The dynamics of chromosome evolution in birds and mammals

David W. Burt*, Charlotte Bruley*, Ian C. Dunn*, Cheryl T. Jones*, Anne Ramage*, Andy S. Law*, David R. Morrice*, Ian R. Paton*, Jacqueline Smith*, Dawn Windsor*, Alexei Sazanov†, Ruedi Fries† & David Waddington*

** Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, UK*
† Lehrstuhl fur Tierzucht, TUM, Freising, Weihenstephan, Munich D-85350, Germany

Comparative mapping, which compares the location of homologous genes in different species, is a powerful tool for studying genome evolution[1]. Comparative maps suggest that rates of chromosomal change in mammals can vary from one to ten rearrangements per million years[1–4]. On the basis of these rates we would expect 84 to 600 conserved segments in a chicken comparison with human or mouse. Here we build comparative maps between these species and estimate that numbers of conserved segments are in the lower part of this range. We conclude that the organization of the human genome is closer to that of the chicken than the mouse and by adding comparative mapping results from a range of vertebrates, we identify three possible phases of chromosome evolution. The relative stability of genomes such as those of the chicken and human will enable the reconstruction of maps of ancestral vertebrates.

To examine the dynamics of chromosome rearrangement during vertebrate evolution, we constructed comparative maps of chicken with human and mouse (see Supplementary Information), species that share a common ancestor 300 million years ago. We have used the recommendations of the Human Genome Organisation (HUGO) Comparative Genome Organisation for definitions of gene homologies, orthologies and conserved segments[1]. Over 2,000 orthologous genes ('orthologues') have already been mapped in a mouse–human comparison[2] and these define 195 autosomal conserved segments. Given the large number of genes, we have assumed that the actual number is close to this total. In the chicken, we mapped 223 genes which define 81 autosomal conserved segments in the human comparison and 100 in the mouse comparison. We predict the total number of autosomal conserved segments in the chicken–human comparison to be 96 (95% confidence limits: lower 79; upper 117) and in the chicken–mouse (ref. 5) comparison to be 152 (118, 198). These totals are the sum of the number of autosomes in the last common ancestor (assumed to be 24; ref. 6) and the number of chromosome rearrangements since divergence[3]. Therefore we predict 72 (55, 93) chromosome rearrangements for the chicken–human comparison; this is less than for mouse–chicken (128, (94, 174)) and mouse–human (at least 171). We conclude that the organization of the human genome is closer to that of the chicken than the mouse.

A relative rate test[7] was used to compare the rates of chromosomal change in human and mouse lineages, using the chicken lineage as an outgroup. The rate in the mouse lineage was twice as fast as that in the human. This difference is supported if we compare the sizes of conserved segments that are common to both our chicken–mouse and chicken–human comparative maps. Of 14 segments that differ in size, 13 are larger in the chicken–human comparison (sign-test $P < 0.001$). Rates for human and mouse lineages are 0.58 and 1.14 rearrangements per million years, respectively, since divergence 100 Myr ago. There is no outgroup for chicken, so it is not possible to estimate the number of rearrangements in this lineage; however, the maximum number compatible with the above upper limits for human and mouse is 48. This gives a maximum rate of 0.16 rearrangements per million years since divergence 300 Myr ago, which is considerably less than the minimum estimate for the mouse lineage (0.9 per Myr).

To reveal the global pattern of chromosome evolution we have combined results from genetic mapping and chromosome painting (Zoo-FISH) with vertebrate phylogeny[6–10]. A summary is shown in Fig. 1, from which we identify three phases of chromosome evolution. In phase I (100–300 Myr ago), the rate of chromosome rearrangement was slow, less than 0.2 per million years in both avian and mammalian lineages. In phase II (65–100 Myr ago), the rate increased to over 1.1 per million years in both mouse and non-rodent lineages. Recent comparisons of rat and human[11] suggest

