Three Periods of Regulatory Innovation During Vertebrate Evolution

Craig B. Lowe, *et al.*

Science **333**, 1019 (2011);
DOI: 10.1126/science.1202702

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this infomation is current as of August 18, 2011):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/content/333/6045/1019.full.html

Supporting Online Material can be found at:
http://www.sciencemag.org/content/suppl/2011/08/17/333.6045.1019.DC1.html

This article **cites 48 articles**, 28 of which can be accessed free:
http://www.sciencemag.org/content/333/6045/1019.full.html#ref-list-1

This article has been **cited by** 1 articles hosted by HighWire Press; see:
http://www.sciencemag.org/content/333/6045/1019.full.html#related-urls

This article appears in the following **subject collections:**
Evolution
http://www.sciencemag.org/cgi/collection/evolution

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2011 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.

cantly associated with increased likelihood of an application being scored ($P < .001$). Marginal effects for whites, Asians, and Hispanics are not different from the full sample. However, marginal effects for applications from blacks were significantly different from the full sample ($P < .05$): For blacks only, NIH review committee experience ($P < .001$) and citation count ($P < .01$) were significantly correlated with receiving a priority score. Together, these results suggest that previous research and affiliation do not have the same impact across racial and ethnic applicant groups.

Throughout the education pipeline, blacks are less likely to graduate from high school, attend college and major in biomedical science, and obtain a Ph.D. in biomedical science. Nevertheless, upon entering the biomedical academic career track, black and white faculty members are equally likely to be tenured at institutions that grant doctorates and at Research I institutions. (*3*). Given our previous results, we expected to find that black scientists who made it to the stage of principal investigator would have similar chances of obtaining NIH funding, all other things being equal. We find it troubling that the typical measures of scientific achievement—NIH training, previous grants, publications, and citations—do not translate to the same level of application success across race and ethnic groups. Our models controlled for demographics, education and training, employer characteristics, NIH experience, and research productivity, yet they did not explain why blacks are 10 percentage points less likely to receive R01 funding compared with whites.

Although our models do not fully explain the funding gap, the greatest differences between blacks and whites that we observed were in the effect of previous training and the probability of receiving a priority score. Although more research is needed to discern the basis for the award differences, it is possible that cumulative advantage may be involved (*15*). Small differences in access to research resources and mentoring during training or at the beginning of a career may accumulate to become large between-group differences. This suggests that more analysis on the impact of NIH training may be warranted. In addition, further research into the review process could help to understand why variables that increased the likelihood of an application receiving a priority score for the full sample did not have the same impact for applications from black investigators.

References and Notes

1. D. J. Nelson, *A National Analysis of Minorities in Science and Engineering Faculties at Research Universities* (University of Oklahoma, Norman, OK, 2007); http://chem.ou.edu/~djn/diversity/Faculty_Tables_FY07/FinalReport07.html.
2. National Science Foundation, Division of Science Resources Statistics, *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2009* (NSF 09-305, Arlington, VA, 2009).
3. D. K. Ginther *et al.*, *Diversity in Academic Biomedicine: An Evaluation of Education and Career Outcomes with Implications for Policy* (Social Science Research Network, Rochester, NY, 2009); available online at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1677993.
4. Committee on Science, Engineering, and Public Policy, *Expanding Underrepresented Minority Participation: America's Science and Technology Talent at the Crossroads* (National Academies Press, Washington, DC, 2010).
5. IMPAC II, www.tfgov.com/data/NIH.gov/Pages/era.nih.gov%5Dimpacii%5Dindex.cfm
6. Data sources and methods are described more fully in the supporting material on *Science* Online.
7. Sometimes applicants report different race/ethnicity in IMPAC II and the DRF. In that case, the most frequently reported race/ethnicity was used.
8. We investigated whether attending an undergraduate Historically Black College or University had an impact on R01 award probability compared with the balance of the Ph.D. researcher sample and we found no statistically significant relationship.
9. American Association of Medical Colleges Faculty Report, U.S. Medical School Faculty (AAMC, Washington, DC, 2008); www.aamc.org/data/facultyroster/usmsf08/start.htm.
10. C. R. Bollinger, *Rev. Econ. Stat.* **85**, 578 (2003).
11. D. Lubotsky, M. Wittenberg, *Rev. Econ. Stat.* **88**, 549 (2006).
12. R. K. Merton, *Science* **159**, 56 (1968).
13. Funding data are available from RePORTER at http://report.nih.gov/award/trends/FindOrg.cfm.
14. We combined two categories of citations (the third and fourth quartiles) into a single category (>24 citations / citations above the median) for Fig. 3 for ease of presentation.
15. T. A. DiPrete, G. M. Eirich, *Annu. Rev. Sociol.* **32**, 271 (2006).

