

thirds of NAW, AIW and Arctic Ocean Deep Water (AODW). Since there is evidence that the balance of vertical and horizontal exchange has altered within the Nordic seas^{17,18}, it is possible—even likely—that the DSOW recipe is not fixed or definable but is itself time-dependent. Nevertheless, the full range of modern DSOW recipes all contain a sufficient contribution of NAW to justify the idea that changes in DSOW characteristics may be linked to fluctuations in the NAW at some appropriate remove in time and space.

Second, though the available records are uncomfortably short and 'gappy', we suggest there is evidence that the ocean temperature series at either end of this circuit are correlated, once an appropriate advective–diffusive time-lag is interposed. In Fig. 4, we compare the 50–500 m mean temperature of the West Spitzbergen Current in Fram Strait (the northward extension of the Norwegian Atlantic Current) with a time-series of current meter thermistor records from the Denmark Strait overflow at 2,000 m depth off Angmagssalik, some 3 years later and some 2500 km further south. The main features of the two series—the long decline in temperature in the mid-to-late 1980s, the temperature maximum in the mid-to-late 1990s, and the recent rapid cooling associated with the brief recurrence of extreme negative NAO conditions in 1996—seem to correspond at each site.

The three-year advective–diffusive time-lag between the two sites remains to be tested. At present, we have little more basis for this than the fact that another significant ocean climate signal, the "Great Salinity Anomaly"¹⁹, took about this time to spread the equivalent distance from the Faroe–Shetland Channel (1976) to Fram Strait (1979). However, we may soon determine the rate of the Nordic seas circulation directly using two novel tracers: Sellafield ⁹⁹Tc spreading around the margins in the Norwegian Atlantic Current, and SF₆ purposely released into the convective centre of the Greenland Sea by the ESOP-II programme of EC-MAST. An experimental bottom-mounted whole-water sampler has been moored at 2,000 m depth in the core of the overflow to record their time of arrival.

Although the demonstration of some physical linkage between Fram Strait and Denmark Strait is not irrelevant to the prognosis, and the lagged correlation in temperature between the two sites is encouraging, these two results fall far short of demonstrating whether (and to what extent) the hydrographic character of DSOW may be set by events upstream (and hence be predictable). The most formidable objection is evident in Fig. 4—that however well-correlated their time-series may be, the transfer of an ocean-climate anomaly from Fram Strait to Denmark Strait must somehow survive the water-mass transformations implied by a change in mean temperature (from 3 °C in Fram Strait through 0 °C at the Denmark Strait sill, to 2 °C at depth off Angmagssalik). The modelling of water-mass mixing and exchange processes around this circuit is still insufficiently advanced to comment usefully on variability in either speed or hydrography. We note merely that in the model runs described above, about half of the imposed 1 °C temperature change survived the process of entrainment south of the Denmark Strait sill, which must represent the most intensive process of water-mass transformation on the circuit from Fram Strait. So for the present, the important aim of predicting overflow variability is likely to be led by observations rather than models. □

Received 2 July; accepted 16 October 1998.

- Dickson, R. R., Gmitrowicz, E. M. & Watson, A. J. Deep water renewal in the northern North Atlantic. *Nature* **344**, 848–850 (1990).
- Dickson, R. R. & Brown, J. The production of North Atlantic Deep Water: sources, rates and pathways. *J. Geophys. Res.* **99**, 12319–12341 (1994).
- Jungclaus, J. H. & Backhaus, J. O. Application of a transient, reduced-gravity, plume model to the Denmark Strait Overflow. *J. Geophys. Res.* **99**, 12375–12396 (1994).
- Cayan, D. R. Latent and sensible heat flux anomalies over the northern oceans: the connection to monthly atmospheric circulation. *J. Clim.* **5**, 354–369 (1992).
- Hurrell, J. W. Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. *Science* **269**, 676–679 (1995).
- Barnston, A. G. & Livezey, R. E. Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon. Weath. Rev.* **115**, 1083–1126 (1987).