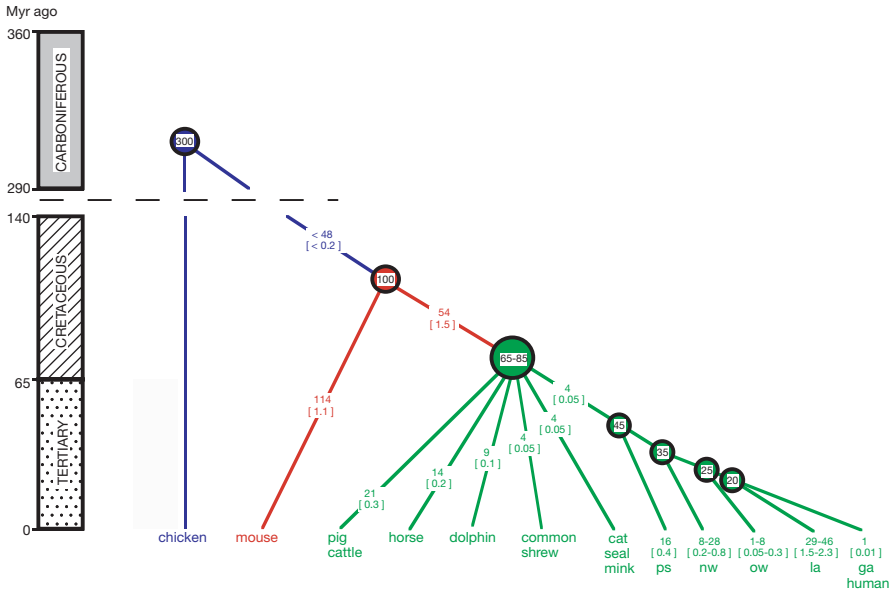

Figure 1 Dynamics of chromosome evolution in birds and mammals. Circles represent estimated times (in Myr) of divergence of common ancestors. The three phases of chromosomal evolution discussed in the text are shown in blue (phase I), red (phase II) and green (phase III). The estimated number of chromosome rearrangements is shown along each lineage, with the rates of chromosomal rearrangement per million years (in brackets): ps, prosimians (lemur); nw, New World monkeys (six species); ow, Old World monkeys (four species); la, lesser apes (four species); ga, great apes (six species). For more details see Methods and Supplementary Information.

that rates of chromosome change may have been high in mice (1.1 per Myr), rats (0.7 per Myr) and possibly other rodents. In phase III (during the past 65–85 million years) the rate has been variable in non-rodent mammals, with the slowest rate (less than 0.1 per Myr) for human, carnivores and common shrew, a higher rate (0.1–0.3 per Myr) for pig, cattle, horse and dolphin, and the highest rate in the lesser apes (1.5–2.3 per Myr). These rates of chromosome evolution differ from earlier estimates[12], measured from changes in chromosome and chromosome arm number, averaged over genera within major taxonomic groups. The differences we find within specific lineages cannot be detected by this measure.

How can we explain these patterns of chromosome evolution, and how are they explained by current models of genome evolution? Many chromosome changes are likely to be deleterious and will therefore be rapidly lost by natural selection; they will thus not contribute to chromosome diversity. Some slightly deleterious mutations are associated with reduced fitness, for example, owing to reduced fertility caused by abnormal segregation of chromosomes during meiosis. These are unlikely to be fixed unless the mutation rate is high or the population size is small.

When the population size becomes small (such as after a natural disaster or at the time of speciation), then genetic drift will predominate and chromosomal rearrangements with reduced fitness will have a greater chance of fixation[12]. As the population expands, natural selection becomes increasingly influential. Then the rate of chromosomal evolution will diminish to a basal rate of change, possibly representing rare, advantageous mutations. Under natural selection, these mutations will be fixed at a rate that is largely controlled by how much the environment changes, which would depend on chronological time. This rate will tend to be equal in all lineages. Fluctuations in population size may have been a major factor leading to the wide variation in the rates of chromosome rearrangement in non-rodent mammals in phase III (Fig. 1). Such changes would be expected 65–85 Myr ago, a time of rapid speciation in mammals.

When the mutation rate is the primary cause of differences in the rates of chromosomal rearrangement between lineages, then mutation mechanisms must differ. One possible mechanism is chromosome mispairing (and subsequent translocation) caused by homologous sequences at different sites in the genome[6]. Then changes in the frequency of mispairing may alter the mutation rate. The frequency of at least three types of homologous sequence differs between mammalian and avian genomes. Dispersed repetitive sequences make up more than 50% of a mammalian genome but less than 15% of a bird genome. Also, mammals have larger gene families than birds, often with many pseudogenes. Multiple copies of retroviral elements are more frequent in mammals, and these may play a role in some rearrangements[13]. Consequently, the potential for chromosome mispairing may be low in birds and this may explain the low rate of chromosomal rearrangement within the chicken lineage. It may also have led to a low rate in ancestral mammals, 100–300 Myr ago (phase I, Fig. 1). Perhaps the accumulation of homologous sites in rodent and non-rodent lineages increased the rate of chromosome change in phase II (Fig. 1). Our findings on rates of chromosomal rearrangement show similarities to the comparison of nucleotide substitution rates[14,15] between mouse and human. Both rates appear to be more dependent on generation time than chronological time, with the mouse rates being at least twice those found in human during the last 100 million years. This is an inevitable consequence of different generation times between mouse and human, if we assume equal mutation rates per DNA replication[16]. Differences in generation time will also contribute to the variation in chromosomal evolution within non-rodent mammals in phase III (Fig. 1).

