

Methods

Two populations of *C. floridanum* were established in the laboratory from field-collected broods obtained in the southern (Georgia) and northern (Wisconsin) US. Each culture was maintained separately as large randomly mating populations on the host *Trichoplusia ni*. Host larvae were reared on an artificial diet at 27 °C and in a 16 h light/8 h dark photoperiod as previously described²⁷. Experimental conditions were manipulated to favour production of all-male and all-female broods. Ovipositions were observed to determine that wasps laid one male or female egg to produce all-female or all-male broods¹⁷. Identifying the composition of broods in advance was possible, because females exhibit different behaviours when laying a fertilized or an unfertilized egg¹⁴. Hosts containing full sisters were produced by mating clonal females from one brood with a male from another. Hosts containing brothers were produced by allowing unmated females from the same brood to oviposit.

Male or female polymorulae containing on average 500 embryos were collected from fourth instar hosts and labelled with carboxyfluorocin diacetate succinimidyl ester (CFSE) using previously established methods^{14,16}. These previous studies confirmed that soldiers that attack CFSE-labelled tissues ingest the label which is then clearly visible when larvae are examined by epifluorescent microscopy. In experiment 1, CFSE-labelled sister, brother, non-relative Georgia female, non-relative Wisconsin female and non-relative Wisconsin male polymorulae were injected into fifth instar hosts containing a female brood from the Georgia population using a glass needle mounted on a micromanipulator. As a control, testes from fourth instar *T. ni*, which are similar in size to a polymorula, were labelled and individually injected. Hosts were then dissected 24 h after injection and the number of soldiers with label in their gut were determined. *In vitro* assays were conducted in 1-ml culture wells containing TC-100 medium (Sigma)^{14,16}. Soldiers and polymorulae from relative and non-relative broods were collected from fourth instar hosts immediately before the experiment. One soldier larva was placed in the culture well with a full sister, brother, non-relative female or non-relative male polymorula. The proportion of each type of polymorula that was attacked during a 2-h bioassay period was then recorded. An attack was defined as the larva gripping a polygerm with its mandibles for more than 1 min. When this occurred, consumption of tissue by the soldier was readily visible.

In experiment 2, hosts containing female broods were starved by removing them on the first day of the fifth instar from the food source to cups containing moist cotton wool. Each larva was returned to food 24 or 48 h after starvation and reared until completion of *C. floridanum* development. The average maximum size of hosts treated in this manner declined linearly from 507 ± 56 mg (mean ± s.e.m., *N* = 20) for unstarved controls to 398 ± 42 mg (*N* = 20) and 286 ± 38 mg (*N* = 20) for hosts starved for 24 or 48 h, respectively. Previous studies indicated that the total number of wasp progeny produced per polymorula is established at the end of the host fourth instar^{11,18}. Thus, any reduction in the number of progeny produced in starved, fifth instar hosts is due to mortality caused by increased competition for more limited host resources rather than a change in the development of the polymorula itself. To assess the effect of host starvation on the resident clone, parasitized hosts containing a female brood were starved for 48 h, returned to food and then reared until completion of development. The total number of wasp progeny per host was then counted. For comparisons of soldier aggression in starved and unstarved hosts, one cohort of hosts was starved for 48 h, returned to food and then injected 24 h later with a CFSE-labelled sister, brother, non-relative female (Wisconsin) or non-relative male polymorula. The second, unstarved cohort was treated identically except that hosts were continuously provided with food. Hosts in both cohorts were dissected 24 h after injection and the number of hosts containing soldiers with label in their gut was determined as described in experiment 1.

The association between soldier aggression, relatedness, resource competition or gender of the competitor were analysed by likelihood ratio chi-square tests using the JMP, v.3.0 statistical package.

Received 30 March; accepted 7 June 2004; doi:10.1038/nature02721.