- Hurrell, J. W. & van Loon, H. Decadal variations in climate associated with the North Atlantic Oscillation. *Clim. Change* **36**, 301–326 (1997).
- Dickson, R. R. *et al.* The Arctic Ocean response to the North Atlantic Oscillation. *J. Clim.* (submitted).
- Swift, J. H. & Aagaard, K. Seasonal transitions and water mass formation in the Iceland and Greenland Seas. *Deep-Sea Res.* **28**, 1107–1129 (1981).
- Smethie, W. M. & Swift, J. H. The tritium: krypton-85 age of Denmark Strait Overflow Water and Gibbs Fracture Zone Water just south of Denmark Strait. *J. Geophys. Res.* **94**, 8265–8275 (1989).
- Swift, J. H. & Koltermann, K.-P. The origin of Norwegian Sea Deep Water. *J. Geophys. Res.* **93**, 3563–3569 (1988).
- Aagaard, K., Fahrbach, E., Meincke, J. & Swift, J. H. Saline outflow from the Arctic Ocean: its contribution to the deep waters of the Greenland, Norwegian and Icelandic seas. *J. Geophys. Res.* **96**, 20433–20441 (1991).
- Mauritzen, C. Production of dense overflow waters feeding the North Atlantic across the Greenland–Scotland Ridge: Part 1. Evidence for a revised circulation scheme. *Deep-Sea Res.* **43**, 769–806 (1996).
- Strass, V. H., Fahrbach, E., Schauer, U. & Sellmann, L. Formation of Denmark Strait Overflow Water by mixing in the East Greenland Current. *J. Geophys. Res.* **98**, 6907–6919 (1993).
- Swift, J. H. Dissolved Oxygen in the Arctic Ocean and Nordic Seas. *Proc. ACSYS Conf.* **2**, 259–261 (1997).
- Rudels, B., Quadfasel, D. & Friedrich, H. J. The Arctic Circumpolar Boundary Current. *Deep-Sea Res.* (submitted).
- Meincke, J., Rudels, B. & Friedrich, H. J. The Arctic Ocean–Nordic Seas thermohaline system. *Int. Council Explor. Sea J. Mar. Sci.* **54**, 283–299 (1997).
- Dickson, R. R., Lazier, J., Meincke, J., Rhines, P. & Swift, J. Long-term coordinated changes in the convective activity of the North Atlantic. *Prog. Oceanogr.* **38**, 241–295 (1996).
- Dickson, R. R., Meincke, J., Malmberg, S.-A. & Lee, A. J. The "Great Salinity Anomaly" in the Northern North Atlantic, 1968–82. *Prog. Oceanogr.* **20**, 103–151 (1988).

Acknowledgements. This work was funded in part by the EC MAST-III VEINS programme, and in part by the SIO/LDEO Consortium of the NOAA Global Change Program.

Correspondence and requests for materials should be addressed to R.D. (e-mail: r.r.dickson@cefes.co.uk).

Mesozoic subducted slabs under Siberia

Rob Van der Voo^{*,†}, Wim Spakman[†] & Harmen Bijwaard[†]

^{*} Department of Geological Sciences, the University of Michigan, Ann Arbor, Michigan 48109-1063, USA

[†] Vening Meinesz School of Geodynamics, Institute of Earth Sciences, Utrecht University, Budapestlaan 4, 3584 CD Utrecht, the Netherlands

Recent results from seismic tomography demonstrate that subducted oceanic lithosphere can be observed globally as slabs of relatively high seismic velocity in the upper as well as lower mantle^{1,2}. The Asian mantle is no exception, with high-velocity slabs being observed downwards from the west Pacific subduction zones under the Kurile Islands, Japan and farther south^{3–5}, as well as under Asia's ancient Tethyan margin. Here we present evidence for the presence of slab remnants of Jurassic age that were subducted when the Mongol–Okhotsk and Kular–Nera oceans closed between Siberia, the combined Mongolia–North China blocks and the Omolon block^{6–8}. We identify these proposed slab remnants in the lower mantle west of Lake Baikal down to depths of at least 2,500 km, where they join what has been interpreted as a 'graveyard' of subducted lithosphere at the bottom of the mantle. Our interpretation implies that slab remnants in the mantle can still be recognized some 150 million years or more after they have been subducted and that such structures may be useful in associating geodynamic to surface-tectonic processes.

