

# The genome of the green anole lizard and a comparative analysis with birds and mammals

Jessica Alfoldi<sup>1\*</sup>, Federica Di Palma<sup>1\*</sup>, Manfred Grabherr<sup>1</sup>, Christina Williams<sup>2</sup>, Lesheng Kong<sup>3</sup>, Evan Mauceli<sup>1</sup>, Pamela Russell<sup>1</sup>, Craig B. Lowe<sup>4</sup>, Richard E. Glor<sup>5</sup>, Jacob D. Jaffe<sup>1</sup>, David A. Ray<sup>6</sup>, Stephane Boissinot<sup>7</sup>, Andrew M. Shedlock<sup>8</sup>, Christopher Botka<sup>9</sup>, Todd A. Castoe<sup>10</sup>, John K. Colbourne<sup>11</sup>, Matthew K. Fujita<sup>3,12</sup>, Ricardo Godinez Moreno<sup>12</sup>, Boudewijn F. ten Hallers<sup>13</sup>, David Haussler<sup>14</sup>, Andreas Heger<sup>3</sup>, David Heiman<sup>1</sup>, Daniel E. Janes<sup>12</sup>, Jeremy Johnson<sup>1</sup>, Pieter J. de Jong<sup>13</sup>, Maxim Y. Koriabine<sup>13</sup>, Marcia Lara<sup>1</sup>, Peter A. Novick<sup>15</sup>, Chris L. Organ<sup>12</sup>, Sally E. Peach<sup>1</sup>, Steven Poe<sup>16</sup>, David D. Pollock<sup>10,17</sup>, Kevin de Queiroz<sup>18</sup>, Thomas Sanger<sup>12</sup>, Steve Searle<sup>19</sup>, Jeremy D. Smith<sup>6</sup>, Zachary Smith<sup>11</sup>, Ross Swofford<sup>1</sup>, Jason Turner-Maier<sup>1</sup>, Juli Wade<sup>20</sup>, Sarah Young<sup>1</sup>, Amonida Zadissa<sup>19</sup>, Scott V. Edwards<sup>12</sup>, Travis C. Glenn<sup>21</sup>, Christopher J. Schneider<sup>22</sup>, Jonathan B. Losos<sup>12</sup>, Eric S. Lander<sup>1</sup>, Matthew Breen<sup>2,23,24</sup>, Chris P. Ponting<sup>3</sup> & Kerstin Lindblad-Toh<sup>1,25</sup>

**The evolution of the amniotic egg was one of the great evolutionary innovations in the history of life, freeing vertebrates from an obligatory connection to water and thus permitting the conquest of terrestrial environments<sup>1</sup>. Among amniotes, genome sequences are available for mammals and birds<sup>2–4</sup>, but not for non-avian reptiles. Here we report the genome sequence of the North American green anole lizard, *Anolis carolinensis*. We find that *A. carolinensis* microchromosomes are highly syntenic with chicken microchromosomes, yet do not exhibit the high GC and low repeat content that are characteristic of avian microchromosomes<sup>2</sup>. Also, *A. carolinensis* mobile elements are very young and diverse—more so than in any other sequenced amniote genome. The GC content of this lizard genome is also unusual in its homogeneity, unlike the regionally variable GC content found in mammals and birds<sup>5</sup>. We describe and assign sequence to the previously unknown *A. carolinensis* X chromosome. Comparative gene analysis shows that amniote egg proteins have evolved significantly more rapidly than other proteins. An anole phylogeny resolves basal branches to illuminate the history of their repeated adaptive radiations.**

The amniote lineage divided into the ancestral lineages of mammals and reptiles ~320 million years ago. Today, the surviving members of those lineages are mammals, comprising ~4,500 species, and reptiles, containing ~17,000 species. Within the reptiles, the two major clades diverged ~280 million years ago: the lepidosaurs, which contains lizards (including snakes) and the tuatara; and the archosaurs, containing crocodylians and birds (the position of turtles remains unclear)<sup>6</sup>. For simplicity, we will refer here to lepidosaurs as lizards (Fig. 1).

The study of the major genomic events that accompanied the transition to a fully terrestrial life cycle has been assisted by the sequencing of several mammal (K.L.-T. *et al.*, manuscript submitted) and three bird genomes<sup>2–4</sup>. The genome of the lizard *A. carolinensis* thus fills an important gap in the coverage of amniotes, splitting the long branch between mammals and birds and allowing more robust evolutionary analysis of amniote genomes.