Acknowledgments: D.K.G. acknowledges financial support from NSF grant SES-0353703 and NIH grant 1R01AG36820-01. D.K.G. and Discovery Logic, a Thomson Reuters company, acknowledge support from NIH contracts HHSN276200800458U and HHSN276200900100U, made possible through an NIH Evaluation Set-Aside Award to W.T.S. (07-6008 OD OER). We thank C. Bollinger, S. Kahn, and W. McGarvey for helpful discussions. The analytical files used in the studies described here contain personal information from individuals who have submitted applications and in some cases have received awards from NIH. Many of these application records have been matched to records included in the Survey of Earned Doctorates as maintained by the National Science Foundation and to records included in the Faculty Roster maintained by the Association of American Medical Colleges. The information is therefore protected by the Privacy Act of 1974 as amended (5 U.S.C. 552a) and the National Science Foundation Act of 1950 as amended (42 U.S.C. 1873(i). More complete information can be found in the NSF/SRS Restricted-Use Data Procedures Guide available at www.nsf.gov/statistics/license/forms/pdf/srs_license_guide_august_2008.pdf and the NSF Data and Tools Web site at www.nsf.gov/statistics/database.cfm. Researchers interested in access should call W.T.S. at 301-402-2725 to discuss the security clearance requirements necessary for using the data. A de-identified version of the data files will be posted at http://report.nih.gov/investigators_and_trainees/index.aspx and archived by *Science*.

Supporting Online Material

www.sciencemag.org/cgi/content/full/333/6045/1015/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 to S15
References

20 August 2010; accepted 24 June 2011
10.1126/science.1196783

Three Periods of Regulatory Innovation During Vertebrate Evolution

Craig B. Lowe,[1,2,3] Manolis Kellis,[4,5] Adam Siepel,[6] Brian J. Raney,[1] Michele Clamp,[5] Sofie R. Salama,[1,3] David M. Kingsley,[2,3] Kerstin Lindblad-Toh,[5,7] David Haussler[1,3]*

The gain, loss, and modification of gene regulatory elements may underlie a substantial proportion of phenotypic changes on animal lineages. To investigate the gain of regulatory elements throughout vertebrate evolution, we identified genome-wide sets of putative regulatory regions for five vertebrates, including humans. These putative regulatory regions are conserved nonexonic elements (CNEEs), which are evolutionarily conserved yet do not overlap any coding or noncoding mature transcript. We then inferred the branch on which each CNEE came under selective constraint. Our analysis identified three extended periods in the evolution of gene regulatory elements. Early vertebrate evolution was characterized by regulatory gains near transcription factors and developmental genes, but this trend was replaced by innovations near extracellular signaling genes, and then innovations near posttranslational protein modifiers.

The gain, loss, and modification of gene regulatory elements has led to many phenotypic changes during animal evolution, including pigmentation changes in dogs, fish, and flies (*1*–*3*); bristle patterns on flies (*4*); and skeletal differences in fish (*5*, *6*). A recent analysis of published genome-wide association studies also noted a strong enrichment for regulatory regions to be in linkage with trait/disease-associated single nucleotide polymorphisms (*7*). Mutations in regulatory modules can avoid the pleiotropic effects that often result from protein-coding mutations and, hence, can provide an exceptionally flexible source of evolutionary change (*8*).

---

[1]Center for Biomolecular Science and Engineering, University of California, Santa Cruz, CA 95064, USA. [2]Department of Developmental Biology, Stanford University, Stanford, CA 94305, USA. [3]Howard Hughes Medical Institute, University of California, Santa Cruz, CA 95064, USA. [4]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA. [5]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [6]Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853, USA. [7]Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden.

*To whom correspondence should be addressed. E-mail: haussler@soe.ucsc.edu

Computational methods can identify strong candidates for gene regulatory elements by detecting regions of the genome that show evolutionary conservation, yet do not appear in any coding or noncoding mature transcript. Though many noncoding RNAs or noncoding portions of protein-coding transcripts may serve a regulatory purpose, we exclude these regions to focus on cis-regulatory elements that are functional at the DNA level. For this reason, we use conserved nonexonic elements (CNEEs) to describe the set of putative regulatory regions. Strong conservation indicates a sustained purifying selection against mutations, suggesting that the genomic region confers selective advantage and, hence, is functional. This procedure is independent of the tissue type, developmental stage, or environmental circumstance in which these elements become relevant. An experimental survey of CNEEs found that 50% of the 437 CNEEs tested in an in vivo mouse enhancer assay drove reproducible expression patterns during mouse development (*9*). Because only a single time-point during development was tested, many of the CNEEs that did not drive expression in the assay may act as transcriptional enhancers at other developmental stages. CNEEs have also been shown to act as repressors and insulators (*10*–*12*).

Early genome-wide studies of CNEEs not only noted their role in regulating the transcription of nearby genes, but also observed that CNEEs tend to be located near genes acting as developmental regulators, including many