1. Hamilton, W. D. The evolution of altruistic behaviour. *Am. Nat.* **97**, 354–356 (1963).
2. Hamilton, W. D. The genetical evolution of social behaviour, I & II. *J. Theor. Biol.* **7**, 1–52 (1964).
3. Wilson, D. S., Pollock, G. B. & Dugatkin, L. A. Can altruism evolve in a purely viscous population? *Evol. Ecol.* **6**, 331–341 (1992).
4. Taylor, P. D. Altruism in viscous populations—an inclusive fitness model. *Evol. Ecol.* **6**, 352–356 (1992).
5. Taylor, P. D. Inclusive fitness in a homogeneous environment. *Proc. R. Soc. Lond. B* **249**, 299–302 (1992).
6. Queller, D. C. Genetic relatedness in viscous populations. *Evol. Ecol.* **8**, 70–73 (1994).
7. Grafen, A. in *Behavioural Ecology: An Evolutionary Approach* (eds Krebs, J. R. & Davies, N. B.) 62–84 (Blackwell Scientific, Oxford, 1984).
8. West, S. A., Murray, M. G., Machado, C. A., Griffin, A. S. & Herre, E. A. Testing Hamilton's rule with competition between relatives. *Nature* **409**, 510–513 (2001).
9. West, S. A., Penn, I. & Griffin, A. S. Cooperation and competition between relatives. *Science* **296**, 72–75 (2002).
10. Strand, M. R. in *Encyclopedia of Insects* (eds Cardé, R. & Resch, V.) 928–932 (Academic, San Diego, 2003).
11. Strand, M. R. & Grbic, M. The development and evolution of polyembryonic insects. *Curr. Top. Dev. Biol.* **35**, 121–160 (1997).
12. Donnell, D., Corley, L. S., Chen, G. & Strand, M. R. Caste determination in a polyembryonic wasp involves inheritance of germ cells. *Proc. Natl Acad. Sci. USA* (in the press).
13. Cruz, Y. P. A sterile defender morph in a polyembryonic hymenopterous parasite. *Nature* **294**, 446–447 (1981).
14. Harvey, J. A., Corley, L. S. & Strand, M. R. Competition induces adaptive shifts in caste ratios of a polyembryonic wasp. *Nature* **406**, 183–186 (2000).
15. Strand, M. R. Oviposition behavior and progeny allocation by the polyembryonic wasp *Copidosoma floridanum*. *J. Insect Behav.* **2**, 355–369 (1989).
16. Grbic, M., Ode, P. J. & Strand, M. R. Sibling rivalry and brood sex ratios in polyembryonic wasps. *Nature* **360**, 254–256 (1992).

17. Ode, P. J. & Strand, M. R. Progeny and sex allocation decisions of the polyembryonic wasp *Copidosoma floridanum*. *J. Anim. Ecol.* **64**, 213–224 (1995).
18. Grbic, M., Nagy, L., Carroll, S. B. & Strand, M. R. Development and pattern formation in the polyembryonic wasp, *Copidosoma floridanum*. *Development* **122**, 795–804 (1996).
19. Ode, P. J. & Hunter, M. S. in *Sex ratios: Concepts and Research Methods* (ed. Hardy, I. C. W.) 218–234 (Cambridge Univ. Press, Cambridge, 2002).
20. Giron, D. & Strand, M. R. Host resistance and the evolution of kin recognition in polyembryonic wasps. *Proc. R. Soc. Lond. (Suppl.) Biol. Lett.* published online 17 June 2004 (DOI:10.1098/rsbl/2004.0205).
21. Queller, D. C. Relatedness and the fraternal major transitions. *Phil. Trans. R. Soc. Lond. B* **355**, 1647–1655 (2000).
22. Crozier, R. H. Genetic clonal recognition abilities in marine invertebrates must be maintained by selection for something else. *Evolution* **40**, 1100–1101 (1986).
23. Buss, L. W. *The Evolution of Individuality* (Princeton Univ. Press, Princeton, 1987).
24. Grosberg, R. K. The evolution of allorecognition specificity in clonal invertebrates. *Q. Rev. Biol.* **63**, 377–412 (1988).
25. Strassmann, J. E., Zhu, Y. & Queller, D. C. Altruism and social cheating in the social amoeba *Dictyostelium discoideum*. *Nature* **408**, 965–967 (2000).
26. Abbot, P., Withgott, J. H. & Moran, N. A. Genetic conflict and conditional altruism in social aphid colonies. *Proc. Natl Acad. Sci. USA* **98**, 12068–12071 (2001).
27. Strand, M. R. Development of the polyembryonic parasitoid *Copidosoma floridanum* in *Trichoplusia ni*. *Entomol. Exp. Appl.* **50**, 37–46 (1989).

Acknowledgements This work was supported in part by the Natural Environment Research Council (UK), the National Science Foundation (US), the University of Georgia Experiment Station, and the Conseil General de la Region (France).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.R.S. (mrstrand@bugs.ent.uga.edu).