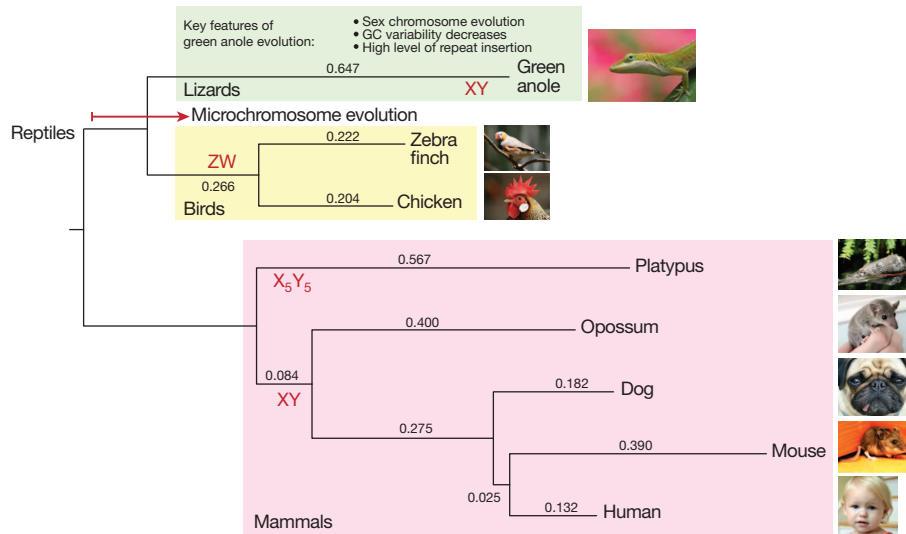
For instance, almost all reptilian genomes contain microchromosomes, but these have only been studied at a sequence level in birds<sup>2,7</sup>, raising the question as to whether the avian microchromosomes' peculiar sequence features are universal across reptilian microchromosomes<sup>8</sup>. Another example is the study of sex chromosome evolution. Nearly all placental and marsupial mammals share homologous sex chromosomes (XY)<sup>9</sup> and all birds share ZW sex chromosomes. However, lizards exhibit either genetic or temperature-dependent sex determination<sup>10</sup>. Characterization of lizard sex chromosomes would allow the study of previously unknown sex chromosomes and comparison of independent sex chromosome systems in closely related species.

*Anolis* lizards comprise a diverse clade of ~400 described species distributed throughout the Neotropics. These lizards have radiated, often convergently, into a variety of ecological niches with attendant morphological adaptations, providing one of the best examples of adaptive radiation. In particular, their diversification into multiple replicate niches on diverse Caribbean islands via interspecific competition and natural selection has been documented in detail<sup>11</sup>. *A. carolinensis* is the only anole native to the USA and can be found from Florida and Texas up to North Carolina. We chose this species for genome sequencing because it is widely used as a reptile model for experimental ecology, behaviour, physiology, endocrinology, epizootics and, increasingly, genomics.

The green anole genome was sequenced and assembled (AnoCar 2.0) using DNA from a female *A. carolinensis* lizard (Supplementary Tables 1–4). Fluorescence *in situ* hybridization (FISH) of 405 bacterial artificial chromosome (BAC) clones (from a male) allowed the assembly scaffolds to be anchored to chromosomes (Supplementary Table 5 and Supplementary Fig. 1). The *A. carolinensis* genome has been reported to have a karyotype of  $n = 18$  chromosomes, comprising six pairs of large macrochromosomes and 12 pairs of small microchromosomes<sup>12</sup>. The draft genome sequence is 1.78 Gb in size (see Supplementary Table 3 for assembly statistics) and represents an intermediate

<sup>1</sup>Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. <sup>2</sup>Department of Molecular Biomedical Sciences, College of Veterinary Medicine, North Carolina State University, Raleigh, North Carolina 27606, USA. <sup>3</sup>MRC Functional Genomics Unit, University of Oxford, Department of Physiology, Anatomy and Genetics, Oxford OX1 3QX, UK. <sup>4</sup>Stanford University School of Medicine, Department of Developmental Biology, Stanford, California 94305, USA. <sup>5</sup>University of Rochester, Rochester, New York 14607, USA. <sup>6</sup>Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State, Mississippi 39762, USA. <sup>7</sup>Department of Biology, Queens College, the City University of New York, New York, New York 11367, USA. <sup>8</sup>Biology Department and Graduate Program in Marine Genomics, College of Charleston, Charleston, South Carolina 29424, USA. <sup>9</sup>Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>10</sup>Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, Colorado 80045, USA. <sup>11</sup>The Center for Genomics and Bioinformatics, Indiana University, Bloomington, Indiana 47405, USA. <sup>12</sup>Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA. <sup>13</sup>Children's Hospital Oakland Research Institute, Oakland, California 94609, USA. <sup>14</sup>Center for Biomolecular Science and Engineering, University of California, Santa Cruz, California 95064, USA. <sup>15</sup>Department of Biological Sciences and Geology, Queensborough Community College, Bayside, New York 11364, USA. <sup>16</sup>Department of Biology, University of New Mexico, Albuquerque, New Mexico 87131, USA. <sup>17</sup>Program in Computational Bioscience, University of Colorado School of Medicine, Aurora, Colorado 80045, USA. <sup>18</sup>Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, District of Columbia 20560, USA. <sup>19</sup>Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. <sup>20</sup>Departments of Psychology and Zoology, Program in Neuroscience, Michigan State University, East Lansing, Michigan 48824, USA. <sup>21</sup>Department of Environmental Health Science and Georgia Genomics Facility, University of Georgia, Athens, Georgia 30602, USA. <sup>22</sup>Biology Department, Boston University, Boston, Massachusetts 02215, USA. <sup>23</sup>Center for Comparative Medicine and Translational Research, North Carolina State University, Raleigh, North Carolina 27695, USA. <sup>24</sup>University of North Carolina Lineberger Comprehensive Cancer Center, Chapel Hill, North Carolina 27514, USA. <sup>25</sup>Science for Life Laboratory Uppsala, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala 751 23, Sweden.