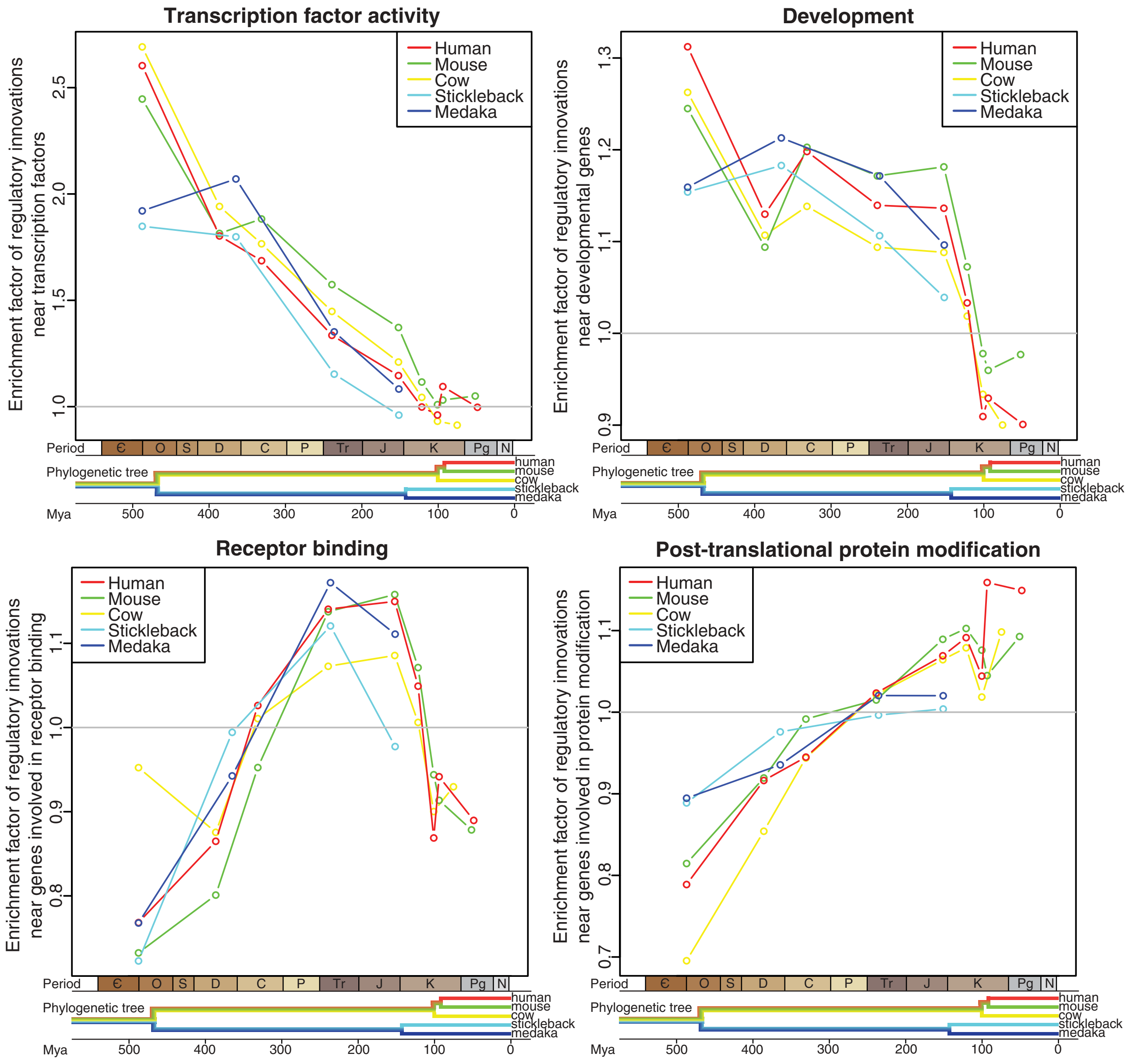


Fig. 1. Regulatory innovation. Each panel shows data for the frequency of regulatory innovations near genes in a different GO category (see panel titles). Colors indicate the five lineages that we studied. Each data point (colored circle) represents the relative frequency of regulatory innovations on a specific lineage as determined by analysis using the reference genome for that lineage. The relative frequency (enrichment factor) for a specific GO category is defined as the frequency of innovations in genes of this GO category as compared to what would be expected by selecting genomic regions at random (denoted by the horizontal line at the relative frequency of 1.0). Each data point is an estimate from at least 2800 putative regulatory innovations. The time associated with each data point, indicated on the *x* axis, is the midpoint of the branch of the phylogenetic tree on which these innovations are inferred to have occurred by comparative genome analysis (fig. S3 and tables S1 and S2). The *x* axis is annotated with both geologic time periods and speciation events for the lineages that we analyzed. Є, Cambrian; O, Ordovician; S, Silurian; D, Devonian; C, Carboniferous; P, Permian; Tr, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene; N, Neogene; Mya, million years ago.

transcription factors, to finely control vertebrate development (*13*–*16*). Only a handful of vertebrate CNEEs date back to our chordate ancestor (*17*), suggesting that the vast majority of regulatory elements in the genomes of living vertebrates have arisen since their common ancestor 650 million years ago (Ma).

To better understand this gain of regulatory sequence on vertebrate lineages, we identified genome-wide sets of CNEEs for the human genome as well as mouse, cow, medaka, and stickleback. Mouse and cow have well-assembled and annotated genomes that can leverage the densely sampled clade of placental mammals to identify and date CNEEs in their genomes. Ray-finned fish are the other clade of vertebrates with enough well-assembled genomes to identify and date CNEEs. Because the ray-finned fish diverged from placental mammals ~425 Ma, their lineage offers a largely independent analysis of regulatory trends in vertebrate evolution.

Using multiple alignments of vertebrate genomes, we employed a phylogenetic hidden Markov model to determine CNEEs (*18*). For each CNEE, we used genomic proximity to predict the gene being regulated. We assigned each CNEE to the gene with the closest transcription start site (*18*) and then inferred the evolutionary branch on which each CNEE came under selective constraint. To do this, for each CNEE we determined the most recent common ancestor of all species in the vertebrate-wide multiple alignment that have an orthologous piece of DNA aligning to at least one-third of the CNEE (fig. S1). For more than 99.97% of CNEEs, we used this most recent common ancestor as the time of origin. In less than 1000 rare cases, there is evidence that a marked onset of constraint occurred after the most recent common ancestor of all species in the alignment (*19*). In these rare cases, we used the branch showing the onset of constraint as the origin of the CNEE (fig. S1). The onset of constraint must be more significant than the rejection of the neutral model by the alignment as a whole (*18*). Finally, we examined the