High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome

Dee R. Denver¹, Krystalynne Morris², Michael Lynch¹, & W. Kelley Thomas²

¹Department of Biology, Indiana University, Bloomington, Indiana 47405, USA
²Hubbard Center for Genome Studies, University of New Hampshire, Durham, New Hampshire 03824, USA

Mutations have pivotal functions in the onset of genetic diseases and are the fundamental substrate for evolution. However, present estimates of the spontaneous mutation rate and spectrum are derived from indirect and biased measurements. For instance, mutation rate estimates for *Caenorhabditis elegans* are extrapolated from observations on a few genetic loci with visible phenotypes and vary over an order of magnitude¹. Alternative approaches in mammals, relying on phylogenetic comparisons of pseudogene loci² and fourfold degenerate codon positions³, suffer from uncertainties in the actual number of generations separating the compared species and the inability to exclude biases associated with natural selection. Here we provide a direct and unbiased estimate of the nuclear mutation rate and its molecular spectrum with a set of *C. elegans* mutation-accumulation lines that reveal a mutation rate about tenfold higher than previous indirect estimates and an excess of insertions over deletions. Because deletions dominate patterns of *C. elegans* pseudogene variation^{4,5}, our observations indicate that natural selection might be significant in promoting small genome size, and challenge the prevalent assumption that pseudogene divergence accurately reflects the spontaneous mutation spectrum.

Table 1 Insertion and deletion mutations in the MA lines

Chr.	Chr. Pos.	Cos./YAC	Mut.	Context	Line(s)	Cod.
I	8,738,889	C36B1	+A	TCGTT → TCGATT	23	IN
I	8,739,217	C36B1	-C	TTCT → TTCT	73	IN
I	8,952,991	F53B6	+C	CCTTT → CCTCT	39	IN
I	12,826,513	B0019	+G	CCGTT → CCGGTT	12	IN
II	3,170,985	Y25C1A	+G	TAGCG → TAGGCG	95	IN
II	5,629,513	C17G10	+C	AGCAA → AGCCAA	13	IG
II	5,924,746	F59G1	+G	TGCAG → TGGCAG	2	EX
II	5,924,746	F59G1	+C	TGCAG → TGCCAG	14	EX
II	13,314,289	Y48C3A	+CCC	AGCCCGC → AGCCCCCGC	80	EX
II	13,314,321	Y48C3A	+GAT	TGGATCA → TGGATGATCA	80	EX
II	13,314,343	Y48C3A	+GA	CGAGCA → CGAGAGCA	80	EX
III	12,353,930	Y75B8A	-11 bp	(T ₇ GGTC) ₂ → (T ₇ GGTC) ₁	90	IN
IV	~1,564,830*	Y41D4B	+ ~500 bp	ND	46	IN
IV	17,277,013	F26D10	+T	ACTTG → ACTTTG	89	EX
V	11,074,113	T04C12	+G	ATGGCG → ATGGGCG	39	IG
X	10,021,926	F19C6	-66 bp	(17 bp) (49 bp) (17 bp) → (17 bp)	29	IN
X	10,172,717	C07A4	-A	GC(A) ₅ TA → GC(A) ₄ TA	2	EX

Chr. indicates the chromosome on which the mutation was found, and Chr. Pos. refers to the chromosomal position of the mutation. Cos./YAC refers to the sequenced *C. elegans* cosmid or yeast artificial chromosome in which the mutation was found. Mut. refers to the observed mutation. Line(s) indicates the MA line number(s) where the mutation was detected. Cod. refers to the type of coding sequence in which the mutation was detected: EX is exon, IG is intergenic, and IN is intron.

*A large insertion mutation whose sequence was not determined (Supplementary Fig. S2).

To obtain a direct estimate of the spontaneous mutation rate and spectrum, we sequenced more than 4 megabases of nuclear DNA from a set of *C. elegans* mutation-accumulation (MA) lines. The MA lines were simultaneously initiated from a single N2 (laboratory strain) ancestor and propagated in a benign environment across generations by single-progeny descent, ensuring that all but the most deleterious mutations accumulated over time in an effectively neutral fashion, and resulting in marked declines⁶ in fitness. We sequenced 29,561 base pairs (bp) from each of 72 MA lines at 280 generations, 14,550 bp from each of 68 MA lines at 353 generations, and 18,718 bp from each of 58 MA lines at 396 generations (Supplementary Table S1). All mutations were visually confirmed on electropherogram data from both strands of directly sequenced polymerase chain reaction (PCR) products, ensuring a very low error rate (see Methods). The same loci were sequenced from a set of natural isolates of *C. elegans* for a direct comparison of the mutation spectrum in the MA lines with natural substitution patterns⁷.