\*These authors contributed equally to this work.



**Figure 1 | Amniote phylogeny based on protein synonymous sites showing major features of amniote evolution.** Major characteristics of lizard evolution including homogenization of GC content, high sex chromosome turnover and

between genome assemblies of birds (0.9–1.3 Gb) and mammals (2.0–3.6 Gb).

We find that few chromosomal rearrangements occurred in the 280 million years since anole and chicken diverged, as had been hinted at by previous comparisons using *Xenopus* and chicken<sup>13</sup>. There are 259 syntenic blocks (defined as consecutive syntenic anchors that are consistent in order, orientation and spacing, at a resolution of 1 Mb) between lizard and chicken (Supplementary Table 6 and Supplementary Fig. 2). Interestingly, 19 out of 22 anchored chicken chromosomes are each syntenic to a single *A. carolinensis* chromosome over their entire lengths (Fig. 2a); by contrast, only 6 (of 23) human chromosomes are syntenic to a single opossum chromosome over their entire lengths, even though the species diverged only 148 million years ago<sup>14</sup>. Segmental duplications follow trends seen in other amniote genomes (Supplementary Note, Supplementary Table 7 and Supplementary Fig. 3).

Approximately 30% of the *A. carolinensis* genome is composed of mobile elements, which comprise a much wider variety of active repeat families than is seen for either bird<sup>2</sup> or mammalian<sup>15</sup> genomes. The most active classes are long interspersed (LINE) elements (27%) and short interspersed (SINE) elements (16%)<sup>16</sup> (Supplementary Table 8). The majority of LINE repeats belong to five groups (L1, L2, CR1, RTE and R4) and seem to be recent insertions based on their sequence similarity (divergence ranges from 0.00–0.76%; ref. 17). This contrasts with observations of mammalian genomes, where only a single family of LINEs—L1—has predominated over tens of millions of years. The DNA transposons comprise at least 68 families belonging to five superfamilies: hAT, Chapaev, Maverick, Tc/Mariner and Helitron<sup>18</sup>. As with retrotransposons, the majority of DNA transposon families seem to be relatively young in contrast to the extremely few recently active DNA transposons found in other amniote genomes (Supplementary Table 9). Overall, *A. carolinensis* mobile elements feature significantly higher GC content (43.5%,  $P < 10^{-20}$ ) than the genome-wide average of 40.3%. In addition to mobile elements, *A. carolinensis* exhibits a high density (3.5%) of tandem repeats, with length and frequency distributions similar to those of human microsatellite DNA<sup>15</sup>. We now know that amniote genomes come in at least three types: mammalian genomes are enriched for L1 elements and have a high degree of mobile element accumulation, bird genomes are repeat poor with very little mobile element activity, while the lizard genome contains an extremely wide diversity of active mobile element families but has a low rate of accumulation, which is reminiscent of the mobile element profile of teleostean fishes<sup>19</sup>.

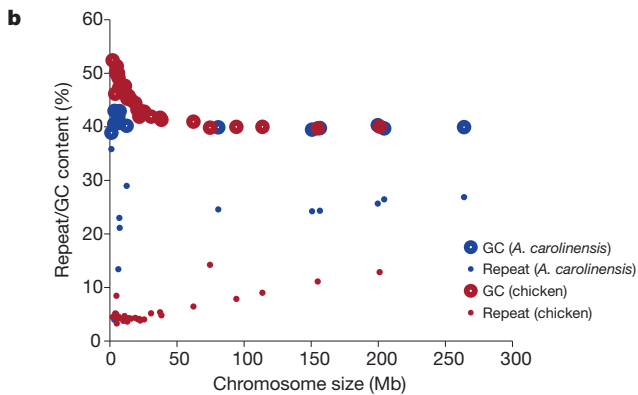
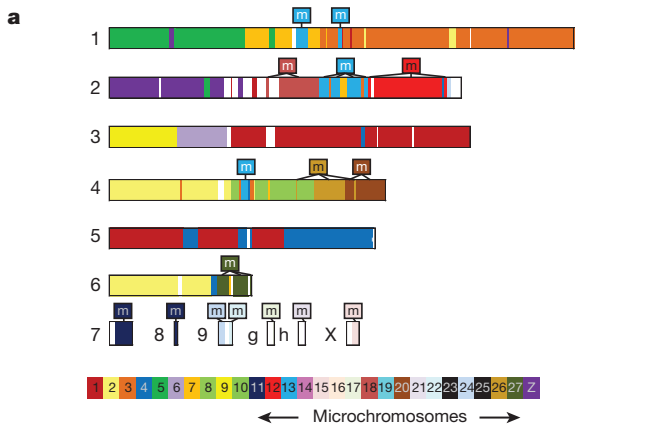
high levels of repeat insertion are featured. Sex chromosome inventions are indicated in red. Branch length is proportional to dS (the synonymous substitution rate); dS of each branch is indicated above the line.