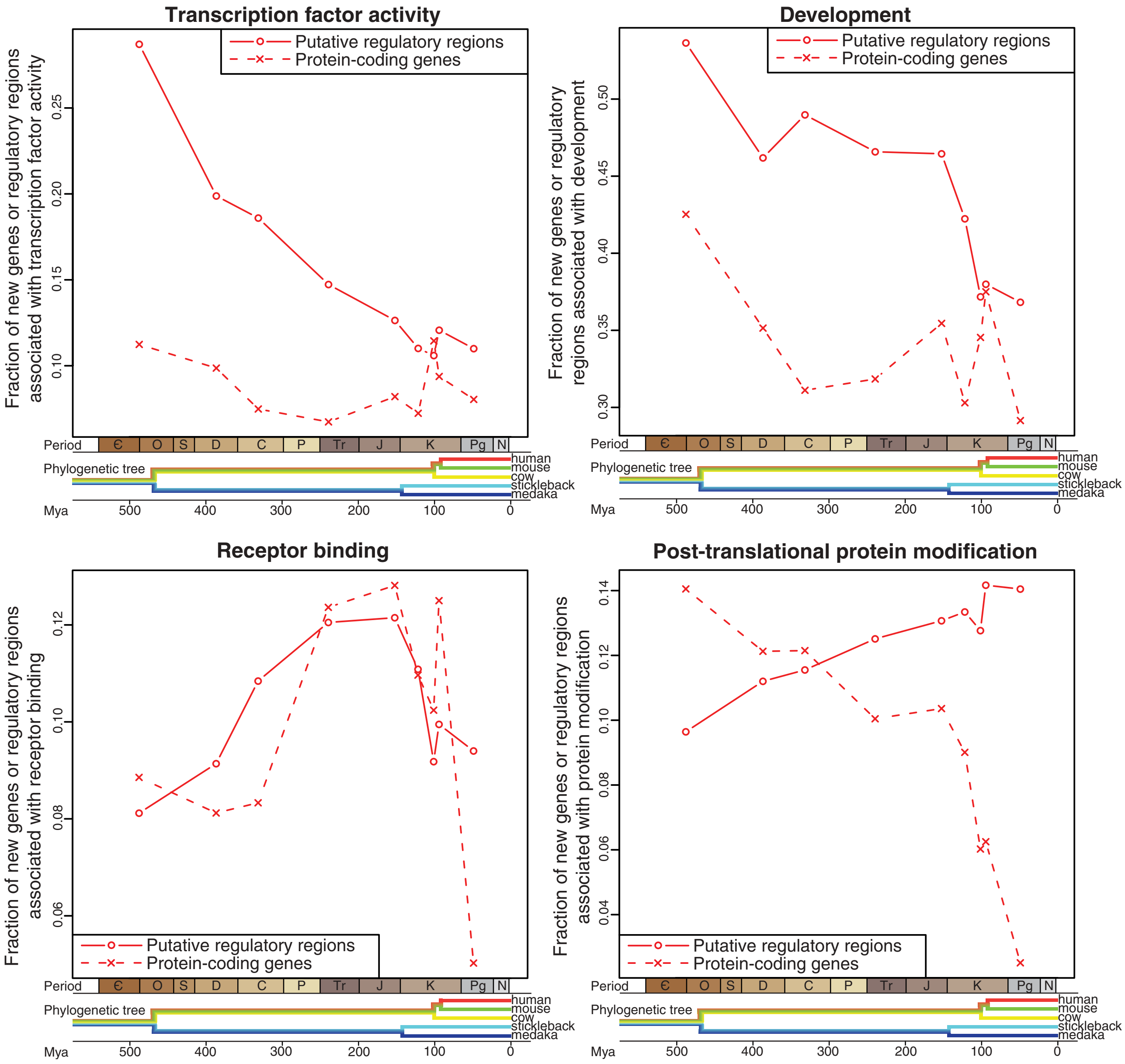


Fig. 2. Comparative history of innovations in regulatory regions and genes. Each panel shows data for the frequency of regulatory innovations near genes in a different GO category, as well as the frequency of genic innovations for that GO category. Each data point for regulatory innovations (red circles) represents the percentage of regulatory innovations appearing on a branch in the human lineage that are associated with a gene annotated with the given GO category. Each data point for protein-coding genes (red crosses) represents the percentage of protein-coding genes appearing on a branch that are associated with the given GO category. The time associated with each data point is as described in Fig. 1.

functional classes of genes that acquired the most regulatory innovations during each evolutionary epoch since the common ancestor of living vertebrates.

For our study of epoch-specific selection in the human lineage, we created alignments of 40 vertebrate genomes (including 31 mammals, 2 birds, 1 lizard, 1 amphibian, and 5 fish) projected onto the human genome (*18*). From this alignment, we found 2,964,909 human CNEEs, averaging 28 bases in length and covering 2.9% of the genome.

To ensure that the CNEEs are enriched for regions that have been under selection, we examined the derived allele frequency spectrum of segregating single nucleotide polymorphisms found by the HapMap Consortium (*20*) in the Yoruban population (*18*). The frequency of derived alleles in the Yoruban population is shifted toward lower frequencies in the set of CNEEs compared with intronic regions ($P < 10^{-90}$), as is characteristic of sites under purifying selection, where the majority of mutations are deleterious and rarely progress to higher frequencies. Although the shift in derived allele frequencies for the set of CNEEs is not as strong as that for nonsynonymous changes in coding regions, these results demonstrate that this set is strongly enriched for regions evolving under purifying selection in humans (fig. S2).