We detected 30 mutations (Tables 1 and 2) in the MA lines, yielding a mutation rate estimate of 2.1×10^{-8} mutations per site per generation (standard error of the mean, s.e.m. = $\pm 6.7 \times 10^{-9}$) or, based on an average generation time of four days, 2.0×10^{-6} mutations per site per year ($\pm 6.1 \times 10^{-7}$). The total haploid genomic mutation rate (U_1) is ~ 2.1 mutations per genome per generation, given that the size of the *C. elegans* genome is $\sim 10^8$ bp. The distribution of mutations in each MA line and each locus were very close to Poisson expectations (see Methods). The nuclear mutation rate estimate reported here is about one order of magnitude lower than the mitochondrial rate detected in the same *C. elegans* MA lines (1.6×10^{-7} mutations per site per generation)⁸ but is nearly one order of magnitude higher than the previous average gene-specific nuclear rate estimate for *C. elegans*¹. This observation is consistent with a recent study showing laboratory-based mutation rate estimates about tenfold higher than phylogenetic estimates for *Escherichia coli* and *Salmonella enterica*⁹.

More than half of the mutations (17 of 30) observed in the MA lines were insertion-deletion (indel) mutations (Table 1); among the indels, insertions were dominant (13 of 17). The latter observation differs markedly from patterns of pseudogene evolution in natural populations of both *Caenorhabditis* and *Drosophila*, in which deletions outnumber insertions^{4,5,10}. Among *C. elegans* transposon pseudogenes, for instance, deletions are 2.8-fold more frequent than insertions⁵. This apparent bias towards deletions inferred from patterns of pseudogene evolution is often invoked as a mutation-based mechanism responsible for the maintenance of small genome size in metazoans such as *Drosophila* and *C. elegans*^{10,11}. However, several studies indicate that pseudogenes might

be subject to selective constraints: the makorin1 p1 mouse pseudogene is transcribed and regulates the messenger RNA stability of its homologous coding gene¹², and transcriptional activity is reported for $\sim 20\%$ of *Arabidopsis* sequences annotated as pseudogenes¹³. Furthermore, many pseudogenes in a variety of species display codon bias, retention of reading frame, and very low ratios of nonsynonymous to synonymous substitution rates (K_a/K_s) (ref. 14). The predominance of insertions observed in the MA lines coupled with the bias towards deletions in *C. elegans* pseudogenes indicates that long-term pseudogene evolution in *C. elegans* might not accurately reflect the baseline spontaneous mutation spectrum and that natural selection might favour deletions and/or prevent insertions in otherwise neutral regions of the *C. elegans* genome. If this interpretation is correct, the evolution of the compact *C. elegans* genome might be driven largely by selective rather than mutational forces¹⁵.

Although on the surface it might seem unlikely that natural selection can act on small insertions in presumably non-functional regions of the genome, selection is much more efficient in invertebrate species with large effective population sizes, such as *C. elegans*, than in those with relatively small effective population sizes, such as vertebrates, as reflected in many aspects of genome evolution, including numbers and sizes of introns and the abundances of mobile elements¹⁶. Furthermore, natural selection prevents the accumulation of other types of mutation with seemingly mild effects, such as unpreferred base substitutions at silent sites in highly transcribed genes ('codon bias')¹⁷ and spurious transcription-factor binding sites¹⁸. In species with very small effective

Table 2 Base substitution mutations in the MA lines

Chr.	Chr. Pos.	Cos./YAC	Mut.	Context	Line(s)	Cod.
I	6,816,148	K02F2	C → T	TACTA → TATTA	64, 78	IN
I	8,738,859	C36B1	A → G	GGATA → GGGTA	90	IN
II	6,537,060	C56E6	G → A	AAGCC → AAACC	97	IG
II	10,851,146	M106	G → C	ACGTC → ACCTC	13	IG
III	1,933,779	F53A3	G → C	GAGAT → GACAT	77	IN
III	1,933,788	F53A3	C → G	AACCT → AAGCT	77	IN
III	12,353,902	Y75B8A	T → C	AATTT → AACTT	26	IN
IV	8,001,467	Y42H9B	G → A	TAGAG → TAAAG	67	EX (N: S → F)
IV	10,332,063	K11E8	C → T	GCCGT → GCTGT	61	IN
V	11,724,218	B0240	C → A	GTCAA → GTAAA	84	IG
X	10,173,063	C07A4	G → A	TCGGC → TCAGC	94	EX (S)
X	10,609,426	F59F5	A → T	CAACA → CATCA	98	EX (N: K → R)

Chr. indicates the chromosome on which the mutation was found, and Chr. Pos. refers to the chromosomal position. Cos./YAC refers to the sequenced *C. elegans* cosmid or yeast artificial chromosome in which the mutation was found. Mut. refers to the observed mutation. Line(s) indicates the MA line number(s) where the mutation was detected. Cod. refers to the coding context of the sequence in which the mutation was found: EX is exon, IG is intergenic, and IN is intron. For exon mutations, S indicates a synonymous mutation and N indicates a nonsynonymous mutation (amino acid change is indicated after N).

population sizes, such as humans, in which natural selection is less efficient, pseudogene variation might more accurately reflect the spontaneous mutation spectrum.

