Published in final edited form as:

Nat Genet. 2008 February ; 40(2): 158–160. doi:10.1038/ng.2007.55.

Ultraconservation identifies a small subset of extremely constrained developmental enhancers

Axel Visel¹, Shyam Prabhakar¹, Jennifer A Akiyama¹, Malak Shoukry¹, Keith D Lewis¹, Amy Holt¹, Ingrid Plajzer-Frick¹, Veena Afzal¹, Edward M Rubin^{1,2}, and Len A Pennacchio^{1,2}

¹Genomics Division, MS 84-171, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

²United States Department of Energy Joint Genome Institute, Walnut Creek, California 94598, USA.

Abstract

Extended perfect human-rodent sequence identity of at least 200 base pairs (ultraconservation) is potentially indicative of evolutionary or functional uniqueness. We used a transgenic mouse assay to compare the embryonic enhancer activity of 231 noncoding ultraconserved human genome regions with that of 206 extremely conserved regions lacking ultraconservation. Developmental enhancers were equally prevalent in both populations, suggesting instead that ultraconservation identifies a small, functionally indistinct subset of similarly constrained *cis*-regulatory elements.

The last common ancestor of human and rodents lived ~75 million years ago¹, and yet the human genome contains 256 non-exonic ultraconserved elements of ≥ 200 bp that are perfectly conserved in mouse and rat as a result of extreme purifying selection^{2,3}. Their depletion in segmental duplications and copy number variable regions⁴ and their strong bias toward rare, derived alleles in humans^{3,5} further point toward a pivotal functional role for these elements. In contrast, the identity of ultraconserved sequences as a distinct class of conserved genomic elements has been challenged by the observation that more rigorous comparative genomic methods (for example, refs. ^{6,7}) identify additional sequences with similar conservation properties, but lacking extended perfect sequence conservation. Another feature of ultraconserved elements that undermines their relevance as a possible distinct category of conservation is that they are almost invariably embedded in larger blocks of constrained sequence, suggesting that they exist not as independent units of biological function, but as somewhat arbitrary fragments of larger functional modules. In the absence of comprehensive experimental data, it remains unclear whether the absolute sequence conservation of ultraconserved elements is indicative of a unique role, or whether they are merely a functionally indistinct fraction of a much larger set of extremely constrained elements.

To explore the possible uniqueness of noncoding ultraconserved elements, we identified a large number of human-rodent conserved elements that are under similar evolutionary constraint as regions containing ultraconservation. We compared the entire set of these elements, the majority of which lack perfectly conserved regions of ≥ 200 bp (that is, ultraconserved regions), to the small subset that overlapped ultraconserved elements in regard to their degree of constraint in other mammalian species and their enrichment near genes with certain functions. Moreover, we compared the ability of a genome-wide set of non-exonic ultraconserved elements and more than 200 extremely constrained human-rodent elements to drive tissue-

Correspondence should be addressed to L.A.P. (LAPennacchio@lbl.gov).

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions>

specific expression *in vivo*, a property that has previously been observed to be a predominant function associated with noncoding ultraconservation^{8–12}.

In an initial comparative genomic assessment of ultraconservation, we found substitutions in 79% of these elements in other mammalian species (Fig. 1a), indicating that their absolute conservation between human and rodents is at least partially a matter of ascertainment bias, rather than absolute intolerance of nucleotide substitutions. This finding is consistent with the limited overlap among human-rodent, human-mouse-dog and human-chicken ultraconserved elements⁴ and supports the possibility that they represent only a subset of a larger group of elements with similar properties. We therefore used a statistical approach (Gumby13) with scoring parameters optimized through multiple genome-wide scans (**Supplementary Methods** and Supplementary Table 1 online) to identify a more general set of noncoding elements marked by extreme human-mouse-rat constraint. Conservation scores for individual elements were derived from log-transformed Gumby *P* values and reflect their length and constraint relative to the local neutral substitution rate. When we compared these elements to the distribution of non-exonic ultraconserved elements, we found that the constraint scores of regions with ultraconservation were distributed over a wide range, and that a much larger number of elements seem similarly constrained (Fig. 1b). We identified a population of 2,614 human-rodent constrained elements that overlap or include 234 (91%) of all 256 non-exonic ultraconserved elements (Supplementary Table 2 online). To quantify the extreme conservation of these elements independently from the scoring scheme used for their identification, we determined their branch length and rejected substitution counts⁶ in human, rodents and five additional mammalian species (Supplementary Fig. 1 online). We found that extremely constrained elements that contain or do not contain regions of ultraconservation have similar characteristics by these two widely used comparative genomic measures, indicating their ‘ultra-like’ conservation. Although an order of magnitude more numerous than non-exonic ultraconserved elements and located in the vicinity of a fivefold larger number of genes, the highly constrained noncoding regions identified here are enriched near genes belonging to a small subset of functional categories. As for ultraconserved elements, these functions include transcriptional regulation and development² and, in particular, development of the nervous system (Fig. 1c; see Supplementary Table 3 online for a list of all significantly enriched functions). In summary, comparative analyses, as well as the genome-wide distributions, suggest that ultraconservation merely defines a subset of genome regions that are under similar constraint and that are enriched near genes with similar functional properties.