Conservation of genome organization throughout the vertebrate kingdom has important practical and evolutionary implications. First, we can be optimistic about the use of vertebrate comparative maps to predict candidate genes for phenotypes mapped in species as diverse as chicken and human. Second, we can now start to use comparative gene maps derived from selected species for the systematic reconstruction of ancestral vertebrate genomes. These will enable us to establish the dynamics of chromosomal reorganization. This will be made possible by the relative stability of some of the genomes being mapped, such as chicken and human, spanning 300 million years of vertebrate evolution. □

Methods

Chicken maps are recorded in Arkdb-CHICK (http://www.ri.bbsrc.ac.uk/chickmap/), and at World-Wide Web sites in Wageningen (http://www.zod.wau.nl/vf/chickensite/chicken.html) and East Lansing (http://poultry.mph.msu.edu/chickmap.html). Human and mouse maps were taken from GDB (http://gdbwww.gdb.org/gdb), Gene Map'98 (http://www.ncbi.nlm.nih.gov) and MGD (http://www.informatics.jax.org, March 1999).

The East Lansing map[17] was used as a reference on which to map all genes. For each chromosome, common markers were linked by a non-parametric regression of genetic distances from other crosses onto East Lansing distances, using cubic splines[18]. The degree of smoothing was selected by minimizing the Akaike information criterion[19]. When there were two or three markers, a linear or quadratic regression was fitted, as appropriate. To estimate the length of the chicken genome, map lengths were converted to genetic distances using the Kosambi mapping function[20] and adjusted for failure to sample telomeric regions[21]. From this approach the total genetic length was estimated to be 3900 cM.

The method of Waddington *et al.*[5] was used to estimate the underlying total number of conserved segments, using as data both the counts of genes in a syntenic block and distances between them. Only genes mapped at random (134 out of 223 mapped genes), that is, with no *a priori* knowledge of map position[5], were included in our analysis. Syntenic blocks defined by a single gene were accepted only when sequence comparisons could be made to minimize orthology errors. Many were confirmed by linkage to other genes not mapped at random or to a confirmed orthology in a human–mouse comparison. See Supplementary Information for a detailed summary of the observed/predicted numbers of conserved segments in the chicken comparisons.

Divergence times for birds and mammals were obtained from published sources[8–10], but definite dates were difficult to obtain owing to conflicts between sequence and fossil data[10]. In non-rodent mammals this is acknowledged to Fig. 1 by a star phylogeny radiating from a common ancestor 65–85 Myr ago[8–10]. Using values at the extremes of this range alters our conclusions about rates of change very little. Phylogenetic studies generally, but not always, place mice before the divergence of all other mammals[9]. We estimate 58 chromosome rearrangements in the human lineage, since divergence from a common ancestor with mice, but only 8 since divergence from the cat[4]. The difference is consistent with mice being placed in the more ancient branch, 100 Myr ago[10].

The resolution of Zoo-FISH is limited to 10 Mb (ref. 22) and so the observed number of conserved segments may be underestimated. However, assuming random chromosome breaks throughout the genome[5], we would expect few small segments in lineages with few chromosome rearrangements and therefore this bias should be small. Most Zoo-FISH data used in the construction of Fig. 1 were based on comparisons with human and so it was not possible to calculate the exact number of chromosome rearrangements in each lineage by solving simultaneous equations. However, since the number of rearrangements between human and cat was small[4], we assumed that the rate was equal in these two lineages. With this approximation, the numbers of rearrangements in the other lineages were derived by subtraction.

Received 5 May; accepted 30 September 1999.

