

※当資料は、8月19日（木）、文部科学記者会、科学記者会、神奈川県政記者クラブ、平塚記者クラブ、秦野記者クラブにて資料配布を行っております。

2021年8月19日

報道関係各位

東海大学
横浜市立大学
慶應義塾大学

北東アジアの言語族間の遺伝的および文化的多様性の相関関係を解明 ～文法が遺伝的歴史の指標である可能性を示唆。文化的な関係と遺伝的な関係の違い明確に～

東海大学〔伊勢原キャンパス〕（所在地：神奈川県伊勢原市下糟屋 143、学長：山田 清志〔やまだ きよし〕）医学部医学科基礎医学系分子生命科学 助教の松前ひろみ、および慶應義塾大学環境情報学部（所在地：神奈川県藤沢市遠藤 5322、塾長：伊藤 公平〔いとう こうへい〕）准教授の Patrick E. Savage、横浜市立大学木原生物学研究所（所在地：神奈川県横浜市戸塚区舞岡町 641-12、学長：相原道子〔あいはら みちこ〕）客員教授の清水健太郎（スイス・チューリッヒ大学 教授兼任）らの研究グループは、北東アジアとその周辺地域にまたがる 11 の言語族の関係性について、言語（文法、音韻、語彙）、音楽、ゲノムの 5 要素からその相関を検証した結果、文法と遺伝的な歴史の間で統計的に有意に相関していることを明らかにしました。この研究をまとめた論文が、2021 年 8 月 18 日（水）に『Science Advances』へ掲載されました。

＜本件のポイント＞

- ①北東アジアとその周辺地域にまたがる 11 の言語族の言語（文法、音韻、語彙）、音楽、ゲノムを比較・分析し、相関関係を検証。語族を超えた解析は初。
- ②文法が遺伝的歴史の文化的な指標である可能性を示唆。文化的な関係と遺伝的な関係の違いが明確になり、人類の歴史の複雑さが浮き彫りに。
- ③言語や文化のバリエーションと進化の過程を明らかにすることは、ヒト特有の行動と社会からの影響について分析する上で重要。

■研究の背景

文化はヒト特有の精神活動や行動が表出したものであり、ヒトの進化や多様性を理解する上で鍵となる指標です。文化はゲノムと似ているようで似ていない進化を遂げます。特に、言語や音楽（歌）は、知られている限り全ての民族集団がもつ文化で、ヒトのコミュニケーションに必須であるため、ヒト特有の進化の痕跡を調べる上で、重要な要素と考えられてきました。また、言語などの高次機能に関連した一部の障害・疾患は一部の遺伝的因子が知られていますが、マウスモデルでは表現型と遺伝型の関連を明らかにすることが難しい場合もあり、進化研究からのアプローチも検討されています。

ダーウィンは今からちょうど 150 年前に著書『人間の由来』の中で、言語も生物種と同じように漸進的な変化をしてきたのではないかと論じました。そして、遺伝学の誕生により、ヒトのゲノムを解読することで、ヒトの歴史が明らかになってきました。人類遺伝学者のルイジ・ルーカ・カヴァッリ＝スフォルツァは、1980 年代に当時利用可能になっていた 100 程度のヒトゲノム上の座位を調べることで、民族間の遺伝的な系統関係と、言語の分類体系が類似していることを指摘した最初の研究者

の1人でした (Cavalli-Sforza et al, 1988, PNAS)。しかし、言語の分類体系は定性的で、多様な文化データを用いて類似性を定量的に検証した研究はこれまでほとんどありませんでした。