Eurasia is, without doubt, the next supercontinent in the making, as continental elements such as Australia and even the Americas are steadily approaching it and seem destined to join it in the future. The growth of Asia started, however, a long time ago, when significant additions to its Siberian core began to occur through continent–continent collisions and accretion of smaller continental blocks⁶. The Mongolia and North and South China blocks, having amalgamated in the Permian–Triassic periods¹⁰ (or possibly as late as the Jurassic period¹¹) joined Siberia in Late Jurassic¹² or Early Cretaceous times⁸ (Fig. 1). We will call the former ocean between Siberia and Mongolia the Mongol–Okhotsk Ocean, but it has also been called the Khangai–Khanatey Ocean⁶. The northeastern-most Siberian block (Omolon) accreted to the growing Asian continent in

the Late Jurassic, closing the Kular–Nera Ocean⁶, and causing the Verkhoyansk Mountains¹³ (Fig. 1).

Improvements in tomographic techniques in the past few years have allowed more detailed visualization of deep fossil slabs, such as those that appear to have resulted from subduction of the Pacific and Tethyan oceans^{1,14}. Tomography parametrizes the mantle either by spherical harmonic functions or by using local cell slowness functions. The sizes of these cells in the latter parametrization have steadily decreased from early values¹⁵ of about 10° to recent values^{1,14,16} of about 2°–0.6°. Improvements in the travel-time data set¹⁷, the use of a large number of composite rays, and variably sized cells that minimize the difference in cell hit count¹⁸ have now¹⁴ provided tomographic images with great detail. Examples of the resolution, data fit, and sensitivity tests for the tomographic results we interpret here are more fully described elsewhere¹⁴.

Figures 2 and 3 show map views at various lower-mantle depths

and a cross-section under Siberia. The Mongol–Okhotsk and Verkhoyansk sutures are marked by a P-wave velocity contrast in the crust and uppermost mantle (not shown), also visible in surface wave studies^{19,20}. Clear anomalies, well resolved both vertically and horizontally, are seen at depths greater than 1,200 km (Fig. 2). In the crust, the Mongol–Okhotsk–Verkhoyansk suture runs from near the northern shore of Lake Baikal to the Sea of Okhotsk (for locations, see Fig. 2), and from there northwards under the Verkhoyansk Mountains. At depths greater than 1,500 km, the velocity anomalies are pronounced; amplitude variance tests indicate that the variation (of about +0.3% or higher) is well above amplitude noise level in the lower mantle¹⁴. At 1,500 km depth, the anomalies form a ‘hook’, running first west–northwestwards from Mongolia and then northwards towards the Siberian Arctic coast. At depths of 1,900–2,300 km, the anomalies are gradually more displaced to the west, and display an overall ‘Z’-shaped feature at 2,300 km (see curved yellow line in Fig. 2). At 2,500 km and below, the high-velocity anomalies merge with a broad east Asian high-velocity area in the lowermost mantle (Figs 2 and 3), which has been called a ‘graveyard’ of slabs⁹, an interpretation corroborated by our results.

In the cross-section (Fig. 3, top), this feature (M) is illustrated side-by-side with the currently active subduction zone (P) under northern Japan. Below 1,400 km, we see the southern leg of the hook to the northwest of Mongolia portrayed all the way down to the core–mantle boundary. Figure 3 also illustrates, for this same cross-section, how a simulated input of a layer-cake model¹⁴ is resolved in our inversion. The figure shows that below depths of 1,500 km, the M and P anomalies are both well resolved. The overall amplitude and spatial characteristics of the high P-wave velocity anomalies in the deep mantle under Siberia are comparable to those observed around the Pacific and under the Mediterranean–Himalaya–Indonesian belt and warrant a similar interpretation: as remnants of subducted slabs. Some positive anomalies are imaged between 400 and 1,200 km depth, but these are spatially scattered, of low amplitude, and not resolved. For instance, the ‘connection’ between upper and mid-mantle (>1,200 km) anomalies in Fig. 3 (top) is rather local and unresolved.

