

ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes

Ivan Ovcharenko^{1,2,*}, Marcelo A. Nobrega³, Gabriela G. Loots¹ and Lisa Stubbs¹

¹Genome Biology Division and ²Energy, Environment, Biology and Institutional Computing, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA and ³Department of Genome Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

Received February 11, 2004; Accepted February 25, 2004

ABSTRACT

With an increasing number of vertebrate genomes being sequenced in draft or finished form, unique opportunities for decoding the language of DNA sequence through comparative genome alignments have arisen. However, novel tools and strategies are required to accommodate this large volume of genomic information and to facilitate the transfer of predictions generated by comparative sequence alignment to researchers focused on experimental annotation of genome function. Here, we present the ECR Browser, a tool that provides easy and dynamic access to whole genome alignments of human, mouse, rat and fish sequences. This web-based tool (<http://ecrbrowser.dcode.org>) provides the starting point for discovery of novel genes, identification of distant gene regulatory elements and prediction of transcription factor binding sites. The genome alignment portal of the ECR Browser also permits fast and automated alignments of any user-submitted sequence to the genome of choice. The interconnection of the ECR Browser with other DNA sequence analysis tools creates a unique portal for studying and exploring vertebrate genomes.

INTRODUCTION

Several vertebrate genomes, including those of human, mouse, rat and several fish, have recently been sequenced and assembled, and with the exponential increase of sequencing performance and capabilities, DNA sequences of several other vertebrate genomes are expected to emerge in the near future. A number of studies have underscored the value of comparative genome alignments in the functional annotation of such complex genomes, demonstrating clearly that DNA sequence

conservation can serve as a faithful guide to the identification of sequence elements with critical biological functions. This strategy has been validated with the identification of both novel genes and functional noncoding elements (1–3). While comparisons between human and rodent sequences yield informative results in many cases and have been exploited extensively (4–6), many genomic segments are not usefully annotated in such comparisons due to the non-uniform structure and evolutionary rate across vertebrate genomes (7). Moreover, several cardinal features in the human genome are likely to have been acquired or shaped more recently than the human–mouse evolutionary separation (8,9). These examples underscore the need for alternative comparative strategies that can accommodate the evolutionary asymmetries and architectural uniqueness of the human genome.

Several strategies have been devised recently to overcome these difficulties. In particular, the use of multiple species sequence comparisons has been proposed as an alternative to standard pairwise comparisons, aiming at the identification of a subset of sequences that are conserved in multiple species. Using this premise, a new method of multiple comparative sequence analysis was developed (10) based on the identification of an optimal dataset of species to compare that results in the best correlation between multiple conserved sequences (MCSs) and biologically relevant regions. A similar prioritization strategy, using comparisons between human sequence and that of a single distantly related species, the puffer fish, was recently used to identify evolutionarily conserved regions (ECRs) corresponding to critical regulatory elements in large (>1 Mb), highly conserved gene desert regions flanking the human *DACH* gene (11). This and other similar studies have emphasized the identification of ECRs that arose prior to the divergence of fish and primate lineages and have been conserved since that time (11–13). Another strategy, dubbed phylogenetic shadowing, has been developed to detect and identify more recently or rapidly evolving ECRs including primate-specific functional elements (14), which would not be detected in sequence comparisons of humans and rodents

*To whom correspondence should be addressed. Tel: +1 925 422 5035; Fax: +1 925 422 2099; Email: ovcharenko1@llnl.gov
Correspondence may also be addressed to Lisa Stubbs. Email: stubbs5@llnl.gov

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated.

or more distant vertebrates. Taken together, these studies illustrate the fact that no single comparative genomic strategy suffices for genome-wide comparative studies. Rather, there is a pressing need for the development of tools that can integrate sequences of multiple genomes in a custom-made fashion, allowing for a dynamic overlay of orthologous sequences from a selected number of species, as deemed necessary on a case-by-case basis.

To fulfill this need, we have created a genome browser displaying multiple alignments of genomic sequence of various sequenced species including human, rodents and fish. This tool, called the ECR Browser, presents a dynamic representation of sequence comparisons, allowing the user to specify optimal parameters for alignment and display in analysis of genomic regions with different divergence rates. Two main goals have driven the creation of the ECR Browser: (i) to permit genome alignments to be generated, retrieved and displayed quickly, (ii) to provide maximum flexibility in genome alignments by allowing the user to dynamically adjust alignment and display parameters. These parameters include the number and types of species to be included in the comparison, the sequence to be used as the 'base' against which other genomes are compared, types of annotation to be displayed, thresholds to define significantly conserved sequence elements (e.g. sequence lengths and percentage identity with comparable sequences in the base genome) and other features that permit the user to tailor comparisons specifically to regions characterized by different evolutionary rates. Furthermore, the ECR Browser is designed to permit the incorporation of novel genomes immediately as their sequences become available in public databases.