1. Andersson, L. *et al.* The first international workshop on comparative genome organisation. *Mamm. Genome* **7,** 717–734 (1997).
2. Davisson, M. T., Bradt, D. W., Merriam, J. J., Rockwood, S. F. & Eppig, J. T. The mouse gene map. *Inst. Lab. Anim. Res. J.* **39,** 96–131 (1998).
3. Nadeau, J. H. & Taylor, B. A. Lengths of chromosomal segments conserved since divergence of man and mouse. *Proc. Natl. Acad. Sci. USA* **81,** 814–818 (1984).
4. O'Brien, S. J., Wienberg, J. & Lyons, L. A. Comparative genomics: lessons from cats. *Trends Genet.* **13,** 393–399 (1997).
5. Waddington, D., Springbett, A. J. & Burt, D. W. A chromosome based model to estimate the number of conserved segments between pairs of species from comparative genetic maps. *Genetics* (in the press).
6. Morizot, D. C. Reconstructing the gene map of the vertebrate ancestor. *Anim. Biotech.* **5,** 113–122 (1994).
7. Sarich, V. M. & Wilson, A. C. Immunological time scale for hominid evolution. *Science* **158,** 1200–1203 (1994).
8. Novacek, M. J. Mammalian phylogeny: shaking the tree. *Nature* **356,** 121–125 (1992).
9. de Jong, W. W. Molecules remodel the mammalian tree. *Trends Ecol. Evol.* **13,** 270–275 (1998).
10. Kumar, S. & Hedges, B. A molecular time-scale for vertebrate evolution. *Nature* **392,** 917–920 (1998).
11. Watanabe, T. K. *et al.* A radiation hybrid map of the rat genome containing 5,255 markers. *Nature Genet.* **22,** 27–36 (1999).
12. Bush, G. L., Case, S. M., Wilson, A. C. & Patton, J. L. Rapid speciation and chromosomal evolution in mammals. *Proc. Natl. Acad. Sci. USA* **74,** 3942–3946 (1977).
13. O'Neill, R. J. W., O'Neill, M. J. & Graves, J. A. M. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* **393,** 68–72 (1998).
14. Li, W.-H., Ellsworth, D. L., Krushkal, J., Chang, B. H.-J. & Hewett-Emmett, D. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Mol. Phylogenet. Evol.* **5,** 182–187 (1996).

15. Ohta, T. An examination of the generation-time effect on molecular evolution. *Proc. Natl. Acad. Sci. USA* **90,** 10676–10680 (1993).
16. Vogel, F., Kopun, M. & Rathenberg, R. in *Molecular Anthropology* (eds Goodman, M. & Tashian, R. E.) 13–33 (Plenum, New York, 1976).
17. Burt, D. W. & Cheng, H. H. The chicken gene map. *Inst. Lab. Anim. Res. J.* **39,** 229–236 (1998).
18. Hastie, T. J. & Tibshirani, R. J. *Generalised Additive Models* (Chapman and Hall, London, 1990).
19. Hurvich, C. M., Simonoff, J. S. & Tsai, C.-L. Smoothing parameter selection in non-parametric regression using an improved Akaike information criterion. *J. R. Statist. Soc.* **60,** 271–293 (1998).
20. Kosambi, D. D. The estimation of map distance from recombination values. *Ann. Eugen.* **12,** 172–175 (1944).
21. Morton, N. E. Parameters of the human genome. *Proc. Natl. Acad. Sci. USA* **88,** 7474–7476 (1991).
22. Wienberg, J. & Stanyon, R. Comparative chromosome painting of primate genomes. *Inst. Lab. Anim. Res. J.* **39,** 77–91 (1998).

Supplementary information is available on *Nature*'s World-Wide Web site (http://www.nature.com) or as paper copy from the London Editorial office of *Nature*.

Acknowledgements

We thank members of the EC ChickMAP project and H. Cheng for providing data prior to publication, and J. Burt for helpful comments. Genome research at the Roslin Institute is supported by the Ministry of Agriculture, Fisheries and Food, the Biotechnology and Biological Sciences Research Council and the Commission of the European Communities.

Correspondence and requests for materials should be addressed to D.W.B. (e-mail: Dave-Burt@bbsrc.ac.uk).

Large-scale analysis of the yeast genome by transposon tagging and gene disruption

Petra Ross-Macdonald*, Paulo S. R. Coelho*, Terry Roemer*, Seema Agarwal*, Anuj Kumar*, Ronald Jansen†, Kei-Hoi Cheung‡, Amy Sheehan*, Dawn Symoniatis*, Lara Umansky*, Matthew Heidtman*, F. Kenneth Nelson*, Hiroshi Iwasaki*, Karl Hager§, Mark Gerstein†, Perry Miller‡, G. Shirleen Roeder* & Michael Snyder*†

** Department of Molecular, Cellular and Developmental Biology, Yale University, PO Box 208103, New Haven, Connecticut 06520-8103, USA*
† Department of Molecular Biophysics and Biochemistry, Yale University, PO Box 208114, New Haven, Connecticut 06520-8114, USA
‡ Center for Medical Informatics, Department of Anesthesiology, Yale University School of Medicine, New Haven, Connecticut 06510, USA
§ Keck Foundation Biotechnology Resource Laboratory, 295 Congress Avenue, Yale University, New Haven, Connecticut 06520, USA

