae.

Estimate data	Palms	Rest of Commelinidae	Difference
Third sites	0.0369	0.1842	4.9872
First, second, and third sites	0.0331	0.1284	3.8808
First and second sites	0.0313	0.1007	3.2180
Amino acids	0.0464	0.0993	2.1379
All genes/species	0.1363	0.4350	3.1915

for identifying plant species from short DNA sequences [reviewed in (33)]. We predict that the chloroplast genes proposed as universal barcode loci will be most successful in resolving herbaceous species and may be incapable of confidently distinguishing closely related woody species.

Finally, our studies underscore the need for better and more accessible information on the underlying drivers of rates of molecular evolution. In addition to data on generation times, we need better knowledge of effective population sizes. Past analyses (e.g., in mammals) have assumed that larger, longer-lived organisms have smaller population sizes, but this may be reversed in plants, where tropical trees often appear to have large population sizes (31). Our analyses imply that somatic mutation has not counteracted the influence of generation time on rates of evolution, but more data are needed on the rate and fate of such mutations (8). In any event, our analyses demonstrate a general pattern that must now be taken into account in evolutionary studies and whose existence demands the elaboration of a cohesive causal explanation.

References and Notes

1. A. P. Martin, S. R. Palumbi, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4087 (1993).
2. A. O. Mooers, P. H. Harvey, *Mol. Phylogenet. Evol.* **3**, 344 (1994).
3. L. Bromham, A. Rambaut, P. H. Harvey, *J. Mol. Evol.* **43**, 610 (1996).
4. J. F. Gillooly, A. P. Allen, G. B. West, J. H. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 140 (2005).
5. K. Andreasen, B. G. Baldwin, *Mol. Biol. Evol.* **18**, 936 (2001).
6. J. Laroche, P. Li, L. Maggia, J. Bousquet, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5722 (1997).
7. K. M. Kay, J. B. Whittall, S. A. Hodges, *BMC Evol. Biol.* **6**, 36 (2006).
8. E. J. Klekowski, P. J. Godfrey, *Nature* **340**, 389 (1989).
9. C. A. Whittle, M. O. Johnston, *J. Mol. Evol.* **56**, 223 (2003).
10. K. Andreasen, *Conserv. Genet.* **6**, 399 (2005).
11. B. S. Gaut, B. R. Morton, B. C. McCaig, M. T. Clegg, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10274 (1996).
12. R. C. Edgar, *Nucleic Acids Res.* **32**, 1792 (2004).
13. Materials and methods are available as supporting material on Science Online.
14. M. M. McMahon, M. J. Sanderson, *Syst. Biol.* **55**, 818 (2006).
15. A. Stamatakis, *Bioinformatics* **22**, 2688 (2006).
16. M. J. Sanderson, *Mol. Biol. Evol.* **14**, 1218 (1997).
17. T. Britton, C. L. Anderson, D. Jacquet, S. Lundqvist, K. Bremer, *Syst. Biol.* **56**, 741 (2007).
18. K. Bremer, E. M. Friis, B. Bremer, *Syst. Biol.* **53**, 496 (2004).
19. C. D. Bell, M. J. Donoghue, *Am. J. Bot.* **92**, 284 (2005).