Experimental methods such as chromatin immunoprecipitation sequencing (ChIP-seq) (*21*) and DNase hypersensitivity (*22*), though still limited by the tissue-type and time point of the experiment, offer an alternative to in vivo mouse models that allows the regulatory potential of CNEEs to be assessed on a much larger scale. We examined the overlap of the CNEEs with more than 100,000 DNAase hypersensitivity sites identified in human CD4$^+$ cells, indicating regions of open chromatin (*23*). There is a 1.7-fold enrichment in these DNase hypersensitivity sites compared with a random control for overlap with the set of CNEEs ($P < 10^{-800}$) (*18*). More specific to being a regulatory region, although restricted to a single protein of interest, are locations of protein-DNA interactions discovered by the use of ChIP-seq. We examined the intersection of CNEEs with ChIP-seq data sets for serum response factor (SRF), growth-associated binding protein (GABP), and neuron-restrictive silencing factor (NRSF) in human Jurkat cells (*24*). There is a 2.5-fold enrichment for SRF binding sites ($P < 10^{-142}$), a 3.5-fold enrichment for GABP binding sites ($P < 10^{-210}$), and a 4.7-fold enrichment for NRSF binding sites ($P < 10^{-220}$). This indicates that our set of CNEEs is strongly enriched for functional elements. Our set of CNEEs typically covers only 1 out of every 10 putative regulatory elements identified by these experimental approaches. However, evolutionary conservation will only detect those binding sites under considerable purifying selection in multiple species and not lineage-specific or recently created regulatory elements.

To assess the functions of genes putatively regulated by the CNEEs, we assigned each CNEE to the gene with the closest transcription start site and determined the Gene Ontology (GO) (*25*) terms associated with that gene. We tested for enrichment against the assumption that CNEEs are uniformly distributed throughout the genome, allowing for differences in genic and intergenic sizes among classes of genes (*18*).

Although the set of human CNEEs, when treated as a whole, is enriched in locations where the closest transcription start site is a "trans-dev" gene ("transcription factor activity" $P < 10^{-2000}$, "development" $P < 10^{-3000}$), these results showed that this enrichment is due to the subset of human CNEEs that came under selection before the boreoeutherian (human-cow) ancestor (Fig. 1, fig. S3, and tables S1 and S2). Despite a dramatic enrichment for regulatory innovations near trans-dev genes in the earliest period of vertebrate evolution through the radiation of tetrapods onto land and, to a lesser extent, in our early mammalian ancestors, we found a sharp decrease from that enrichment to a rate expected by chance, or even less, since the ancestor of placental mammals, approximately 100 Ma.

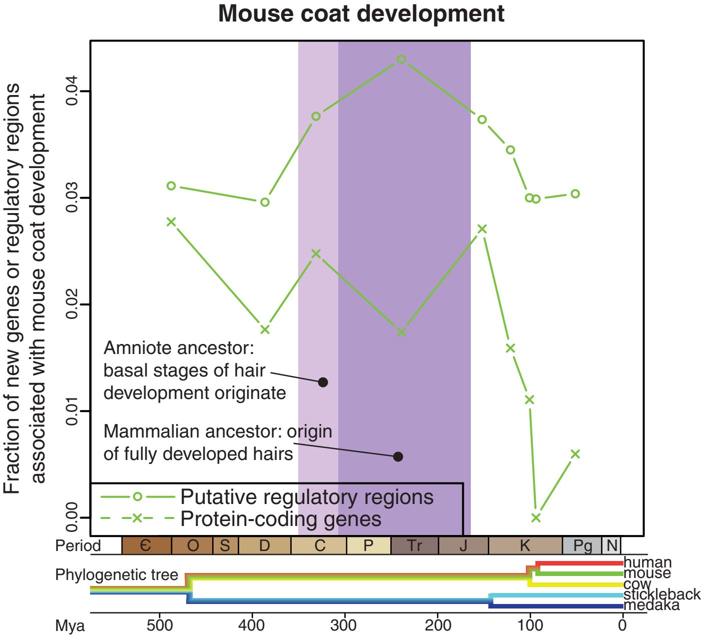
Separating developmental genes and transcription factors shows that developmental genes maintained a moderate enrichment for acquisition of regulatory elements until a sharp decline in our early placental ancestor. Transcription factors were dramatically enriched for acquisition of regulatory elements in our early vertebrate ancestor, but this enrichment has consistently declined until reaching random expectation in our placental ancestor (Fig. 1 and fig. S4).

We repeated the above study using the mouse, cow, medaka, and stickleback species as reference species, each time starting with a new multiple alignment of other vertebrates to the chosen reference species. The analysis of each of these additional lineages gave similar results (Fig. 1).

Creating genome-wide alignments for vertebrate species is still an active area of research. We used alignments to infer the branch on which CNEEs came under selection, so inaccuracies in the alignment may result in inaccurate inferences as to when some CNEEs came under selection. False-positive alignments to distantly related species may cause CNEEs to appear more ancient. False negatives in the alignment process, combined with CNEEs missing from assemblies because of low-coverage sequencing or deletions, may cause CNEEs to appear more recent. However, to explain the trends described, there would need to be a systematic bias that treated the CNEEs associated with genes of one function differently than those associated with genes of other functions.

