View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Digital.CSIC

Tandem repeat variation in human and great ape populations and its impact on gene expression divergence

Tugce Bilgin Sonay[1, 2+], Tiago Carvalho[3+], Mark D. Robinson[2,4], Maja P. Greminger[5], Michael Krützen[5], David Comas[3], Gareth Highnam[6], David Mittelman[6], Andrew J. Sharp[7], Tomàs-Marques Bonet[3,8+], Andreas Wagner[1,2,9+]

+ equal contribution

1 Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland

2 The Swiss Institute of Bioinformatics, Lausanne, Switzerland

3 Institute of Evolutionary Biology (CSIC-UPF), Department of Experimental and Health Sciences, Universitat Pompeu Fabra, 08003 Barcelona, Spain

4 Institute of Molecular Life Sciences, University of Zurich, 8057 Zurich, Switzerland

5 Evolutionary Genetics Group, Anthropological Institute and Museum, University of Zurich, Zurich

6 Department of Biological Science and Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA 24061, United States of America

7 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai School, New York, New York, United States of America

8 Centro Nacional de Análisis Genómico (CNAG), PCB, Barcelona, Catalonia 08028, Spain

9 The Santa Fe Institute, Santa Fe, New Mexico, United States of America

Introduction

Tandem repeats (TR) are stretches of DNA that are **highly variable in length and mutate rapidly**, and thus an important source of genetic variation. This variation is **highly informative for population and conservation genetics**, and has also been associated with several **pathological conditions and with gene expression regulation**. However, genome-wide surveys of TR variation in humans and close species have been scarce due to the technical difficulties derived from short-read technology.

Methodology

We genotyped TRs (repeat unit length 1-5 base pairs) in a panel of **83 human and nonhuman great ape genomes**, from 6 different species, by running **repeatseq** (Highnam et. al 2013) on a catalog of TRs identified in the human genome reference with **Tandem Repeat Finder** (TRF). We also used TRF to identify larger repeats (repeat unit length 2-50 base pairs) located in orthologous genes in human, chimp and macaque. We then checked whether gene repeat content was associated with gene expression divergence between species.

Results

Population structure and diversity

Analysis of **TR genotype data** for human and non-human great ape samples **recapitulate results from previous studies**, using millions of SNPs, regarding the genetic distance between species, their heterozygosity estimates (Fig. 1), and population structure (not shown).

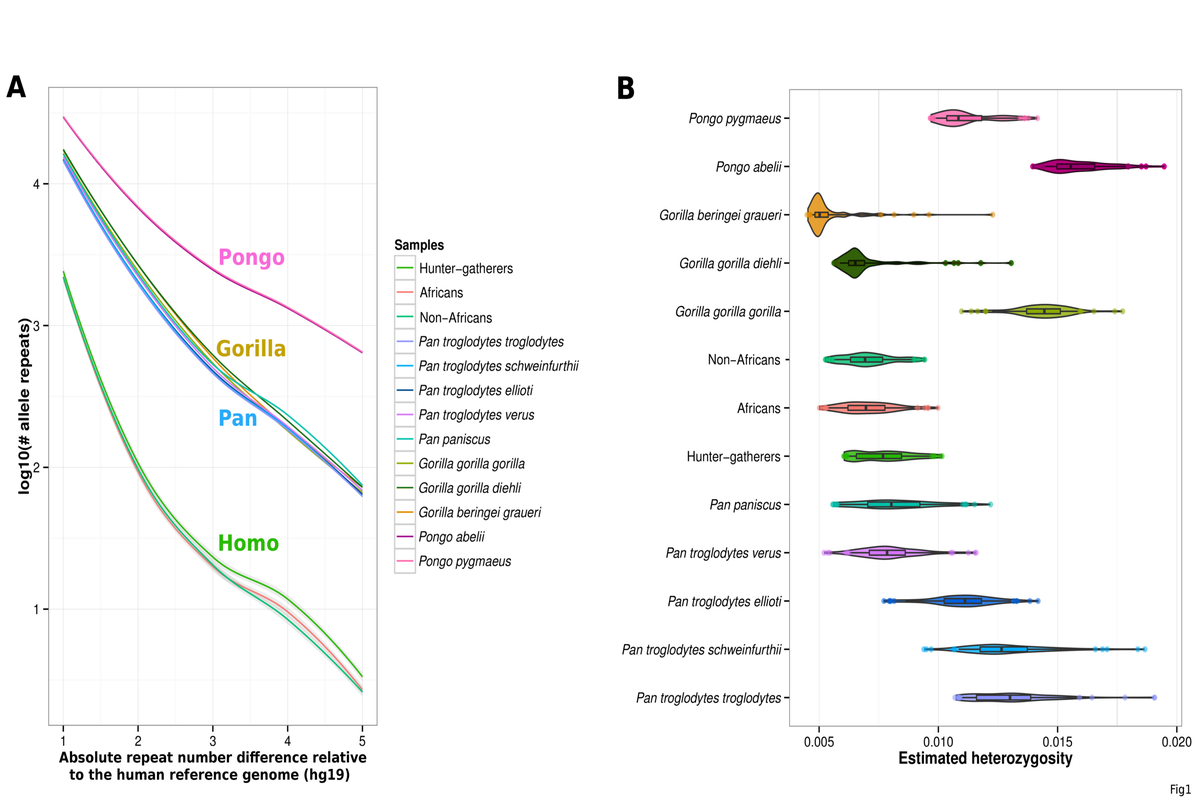

Fig 1. Population diversity estimates for several human groups and great ape subspecies and species A) Number of events in log10 scale (x-axis) of different categories of absolute repeat number difference relative to the human genome reference (y-axis) **B)** Heterozigosity estimates

The impact of TR on gene expression

For the larger repeats, we found that genes with promoters in their repeats showed more expression divergence for all tissues analysed (Fig. 2A), compared to those without. The same trend, but with different intensity, was observed when considering different repeat locations within the gene or at nearby regions (Fig. 2B).