20. N. J. C. Zerega, W. L. Clement, S. L. Datwyler, G. D. Weiblen, *Mol. Phylogenet. Evol.* **37**, 402 (2005).
21. T. Janssen, K. Bremer, *Bot. J. Linn. Soc.* **146**, 385 (2004).
22. J. Bousquet, S. H. Strauss, A. H. Doerksen, R. A. Price, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 7844 (1992).
23. R. J. Petit, A. Hampe, *Annu. Rev. Ecol. Syst.* **37**, 187 (2006).
24. M. Verdú, *Evol. Int. J. Org. Evol.* **56**, 1352 (2002).
25. P. Hietz, J. Ausserer, G. Schindler, *J. Trop. Ecol.* **18**, 177 (2002).
26. S. S. Renner, R. E. Ricklefs, *Evol. Int. J. Org. Evol.* **82**, 596 (1995).
27. T. J. Davies, V. Savolainen, *Evol. Int. J. Org. Evol.* **60**, 476 (2006).
28. D. E. Soltis, P. S. Soltis, P. K. Endress, M. W. Chase, *Phylogeny and Evolution of Angiosperms* (Sinaur Associates, Sunderland, MA, 2005).
29. P. F. Stevens, Angiosperm Phylogeny Website, www.mobot.org/MOBOT/research/APweb (2008).
30. M. W. Chase *et al.*, *Aliso* **22**, 63 (2006).
31. T. Ohta, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10676 (1993).
32. M. J. Sanderson, *Mol. Biol. Evol.* **19**, 101 (2002).
33. A. J. Drummond, S. Y. W. Ho, M. J. Phillips, A. Rambaut, *PLoS Biol.* **4**, e88 (2006).
34. R. Lahaye, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2923 (2008).
35. T. G. Barraclough, V. Savolainen, *Evol. Int. J. Org. Evol.* **55**, 677 (2001).
36. E. W. Sinnott, *Am. Nat.* **50**, 466 (1916).
37. F. Forest *et al.*, *Nature* **445**, 757 (2007).
38. We thank D. Tank, J. Beaulieu, S. Stearns, D. Weinreich, C. Dunn, and E. Edwards for helpful discussions. Funded by the NSF through the Cyberinfrastructure for Phylogenetic Research (EF-0331654) and the angiosperm Assembling the Tree of Life (EF-0431258) projects.

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5898/86/DC1
Materials and Methods
Fig. S1 to S8
Tables S1 and S2
References

14 July 2008; accepted 5 September 2008
10.1126/science.1163197

Chemokine Signaling Controls Endodermal Migration During Zebrafish Gastrulation

Sreelaja Nair and Thomas F. Schilling*

Directed cell movements during gastrulation establish the germ layers of the vertebrate embryo and coordinate their contributions to different tissues and organs. Anterior migration of the mesoderm and endoderm has largely been interpreted to result from epiboly and convergent-extension movements that drive body elongation. We show that the chemokine Cxcl12b and its receptor Cxcr4a restrict anterior migration of the endoderm during zebrafish gastrulation, thereby coordinating its movements with those of the mesoderm. Depletion of either gene product causes disruption of integrin-dependent cell adhesion, resulting in separation of the endoderm from the mesoderm; the endoderm then migrates farther anteriorly than it normally would, resulting in bilateral duplication of endodermal organs. This process may have relevance to human gastrointestinal bifurcations and other organ defects.

A crucial feature of vertebrate embryogenesis is the coordinated morphogenesis of germ layers (endoderm, mesoderm, and ectoderm) during gastrulation (1). Interactions between the endoderm and mesoderm specify organ locations and symmetries (2). Defects in the endoderm alter the morphogenesis of mesodermal organs (e.g., heart, kidneys,

and blood), whereas mesodermal defects disrupt the locations of the liver and pancreas (2–5). Morphogenesis is regulated by Wnt (6) and Nodal signaling (7) when cells are intermingled in a bipotential “mesendoderm” (8). However, relatively little is known about germ layer-specific pathways that establish organ rudiments. In zebrafish, mesendodermal organ progenitors

involute at the gastrula margin (blastopore) and move anteriorly toward the animal pole (future head) while converging toward the midline (convergent extension).

The chemokine receptor CXCR4 controls directional migration in many contexts and is expressed in the endoderm. It is up-regulated by the endodermal determinants Mixer and Sox17 β (9–13) and is required for gastrointestinal vascularization (14). Of the two closely related zebrafish Cxcr4s, Cxcr4b regulates the migration of many cell types (12, 15–18), but no roles have been reported for Cxcr4a during embryogenesis.

Zebrafish embryos deficient in Cxcr4a or Cxcl12b, generated by injection with antisense morpholino oligonucleotides (MO), appeared morphologically normal (Fig. 1, A to C, and fig. S1). However, analysis of *Tg(gutGFP)*⁸⁵⁴ transgenic embryos in which the entire gut fluoresces [(19); GFP, green fluorescent protein] revealed duplications of endodermal organs at 56 hours post-fertilization (hpf) (Fig. 1, D to F, and fig. S2), including the pancreas (normally on the right; fig. S3, A to F) and liver (normally

Department of Developmental and Cell Biology, University of California, Irvine, CA 92697, USA.

*To whom correspondence should be addressed. E-mail: tschilli@uci.edu