Identification of the Pacific and Tethyan subduction zones in the maps and section is rather straightforward^{1,2,14}; thus, we are confident that the high-velocity zone we have described under Siberia is not related to Cenozoic subduction. However, elimination of other, Mesozoic, subduction zones under Asia as candidates for the deep-mantle portions of the Siberian slab is more difficult: this is because the Asian lithosphere may have drifted significantly with respect to the deeper mantle formerly underlying it. Because of the increasing certainty that the hotspots have moved significantly with respect to each other, they do not provide a suitable framework to assess this issue^{21,22}. However, the western and eastern Pacific subduction zones and their deep slabs provide an alternative approach. The western Pacific subducted slab dips westwards and then sinks down vertically in the lower mantle, without large apparent displacements from where one would expect to find it (Figs 2, 3). In contrast, the deep-mantle slabs resulting from east Pacific subduction under North America in late Mesozoic–Palaeogene times can be recognized under the Great Lakes area and east of there, as far away as New England^{21,14}. The North American lithosphere has clearly moved far to the west with respect to these deep-mantle slabs. From (1) these relative displacements of the two slabs with respect to the present lithospheric locations where they started their downward journeys, and (2) the displacements of the continents during the opening of the Atlantic Ocean, we can deduce that the Siberian lithosphere has moved very little in an east–west direction with respect to the deep-mantle slabs; we can also deduce that the Atlantic Ocean has spread primarily by westward drift of North America, which has overridden the slabs under it to the extent of some 3,500 km or more. Moreover, other deep slab remnants under

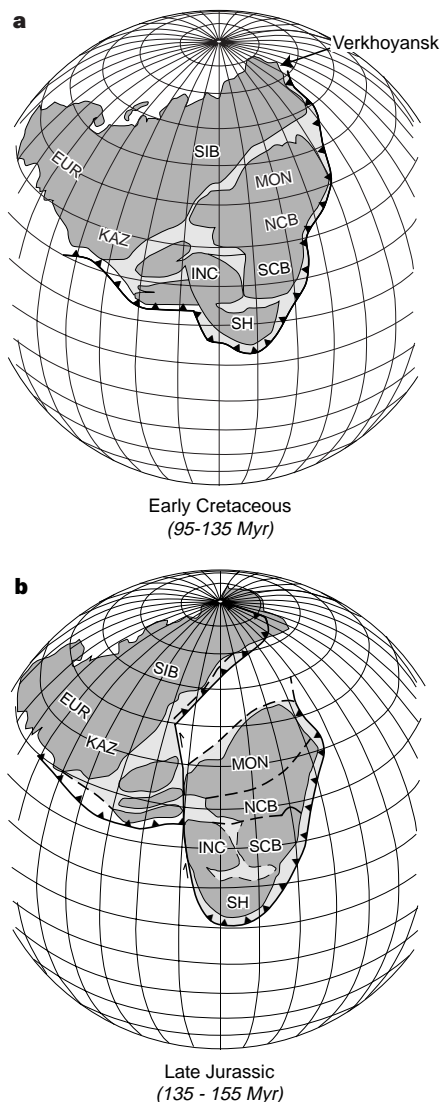


Figure 1 Early Cretaceous and Late Jurassic palaeoreconstructions of Siberia (SIB) and adjacent Asian blocks⁸. (Redrawn from ref. 8 with permission from the authors, R. Enkin *et al.*) **a**, Early Cretaceous; **b**, Late Jurassic. Late Jurassic subduction of the Mongol–Okhotsk Ocean is shown as occurring under the southeastern and northeastern (Verkhoyansk) margins of Siberia. Subduction probably also occurred under the Mongolian margin⁶, but this is not shown here. Abbreviations of block names are as follows: EUR, Europe; KAZ, Kazakhstan; MON, Mongolia; NCB and SCB, North and South China blocks; INC, Indochina; and SH, Shan–Thai block⁸.