Our observations are also based on organisms propagated in the laboratory, whereas previous comparative phylogenetic approaches to estimating the mutation spectrum, although indirect, are derived from sequences evolving in nature. Although we cannot exclude the possibility that mutation processes might differ between laboratory-reared organisms and those occurring in nature, observations in prokaryotes indicate that certain mutation processes (the bias towards transitions over transversions, for instance) do not significantly differ between bacteria raised in the laboratory and those evolving in nature⁹. Furthermore, the rate of sex-linked lethal mutations is highly similar in wild-caught and laboratory populations of *Drosophila*¹⁹.

Five frameshift-causing indel mutations were observed in exon sequence, each resulting in a premature stop codon (Table 1). Two of these indels occurred independently in MA2 and MA14 at the same precise location in the final exon of the *vps-35* gene (F59G1.3). Length variation was also observed at this precise location in the *C. elegans* natural isolates⁷, indicating that this might be a hotspot for insertion mutations. Three short insertions were observed within a 60-bp segment of the fourth exon of the *mac-1* gene (Y48C3A.7) in MA80 (Supplementary Fig. S1), an insertion of a single base pair was detected in the second exon of the *hsp-1* gene, and a single base-pair deletion occurred in the third exon of an unnamed gene (C07A4.2). Although the functional consequences of these specific mutations remain uncertain, *C. elegans* RNA-mediated interference (RNAi) studies²⁰ provide some insights: for the *mac-1* and C07A4.2 genes RNAi treatment resulted in no detectable phenotypes, but sterility was observed for *hsp-1* and *vps-35*. However, the frameshift mutations observed at *hsp-1* and *vps-35* in the MA lines might result in the expression of truncated protein products and result in less severe phenotypes than what is observed in RNAi studies. The largest indel observed was a ~500-bp insertion in MA46 at locus Y41D4B (Supplementary Fig. S2). Although exhaustive efforts to determine the sequence of this mutation were unsuccessful, the presence of five copies of the *C. elegans* repetitive element Rc123 (ref. 21) at this locus in wild-type (N2) DNA indicates that there might have been an expansion of these elements at this locus in MA46. Size variation at this locus between the *C. elegans* natural isolates is due to variation in Rc123 copy number.

Base substitutions are about fourfold more frequent than indels in the *C. elegans* natural isolates⁷, and this observation is often used to justify the use of base-substitution-inducing mutagens such as ethylmethane sulphate to mimic spontaneous mutation processes^{22,23}. However, our observation that the mutation rate to indels is higher than the rate for base substitutions in the MA lines raises questions about the validity of this assumption. It is unlikely that any single mutagen would induce the diverse assortment of spontaneous mutations observed in the MA lines.

Transitions are consistently more common than transversions in natural patterns of nuclear variation in metazoans, but it remains unclear whether this pattern is driven by biased mutation and/or selection. Among the 13 base substitutions observed in the MA lines (Table 2), 8 were transitions and 5 were transversions, yielding a transition/transversion (Ts/Tv) ratio (1.6) that is strikingly similar to that observed at the same nuclear loci in the *C. elegans* natural isolates (1.7) (ref. 7). Even higher transition biases were found in the mitochondrial genomes of both the *C. elegans* MA lines (Ts/Tv = 4.3)⁸ and natural isolates (Ts/Tv = 2.6)⁷. These observations suggest that underlying mutational forces drive the widely observed transition-bias phenomenon in both the nuclear and mitochondrial genomes of *C. elegans*.