These findings are not due to biases in length, rate of evolution, or rate of turnover between CNEEs near various functional classes of genes. We have greater statistical power to align both longer- and slower-evolving CNEEs over large evolutionary distances, And we also have greater power to detect evolutionary conservation in longer- and slower-evolving CNEEs. Both of these factors contribute to our sets of very ancient and very recent CNEEs being enriched for longer and more slowly evolving elements. However, there is not a consistent trend for CNEEs associated with trans-dev genes or genes with any particular GO term to have different lengths or rates of evolution (figs. S5 and S6). For this reason, it is unlikely that our results are caused by either of these biases in creating alignments and detecting conservation. To show that the results are not due to different rates of turnover for CNEEs near trans-dev genes, we identified

Fig. 3. Enrichment for regulatory regions originating near genes involved in hair development during the time that hair originated in evolution. For each branch, we plot the percentage of genes or regulatory elements created on that branch that are associated with hair development.


Mouse coat development

CNEEs in the human-, mouse-, and cow-referenced alignments that have clear orthologs in other well-assembled mammalian genomes, but are not present in either the human and rhesus, mouse and rat, or dog and horse genome assemblies (*18*). These CNEEs are likely to have been lost in one of the mammalian lineages after having been present in the ancestor. We counted the number of these lost CNEEs near trans-dev genes versus other types of genes and found no consistent difference (fig. S7). This indicates that the results are unlikely to be attributable to different rates of turnover.

Neither are the results due to a bias in dating CNEEs whose time of origin is uncertain. The same trends seen in the entire set of CNEEs are present in the subset that have a clear point of origin (i.e., that exist precisely in all species descendant from a common ancestor and in no additional species) (fig. S8).

To ensure that the changes we see in enrichments over time are robust against the alignment methods and against choices in what species are included in the analysis, we have performed our analysis on a separate human-referenced alignment using only deeply sequenced and well-assembled genomes along with stringent alignment parameters (fig. S9). The results were similar. To ensure that our results are robust against the choice of gene set, CNEE to gene assignment algorithm, and GO term to gene mapping, we performed our enrichment analysis for the human lineage using a completely independent method (fig. S9) (*26*). Again, the results were similar. Hence, our conclusions are robust to a large number of variations in methodological approach.

Given the changes observed for trans-dev genes, we extended our analysis to each of the ~13,000 GO terms found in vertebrate genomes. We determined the approximate times of all regulatory innovations associated with each GO term and ranked all terms by their increase or decline over time, based on the slope of a linear model fit to time versus percent of CNEEs associated with the GO term. At one end of the spectrum, the top 40 fastest-declining gene categories were development, transcription factor activity, and related GO terms, irrespective of the choice of human, mouse, cow, medaka, or stickleback as the reference genome (tables S3 to S7).

At the other end of the spectrum, we found increases in the accumulation of regulatory innovations for several GO terms, suggesting that the decrease in regulatory innovations near trans-dev genes has been accompanied by an increase in innovations near genes of other functions. These include genes annotated with posttranslational protein modification, organelle membrane, and other GO categories related to intracellular signaling ($P < 10^{-110}$, given the dating of CNEEs and the annotation of genes) (tables S8 to S10).

This set of GO terms does not show a significant enrichment in the medaka or stickleback lineages; however, the fish lineages do not have the dense tree of closely related species that is required to identify recent regulatory innovations. Instead, the terms showing the sharpest increases on both fish lineages are receptor binding, plasma membrane, signaling, and other terms related to extracellular rather than intracellular signaling (tables S11 and S12). Upon analyzing extracellular membrane signaling genes in the mammalian lineages, we found that mammalian ancestors also once had a high rate of regulatory innovations near this same set of genes, but such innovations have now returned to random expectation (Fig. 1). In fact, receptor-binding genes show a peak of regulatory innovation between the time of our amniote ancestor and our placental mammalian ancestor, relative to the rate of innovation before and after this time period ($P < 10^{-250}$, given the dating of CNEEs and the annotation of genes).

For some functional categories, the trend in regulatory innovations appears to be correlated with an increased or decreased appearance of the genes themselves. The proportion of newly appearing genes that are associated with intercellular signaling closely mirrors the proportion of regulatory innovations near genes with this function (Fig. 2). During the time from the amniote ancestor to the placental mammalian ancestor, there was an enrichment for both gain of genes and gain of regulatory elements associated with intercellular signaling. However, for transcription factors, which show the most dramatic change in their proportion of regulatory innovations, the proportion of gene births has stayed relatively constant. This hints that the selective pressures on novel regulatory elements and genes may be similar for some classes of genes and different for others. It could be that evolutionary advances in some cell functions may be realized both through novel genes and regulatory elements, whereas advances in other cell functions may preferentially happen through one of the two methods.

As regulatory innovations near genes involved in posttranslational protein modification have become increasingly common in placental mammals, the appearance of these genes themselves has become increasingly rare. This emphasizes the fact that many regulatory innovations are associated with ancient genes. For example, we have identified 10 regulatory innovations near protein kinase D1 (PRKD1, PKD1) since the human lineage split with cow. This is the highest number of regulatory innovations of any posttranslational protein modification gene during this recent time period, yet PRKD1 dates to at least our tetrapod ancestor. This enrichment for recent CNEE gains near genes involved in posttranslational protein modification is consistent across humans, cows, and, to a lesser degree, mice, even when looking only at independent events after these species diverged (tables S8 to S10).