Most reptile genomes contain microchromosomes, but the numbers vary among species; the *A. carolinensis* genome contains 12 pairs of microchromosomes<sup>12</sup>, whereas the chicken genome contains 28 pairs. Bird microchromosomes have very distinctive properties compared to bird macrochromosomes, such as higher GC and lower repeat contents<sup>2</sup>, whereas lizard microchromosomes do not exhibit these features (Fig. 2b). Remarkably, all sequence anchored to microchromosomes in *A. carolinensis* also aligns to microchromosomes in the chicken genome, and all but one *A. carolinensis* microchromosome is syntenic to only a single corresponding chicken microchromosome (Fig. 2a). Microchromosomes conserved between *A. carolinensis* and chicken thus could have arisen in the reptile ancestor, whereas the remaining chicken microchromosomes could be derived in the bird lineage. Alternatively, the remaining chicken microchromosomes could have been present in the reptile ancestor but fused to form macrochromosomes in the lizard lineage.

The *A. carolinensis* genome has surprisingly little regional variation of GC content, substantially less than previously observed for birds and mammals; it is the only amniotic genome known whose nucleotide composition is as homogenous as the frog genome<sup>5</sup> (Supplementary Figs 4 and 5). Figure 3 illustrates how local GC content is evolutionarily conserved between human chromosome 14 and chicken chromosome 5, but to a much lesser degree with *A. carolinensis* chromosome 1. As all sequenced amniote genomes other than *A. carolinensis* contain these homologous varying levels of GC content ('isochores')<sup>20</sup>, the ancestral amniote GC heterogeneity is likely to have eroded towards homogeneity in this lizard's lineage. It has been proposed that isochores with high GC content are a consequence of higher rates of GC-biased gene conversion in regions of higher recombination<sup>2</sup>. The greater GC homogeneity in the anole genome may thus reflect more uniform recombination rates, or else a substantially reduced bias towards GC during the resolution of gene conversion events in the *A. carolinensis* lineage (for a discussion, see ref. 5).

Both temperature-dependent sex determination and XY genetic sex determination have been found in Iguania<sup>10</sup>. Within the genus *Anolis*, there are species with heteromorphic XY chromosomes (including those with multiple X and Y chromosomes), and others with entirely homomorphic chromosomes<sup>12</sup>. *A. carolinensis* is known to have genetic sex determination<sup>21</sup>, but the form of its sex chromosomes (ZW or XY) has thus far been unknown owing to a lack of obviously heteromorphic chromosomes.

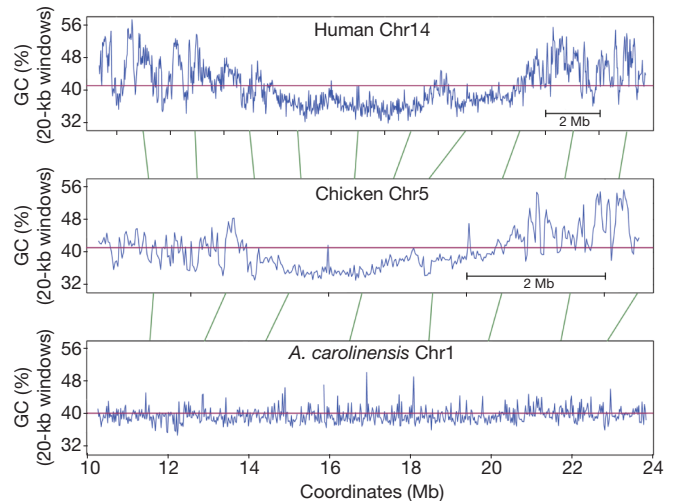
In depth examination of male and female cells using FISH allowed us to identify the microchromosome previously designated as 'b' as the



**Figure 2** | *A. carolinensis*–chicken synteny map reveals synteny of reptile microchromosomes but dissimilar GC and repeat content. **a**, Very few rearrangements have occurred in the 280 million years since *A. carolinensis* and chicken diverged. *A. carolinensis* microchromosomes are exclusively syntenic to chicken microchromosomes. Horizontal coloured bars depict the six *A. carolinensis* macrochromosomes (1–6) and the six (of 12) *A. carolinensis* microchromosomes that have sequence anchored to them that is syntenic to the chicken genome (7, 8, 9, X, LGg, LGh). Chromosomes that could be ordered by size were assigned a number; the smaller microchromosomes that could not be distinguished by size were assigned a lowercase letter. Each colour corresponds to a different chicken chromosome as indicated in the key. Any part of an *A. carolinensis* chromosome that is syntenic to a chicken microchromosome is indicated by ‘m’. **b**, Chicken microchromosomes have both higher GC content and lower repeat content than chicken macrochromosomes, whereas *A. carolinensis* chromosomes do not vary in GC or repeat content by chromosome size. Large circles designate the GC percentage of each chromosome in the chicken and lizard genomes with greater than 100 kb of sequence anchored to it. Small circles designate the percentage of the genome made up of repetitive sequence of each chromosome in the chicken (blue circles) and lizard (red circles) genomes.