To test whether such apparent equivalence at the sequence level is also associated with similar functional properties, we focused on transcriptional enhancer activity during embryonic development. We used a transgenic mouse assay to determine the embryonic enhancer activities of 155 human genome regions that contain non-exonic ultraconserved elements and combined these data with a previously reported smaller dataset¹⁰ to establish a genome-wide compendium of their enhancer activities. A total of 231 transgenic assays was considered, in which the tested human genome fragments included 245 of all 256 non-exonic ultraconserved elements (Supplementary Table 4 online). We found that half (115/231) of the ultraconserved regions drove reproducible reporter gene expression in various tissues of the developing mouse embryo, often in a tight spatially restricted manner and with subregions of the central nervous system among the most frequently targeted structures (Fig. 2a).

To determine whether such an enrichment in embryonic enhancers is specifically associated with the presence of ultraconserved regions, we also tested the enhancer activities of 206 extremely constrained human-rodent noncoding sequences that lack regions of ultraconservation. Of note, these regions were selected blind to evolutionary conservation depth in nonmammalian species, and purely based on their human-rodent constraint scores. Using identical scoring criteria as before, we found that 102 of these 206 elements (50%) acted

as tissue-specific enhancers at embryonic day 11.5 (Supplementary Table 5 online). We did not observe significant differences between the ultraconserved and non-ultraconserved elements regarding the overall distribution of the targeted anatomical structures (Fig. 2a). We observed multiple cases of ultraconserved and non-ultraconserved elements driving virtually identical patterns when scrutinized at higher resolution (Fig. 2b), as well as dozens of patterns driven by non-ultraconserved elements for which no counterpart was found among ultraconserved elements (Supplementary Fig. 2 online). Our findings indicate that extreme human-rodent constraint identifies genome regions that are, in their entirety, highly enriched in embryonic enhancers, whereas the ultraconserved subset within this population was neither found to be enriched in enhancers targeting specific tissues nor found to be generally more enriched in developmental enhancers.

Ultraconserved elements seem to have remained practically ‘frozen’ during mammalian evolution², and their perfect, uninterrupted sequence identity between human and rodents has suggested that they might represent the pinnacle of extreme noncoding sequence conservation in mammals. In contrast to this proposal, and consistent with findings based on alternative comparative metrics^{6,7}, our results support the notion that the relatively small number of ultraconserved elements may more likely be due to their definition by a simple percent-identity-plot approach¹⁴ than to a uniquely high degree of constraint of the conserved regions in which they are located. If the enrichment in enhancer activity observed in our *in vivo* testing of over 400 distinct genome fragments is considered as a measure, ultraconserved elements do not represent the very tail of the continuum of human-rodent conservation, but are merely a subset of a tenfold larger population of elements that are under similar constraint and have apparently equivalent regulatory function. The possibility of functional redundancy within this much larger population of conserved elements may also provide a partial explanation for the observation that some ultraconserved noncoding elements are dispensable for viability in mice¹⁵. The elements identified in this study are defined independent of their conservation in nonmammalian vertebrate species. We therefore expect that, of the hundreds of additional tissue-specific enhancers that remain to be discovered in this category of extreme conservation, some will be unique to mammals. Although subsets of extremely conserved noncoding elements undoubtedly have other molecular functions, our results indicate that a large proportion of these elements choreograph the transcription of key genes during mammalian development, regardless of whether they are ultraconserved.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank I. Dubchak, A. Poliakov and S. Minovitsky for help with genome alignments and database development; S. Bhardwaj and S. Phouanavong for technical assistance; and N. Ahituv, M. Nobrega, J. Noonan and members of the Pennacchio and Rubin laboratories for discussion and critical comments on the manuscript. L.A.P. was supported by grant HL066681, Berkeley-Program for Genomic Applications, under the Programs for Genomic Applications, funded by National Heart, Lung, and Blood Institute, and HG003988 funded by National Human Genome Research Institute. Research was done under Department of Energy Contract DE-AC02-05CH11231, University of California, E.O. Lawrence Berkeley National Laboratory. A.V. was supported by an American Heart Association postdoctoral fellowship.

