An investigation of the worldwide distribution of the READ1 element in *DCDC2* and its relationship with variation of linguistic features across languages

DeMille, M.M.C.[1]; Frijters, J.C.[3]; Tang, K.; Geissler, C.[3]; Powers, N.R.[2]; Bowen, B.M.[2]; Mehta, C.[1]; Malins, J.G.[5]; Adams, A.K.[2]; Truong, D.T.[1]; Gruen, J.R.[1]

Affiliations: Departments of [1]Pediatrics and [2]Genetics, Yale Medical School, New Haven, CT ; [3]Dept of Pediatrics, Child and Youth Studies, Brock University; [4]Department of Linguistics, Yale University, New Haven, Connecticut; [5]Haskin Laboratories

Abstract

We compared the evolution and worldwide distribution of READ1 (regulatory element associated with dyslexia encoded in *DCDC2)* with worldwide variation in language properties. While *DCDC2* was identified as a reading disability gene, it may also influence normal variation in language. We estimated the mutational history of the READ1 element by examining primate and archaic hominin sequences. We tested the association of READ1 functional groups with numbers of consonants and vowels in 57 population samples from across the world. We found an association between numbers of consonants and the RU1-1 functional group of READ1.

Introduction

The DYX2 locus (6p21.3), containing dyslexia candidate genes *DCDC2* and *KIAA0319*, is the most replicated locus for reading disability (RD) and language impairment (LI) worldwide. READ1, located in intron 2 of *DCDC2*, is a multi-unit transcriptional control element with approximately 40 known alleles[1]. Previous research has shown that repeat unit (RU) length groupings of alleles of READ1 (RU1-1 and RU2-long) are associated with RD and LI[1.2].

Central Question

Is there a relationship between READ1 and linguistic features across world languages?

57 Populations: 4 Continental Groups

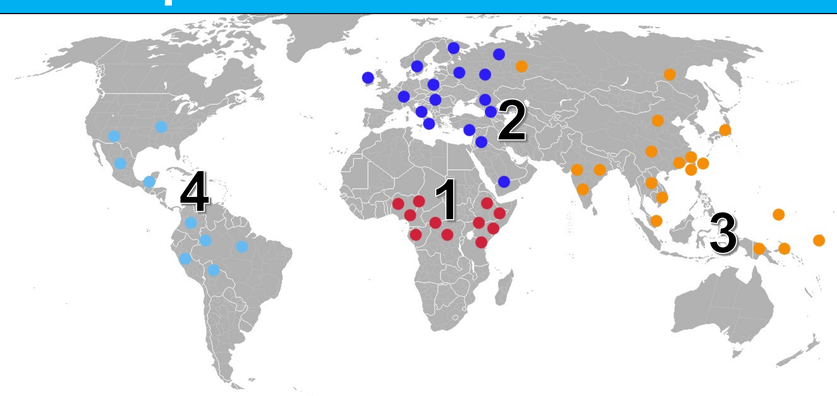


Figure 1: We obtained 57 population samples from across the globe from the Kidd laboratory. We used measures of genetic relatedness between populations to cluster them into five continental groups to account for variation between regions of the world.

READ1 Phylogeny

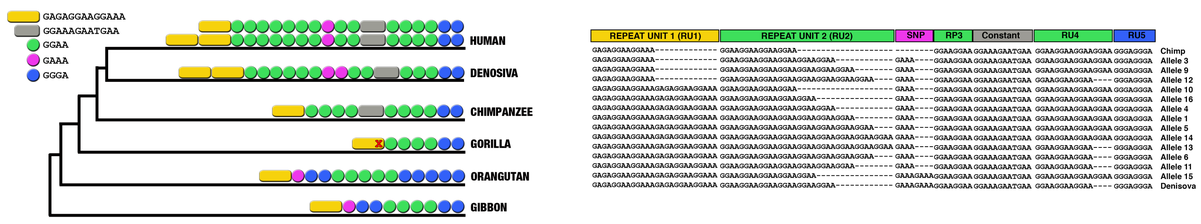
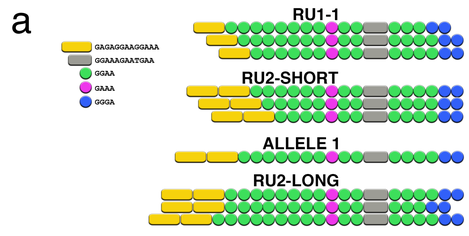


Figure 2: (a) READ1 sequenced in primates aligned in a phylogenetic tree reflecting the ancestral relationships between species. Branch lengths do not reflect evolutionary time. Sequence motifs are depicted as colored dots for 4 bp motifs and rectangles for longer motifs. Over 40 different READ1 alleles have been observed in *Homo sapiens*, the most common allele of each RU1 type are depicted here. The Denisova allele observed in sequence data is similar to human allele 34 (only differs by the number of RU2 repeats, Allele 34 has seven and the Denosiva allele has six). Chimpanzee (*Pan paniscus* and *P. trogdolytes*) has two alleles that differ by one GGAA repeat, the longest is represented here. Gorilla lacks the final "A" in the first motif, represented by the red X at the end of the motif. (b) Shown is the alignment of chimpanzee, Denisova, and 13 of the most common human READ1 sequences. Repeat motifs are labeled with the same color used in (a).