カヴァッリ＝スフォルツァの研究から 30 年以上が経った現在、ヒトゲノムが解読され、民族間の遺伝情報が全ゲノムスケールで解析できるようになりました。言語学では、祖先となる語彙を共有する「語族」という分類で言語をまとめ、その語族内の関係は推定できるようになりました。世界には 7000 以上の言語が知られており、約 400 の語族に分類されます。例えば英語、ドイツ語、フランス語など、ヨーロッパで話される言語のほとんどは、インド・ヨーロッパ語族という大きな語族に属します。インド・ヨーロッパ語族のような同一語族の語彙に基づいた系統解析¹が進みましたが、一方で、語彙で遡れる進化的関係性には限界がありました。しかし、特に北東アジアなど東ユーラシアでは、多様な言語族が存在するため、語族を超えた言語の関係は語彙を用いて定量解析ができないという課題がありました。語彙以外の言語の特徴としては、音素²は語族に縛られずに解析ができますが、言語接触によって隣り合う言語同士の音は似てくることも知られています。文法は、比較言語学・言語類型論という分野で体系化されてきましたが、データベースが限られていました。今回の論文の共著者である Balthasar Bickel らは、2017 年に文法と音素に着目した統合データベースを発表しました (Bickel et al, 2017)。本研究では、言語解析にこのデータベースを利用しました。

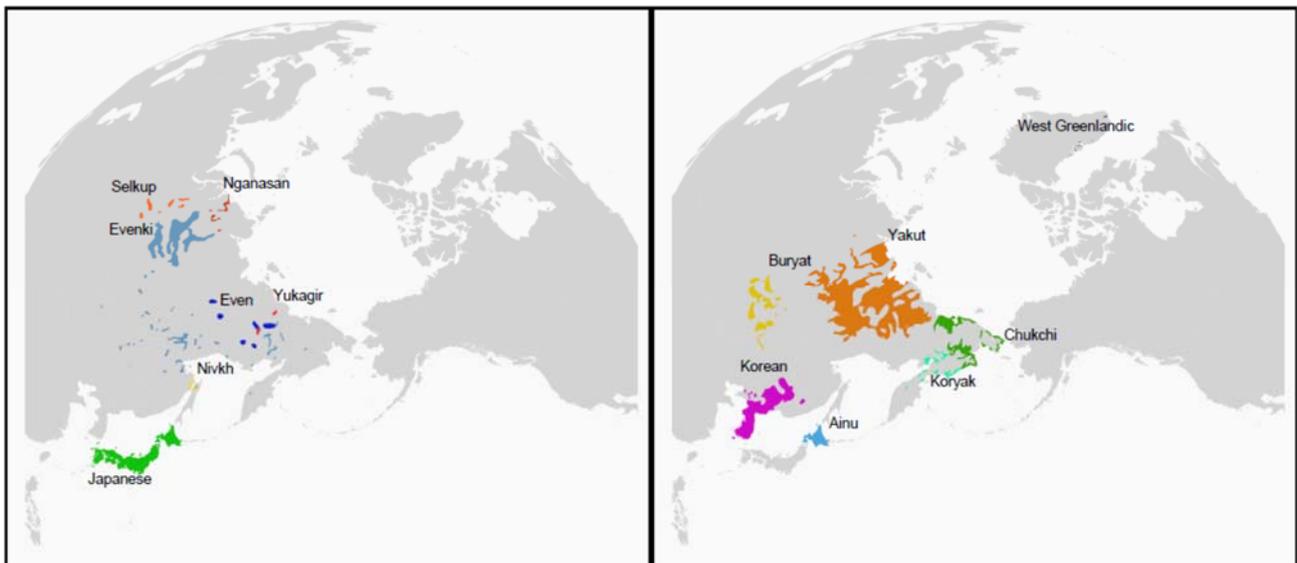
また、カヴァッリ＝スフォルツァの研究に刺激を受けた音楽学者の Steven Brown と Patrick E. Savage (慶應義塾大学環境情報学部) らは、ヒトゲノム同様に、音楽 (歌) を定量化する方法を構築してきました (Savage et al, *Anal. Approaches to World Music* 2012)。そしてこれまで北東アジアの歌の特徴 (Savage et al, 2015) や、台湾先住民の歌の類似性と語彙、遺伝的類似性について明らかにしてきました (Brown et al, 2014, *Proc. R. Soc. B-Biological Sci*)。しかし、語族を超えた解析はこれまでできていませんでした。