References

1. Waterston RH, et al. Nature 2002;420:520–562. [PubMed: 12466850]
2. Bejerano G, et al. Science 2004;304:1321–1325. [PubMed: 15131266]
3. Katzman S, et al. Science 2007;317:915. [PubMed: 17702936]
4. Derti A, Roth FP, Church GM, Wu CT. Nat. Genet 2006;38:1216–1220. [PubMed: 16998490]

5. Drake JA, et al. *Nat. Genet* 2006;38:223–227. [PubMed: 16380714]
6. Cooper GM, et al. *Genome Res* 2005;15:901–913. [PubMed: 15965027]
7. Siepel A, et al. *Genome Res* 2005;15:1034–1050. [PubMed: 16024819]
8. de la Calle-Mustienes E, et al. *Genome Res* 2005;15:1061–1072. [PubMed: 16024824]
9. Poulin F, et al. *Genomics* 2005;85:774–781. [PubMed: 15885503]
10. Pennacchio LA, et al. *Nature* 2006;444:499–502. [PubMed: 17086198]
11. Ghanem N, et al. *J. Neurosci* 2007;27:5012–5022. [PubMed: 17494687]
12. Paparidis Z, et al. *Dev. Growth Differ* 2007;49:543–553. [PubMed: 17661744]
13. Prabhakar S, et al. *Genome Res* 2006;16:855–863. [PubMed: 16769978]
14. Hardison RC, Oeltjen J, Miller W. *Genome Res* 1997;7:959–966. [PubMed: 9331366]
15. Ahituv N, et al. *PLoS Biol* 2007;5:e234. [PubMed: 17803355]

Figure 1.

Ultraconservation identifies a small fraction of elements that are under similar evolutionary constraint. **(a)** We considered nucleotide substitutions in 256 non-exonic human-rodent ultraconserved elements² in five additional placental mammalian genomes (chimpanzee, rhesus, dog, horse and cow). We found that 203 elements (79%) have at least one position substituted in other mammals, 153 (60%) have two or more substituted positions and 43 (17%) have substitutions at five or more positions. We did not consider imperfect sequence conservation due to insertions and deletions. **(b)** We identified more than 2,600 extremely constrained human-rodent elements at a constraint score threshold of ≥ 40 , of which more than 2,300 are not defined as ultraconserved. Of the 500 most human-rodent constrained noncoding elements (score ≥ 74.7), 350 (70%) do not contain or overlap regions of ultraconservation. The graph does not depict overlap with possibly exonic² ultraconserved regions. **(c)** Enrichment near genes involved in transcriptional regulation, general development and nervous system development. We considered the function (Gene Ontology, biological process) of the closest neighboring gene of each conserved element, and compared the observed numbers of genes in each category to the number expected based on all annotated RefSeq genes. Additional significantly enriched categories are listed in Supplementary Table 3.

Figure 2.

Highly constrained enhancers target expression to similar tissues independent of ultraconservation. **(a)** Binning of patterns driven by ultraconserved (top) and highly constrained human-rodent (bottom) enhancers into broad anatomical domains does not show significant differences for any structure (all P values > 0.05 , Fisher's exact test with Bonferroni correction for multiple hypothesis testing). Enhancers targeting expression to more than one region are reported in each respective category. **(b)** Examples of extremely constrained enhancers that contain (left) or do not contain (right) regions of ultraconservation, but drive highly similar expression patterns. Arrows indicate viewing angle of insets. Only one representative embryo per enhancer is shown; all patterns were reproducible in at least two additional embryos resulting from independent transgenic integration events. DRG, dorsal root ganglia. Genomic coordinates for all enhancers are provided in Supplementary Tables 4 and 5.