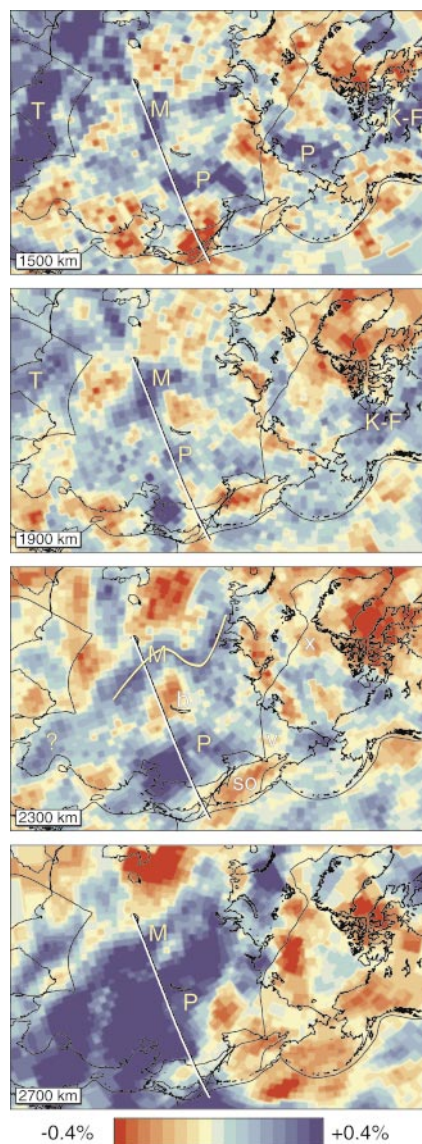


Figure 2 Tomographic P-wave velocity anomaly patterns in the deep mantle under Asia, for four different depths between 1,500 and 2,700 km. The superposed straight line indicates the location of the cross-section of Fig. 3. Velocity anomalies are displayed in percentages with respect to the average P-wave velocity at depth from model ak135²⁸. Capital letters represent our interpretation of the Mongol–Okhotsk (M), Pacific (P), Tethyan (T) and Kula–Farallon (K–F) oceanic lithospheric slabs in all four frames; lower-case letters in the 2,300 km frame represent geographical features: b, Lake Baikal; so, Sea of Okhotsk; v, Verkhoyansk suture; x, present-day North Pole. The question mark denotes a ‘slab’ of uncertain origin.

Eurasia due to Tethyan subduction are not shifted in an east–west sense. It seems justifiable, therefore, to infer that the deep-mantle slab we have described under Siberia is not related to subduction other than that of the Mongol–Okhotsk and Kular–Nera oceans.

In Fig. 4 we show the migration of this slab as a function of depth, and compare it with the palaeolatitudinal location of the lithospheric Mongol–Okhotsk–Verkhoyansk suture zone as a function of geological time. This figure shows that the orientations of the present-day suture zones at the surface parallel a higher-velocity contrast within the lithosphere down to depths of at least 95 km. At depths of 1,200 km or more, the velocity anomalies are rotated anticlockwise from the surface orientation. However, their locations correspond to the (palaeogeographically restored) position of the eastern and southern margin of Siberia in early Tertiary and older