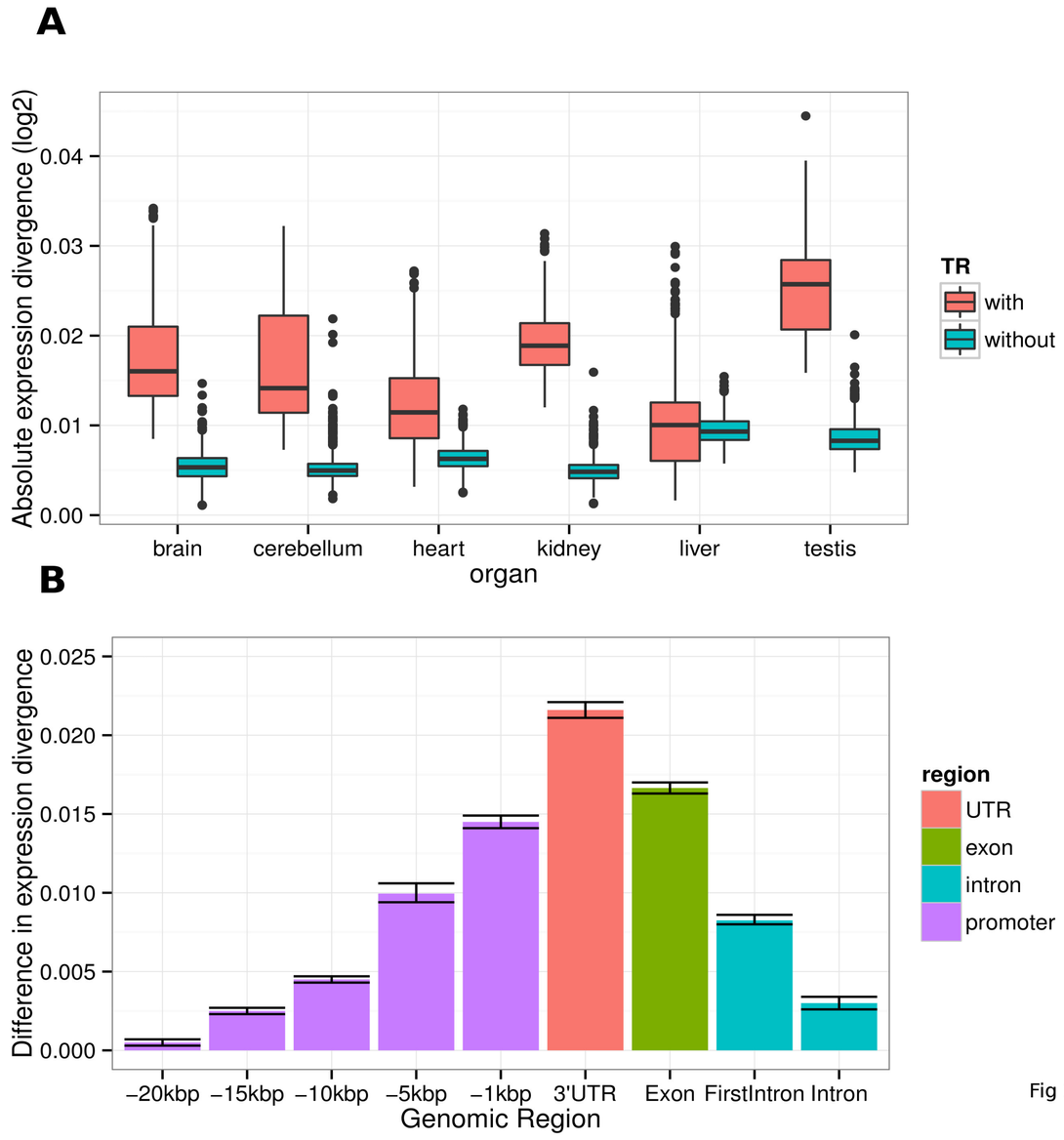

Fig 2

Fig. 2. Relationship between expression divergence (normalized by mean tissue expression divergence) and the presence of repeats in gene promoters and other genic regions. A) Mean expression divergence for genes with (pink) or without (blue) TRs in their promoters. **B)** Mean difference in expression divergence, between genes with or without repeats, according to the repeat location relative to the gene.

Using the TR genotype data from human and chimpanzee samples we also classified genes according to whether both species contained **polymorphic repeats**, **same TR genotype fixed** in all samples, or **no repeats,** in their promoters. We found that the **former showed more expression divergence** for all tissues analyzed compared to the two latter classes (not shown).

Conclusions

Our work shows that TRs, a type of sequence with unusually high mutability, may be an relevant class of regulatory mutations that might contribute to species differences, and highlight the potential contribution of TRs to human evolution through gene regulation.

Bibliography

1. Highnam G, *et al* "Accurate human microsatellite genotypes from high-throughput resequencing data using informed error profiles." Nucleic Acids Res. 2013 41:e32.