ALIGNING GENOMES

Several strategies have recently been developed to analyze large segments of genome sequence, from whole microbial genomes to homology regions in the chromosomes of higher vertebrates (15–18). For the creation of the ECR Browser, we employed a strategy of genome alignment that is based on four consecutive sequence management steps. Briefly, after masking of repetitive elements, all the genomes were mapped pairwise, to establish large-scale syntenic relationships. Subsequently, each syntenic orthologous pair of sequences was aligned. Finally, data were collected and stored in a central database that is then utilized by the ECR Browser to construct conservation profile graphs at the user's specification.

The sequences and annotation data that are utilized by the ECR Browser are taken from the UCSC Genome Browser (17). In addition to the human, mouse and rat sequences obtained from this source, we have augmented the genome dataset with sequences of three fish genomes, namely *Fugu rubripes* (<http://www.jgi.doe.gov/fugu/>), *Tetraodon nigroviridis* (<http://www.genoscope.cns.fr/externe/tetraodon/Ressource.html>) and *Danio rerio* (http://www.sanger.ac.uk/Projects/D_rerio/). Repetitive elements in each genome were identified and masked, using precomputed data downloaded from UCSC where available, or by a local run of the Repeat-Masker program (<http://www.repeatmasker.org>).

Over millions of years of evolution, multiple large-scale rearrangements have altered gene order dramatically in the

genomes of fish, rodent, human and other vertebrate lineages. To identify related syntenic blocks in these divergent species, each genome was mapped to all others in pairwise fashion. The dramatically different evolutionary history that separates primates and rodents from fish, compared with the evolutionary separation between different fish or within the primate and rodent lineages, required the application of different approaches to genome alignments in each type of pairwise mapping. For mapping syntenic homologies between more closely related species, such as humans and rodents, we used a locally installed version of the BLAT tool (19). For comparative mapping of more distantly related species, such as humans and fish (or rodents and fish), the more sensitive but slower blast tool was employed (20). At the final step of synteny mapping, neighboring short hits of similarity were joined into large blocks of synteny (see Supplementary materials for details). Finally, pairs of orthologous sequences from each syntenic block were aligned with the use of the blastz alignment tool, with long alignments being cleaned from non-diagonal spurious hits (Supplementary materials).

From a technical viewpoint, alignments of the human and mouse genomes (as an example) utilize 50 Mb of disk space (i.e. significantly less than the size of the original genome FASTA files) and require less than a week on a P4-processor machine to be created. This is significantly faster than any other genome alignment strategy previously reported (16,18). This scale-up in performance and significant savings of disk space allow us to have multiple genome alignments on hand with a relatively short response time to update the ECR Browser as new assemblies of genomes are released.

VISUALIZATION AND DATA BROWSING SCHEME

The conservation-profile visualization scheme of the ECR Browser tool is based on an idea originally implemented in the PipMaker tool (21) and later adopted by both Vista (22) and zPicture (23). In this model, the base genome sequence is schematically displayed as the horizontal axis of a 2D graph, while the vertical axis represents the percentage identity between the base sequence and the sequence being compared (Figure 1). ECRs are differentiated from the neutrally evolving background and are colored according to their classification as protein coding exons, UTRs, introns, repetitive elements or conserved intergenic regions.

The ECR Browser dynamically constructs graphical conservation profiles for any region in the genome, which can be specified either by a gene name or by absolute genomic coordinates of the region of interest. Depending on user preferences the browser augments the conservation profile with an annotation of different genomic features, such as known genes, gene predictions, repetitive elements and single nucleotide polymorphisms, with annotations downloaded directly from the UCSC Genome Browser. Other browser features such as zooming, shifting and re-centering allow for the rapid conversion of the genomic size and coordinates being analyzed (Figure 1).

To accommodate the non-uniform evolutionary structure of vertebrate genomes, a flexible definition of ECR parameters was implemented in the browser. Display variability allows the user to require high stringency parameters in detecting