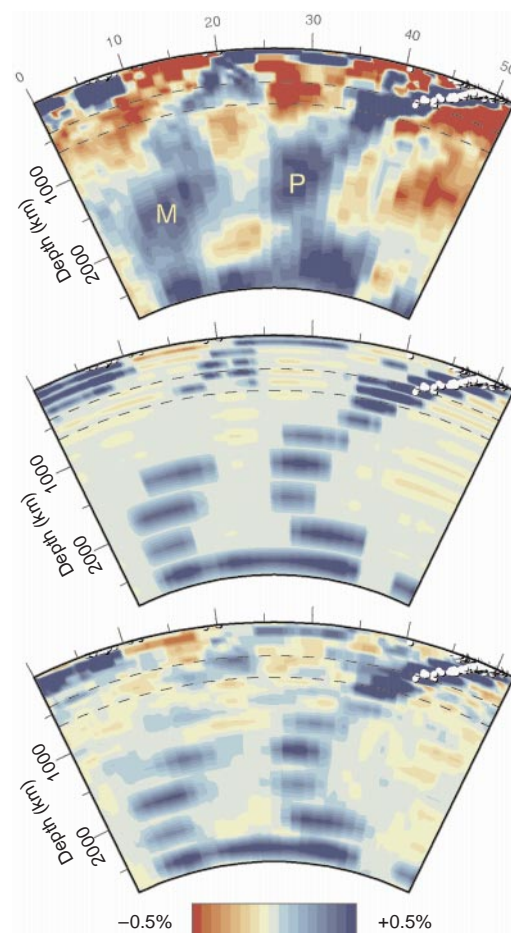


Figure 3 Cross-section through tomographic model. Top, tomographic results displayed in a vertical section along the line indicated in Fig. 2, showing the Pacific (P) subduction zone under Japan (right) and the inferred Mongol–Okhotsk slab (M) discussed here (left) side-by-side. White dots show locations of earthquakes in the west Pacific Benioff zone. Middle, a simulated input of a whole-mantle layer-cake model¹⁴ for this same cross-section, which is resolved and (partially to completely) reproduced in our inversion analysis (bottom).

times, but are further to the west, presumably owing to the angle of generally westward subduction during Late Jurassic times (Fig. 4).

The ‘Z’-shape of the slabs below 2,000 km (Fig. 2) is interpreted as the result of a similarly ‘Z’-shaped surficial subduction pattern of the Kular–Nera and Mongol–Okhotsk oceans not only under Siberia itself⁶ (Fig. 1), but under Siberia’s Jurassic northeastern and southeastern margins, as well as under the northern margin of Mongolia (see Fig. 21.39 in ref. 6).

Given that subduction ceased about 150 Myr ago, the depth of about 1,500 km for the top of the deep slab under Siberia implies an averaged sinking velocity of 1 cm yr^{-1} since the Jurassic period, which is not very different from sinking velocities inferred in some previous studies², although higher rates of $\sim 3 \text{ cm yr}^{-1}$ have also been proposed²³. Subduction velocities are thought to reduce by a factor of about four when slabs begin to penetrate into the higher-viscosity deeper mantle²⁴; a sinking velocity of 1 cm yr^{-1} would then imply a surficial convergence rate of 4 cm yr^{-1} . Even if subduction occurred on only one (for example, the Siberian) side of the Mongol–Okhotsk Ocean, the width of this ocean may have been at least 2,000 km at 200 Myr, which matches plate reconstructions²⁵.

We argue that the high-velocity structures under Siberia are logically interpreted as remnants of oceanic lithosphere that subducted before the Early Cretaceous and that, therefore, subducted lithosphere of Jurassic age can still be recognized after penetration

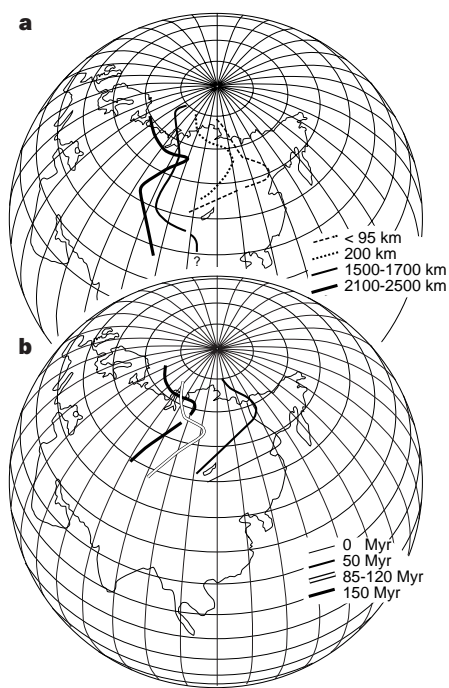


Figure 4 A comparison of the locations of tomographic velocity anomalies and the expected palaeolocations of the Siberian active margins. **a**, Locations of the lithospheric suture and the main axis of the ocean-lithospheric slab remnants as a function of depth, as determined from our tomographic results. **b**, Present-day and palaeogeographically reconstructed locations of the Mongol-Okhotsk-Verkhoyansk suture zones as a function of time (longitudes are arbitrary), using the palaeomagnetic pole determinations for Siberia of Zhao *et al.*²⁹ (81° N, 158.6° E for 50 Myr; 73.8° N, 202.4° E for 85–120 Myr; 70.1° N, 184.3° E for 150 Myr).