*A. carolinensis* X chromosome; it is present in two copies in females and one in males. This chromosome is syntenic to chicken microchromosome 15. Eleven BACs assigned to two scaffolds, 154 (3.3 Mb) and chrUn0090 (1.8 Mb), hybridize via FISH to the p arms of the two X chromosomes in females, and hybridize to the p arm of the single X chromosome in males (Fig. 4 and Supplementary Fig. 1). *A. carolinensis* thereby shows a pattern representative of a male heterogametic system of genotypic sex determination. We have not identified the Y chromosome, but we hypothesize that *A. carolinensis* possesses both X and Y chromosomes, as both male and female cells contain the same number of chromosomes.

The 5.1 Mb of sequence assigned to the X chromosome contains 62 protein-coding genes (Supplementary Table 10); Gene Ontology (GO) terms associated with these genes show no significant enrichment. It is very likely that there is more X chromosome sequence that is currently labelled as unanchored scaffolds in the AnoCar 2.0 assembly. Identification of the *A. carolinensis* sex determination gene will require

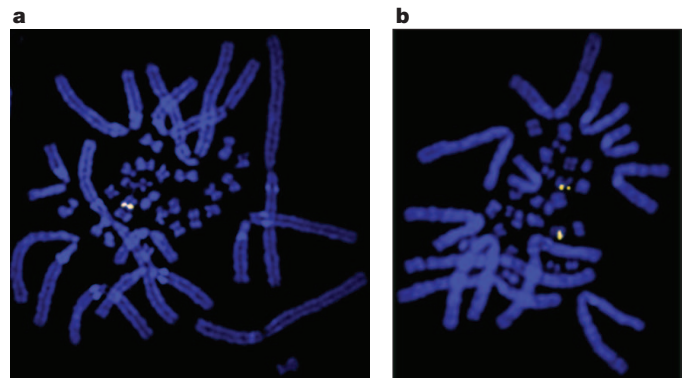


**Figure 3** | The *A. carolinensis* genome lacks isochores. The *A. carolinensis* genome shows only very local variation in GC content, unlike the human and chicken genomes, which also show larger trends in GC variation, sometimes called isochores. Syntenic regions of human chromosome 14, chicken chromosome 5 and *A. carolinensis* chromosome 1 are shown. The human and chicken regions are inverted and rearranged to align with the *A. carolinensis* region. Blue lines depict GC percentage in 20-kb windows. The purple line designates the genome average. Green lines represent examples of syntenic anchors between the three genomes.

considerable functional biology, but we note that the chicken sex determination gene *DMRT1* is located on *A. carolinensis* chromosome 2 and that *SOX3* (the X chromosome paralogue of the therian mammal sex determination gene *SRY*) is located on an unanchored *A. carolinensis* scaffold; these genes are thus unlikely to be the *A. carolinensis* sex determination gene.

All ten *A. carolinensis* individuals (originating from South Carolina and Tennessee) used for FISH mapping showed large pericentromeric inversions in one or more of chromosomes 1–4, with no correlation between different chromosomal inversions or with the sex of the lizard (see Supplementary Note, Supplementary Table 11 and Supplementary Fig. 6).

A total of 17,472 protein-coding genes and 2,924 RNA genes were predicted from the *A. carolinensis* genome assembly (Ensembl release 56, September 2009). We built a phylogeny for all *A. carolinensis* genes and their homologues in eight other vertebrate species (human,



**Figure 4** | The *A. carolinensis* genome contains a newly discovered X chromosome. **a**, **b**, The X chromosome, a microchromosome, is found in one copy in male *A. carolinensis* (**a**) and in two copies in females (**b**). The BAC 206M13 (CHORI-318 BAC library) is hybridized to the p arm of the X chromosome using FISH in both male and female metaphase spreads. 206M13 and ten other BACs showed this sex-specific pattern in cells derived from five male and five female individuals. Original magnification,  $\times 1,000$ .

mouse, dog, opossum, platypus, chicken, zebra finch and pufferfish), allowing us to identify a conservative set of 3,994 one-to-one orthologues, that is, genes that have not been duplicated or deleted in any of these vertebrates since their last common ancestor. These gene phylogenies were also used to identify genes that arose by duplication in the lizard lineage after the split with the avian lineage and, separately, those that were lost in the mammalian lineage after the mammal–reptile split (Fig. 1, Supplementary Note, Supplementary Fig. 7 and Supplementary Table 12).

We found 11 *A. carolinensis* opsin genes that have no mammalian orthologues (but have orthologues in invertebrates, fishes and frog), and thus seem to have been lost during mammalian evolution (Supplementary Table 13). The large repertoire of opsins may contribute to the excellent colour vision of anoles—including the ability to see in the ultraviolet range—and also may contribute to their hyperdiversity by allowing the evolution of diverse, species-specific colouration of the dewlap, which has an important role in sexual selection and species recognition<sup>11</sup>. Similarly, olfactory receptor and  $\beta$ -keratin genes are highly duplicated in *A. carolinensis* (Supplementary Note and Supplementary Fig. 9).

Many reptiles, including green anoles, differ from placental mammals in being oviparous (laying eggs). Vivipary in placental mammals is a derived state, reflected in their loss of some egg-related genes. We used mass spectrometry to identify proteins present in the immature *A. carolinensis* egg, as most egg proteins are produced in the mother's body and then transported into the immature egg. We found that in contrast with mammals, reptiles have lineage-specific gene duplications, including in vitellogenins (VTGs), apovitellenin-1, ovomucin- $\alpha$  and three homologues of ovocalyxin-36, a chicken eggshell matrix protein.