Finally, to further validate this approach, we looked outside of the GO at a particularly well-studied gene set associated with the evolution of body hair, a phenotypic trait characteristic of mammals. There is a long history of genetic studies concerning mouse coat phenotypes that provides a set of almost 400 genes in the Mouse Genome Database for this process (*27*). Hair development begins as an area of epidermal thickening (epidermal placode) and shares the first developmental steps with avian feathers (*28*). The remaining stages in hair development do not appear until the mammalian ancestor, when body hair first originates. There is a slight enrichment for newly introduced regulatory regions to be associated with hair genes during the time of our amniote ancestor, when the basal stages of hair development originated (Fig. 3). Then the enrichment for gains of putative regulatory regions near hair-associated genes peaks in the mammalian ancestor, at the time that body hair first originates. We do not see a clear enrichment for gains of genes, only for gains of regulatory elements.

It appears that at least three broad periods of regulatory innovation can be reliably detected in several vertebrate lineages. The first period, ranging from our vertebrate ancestor until ~300 Ma, when mammals split with birds and reptiles, is dominated by regulatory innovations near transcription factors and the key developmental genes they control. The second period, from ~300 to ~100 Ma, is characterized by a high frequency of regulatory innovations near receptors of extracellular signals and a gradual decline in innovations near trans-dev genes. These two trends occurred independently in tetrapods and ray-finned fish. Finally, at least in placental mammals, we see a third period in which regulatory innovations for trans-dev and receptor genes have dropped to background frequencies, whereas regulatory innovations for genes involved in posttranslational protein modification, including those in intracellular signaling pathways, are on the rise. Further sequencing of additional vertebrate species will make it possible to determine the pervasiveness of these trends and to look for additional functional categories that may be associated with epochs of evolutionary change in particular lineages.

References and Notes

1. E. K. Karlsson *et al.*, *Nat. Genet.* **39**, 1321 (2007).
2. C. T. Miller *et al.*, *Cell* **131**, 1179 (2007).
3. M. Rebeiz, J. E. Pool, V. A. Kassner, C. F. Aquadro, S. B. Carroll, *Science* **326**, 1663 (2009).
4. A. P. McGregor *et al.*, *Nature* **448**, 587 (2007).
5. P. F. Colosimo *et al.*, *Science* **307**, 1928 (2005).
6. Y. F. Chan *et al.*, *Science* **327**, 302 (2010); 10.1126/science.1182213.
7. L. A. Hindorff *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 9362 (2009).
8. G. A. Wray, *Nat. Rev. Genet.* **8**, 206 (2007).
9. A. Visel *et al.*, *Nat. Genet.* **40**, 158 (2008).
10. T. I. Lee *et al.*, *Cell* **125**, 301 (2006).
11. T. H. Kim *et al.*, *Cell* **128**, 1231 (2007).
12. X. Xie *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7145 (2007).
13. G. Bejerano *et al.*, *Science* **304**, 1321 (2004); 10.1126/science.1098119.
14. K. Lindblad-Toh *et al.*, *Nature* **438**, 803 (2005).

15. A. Woolfe *et al.*, *PLoS Biol.* **3**, e7 (2005).
16. A. Siepel *et al.*, *Genome Res.* **15**, 1034 (2005).
17. L. Z. Holland *et al.*, *Genome Res.* **18**, 1100 (2008).
18. Materials and methods are available as supporting material on *Science* Online.
19. K. S. Pollard, M. J. Hubisz, K. R. Rosenbloom, A. Siepel, *Genome Res.* **20**, 110 (2010).
20. International HapMap Consortium, *Nature* **449**, 851 (2007).
21. G. Robertson *et al.*, *Nat. Methods* **4**, 651 (2007).
22. G. E. Crawford *et al.*, *Genome Res.* **16**, 123 (2006).
23. A. P. Boyle *et al.*, *Cell* **132**, 311 (2008).
24. A. Valouev *et al.*, *Nat. Methods* **5**, 829 (2008).
25. M. Ashburner *et al.*, *Nat. Genet.* **25**, 25 (2000).
26. C. Y. McLean *et al.*, *Nat. Biotechnol.* **28**, 495 (2010).
27. C. J. Bult, *Nucleic Acids Res.* **36**, D724 (2008).
28. P. Wu *et al.*, *Int. J. Dev. Biol.* **48**, 249 (2004).

Acknowledgments: This work was supported by the Howard Hughes Medical Institute (C.B.L., S.R.S., D.M.K., D.H.), the NSF (CAREER-0644282 to M.K., DBI-0644111 to A.S.), the NIH (R01-HG004037 to M.K., P50- HG02568 to D.M.K., U54-HG003067 to K.L-T., 1U01-HG004695 to C.B.L., 5P41-HG002371to B.J.R.), the Sloan Foundation (M.K.), and the European Science Foundation (EURYI to K.L-T.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/333/6045/1019/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 to S12
References (*29–49*)

10 January 2011; accepted 24 June 2011
10.1126/science.1202702

Rapid Range Shifts of Species Associated with High Levels of Climate Warming

I-Ching Chen,[1,2] Jane K. Hill,[1] Ralf Ohlemüller,[3] David B. Roy,[4] Chris D. Thomas[1]*

The distributions of many terrestrial organisms are currently shifting in latitude or elevation in response to changing climate. Using a meta-analysis, we estimated that the distributions of species have recently shifted to higher elevations at a median rate of 11.0 meters per decade, and to higher latitudes at a median rate of 16.9 kilometers per decade. These rates are approximately two and three times faster than previously reported. The distances moved by species are greatest in studies showing the highest levels of warming, with average latitudinal shifts being generally sufficient to track temperature changes. However, individual species vary greatly in their rates of change, suggesting that the range shift of each species depends on multiple internal species traits and external drivers of change. Rapid average shifts derive from a wide diversity of responses by individual species.