Last, we note that our sequence-based estimate of U_t (~2.1 mutations per haploid genome per generation) is about two orders

of magnitude higher than the haploid deleterious genomic mutation rate (U_d) estimated by laboratory fitness assays with the same set of *C. elegans* MA lines ($U_d \approx 0.015$ mutations per haploid genome per generation)⁶, indicating that most mutations in the MA lines (more than 99%) might either be neutral or have deleterious effects too mild to be detected in laboratory fitness assays. However, comparative analyses of *Caenorhabditis* genome sequences suggest that the vast majority of nonsynonymous mutations (~94% on the basis of a between-species K_a/K_s ratio of 0.06)²⁴ are sufficiently deleterious for selection to act against them in nature, and fitness-based estimates of U_d are known to be highly dependent on environmental conditions²⁵. If we consider only nonsynonymous base substitution mutations and indels in exon sequence, the mutation rate reported here coupled with the K_a/K_s data for *Caenorhabditis* yields an estimate of 0.48 for U_d (see Methods). This approximation of U_d is probably an underestimate because it does not consider mutations in intron and intergenic regions that can also have negative effects. Nevertheless, the fact that our sequence-based minimal estimate for U_d in *C. elegans* is about 30-fold higher than the estimate based on laboratory fitness assays indicates that most spontaneous mutations in *C. elegans* have very mildly deleterious effects that are realized only in natural contexts. □

Methods

Mutation detection and confirmation

Details about PCR amplification and sequencing of *C. elegans* nuclear loci have been described previously^{7,8}. Most loci sequenced were randomly distributed across *C. elegans* chromosomes by designing PCR primer pairs around chromosomal positions selected by a random-number generator. All amplifications were performed with a large amount of genomic DNA (~25,000 diploid genomes per reaction) and 2 U Taq DNA polymerase (Applied Biosystems) to eliminate artefacts associated with initial amplification from small amounts of genomic DNA. All PCR products were directly sequenced in both directions, and internal primers were used where necessary.

DNA sequence text files were aligned to wild-type (N2) sequences with CLUSTALW²⁶ to identify putative mutations in the MA lines. Putative mutations identified in the alignments were then scrutinized visually on the electropherogram data to eliminate base-caller errors and other sequencing artefacts. All putative mutations showing evidence of wild-type or ambiguous sequence data were resequenced. Putative mutations supported by clean, unambiguous electropherogram data were then evaluated on the opposite strand (sequencing reaction in the opposite direction), with internal primers where necessary. If we conservatively assume that our sequencing error rate was on the high end of that reported for the human genome (~10⁻⁴) (ref. 27) then an error rate of ~10⁻⁸ is expected for our approach where mutations were confirmed on both strands. Assuming this conservative error rate, we would expect only ~0.04 mutations due to sequencing error in the more than 4 megabases of DNA that we surveyed. All mutations reported here received high Phred scores (more than 33) on both strands (corresponding to an error probability of less than 0.05%)²⁸. Finally, starting from genomic DNA we re-amplified and sequenced a subset (10 of 30) of the PCR products containing the initially identified mutations, and in each of the ten instances the mutations were again detected.

Calculation of mutation rates

Individual mutation rates were calculated with the equation $\mu = m/(LnT)$, where μ is the mutation rate (per nucleotide site per generation), m is the number of observed mutations, L is the number of MA lines, n is the number of nucleotide sites and T is the time in generations, as described previously⁸. The standard errors for individual mutation rates were calculated as $[\mu/(LnT)]^{1/2}$, as described previously⁸. Three highly similar independent mutation rates were calculated for each of the three mutation assays: 2.0×10^{-8} mutations per site per generation ($\pm 5.8 \times 10^{-9}$) for the first assay (280 generations), 2.6×10^{-8} ($\pm 8.6 \times 10^{-9}$) for the second assay (353 generations) and 2.1×10^{-8} ($\pm 7.0 \times 10^{-9}$) for the third assay (396 generations). The single rate reported in the text is the average of the three rates, weighted by the inverse of the sampling variance for each rate. To avoid providing an upwardly biased estimate of the mutation rate, we excluded mutations detected at homopolymeric nucleotide runs of 8 bp or longer that are hotspots for indel mutation and were specifically targeted for another study²⁹.

Poisson distribution of mutations

For both the distribution of mutations between individual MA lines and between individual PCR products, our observations were not significantly different from Poisson expectations. Assuming a Poisson distribution and a mean of 0.45 mutations per MA line (with a weighted average of 66 total MA lines), the expected number of MA lines with 0, 1, 2 and 3 mutations are 42, 19, 4 and 1, respectively; we observed 43, 17, 5 and 1 MA lines. Assuming a Poisson distribution and a mean of 0.39 mutations per PCR product locus surveyed (76 total), the expected number of loci with 0, 1, 2 and 3 mutations are 51, 20, 4 and 1, respectively; we observed 54, 15, 5 and 2 loci.