into the lower mantle even after subduction stopped some 150 Myr ago. Whether this visibility is a result of temperature²⁶, composition, pressure, or a combination thereof, is an open question, but what is clear is that the Mongol–Okhotsk subducted lithospheric material has been feeding the ‘graveyard’ of slabs under Asia. Our conclusions also imply that significant downwelling is a characteristic of growing supercontinents for hundred of millions of years²⁷, and that most, if not all, significantly fast anomalies in the deeper mantle appear to be associated with past subduction. This renders tomography an important tool for testing palaeogeographical reconstructions. □

Received 23 June; accepted 27 October 1998.

- van der Hilst, R. D., Widiyantoro, S. & Engdahl, E. R. Evidence for deep-mantle circulation from global tomography. *Nature* **386**, 578–584 (1997).
- Grand, S., van der Hilst, R. D. & Widiyantoro, S. Global seismic tomography: A snapshot of convection in the Earth. *GSA today* **7**, 1–7 (1997).
- van der Hilst, R., Engdahl, R., Spakman, W. & Nolet, G. Tomographic imaging of subducted lithosphere below the northwest Pacific island arcs. *Nature* **353**, 37–43 (1991).
- Richards, M. A. & Engebretson, D. C. Large-scale mantle convection and the history of subduction. *Nature* **355**, 437–440 (1992).
- Wen, L. & Anderson, D. L. The fate of slabs inferred from seismic tomography and 130 million years of subduction. *Earth Planet. Sci. Lett.* **133**, 185–198 (1995).
- Şengör, A. M. C. & Natal'in, B. A. in *The Tectonic Evolution of Asia* (eds Yin, A. & Harrison, M.) 486–640 (Cambridge Univ. Press, 1996).
- Zhao, X. X., Coe, R. S., Zhou, Y., Wu, H. & Wang, J. New paleomagnetic results from northern China: Collision and suturing with Siberia and Kazakhstan. *Tectonophysics* **181**, 43–81 (1990).
- Enkin, R., Yang, Z. Y., Chen, Y. & Courtillot, V. Paleomagnetic constraints on the geodynamic history of the major blocks of China from the Permian to the Present. *J. Geophys. Res.* **97**, 13953–13989 (1992).
- Wyssession, M. E. in *Subduction* (eds Bebout, G. *et al.*) 369–384 (Monogr. 96, Am. Geophys. Union, Washington DC, 1996).
- Yin, A. & Nie, S. Y. in *The Tectonic Evolution of Asia* (eds Yin, A. & Harrison, M.) 442–485 (Cambridge Univ. Press, 1996).
- Gilder, S. & Courtillot, V. Timing of the North-South China collision from new middle to late Mesozoic paleomagnetic data from the North China block. *J. Geophys. Res.* **102**, 17713–17727 (1997).
- Ziegler, A. M. *et al.* in *The Tectonic Evolution of Asia* (eds Yin, A. & Harrison, M.) 371–400 (Cambridge Univ. Press, 1996).
- Parfenov, L. M. Tectonics of the Verkhoyansk-Kolyma Mesozoids in the context of plate tectonics. *Tectonophysics* **199**, 319–342 (1991).