Economical methods by which gene function may be analysed on a genomic scale are relatively scarce. To fill this need, we have developed a transposon-tagging strategy for the genome-wide analysis of disruption phenotypes, gene expression and protein localization, and have applied this method to the large-scale analysis of gene function in the budding yeast *Saccharomyces cerevisiae*. Here we present the largest collection of defined yeast mutants ever generated within a single genetic background—a collection of over 11,000 strains, each carrying a transposon inserted within a region of the genome expressed during vegetative growth and/or sporulation. These insertions affect nearly 2,000 annotated genes, representing about one-third of the 6,200 predicted genes in the yeast genome[1,2]. We have used this collection to determine disruption phenotypes for nearly 8,000 strains using 20 different growth conditions; the resulting data sets were clustered to identify groups of functionally related genes. We have also identified over 300 previously non-annotated open reading frames and analysed by indirect immunofluorescence over 1,300 transposon-tagged proteins. In total, our study encompasses over 260,000 data points, constituting the largest functional analysis of the yeast genome ever undertaken.

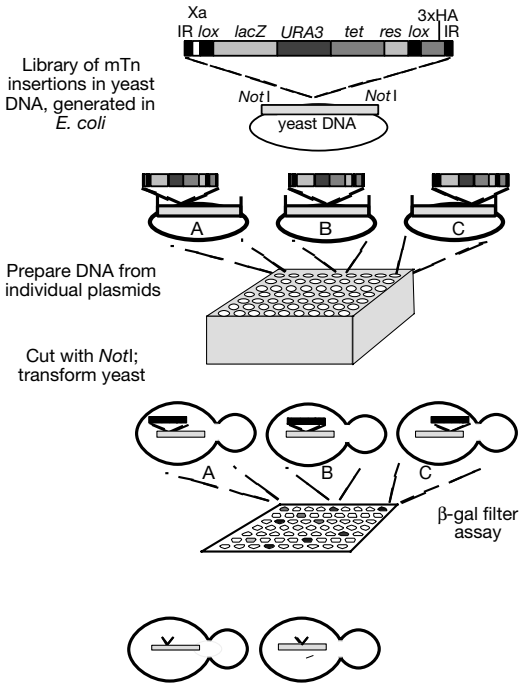

Figure 1 The mTn insertion project. Most steps were performed using a Robbins Hydra 96-channel dispenser; all strains are maintained in a 96-well format.

The ability to sequence entire genomes has resulted in an abundance of raw sequence data[3]; however, relatively few methods exist to assess gene function on a genomic scale[4–8]. We have developed a transposon-based method for the large-scale accumulation of expression, phenotypic and protein localization data in yeast without bias towards previously annotated genes. Our approach utilizes a multipurpose minitransposon (mTn) derived from the bacterial transposable element Tn3 (ref. 9). The minitransposon mTn–3xHA/*lacZ* (Fig. 1) contains a *lacZ* reporter gene lacking an initiator methionine and upstream promoter sequence. Introduction of this transposon into yeast results in production of β-galactosidase (β-gal) if the mTn is present within a transcribed and translated region of the genome, typically corresponding to an in-frame fusion of *lacZ* to the yeast protein-coding sequence. Additionally, mTn–3xHA/*lacZ* contains a *lox* site near each Tn3 end; adjacent to one *lox* site is DNA encoding three copies of a haemagglutinin (3xHA) epitope tag. Production of the Cre recombinase in yeast containing this minitransposon results in recombination of the *lox* sites, reducing the mTn to a 274 base pair (bp) element (the HAT tag)[9]. As five bases of genomic DNA are duplicated during Tn3 insertion, the net result is a 279-bp insertion encoding a 93-codon open reading frame (ORF) which includes the 3xHA sequence. When the mTn's *lacZ* reporter has been fused in-frame to a yeast coding region, creation of the HAT tag allows production of a full-length, epitope-tagged protein from that gene.

We have used mTn–3xHA/*lacZ* in conjunction with high throughput methods to generate a large collection of yeast strains, each containing an insertion at a known location within the genome (Fig. 1). Briefly, a yeast genomic DNA library was mutagenized in *Escherichia coli* with mTn–3xHA/*lacZ*. Using a 96-well format, individual plasmids were prepared, digested with *Not*I and transformed into a diploid yeast strain[10]. By homologous recombination, each fragment should integrate at its corresponding genomic locus, thereby replacing its genomic copy. To verify this process, polymerase chain reaction (PCR) analysis was performed on 48 independent yeast transformants, each carrying one of six different mTn insertion alleles. Using a primer in the mTn and a primer external to