Our results show rapid evolution of egg protein genes among amniotes. Specifically, we found proteins from 276 *A. carolinensis* genes in immature *A. carolinensis* eggs (Supplementary Tables 14 and 15), of which only 50 have been confirmed to be present in chicken eggs by mass spectrometry<sup>22,23</sup>. These genes include VTGs, a lysozyme, vitelline membrane outer layer protein 1 (*VMO1*) paralogues, protease inhibitors, natterin and nothepsin. By aligning genes that are one-to-one orthologues in *A. carolinensis* and chicken, we found that egg proteins evolve significantly more rapidly than non-egg proteins (mean dN/dS values (ratio of the rate of non-synonymous substitutions to the rate of synonymous substitutions) of 0.186 and 0.135, respectively;  $P = 1.2 \times 10^{-5}$ ), which reflects reduced purifying selection and/or more frequent episodes of adaptive evolution.

Using multiple vertebrate genome sequences, we identified three *VMO1* paralogues (which we name  $\alpha$ ,  $\beta$  and  $\gamma$ ) that we infer to have been present in the last common ancestor of all reptiles and mammals. Whereas at least one of *VMO1*- $\alpha$ , *VMO1*- $\beta$  and *VMO1*- $\gamma$  has been lost in all other amniote genomes, the *A. carolinensis* genome contains representatives of all three paralogues. Moreover, the *A. carolinensis*-specific *VMO1*- $\alpha$  family has grown to 13 members and has experienced positive selection of amino acid substitutions within a negatively charged, probably substrate-binding cavity; changes that, presumably, modify its lysozyme-like transferase activity (Supplementary Note, Supplementary Fig. 8 and Supplementary Tables 16 and 17).

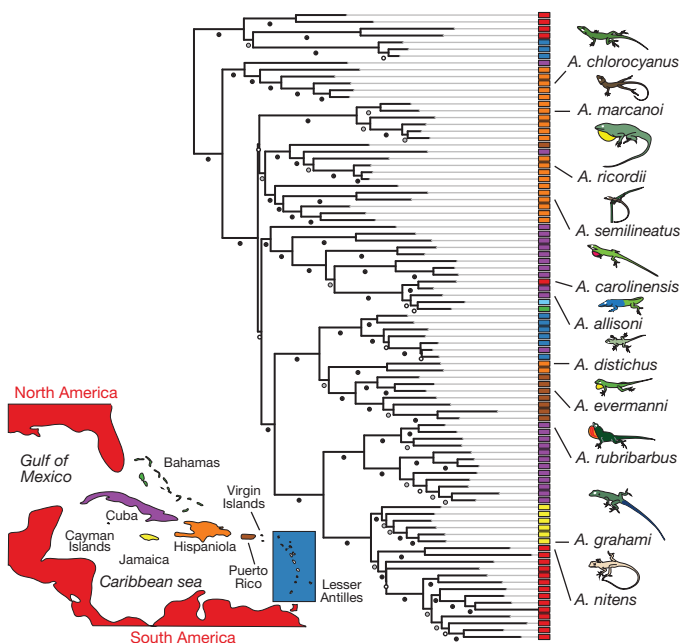
The extensive and active repeat repertoire of *A. carolinensis* has allowed us to discover the origin of several mammalian conserved elements. Through the process of exaptation (a major change in function of a sequence during evolution), certain mobile elements that were active in the amniote ancestor have become conserved, and presumably functional, in mammals, while remaining active mobile elements in *A. carolinensis*. The origin of these conserved mammalian sequences in mobile elements was not recognizable without comparison to a distant and repeat-rich genome sequence<sup>24</sup>. We identified 96 such exapted elements (see Supplementary Table 18) in the human genome tracing back to mobile elements present in the amniote ancestor that are still present in *A. carolinensis*, particularly the CR1, L2 and gypsy families.

Although most exapted elements are non-coding and probably serve a regulatory function, we also identified a protein-coding exon that was exapted from an L2-like LINE, now constituting exon 2 in a mammal-specific N-terminal region of the *MIER1* (mesoderm induction early response 1) protein. This exon is highly conserved across 29 mammals and therefore probably represents a mammalian innovation since the amniote ancestor.

GO terms associated with the transcription start site closest to each exapted element in the human genome show enrichment for neurodevelopmental genes (see Methods), with “ephrin receptor binding”, “nervous system development” and “synaptic transmission” being strongly enriched (all  $P$  values  $< 5 \times 10^{-3}$ ). These enrichments are consistent with adaptive changes in neurodevelopment occurring during the emergence of mammals.

*Anolis* lizards are a textbook case of adaptive radiation, having diversified independently on each island in the Greater Antilles and throughout the Neotropics, producing a wide variety of ecologically and morphologically differentiated species, with as many as 15 found at a single locality<sup>11</sup>. Although anoles are widely used as a model system for phylogenetic comparative studies, it has been difficult to determine the evolutionary relationships among major anole clades owing to rapid evolutionary radiations associated with access to new dimensions of ecological opportunity. Successfully resolving the relatively short branching events associated with such a radiation requires a wealth of data from loci evolving at an appropriate rate.