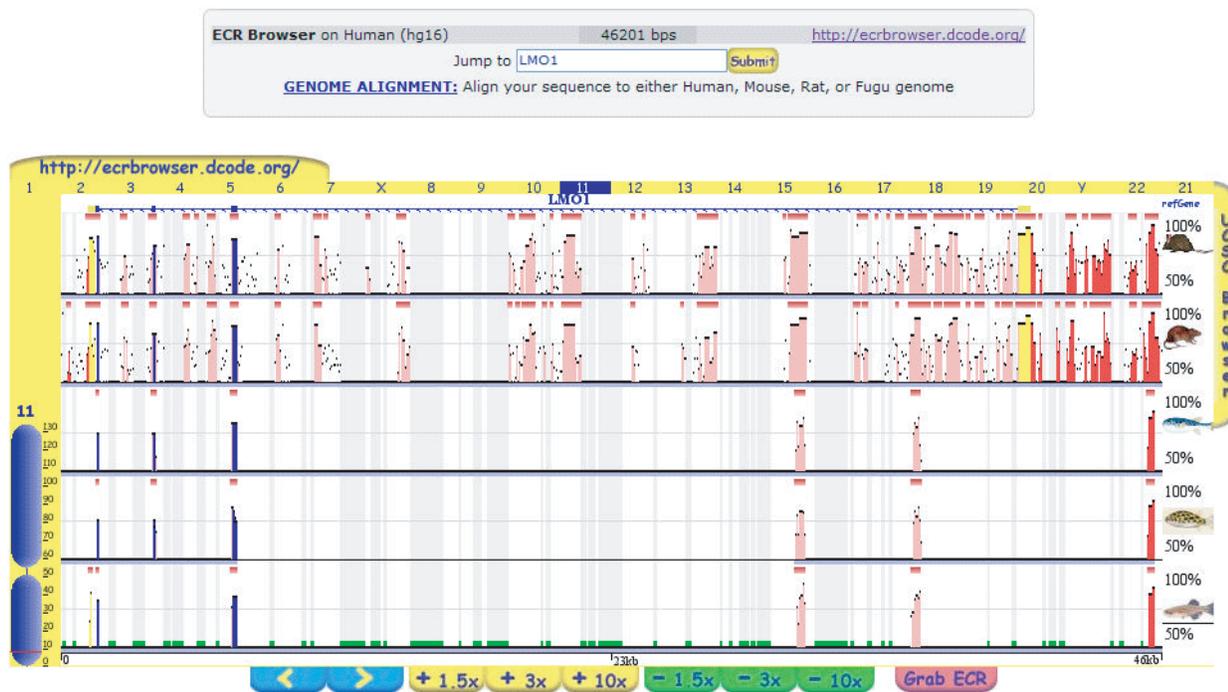


Figure 1. ECR Browser visualization of the Lim Domain Only 1 (*LMO1*) gene 46 kb locus in the human genome (UCSC freeze 16; NCBI Build 34). The conservation profiles of the human region in comparison with the mouse, rat, *Fugu*, tetraodon and zebra fish genomes are shown. The five genomes that were compared to the human region are plotted as horizontal layers of conservation diagrams and the small image icon at the right side of the plot represents the species corresponding to the alignment. Each layer contains a pip-type plot that consists of multiple short horizontal black lines. Each line represents an ungapped alignment, with the vertical height of the line describing the nucleotide identity underlying the alignment. Exons of the *LMO1* gene are depicted in blue and yellow, with the blue bars depicting the exons that are protein coding while yellow bars depict the UTRs. The 5'–3' orientation of the gene is given by arrow lines. A dark red bar on top of every layer provides an overview of the distribution of ECRs and is used to flag underlying regions of alignment. A conserved alignment is blue if it overlaps with a coding exon, yellow, UTR; pink, intron; red, intergenic region. The green bars at the bottom indicate repetitive elements in the base sequence. At the top of the browser display area quick-links to different chromosomes are offered, while the left bar represents a dynamic chromosomal map. The link to the UCSC Genome Browser (17) is also provided on the right side of the browser.

ECRs in slowly evolving genomic regions or less stringent parameters to identify barely distinguishable, short ECRs in other alignments, e.g. in rapidly diverging regions or in comparative analysis of distantly related species. Users can customize the display so that a subset of the available genomes is selected for comparative analysis. Thus, for example, alignments involving sequences from only closely related species might be chosen to analyze rapidly evolving genomic loci. Other custom features include the format of the displayed conservation plot (either a pip-plot or a smooth graph), a selection of different types of gene annotation and selection of picture display parameters (Figure 2).

The ECR Browser readily provides user access to DNA sequences corresponding to the genomic region being displayed. The browser also provides access to the sequences of ECRs detected under a specified set of alignment conditions, and a list of their positions in the displayed region; ECR sequences and positions are readily updated as the user alters the parameters of the alignment and display. To provide ready access to individual ECRs, we have introduced the 'Grab ECR' option, which allows 'one-click' access to any selected ECR in the conservation plot. This option connects to a detailed ECR description page containing ECR sequences from both species in any pairwise comparison, and a display of the underlying DNA sequence alignment. In addition, sequence characteristics such as length, percentage identity,

G+C content and genomic coordinates are listed and accompanied by links to the analysis of potential transcription factor binding sites (TFBS) inside the ECR, through the rVista program (24) (Figure 3). Primer selection and design can be performed for any ECR with the 'primers/oligos' tool (<http://www.primers.dcode.org>) that is integrated into the ECR Browser. This tool selects primer sequences of specified length and GC content for the user to choose, and verifies uniqueness for the designed primers by counting the number of times the primer sequence is encountered in the human and mouse genomes. These tools are designed to facilitate the transfer of data and predictions generated in comparative sequence alignments to the laboratory, where the sequences can be tested experimentally for biological function.