READ1 Functional Groups

Functional Group	Definition	Alleles	Phenotype
RU1-1	One copy of RU1	2, 3, 9, 12, 25, 27	Protective effect for RD in European Cohort [6]
RU2-Long	>= 8 copies of RU2	5, 6, 13, 14, 19, 20, 22, 23	Poor reading performance in European Cohort [6]
RU2-Short	<= 6 copies of RU2	4, 10, 16, 21	Unexpectedly poor comprehension (Li et. al, manuscript in preparation)

Table 1: Description of the three functional groups of READ1 paired with the previously reported associations with reading and language phenotypes



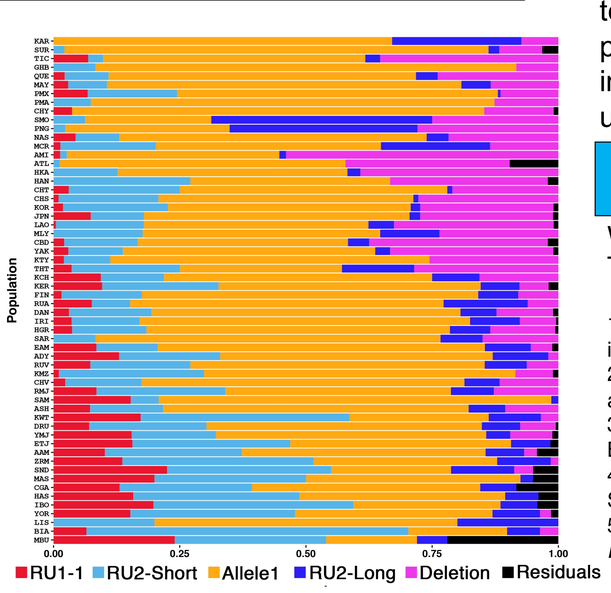


Figure 3: (a) Three functional groups along with Allele 1. Sequence motifs are depicted as colored dots for 4 bp motifs and rectangles for longer motifs. (b) Distribution of READ1 functional groups, Allele1 and the deletion of READ1 in world populations. Populations are ordered in approximate distance out of Africa. Frequencies are shown as percentages of the whole, residuals are the alleles that did not fall into any of the functional groups, Allele 1 or the deletion.

Controlling for General Genetic Relatedness

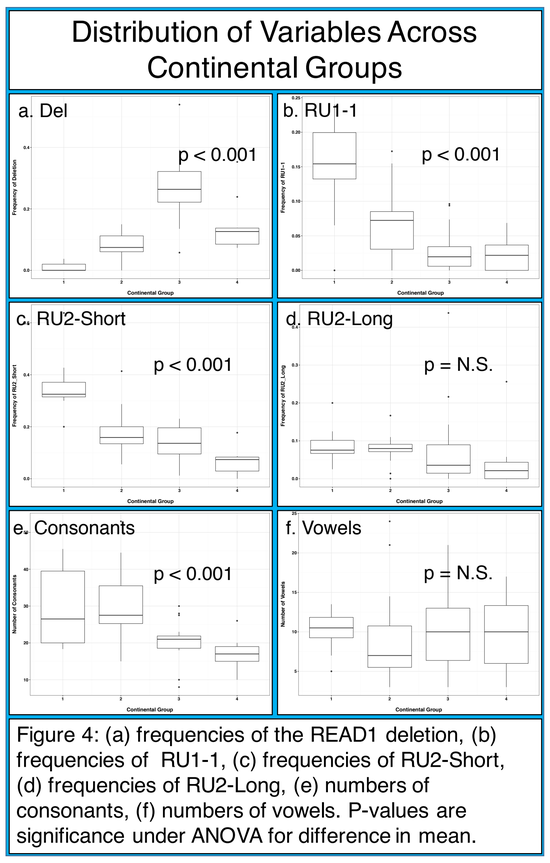


Figure 4: (a) frequencies of the READ1 deletion, (b) frequencies of RU1-1, (c) frequencies of RU2-Short, (d) frequencies of RU2-Long, (e) numbers of consonants, (f) numbers of vowels. P-values are significance under ANOVA for difference in mean.