We used the genome sequence of *A. carolinensis* to develop a new phylogenomic data set comprised of 20 kb of sequence data sampled from across the genomes of 93 species of anoles (Supplementary Tables 19 and 20). Analyses of this data set infer a well-supported phylogeny that reinforces and clarifies the adaptive and biogeographic history of anoles (Fig. 5, details in Supplementary Fig. 10). First, our phylogenomic analysis reaffirms previous molecular and morphological studies indicating that similar anole habitat specialists have



**Figure 5 | A phylogeny of 93 *Anolis* species clarifies the biogeographic history of anoles.** *Anolis* ecomorphs derive from convergent evolution and not from frequent inter-island migration. Using conserved primer pairs distributed across the genome of *A. carolinensis*, we obtain sequences from 46 genomically diverse loci evolving at a range of evolutionary rates and representing both protein-coding and non-coding regions. Maximum likelihood analyses of this new data set of 20 kb aligned nucleotides infer nearly all previously established anole relationships while also partially resolving the basal relationships that have plagued previous studies. Open circles indicate bootstrap (bs) values  $< 70$ ; grey-shaded circles,  $70 < bs < 95$ ; filled circles,  $bs > 95$ .

evolved independently on each of the four large Greater Antillean islands. Second, our analyses suggest a complex biogeographic scenario involving a limited number of dispersal events between islands and extensive *in situ* diversification within islands. The closest relatives of *Anolis* occur on the mainland and the phylogeny confirms the existence of two colonizations, one into the southern Lesser Antilles and the second producing the diverse adaptive radiations throughout the rest of the Caribbean. Within this latter clade, anoles initially diversified primarily on the two larger Greater Antillean islands (although Puerto Rico also seems to have been involved) before subsequently undergoing secondary radiations on all of the islands and eventually returning to the mainland, where this back-colonization has produced an extensive evolutionary radiation. The phylogeny also indicates that very few inter-island dispersal events occurred in Greater Antillean evolution. Rather, the Greater Antillean faunas, renowned for the extent to which the same ecomorphs are found on each island, are primarily the result of convergent evolution<sup>25</sup>.

The genome sequence of *A. carolinensis* allows a deeper understanding of amniote evolution. Filling this important reptilian node with a sequenced genome has revealed derived states in each major amniote branch and has helped to illuminate the amniote ancestor. However, the tree of sequenced reptilian genomes is still extremely sparse, and the sequencing of additional non-avian reptiles would be necessary to fully understand how typical *A. carolinensis* and the sequenced bird genomes are of the entire reptile clade.

In addition to the utility of the *A. carolinensis* genome sequence as a representative of non-avian reptiles, *Anolis* species are a unique resource for the study of adaptive radiation and convergent evolution. With their invasions of and subsequent radiations on Caribbean islands, anoles provide a terrestrial analogue to stickleback and cichlid fish, which underwent adaptive evolution in separate aquatic environments. Just as genomic research in sticklebacks has deepened the study of aquatic ecological speciation, a large-scale genomic phylogenetic survey of the Caribbean anoles would be an opportunity for detailed study of adaptive evolution in a land animal<sup>26</sup>; in particular because anole genomes contain large numbers of active mobile elements that we speculate could form substrates for exaptation of novel regulatory elements.

## METHODS SUMMARY

A full description of methods, including sample collection, sequencing, assembly, anchoring, mass spectrometry and all sequence analysis, can be found in Supplementary Information. All animal experiments were approved by the MIT Committee for Animal Care.

Received 21 April; accepted 27 July 2011.

Published online 31 August 2011.

- Sumida, S. S. & Martin, K. L. M. *Amniote Origins: Completing the Transition to Land* (Academic Press, 1997).
- Hillier, L. W. *et al.* Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
- Warren, W. C. *et al.* The genome of a songbird. *Nature* **464**, 757–762 (2010).
- Dalloul, R. A. *et al.* Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS Biol.* **8**, e1000475 (2010).
- Fujita, M. K., Edwards, S. V. & Ponting, C. P. Unexpected homogeneity in GC content and divergent isochore structure in the *Anolis* genome. *Genome Biol. Evol.* doi:10.1093/gbe/evr072 (in the press).
- Benton, M. J. & Donoghue, P. C. Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* **24**, 26–53 (2007).
- Axelsson, E., Webster, M. T., Smith, N. G. C., Burt, D. W. & Ellegren, H. Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes. *Genome Res.* **15**, 120–125 (2005).