Threats to global biodiversity from climate change (*1-8*) make it important to identify the rates at which species have already responded to recent warming. There is strong evidence that species have changed the timing of their life cycles during the year and that this is linked to annual and longer-term variations in temperature (*9–12*). Many species have also shifted their geographic distributions toward higher latitudes and elevations (*13–17*), but this evidence has previously fallen short of demonstrating a direct link between temperature change and range shifts; that is, greater range shifts have not been demonstrated for regions with the highest levels of warming.

We undertook a meta-analysis of available studies of latitudinal (Europe, North America, and Chile) and elevational (Europe, North America, Malaysia, and Marion Island) range shifts for a range of taxonomic groups (*18*) (table S1). We considered $N = 23$ taxonomic group × geographic region combinations for latitude, incorporating 764 individual species responses, and $N = 31$ taxonomic group × region combinations for elevation, representing 1367 species responses. For the purpose of analysis, the mean shift across all species of a given taxonomic group, in a given region, was taken to represent a single value (for example, plants in Switzerland or birds in New York State; table S1) (*18*).

The latitudinal analysis revealed that species have moved away from the Equator at a median rate of 16.9 km decade^{-1} (mean = 17.6 km decade^{-1}, SE = 2.9, $N = 22$ species group × region combinations, one-sample t test versus zero shift, $t = 6.10$, $P < 0.0001$). Weighting each study by the √(number of species) in the group × region combination gave a mean rate of 16.6 km decade^{-1}. For elevation, there was a median shift to higher elevations of 11.0 m uphill decade^{-1} (mean = 12.2 m decade^{-1}, SE = 1.8, $N = 30$ species groups × regions, one-sample t test versus zero shift, $t = 7.04$, $P < 0.0001$). Weighting elevation studies by √(number of species) gave a mean rate of uphill movement of 11.1 m decade^{-1}.

A previous meta-analysis (*14*) of distribution changes analyzed individual species, rather than the averages of taxonomic groups × regions that we used, and also included data on latitudinal and elevational shifts in the same analysis (*18*). It concluded that ranges had shifted toward higher latitudes at 6.1 km decade^{-1} and to higher elevations at 6.1 m decade^{-1} (*14*), whereas the rates of range shift that we found were significantly greater [$N = 22$ species groups × regions, one-sample t test versus 6.1 km decade^{-1}, $t = 3.99$, $P = 0.0007$ for latitude; $N = 30$ groups × regions, one-sample t test versus 6.1 m decade^{-1}, $t = 3.49$, $P = 0.002$ for elevation (*18*)]. Our estimated mean rates are approximately three and two times higher than those in (*14*), for

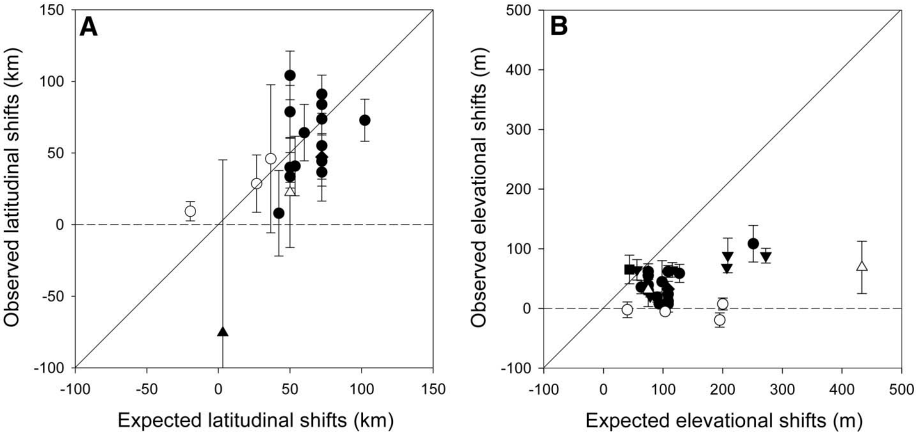


Fig. 1. Relationship between observed and expected range shifts in response to climate change, for (**A**) latitude and (**B**) elevation. Points represent the mean responses (±SE) of species in a particular taxonomic group, in a given region. Positive values indicate shifts toward the pole and to higher elevations. Diagonals represent 1:1 lines, where expected and observed responses are equal. Open circles, birds; open triangles, mammals; solid circles, arthropods; solid inverted triangles, plants; solid square, herptiles; solid diamond, fish; solid triangle, mollusks.

---

[1] Department of Biology, University of York, Wentworth Way, York YO10 5DD, UK. [2] Biodiversity Research Center, Academia Sinica, 128 Academia Road, Section 2, Nankang Taipei 115, Taiwan. [3] School of Biological and Biomedical Sciences, and Institute of Hazard, Risk and Resilience, Durham University, South Road, Durham DH1 3LE, UK. [4] Centre for Ecology & Hydrology, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, UK.

*To whom correspondence should be addressed. E-mail: chris.thomas@york.ac.uk