SYNTENY

As a by-product of the synteny mapping that is required for accurate comparative alignments (described above), the ECR Browser is able to locate and reconstruct syntenic linkage maps, establishing relationships that can be later utilized to navigate between different genomes (Figure 4A). Using the 'Synteny/Alignments' link, the user can jump from the display of a specific locus in one genome directly to a visualization of the syntenically homologous locus in another species

ECR Browser on Human (hg16) Settings

Display alignments with

Mouse (mm3)
 Rat (rn3)
 Fugu (fu3)
 Tetraodon (to6)
 Zebrafish (zf2)

Graph type

Smooth Pip-plot

Display gene features

Genscan RefSeq Twinscan

Number of layers

Layer height

pixels

Detect ECRs
(Evolutionary Conserved Regions)

min length bps
min identity %

[submit & return](#)

Figure 2. ECR Browser settings allow a flexible approach toward analysis of differently evolving genomic regions and dynamic choice of the plot and alignment settings.

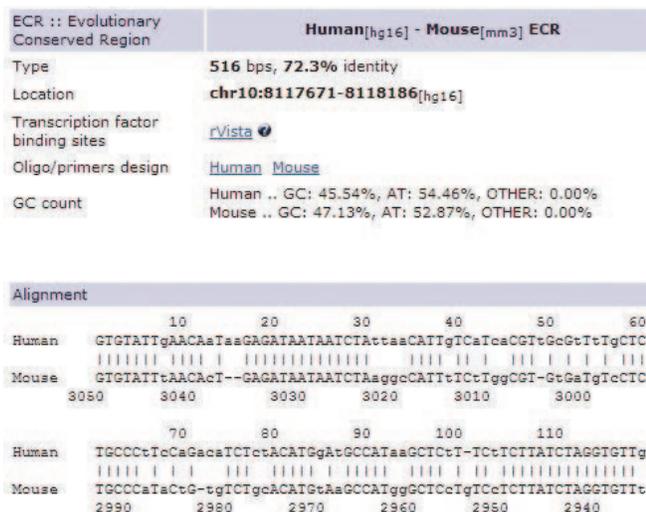


Figure 3. ‘Grab ECR’ feature—an access to the sequence, alignment and sequence analysis tools for a single ECR.

(Figure 4B). This option permits users to compare the size, organization and conserved features at the same locus in divergent genomes. It also permits users to compare ECRs arising in comparisons between human and mouse, for example to those detected in the same genomic locus in rat and mouse genome comparisons.

The identity of the base genome in the display of a particular region can also be readily changed with the use of the ‘Base Genome’ feature. Selecting this option will result in the generation of a new alignment of the same sequences previously displayed, but with a different base genome used as reference.

GENOME ALIGNMENT

While pre-computed alignments of the genomes available in the ECR Browser are sufficient for many tasks, the ability

to add additional sequences to a comparison—e.g. user-generated targeted sequence from additional genomes—can be valuable in a variety of applications. For this purpose, we created custom-defined alignment options within the browser that allow the instantaneous alignment of any user-defined sequence. Such queries may be either submitted directly, in FASTA format, or automatically downloaded from GenBank using the accession number of the sequence to be aligned, which is then forwarded to the ‘Genome Alignment’ portal in the browser. Upon receiving the user-submitted sequence, the ECR Browser will rapidly map this sequence to the selected genome (either human, mouse, rat or *Fugu* genome) using the BLAT tool. When the orthologous region is identified in the selected base genome, it is extracted along with the corresponding RefSeq gene annotation. Blastz alignments of the two sequences are made and a dynamic graphical visualization of the alignments is generated (Figure 5). A dynamic Picture (23) conservation plot, a portal to the rVista tool (24), an alignment dot-plot and a tool for dynamic annotation of ECRs in the alignment are also automatically provided.