- Burt, D. W. Origin and evolution of avian microchromosomes. *Cytogenet. Genome Res.* **96**, 97–112 (2002).
- Potrzebowski, L. *et al.* Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. *PLoS Biol.* **6**, e80 (2008).
- Organ, C. L. & Janes, D. E. Evolution of sex chromosomes in Saurapsida. *Integr. Comp. Biol.* **48**, 512–519 (2008).
- Losos, J. B. *Lizards in an Evolutionary Tree: Ecology and Adaptive Radiation of Anoles* (Univ. of California Press, 2009).
- Gorman, G. C. in *Cytotaxonomy and Vertebrate Evolution* (eds Chiarelli, A. B. & Capanna, E.) Ch. 12 349–424 (Academic, 1973).
- Hellsten, U. *et al.* The genome of the Western clawed frog *Xenopus tropicalis*. *Science* **328**, 633–636 (2010).
- Bininda-Emonds, O. R. *et al.* The delayed rise of present-day mammals. *Nature* **446**, 507–512 (2007).
- Shedlock, A. M. *et al.* Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. *Proc. Natl Acad. Sci. USA* **104**, 2767–2772 (2007).
- Kordis, D. Transposable elements in reptilian and avian (sauropsida) genomes. *Cytogenet. Genome Res.* **127**, 94–111 (2009).
- Novick, P. A., Basta, H., Floumanhaft, M., McClure, M. A. & Boissinot, S. The evolutionary dynamics of autonomous non-LTR retrotransposons in the lizard *Anolis carolinensis* shows more similarity to fish than mammals. *Mol. Biol. Evol.* **26**, 1811–1822 (2009).
- Novick, P. A., Smith, J. D., Floumanhaft, M., Ray, D. A. & Boissinot, S. The evolution and diversity of DNA transposons in the genome of the lizard *Anolis carolinensis*. *Genome Biol. Evol.* **3**, 1–14 (2010).
- Furano, A. V., Duvernell, D. D. & Boissinot, S. L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. *Trends Genet.* **20**, 9–14 (2004).
- Duret, L., Eyre-Walker, A. & Galtier, N. A new perspective on isochore evolution. *Gene* **385**, 71–74 (2006).
- Viets, B. E., Ewert, M. A., Talent, L. G. & Nelson, C. E. Sex-determining mechanisms in squamate reptiles. *J. Exp. Zool.* **270**, 45–56 (1994).
- Mann, K. The chicken egg white proteome. *Proteomics* **7**, 3558–3568 (2007).
- Farinazzo, A. *et al.* Chicken egg yolk cytoplasmic proteome, mined via combinatorial peptide ligand libraries. *J. Chromatogr. A* **1216**, 1241–1252 (2009).
- Lowe, C. B., Bejerano, G., Salama, S. R. & Haussler, D. Endangered species hold clues to human evolution. *J. Hered.* **101**, 437–447 (2010).
- Williams, E. E. in *Lizard Ecology: Studies of a Model Organism* (eds Huey, R. B., Pianka, E. R. & Schoener, T. W.) (Harvard Univ. Press, 1983).
- Colosimo, P. F. *et al.* Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* **307**, 1928–1933 (2005).

Supplementary Information is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** Generation of the *A. carolinensis* sequence at the Broad Institute of MIT and Harvard was supported by grants from the National Human Genome Research Institute (NHGRI); S.P. and R.E.G. were supported by NSF grants DEB-0844624 and DEB-0920892. All sequence data was produced by the Genome Sequencing Platform of the Broad Institute. We would also like to thank the David and Lucile Packard Foundation for their early support of anole genomics, R. Andrews for her advice on lizard egg biology, C. Hickman and B. Temple and the Herpetology group at the University of Georgia's Savannah River Ecology Lab for assistance with sample collection, and L. Gaffney and L. Virnoche for assistance with figure and text preparation.

**Author Contributions** J.A., F.D. and K.L.-T. planned and oversaw the project. M.G., D. Heiman and S.Y. assembled the genome. B.F.T.H., M.Y.K. and P.J.d.J. constructed the BAC library. T.C.G. and J.W. provided tissues for sequencing libraries and FISH analysis. M.B., C.W. and D. Heiman anchored the genome. T.A.C. and D.D.P. assembled the mitochondrial genome. J.K.C. and Z.S. constructed the cDNA library. S.S. and A.Z. annotated the genome. L.K., A.H. and C.P.P. performed the gene repertoire analysis. T.S. aided egg protein experimental design. J.D.J. and S.E.P. performed egg mass spectrometry. M.G. performed genome synteny analysis. E.M. performed segmental duplication analysis. C.W. and M.B. discovered the sex chromosomes and the pericentromeric inversions. P.R. performed the microchromosome and GC analysis. M.K.F. and C.P.P. participated in microchromosome and GC data interpretation. D.A.R. constructed the repeat library. D.A.R., S.B., P.A.N., A.M.S., J.D.S. and C.B. performed the repeat analysis. M.G., J.B.L., R.E.G., S.P., K.d.Q. and R.S. participated in phylogeny data collection. R.E.G., S.P., K.d.Q. and R.S. participated in phylogeny analysis. All authors participated in data discussion and interpretation. J.A., F.D., C.B.L., R.G., D.A.R., S.V.E., C.J.S., J.B.L., E.S.L., M.B., C.P.P. and K.L.-T. wrote the paper with input from other authors.

**Author Information** The *A. carolinensis* whole-genome shotgun project has been deposited in GenBank under the project accession AAWZ00000000.2. All phylogeny sequence data can be found at <http://purl.org/phylo/treebase/phyloids/study/TB2:S11713>. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at [www.nature.com/nature](http://www.nature.com/nature). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.A. ([jalfoldi@broadinstitute.org](mailto:jalfoldi@broadinstitute.org)) or K.L.-T. ([kersll@broadinstitute.org](mailto:kersll@broadinstitute.org)).