- Bijwaard, H., Spakman, W. & Engdahl, R. Closing the gap between regional and global travel time tomography. *J. Geophys. Res.* (in the press).
- Dziewonski, A. M., Hager, B. H. & O’Connell, R. J. Large-scale heterogeneities in the lower mantle. *J. Geophys. Res.* **82**, 239–255 (1977).
- Zhou, H. A high resolution P-wave model for the top 1200 km of the mantle. *J. Geophys. Res.* **101**, 27791–27810 (1996).
- Engdahl, E. R., van der Hilst, R. D. & Buland, R. P. Global teleseismic earthquake relocation with improved travel times and procedures for depth determination. *Bull. Seismol. Soc. Am.* **88**, 722–743 (1998).
- Spakman, W. & Bijwaard, H. Irregular cell parameterization of tomographic problems. *Ann. Geophys.* **16** (suppl. 1), 28 (1998).
- Curtis, C., Dost, B., Trampert, J. & Snieder, R. Eurasian fundamental mode surface wave phase velocities and their relationship with tectonic structures. *J. Geophys. Res.* (in the press).
- Ritzwoller, M. H. & Levshin, A. L. Eurasian surface wave tomography: Group velocities. *J. Geophys. Res.* **103**, 4839–4878 (1998).
- Cande, S. C., Raymond, C. A., Stock, J. & Haxby, W. F. Geophysics of the Pitman Fracture Zone and Pacific–Antarctic plate motions during the Cenozoic. *Science* **270**, 947–953 (1995).
- Tarduno, J. A. & Cottrell, R. D. Paleomagnetic evidence for motion of the Hawaiian hotspot during formation of the Emperor Seamounts. *Earth Planet. Sci. Lett.* **153**, 171–180 (1997).
- Ricard, Y., Richards, M. A., Lithgow-Bertelloni, C. & Le Stunff, Y. A geodynamic model of mantle density heterogeneity. *J. Geophys. Res.* **98**, 21895–21909 (1993).
- Lithgow-Bertelloni, C. & Richards, M. A. The dynamics of Cenozoic and Mesozoic plate motions. *Rev. Geophysics* **36**, 27–78 (1998).
- Scotese, C. R. & Golonka, J. *Paleomap Paleogeographic Atlas* (Paleomap project, Univ. Texas, Arlington, 1992).
- Zhong, S. & Gurnis, M. Mantle convection with plates and mobile, faulted plate margins. *Science* **267**, 838–843 (1995).
- Gurnis, M. Large-scale mantle convection and the aggregation and dispersal of supercontinents. *Nature* **232**, 695–699 (1988).
- Kennett, B. L. N., Engdahl, E. R. & Buland, R. Constraints on seismic velocities in the Earth from travel times. *Geophys. J. Int.* **122**, 108–124 (1995).
- Zhao, X. X., Coe, R. S., Gilder, S. A. & Frost, G. M. Palaeomagnetic constraints on palaeogeography of China: Implications for Gondwanaland. *Austr. J. Earth Sci.* **43**, 643–672 (1996).

Acknowledgements. We thank M. Richards and M. Gurnis for constructive comments. H.B. was supported by the Netherlands Organization for Scientific Research (NWO).

Correspondence and requests for materials should be addressed to R.V.d.V. (e-mail: voo@umich.edu).

Can aposematic signals evolve by gradual change?

Leena Lindström, Rauno V. Alatalo, Johanna Mappes, Marianna Riipi & Laura Vertainen

University of Jyväskylä, Department of Biological and Environmental Science, Konnevesi Research Station, PO Box 35, FIN-40 351 Jyväskylä, Finland

Aposematic species, which signal conspicuously of their unprofitability to predators, have puzzled evolutionary biologists for over a century^{1,2}. Although conspicuousness of unpalatable prey improves avoidance learning by predators^{3–5}, it also involves an evolutionary paradox: with increasing detectability^{4,6–8} the deviant aposematic prey would suffer high predation initially from naive predators. Here we test a neglected idea^{7–11} that aposematic coloration may evolve by gradual change rather than by major mutations. Weak signals did not suffer high initial predation, but predators (great tits, *Parus major*) did not learn to separate them from cryptic palatable prey. Furthermore, enhanced avoidance of more conspicuous signals occurred only if predators had previously encountered relatively strong signals. Thus, the gradual-change hypothesis does not provide an easy solution to the initial evolution of aposematism through predator learning. However, the possibility remains that cost-free step-wise mutations over the range of weak signals could accumulate under neutral selection to produce effective strong signals.

It has been assumed that sudden, pronounced mutations turned unpalatable cryptic prey into highly conspicuous forms. This hypothesis of sudden change may account for why grouping has been suggested to be essential in explaining how brightly coloured insects first evolved^{12–14}. Naive predators encountering a group are likely to leave some of the unpalatable prey untouched, whereas any single solitary individual may not escape being predated. Alternatively, gradual change in prey detectability^{7–9,11}, initially not changing the cryptic type into an overtly conspicuous prey type, may allow evolution of aposematism even in solitary prey. This