¹ 系統解析：生物の遺伝情報(塩基配列またはアミノ酸配列)を用いて、統計的手法により生物が進化してきた歴史を明らかにすること。言語学でも、生物学のアナロジーを用い、語彙を塩基配列に見立て、語彙の歴史を推定することが行われている。

² 音素：ひとつの言語において、意味の違いに関わる最小の音声的な単位

■ 研究内容

今回の論文の責任著者である松前ひろみ、太田博樹（東京大学大学院理学系研究科）らはこれまで、東アジア人や縄文人のゲノム解析により、東ユーラシアの古い遺伝的な歴史を明らかにしてきました（McCull et al Science 2018; Gakuhari et al Comm Biol. 2020）。また、近年では、国際的にも東ユーラシアの基層集団として北東アジアに注目が集まっています（Jeong et al, Nat. Ecol. Evol. 2019）。そこで本研究では、北東アジアとその周辺地域にまたがる 11 の言語族【図 1】の関係に焦点を当て、言語、音楽、ゲノムを比較する分析を行いました。

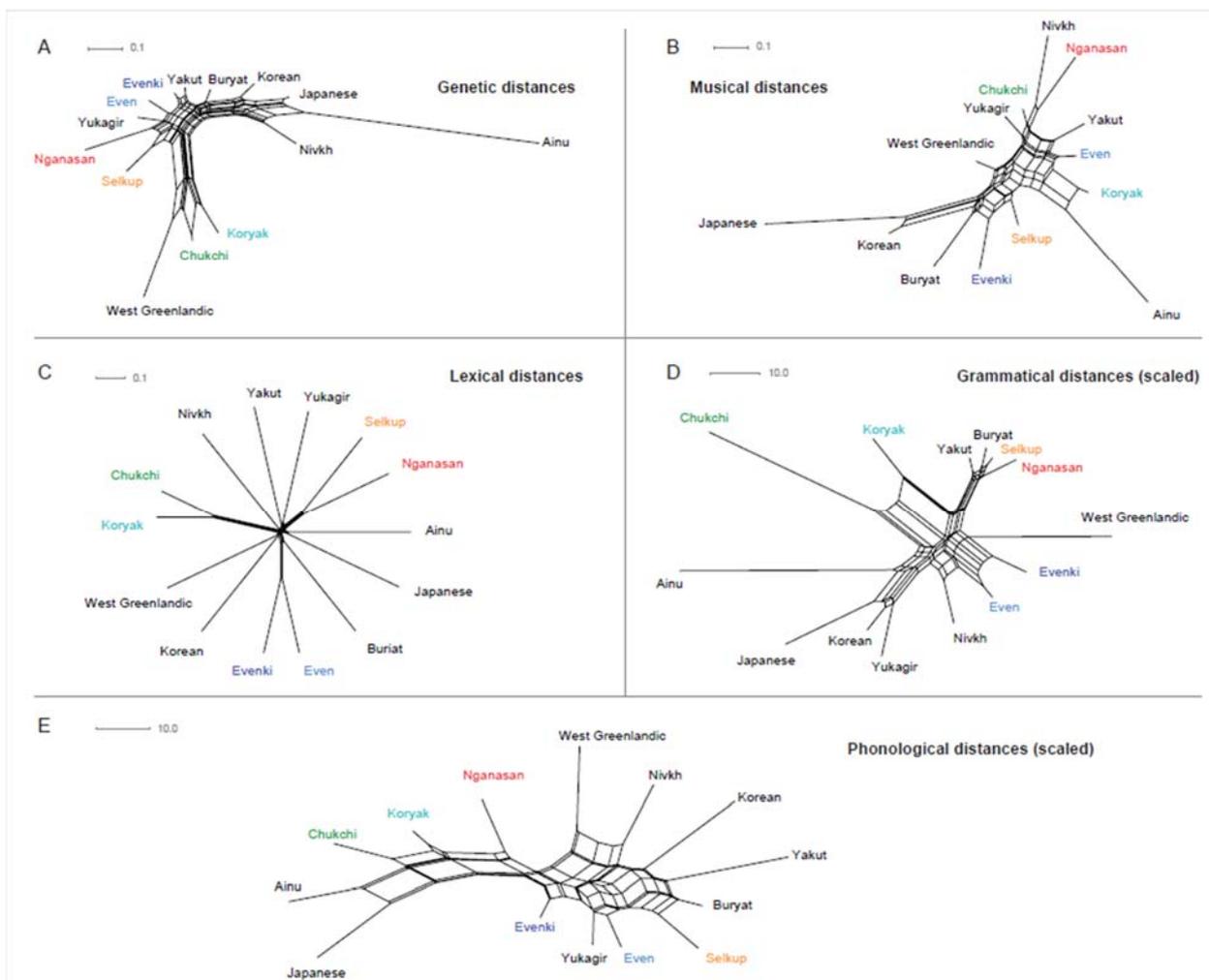