Estimation of the deleterious genomic mutation rate

If we use $1 - (K_d/K_s)$ as a measure of the fraction of mutations in protein-coding genes that are deleterious (for *Caenorhabditis*, $K_d/K_s \approx 0.06$)²⁴ and multiply this by the genomic rate of nonsynonymous base substitution mutations in exon sequences in the MA lines (0.18 nonsynonymous mutations per genome per generation, given that ~26% of the *C. elegans* genome is exon, ~75% of exon nucleotides are nonsynonymous sites, and our total mutation rate of 0.9 base substitutions per genome per generation), we get 0.17 as an estimate of U_d for *C. elegans*. Adding the rate of indel mutations in exon sequence (0.31 indel mutations in exons per genome per generation, given that ~26% of the *C. elegans* genome is exon and our total mutation rate of 1.2 indels per genome per generation) and assuming all exon indels are deleterious, yields 0.48 as an estimate of U_d for *C. elegans*.

Received 3 February; accepted 1 June 2004; doi:10.1038/nature02697.

1. Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. Rates of spontaneous mutation. *Genetics* **148**, 1667–1686 (1998).
2. Nachman, M. W. & Crowell, S. L. Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**, 297–304 (2000).
3. Kumar, S. & Subramanian, S. Mutation rates in mammalian genomes. *Proc. Natl Acad. Sci. USA* **99**, 803–808 (2002).
4. Robertson, H. M. The large *srh* family of chemoreceptor genes in *Caenorhabditis* nematodes reveals processes of genome evolution involving large duplications and deletions and intron gains and losses. *Genome Res.* **10**, 192–203 (2000).
5. Witherspoon, D. J. & Robertson, H. M. Neutral evolution of ten types of mariner transposons in the genomes of *Caenorhabditis elegans* and *Caenorhabditis briggsae*. *J. Mol. Evol.* **56**, 751–769 (2003).
6. Vassilieva, L. L., Hook, A. M. & Lynch, M. The fitness effects of spontaneous mutations in *Caenorhabditis elegans*. *Evolution* **54**, 1234–1246 (2000).
7. Denver, D. R., Morris, K. & Thomas, W. K. Phylogenetics in *Caenorhabditis elegans*: an analysis of divergence and outcrossing. *Mol. Biol. Evol.* **20**, 393–400 (2003).
8. Denver, D. R., Morris, K., Lynch, M., Vassilieva, L. L. & Thomas, W. K. High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science* **289**, 2342–2344 (2000).
9. Ochman, H. Neutral mutations and neutral substitutions in bacterial genomes. *Mol. Biol. Evol.* **20**, 2091–2096 (2003).
10. Petrov, D. A., Lozovskaya, E. R. & Hartl, D. L. High intrinsic rate of DNA loss in *Drosophila*. *Nature* **384**, 346–349 (1996).
11. Petrov, D. A. & Hartl, D. L. Pseudogene evolution and natural selection for a compact genome. *J. Hered.* **91**, 221–227 (2000).
12. Hirotsune, S. *et al.* An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* **423**, 91–96 (2003).
13. Yamada, K. *et al.* Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science* **302**, 842–846 (2003).
14. Balakirev, E. S. & Ayala, F. J. Pseudogenes: are they 'junk' or functional DNA? *Annu. Rev. Genet.* **37**, 123–151 (2003).
15. Charlesworth, B. The changing sizes of genes. *Nature* **384**, 315–316 (1996).
16. Lynch, M. & Conery, J. S. The origins of genome complexity. *Science* **302**, 1401–1404 (2003).
17. Marais, G., Mouchiroud, D. & Duret, L. Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. *Proc. Natl Acad. Sci. USA* **98**, 5688–5692 (2001).
18. Hahn, M. W., Stajich, J. E. & Wray, G. A. The effects of selection against spurious transcription factor binding sites. *Mol. Biol. Evol.* **20**, 901–906 (2003).
19. Langley, C. H. & Ito, K. Spontaneous mutability in *Drosophila melanogaster*, in natural and laboratory environments. *Mutat. Res.* **36**, 385–386 (1976).
20. Gunsalus, K. C., Yueh, W. C., MacMenamin, P. & Piano, F. RNAiDB and PhenoBlast: web tools for genome-wide phenotypic mapping projects. *Nucleic Acids Res.* **32**, D406–D410 (2004).
21. Naclerio, G. *et al.* Molecular and genomic organization of clusters of repetitive DNA sequences in *Caenorhabditis elegans*. *J. Mol. Biol.* **226**, 159–168 (1992).
22. Keightley, P. D. & Ohnishi, O. EMS-induced polygenic mutation rates for nine quantitative characters in *Drosophila melanogaster*. *Genetics* **148**, 753–766 (1998).
23. Davies, E. K., Peters, A. D. & Keightley, P. D. High frequency of cryptic deleterious mutations in *Caenorhabditis elegans*. *Science* **285**, 1748–1751 (1999).
24. Stein, L. D. *et al.* The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol.* **1**, 166–192 (2003).
25. Kondrashov, A. S. & Houle, D. Genotype–environment interactions and the estimation of the genomic mutation rate in *Drosophila melanogaster*. *Proc. R. Soc. Lond. B* **258**, 221–227 (1994).
26. Higgins, D. G., Thompson, J. D. & Gibson, T. J. Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.* **266**, 383–402 (1994).
27. Hill, F., Gemund, C., Benes, V., Ansorge, W. & Gibson, T. J. An estimate of large-scale sequencing accuracy. *EMBO Rep.* **1**, 29–31 (2000).
28. Richterich, P. Estimation of errors in 'raw' DNA sequences: a validation study. *Genome Res.* **8**, 251–259 (1998).
29. Denver, D. R. *et al.* Abundance, distribution, and mutation rates of homopolymeric nucleotide runs in the genome of *Caenorhabditis elegans*. *J. Mol. Evol.* **58**, 584–595 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank L. L. Vassilieva, S. Estes, V. Katju and C. Steding for their respective roles in propagating and maintaining the MA lines over the past 5 years; D. Ash for help with primer sequence design and DNA sequencing; and the *Caenorhabditis* Genetics Center for providing the *C. elegans* natural isolates. This work was supported by a University of Missouri Research Board grant to W.K.T., and an NIH grant to M.L. and W.K.T.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.R.D. (ddenver@bio.indiana.edu).