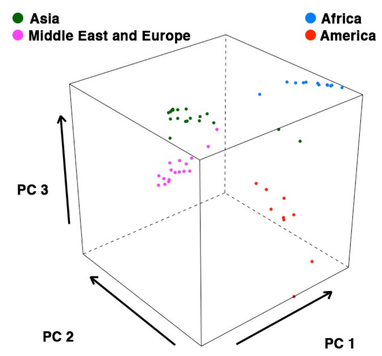


Figure 5: Population sample scores on first three PCs of tau genetic relatedness matrix, colored by continental groups

Controlling for overall genetic relatedness

There are significant associations between continental groups and most of the functional groups of READ1. While there is not a significant association between vowels and continental groups, there is a significant association between consonants and continental groups. To address this concern, we used the first three PCAs in the tau matrix of genetic relatedness between populations as covariates in our multivariate ANCOVA model and included these PCAs as fixed effects and continental grouping as a random effect in our linear mixed effects model.

Tests for Association

a Multivariate ANCOVA				
Variable	Pillai	approx F	p-val	adj p-val
RU2_Short	0.088	2.158	0.127	0.181
RU2_Long	0.085	2.090	0.136	0.181
RU1-1	0.161	4.318	0.019	0.078
Deletion	0.037	0.862	0.429	0.429

b Univariate Responses to RU1-1				
Predictor	Estimate	S.E.	t-stat	p-val
Vowels	-0.648	0.693	-0.936	0.354
Consonants	1.183	0.428	2.761	0.008

c Linear Mixed Effects Model				
Predictor	Estimate	SE	t-stat	p-val
Vowels	-0.002	0.001	-1.936	0.059
Consonants	0.002	6.6e-4	2.590	0.013

Table 2
(a) Results from omnibus MANCOVA analysis examining the effect of READ1 functional groups for vowels and consonants.
(b) Post-hoc examination for the effect of RU1-1 on $\log_{10}$ transformed vowels and consonants separately while controlling for the first three PCs of the tau matrix generated from ancestry specific markers.
(c) Results of a linear mixed effects model using RU1-1 as the response variable while weighting for chromosome numbers, PCs 1-3 as fixed effects and continental grouping as a random effect.

A multivariate ANCOVA examining the relationship between response variables consonants and vowels against the frequency of different READ1 functional groups found an overall significant effect of RU1-1 on the number of vowels and consonants, while covarying for the effects of PCs 1-3 ($p < 0.05$ under Pillai's test). Follow-up univariate analysis evaluating the effects of RU1-1 on vowels and consonants, independently, found that the effects observed in the MANCOVA analysis were driven by variation in consonants, while correcting for general genetic relatedness between populations. Effects of RU1-1 were not associated with number of vowels.

To ensure that differing sample sizes were not skewing the ANCOVA results, we ran a follow up mixed effects model to interrogate RU1-1 frequency as the response variable, weighted by numbers of observed chromosomes and consonants and vowels as fixed effects. The first three PCAs were included as fixed effects and continental grouping as a random effect. The mixed effects model had one significant result and showed a correlation of the number of consonants and the frequency of RU1-1 with a p-value of 0.013, Table 2 c, replicating the multivariate ANCOVA while controlling for sample size.

Conclusions

The goal of this study was to look for association between variation in frequency of alleles of READ1 and variation in core linguistic components used by populations in their languages. We found an association of the READ1 functional group RU1-1 with the number of consonants used by 53 populations. READ1 is a regulatory element in the second intron of *DCDC2* and controls the expression of *DCDC2*. *DCDC2* has been shown to affect neuronal excitability, precision in action potential firing, auditory temporal processing and working memory[3,4,5]. Because of these effects, *DCDC2* may play a role in the determination of perceptual categories that may influence the number of consonants used by a population.

Acknowledgements and References

We would like to thank Judy and Ken Kidd for sharing their DNA samples and expertise.
This work is supported by: The Manton Foundation, NIH/NICHD P50 HD027802

1. Powers, NR, et al. (2013) Alleles of a polymorphic ETV6 binding site in *DCDC2* confer risk of reading and language impairment. Am J Hum Genet. Vol. 93(1):19-28.
2. Powers, NR, *et al.* (2016) The regulatory element READ1 epistatically influences reading and language, with both deleterious and protective alleles. J Med. Genetics Vol. 53(3):163-171.
3. Che, A, *et al.* (2016) Mutation of the Dyslexia-Associated Gene *Dcdc2* Enhances Glutamatergic Synaptic Transmission Between Layer 4 Neurons in Mouse Neocortex. *Cerebral Cortex* **26,** 3705–3718.
4.Centanni, T. M. *et al.* (2016). Knockdown of Dyslexia-Gene Dcdc2 Interferes with Speech Sound Discrimination in Continuous Streams. *J. Neurosci.* **36,** 4895–4906
5. Truong, D. T. *et al.* (2014) Mutation of Dcdc2 in mice leads to impairments in auditory processing and memory ability. *Genes Brain Behav.* **13,** 802–811