INTEGRATION WITH OTHER TOOLS

We intend to maintain the ECR Browser as a constantly updated tool that not only incorporates newly deposited and annotated sequences, but also provides direct connections to the growing set of publicly available external sequence analysis tools. Presently, an extensive annotation of known genes, gene predictions, experimental RNA evidence and many other features is available through the direct interface between the ECR Browser and the Genome Browser at UCSC (17) and the Ensembl Genome Browser (25,26). This portal permits the user to examine any non-genic conservation pattern against the UCSC evidence database on putative novel genes and non-coding RNAs. Also, the ‘Synteny/Alignments’ link of the ECR Browser directs the user to the zPicture analysis web page, described above, offering an easy and fast way to distill a

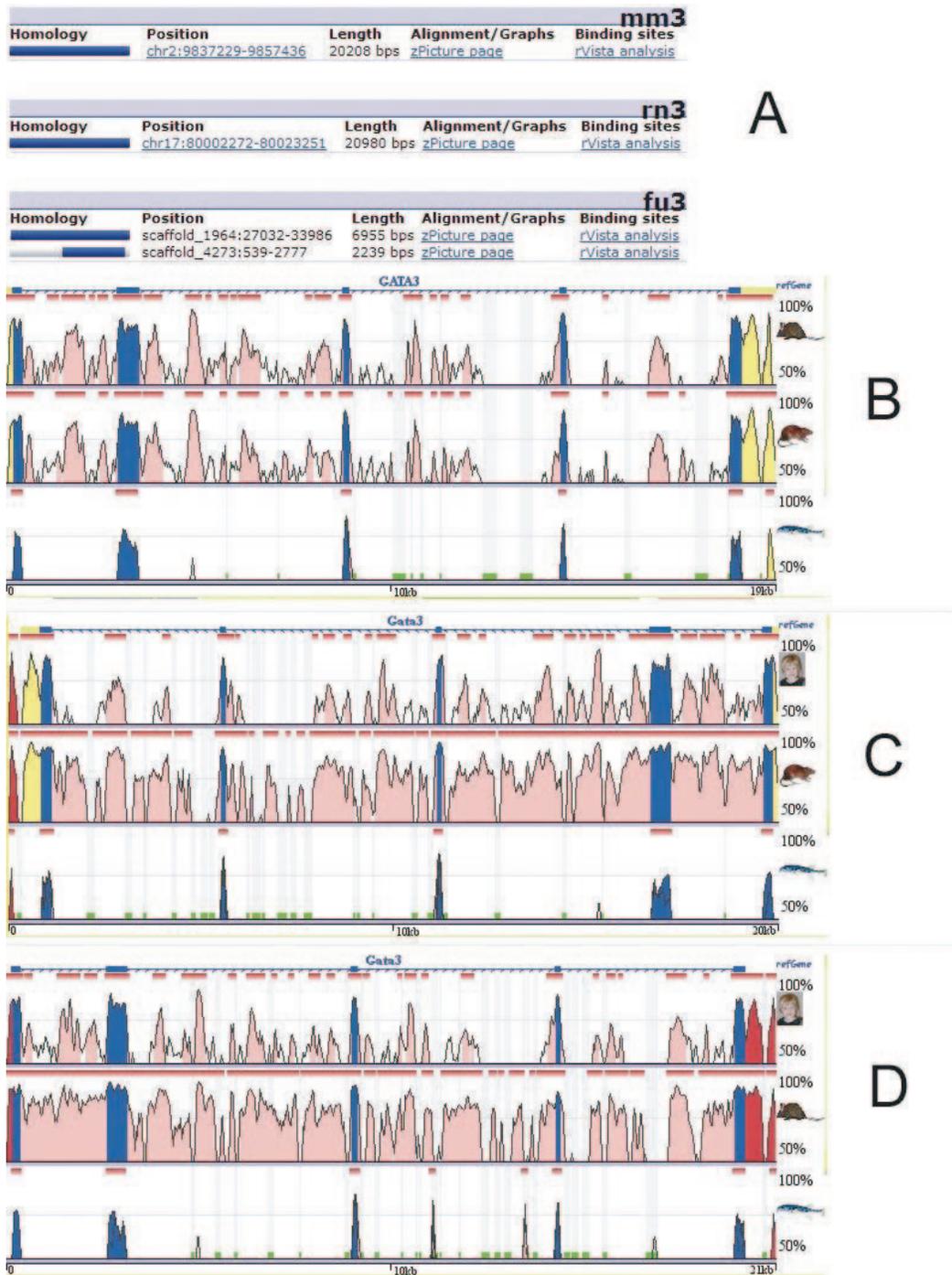


Figure 4. Synteny links in the ECR Browser—the human *GATA3* gene generates syntenic alignments in mouse, rat and *Fugu* genomes (A). The alignments generated using the human (B), mouse (C) and rat (D) genomes as base sequences are displayed. In each case, four-genome alignments (human–mouse–rat–*Fugu*) are obtained, each with a different base genome. Notice the reverse transcriptional orientation of the *GATA3* gene in the mouse genome, compared to that of humans and rats.