Optimal neural population coding of an auditory spatial cue

Nicol S. Harper^{1,2} & David McAlpine¹

¹Department of Physiology and UCL Ear Institute and ²CoMPLEX, University College London, London WC1E 6BT, UK

A sound, depending on the position of its source, can take more time to reach one ear than the other. This interaural (between the ears) time difference (ITD) provides a major cue for determining the source location^{1,2}. Many auditory neurons are sensitive to ITDs^{3,4}, but the means by which such neurons represent ITD is a contentious issue. Recent studies question whether the classical general model (the Jeffress model⁵) applies across species^{6,7}. Here we show that ITD coding strategies of different species can be explained by a unifying principle: that the ITDs an animal naturally encounters should be coded with maximal accuracy. Using statistical techniques and a stochastic neural model, we demonstrate that the optimal coding strategy for ITD depends critically on head size and sound frequency. For small head sizes and/or low-frequency sounds, the optimal coding strategy tends towards two distinct sub-populations tuned to ITDs outside the range created by the head. This is consistent with recent observations in small mammals^{6,7}. For large head sizes and/or high frequencies, the optimal strategy is a homogeneous distribution of ITD tunings within the range created by the head. This is consistent with observations in the barn owl^{8–10}. For humans, the optimal strategy to code ITDs from an acoustically measured distribution depends on frequency; above 400 Hz a homogeneous distribution is optimal, and below 400 Hz distinct sub-populations are optimal.

The ability to localize sound sources has obvious survival value, whether for prey or predator. Many vertebrates, including humans, make use of ITDs to localize sounds in the horizontal plane. These ITDs can be in the order of just a few tens of microseconds. In the Jeffress model⁵, these minute ITDs are encoded by an array of coincidence-detector neurons. Each neuron is tuned for (responds maximally to) an ITD within the range created by the head size (the physiological range), the precise tuning being determined by the difference in axonal conduction time from each ear. The Jeffress model, developed with human spatial hearing in mind, has received extensive support from studies in barn owls^{8–10}, and until recently was presumed to apply to mammals¹¹. However, the preferred ITD tuning of neurons recorded in small mammals that use ITDs for sound localization seems to lie outside the physiological range^{6,7}. This is markedly different to the preferred ITD tuning observed in the barn owl and suggested by the Jeffress model. It is now debated whether a single unifying principle or model can any longer explain the means by which ITD is encoded across the wide range of species in which ITD sensitivity is observed.

We investigated the possibility that different coding strategies observed in different species can be explained by the demand for accurate ITD coding. For simplicity, we considered first how to encode most accurately the ITDs in the ongoing fine structure of a pure tone using a population of ITD-tuned model neurons. In mammals, fine-structure ITD sensitivity is restricted to sound frequencies below ~1,500 Hz. At higher frequencies, some ITD information can be conveyed by the envelope structure of complex sounds¹²; however, this seems to have relatively little impact on localization judgements^{2,13}, and thus we only consider fine-structure ITDs.

Here, we present the optimal coding strategies for four species with different head sizes and/or different sound-frequency ranges