chosen pairwise alignment out of the multiple genome alignments. Using the zPicture tool, various modifications can be applied to the alignment. For example, the zPicture annotation feature permits manual curation of genes and other features that are not annotated in public databases (e.g. incorporating user-generated data), or editing of the public annotation to add features retrieved from other experimental or computational sources. In addition to these external tools, the ECR Browser is

dynamically connected to the GALA annotation database (27), the Rat Genome Browser (http://www.hpc.mcw.edu/mod_perl_gbrowse; when rat is selected as a base genome) and the JGI *Fugu* Genome Browser (28) (when *Fugu* is selected as a base genome). We are also planning to incorporate new methods of analysis and simultaneous scoring of multiple sequence alignments (29,30) into the ECR Browser tool in order to provide a high sensitivity interface toward

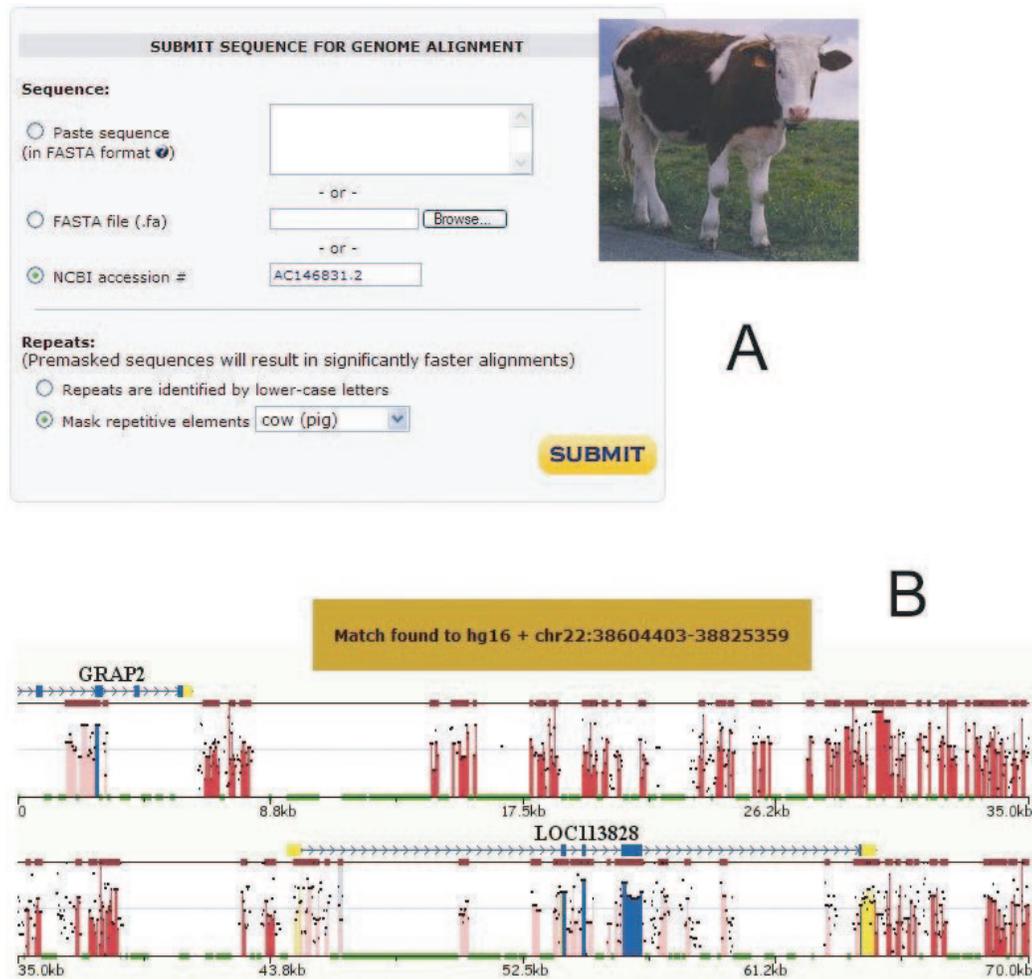


Figure 5. Genome alignment portal of the ECR Browser tool. Genomic sequences from any species that are submitted either as a FASTA file or automatically downloaded from GenBank by an accession number can be mapped and aligned to either human, mouse, rat or *Fugu* genomes (A). (B) A genome alignment of a cow sequence identified by accession number AC14683 with the human genome (hg16 freeze).

identification of functional domains in differentially evolving genomic loci.

As mentioned previously, the ECR Browser is also interconnected with the rVista tool (24). rVista is capable of filtering out up to 95% false positive TRANSFAC (31) predictions of TFBS while preserving high sensitivity of the search. The rVista portal provides a unique opportunity to predict the function of a noncoding element. By identifying evolutionarily conserved TFBS in an ECR, the rVista portal provides a basis for experimental testing and application of the known function of the conserved transcription factors toward understanding the function of a neighboring gene. Any pairwise alignment from the ECR Browser can be automatically submitted for rVista analysis via the 'Synteny/Alignment' link. Also, any ECR retrieved with the use of the 'Grab ECR' function can be submitted directly to rVista for binding-site analysis.

DISCUSSION

The ECR Browser tool is designed to highlight candidate functional coding and noncoding elements and to visualize

their genomic positions relative to the known gene features in the genome. By permitting comparisons of genomic sequence from species representing different, user-selected evolutionary clades, the ECR Browser provides flexibility in assessing evolutionary fates of noncoding sequences, allowing for comparisons that reflect sequence conservation over a wide range of timescales and in species with both shared and lineage-specific biological features. Comparisons between distant organisms, such as primates and fish, will likely uncover the fundamental building blocks shared by all vertebrates, while the comparative sequence analysis with closer comparisons, such as those between mice and rats, can highlight the functional structure of rapidly diverging genomic regions, including those that are specific to certain lineages and dictate lineage-specific traits.

Because it provides links to orthologous regions in other publicly available sequence analysis tools, the ECR Browser offers the user easy, automated access to resources, permitting a thorough annotation of functional elements in the genome (through the portal to the UCSC Genome Browser) and the annotation of TFBS (through the portal to the rVista tool). Because the underlying algorithms and tools that power the

ECR Browser are designed to permit rapid updates, the tool will be constantly updated with new sequence and new links to other relevant sequence analysis sites. These features, the ease with which conservation parameters and included datasets can be changed by the user, and the immediate dynamic display of alignment results make the ECR Browser a powerful new addition to the computational toolkit for annotating functional features in the human sequence and in other genomes sequenced now or in future years.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

The work was performed under the auspices of the US Department of Energy, Office of Biological and Environmental Research, by the University of California, Lawrence Livermore National Laboratory Contract No. W-7405-Eng-48.

REFERENCES

- Loots,G.G., Locksley,R.M., Blankespoor,C.M., Wang,Z.E., Miller,W., Rubin,E.M. and Frazer,K.A. (2000) Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science*, **288**, 136–140.
- Pennacchio,L.A., Olivier,M., Hubacek,J.A., Cohen,J.C., Cox,D.R., Fruchart,J.C., Krauss,R.M. and Rubin,E.M. (2001) An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science*, **294**, 169–173.
- Elnitski,L., Hardison,R.C., Li,J., Yang,S., Kolbe,D., Eswara,P., O'Connor,M.J., Schwartz,S., Miller,W. and Chiaromonte,F. (2003) Distinguishing regulatory DNA from neutral sites. *Genome Res.*, **13**, 64–72.
- Elnitski,L., Li,J., Noguchi,C.T., Miller,W. and Hardison,R. (2001) A negative cis-element regulates the level of enhancement by hypersensitive site 2 of the beta-globin locus control region. *J. Biol. Chem.*, **276**, 6289–6298.
- Waterston,R.H., Lindblad-Toh,K., Birney,E., Rogers,J., Abril,J.F., Agarwal,P., Agarwala,R., Ainscough,R., Alexandersson,M., An,P., *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520–562.
- Hardison,R.C., Roskin,K.M., Yang,S., Diekhans,M., Kent,W.J., Weber,R., Elnitski,L., Li,J., O'Connor,M., Kolbe,D., *et al.* (2003) Covariation in frequencies of substitution, deletion, transposition, and recombination during eutherian evolution. *Genome Res.*, **13**, 13–26.
- Santini,S., Boore,J.L. and Meyer,A. (2003) Evolutionary conservation of regulatory elements in vertebrate Hox gene clusters. *Genome Res.*, **13**, 1111–1122.
- Eichler,E.E., Hoffman,S.M., Adamson,A.A., Gordon,L.A., McCready,P., Lamerdin,J.E. and Mohrenweiser,H.W. (1998) Complex beta-satellite repeat structures and the expansion of the zinc finger gene cluster in 19p12. *Genome Res.*, **8**, 791–808.
- Gardiner,K., Fortna,A., Bechtel,L. and Davison,M.T. (2003) Mouse models of Down syndrome: how useful can they be? Comparison of the gene content of human chromosome 21 with orthologous mouse genomic regions. *Gene*, **318**, 137–147.
- Cooper,G.M., Brudno,M., Green,E.D., Batzoglou,S. and Sidow,A. (2003) Quantitative estimates of sequence divergence for comparative analyses of mammalian genomes. *Genome Res.*, **13**, 813–820.
- Nobrega,M.A., Ovcharenko,I., Afzal,V. and Rubin,E.M. (2003) Scanning human gene deserts for long-range enhancers. *Science*, **302**, 413.
- Ghanem,N., Jarinova,O., Amores,A., Long,Q., Hatch,G., Park,B.K., Rubenstein,J.L. and Ekker,M. (2003) Regulatory roles of conserved intergenic domains in vertebrate Dlx bigene clusters. *Genome Res.*, **13**, 533–543.
- Lettice,L.A., Heaney,S.J., Purdie,L.A., Li,L., de Beer,P., Oostra,B.A., Goode,D., Elgar,G., Hill,R.E. and de Graaff,E. (2003) A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.*, **12**, 1725–1735.
- Boffelli,D., McAuliffe,J., Ovcharenko,D., Lewis,K.D., Ovcharenko,I., Pachter,L. and Rubin,E.M. (2003) Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science*, **299**, 1391–1394.
- Delcher,A.L., Phillippy,A., Carlton,J. and Salzberg,S.L. (2002) Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res.*, **30**, 2478–2483.
- Couronne,O., Poliakov,A., Bray,N., Ishkhanov,T., Ryaboy,D., Rubin,E., Pachter,L. and Dubchak,I. (2003) Strategies and tools for whole-genome alignments. *Genome Res.*, **13**, 73–80.
- Karolchik,D., Baertsch,R., Diekhans,M., Furey,T.S., Hinrichs,A., Lu,Y.T., Roskin,K.M., Schwartz,M., Sugnet,C.W., Thomas,D.J. *et al.* (2003) The UCSC Genome Browser Database. *Nucleic Acids Res.*, **31**, 51–54.
- Schwartz,S., Kent,W.J., Smit,A., Zhang,Z., Baertsch,R., Hardison,R.C., Haussler,D. and Miller,W. (2003) Human–mouse alignments with BLASTZ. *Genome Res.*, **13**, 103–107.
- Kent,W.J. (2002) BLAT—the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.
- Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
- Schwartz,S., Zhang,Z., Frazer,K.A., Smit,A., Riemer,C., Bouck,J., Gibbs,R., Hardison,R. and Miller,W. (2000) PipMaker—a web server for aligning two genomic DNA sequences. *Genome Res.*, **10**, 577–586.
- Mayor,C., Brudno,M., Schwartz,J.R., Poliakov,A., Rubin,E.M., Frazer,K.A., Pachter,L.S. and Dubchak,I. (2000) VISTA: visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics*, **16**, 1046–1047.
- Ovcharenko,I., Loots,G.G., Hardison,R.C., Miller,W. and Stubbs,L. (2004) zPicture: dynamic alignment and visualization tool for analyzing conservation profiles. *Genome Res.*, **14**, 472–477.
- Loots,G.G., Ovcharenko,I., Pachter,L., Dubchak,I. and Rubin,E.M. (2002) rVista for comparative sequence-based discovery of functional transcription factor binding sites. *Genome Res.*, **12**, 832–839.
- Birney,E., Andrews,D., Bevan,P., Caccamo,M., Cameron,G., Chen,Y., Clarke,L., Coates,G., Cox,T., Cuff,J. *et al.* (2004) Ensembl 2004. *Nucleic Acids Res.*, **32**, D468–D470.
- Hubbard,T., Barker,D., Birney,E., Cameron,G., Chen,Y., Clark,L., Cox,T., Cuff,J., Curwen,V., Down,T., *et al.* (2002) The Ensembl genome database project. *Nucleic Acids Res.*, **30**, 38–41.
- Giardine,B., Elnitski,L., Riemer,C., Makalowska,I., Schwartz,S., Miller,W. and Hardison,R.C. (2003) GALA, a database for genomic sequence alignments and annotations. *Genome Res.*, **13**, 732–741.
- Aparicio,S., Chapman,J., Stupka,E., Putnam,N., Chia,J.M., Dehal,P., Christoffels,A., Rash,S., Hoon,S., Smit,A. *et al.* (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science*, **297**, 1301–1310.
- Margulies,E.H., Blanchette,M., Haussler,D. and Green,E.D. (2003) Identification and characterization of multi-species conserved sequences. *Genome Res.*, **13**, 2507–2518.
- Thomas,J.W., Touchman,J.W., Blakesley,R.W., Bouffard,G.G., Beckstrom-Sternberg,S.M., Margulies,E.H., Blanchette,M., Siepel,A.C., Thomas,P.J., McDowell,J.C. *et al.* (2003) Comparative analyses of multi-species sequences from targeted genomic regions. *Nature*, **424**, 788–793.
- Wingender,E., Dietze,P., Karas,H. and Knuppel,R. (1996) TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res.*, **24**, 238–241.