【図 1】北東アジアとその周辺地域にまたがる 11 の言語族に相当する 14 言語の分布。日本語族（日本語）、アイヌ語族（北海道アイヌ語）、朝鮮語族（韓国語）、ツングース語族（エヴェン語・エヴェンキ語）、チュクト・カムチャツカ語族（チュクチ語・コリヤーク語）、テュルク語族（ヤクート語）、モンゴル語族（ブリヤート語）、エスキモー・アリュート語族（西グリーンランドイヌイット語）、フィン・ウラル語族（セルkup語、ガナサン語）、ニブフ語族（ニブフ語）、ユカギール語族（ユカギール語）

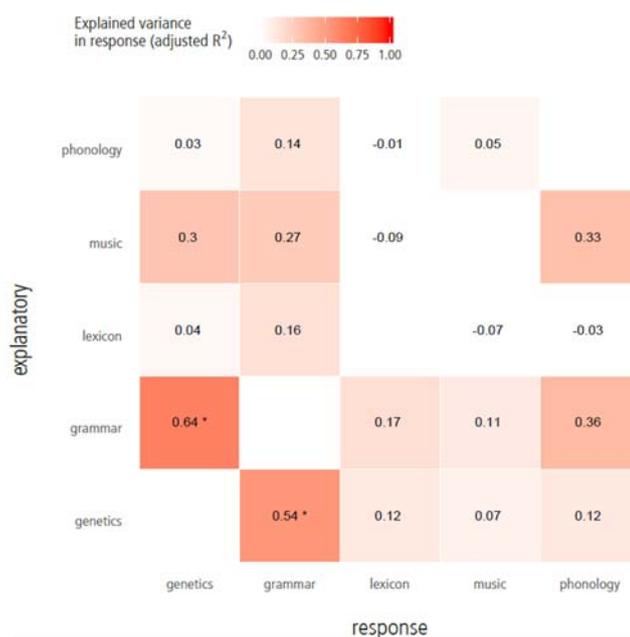
研究グループは、言語（文法、音韻、語彙）、音楽、ゲノムの 5 要素を主成分分析や主座標分析を用いて、同じデータ形式に落とし込みました。そして 5 つの要素の類似性の可視化を行いました【図 2】。

さらに 5 要素間の関係を、冗長性解析³により相関を検証しました。その結果、全ての組み合わせの中で、言語のうち、文法の類似性のみが、遺伝的な歴史と統計的に有意に相関していることがわかりました【図 3】。

³ 冗長性解析（Redundancy Analysis: RDA）：変数セット間の関係（説明変数で説明できる応答変数の変動）を抽出する多変量解析法の 1 つ。ここでは 5 つの因子（遺伝、語彙、文法、音素、音楽）の組み合わせを分析した。



【図2】5つの要素〔言語（文法:grammar、音韻:phonology、語彙:lexicon）、音楽、ゲノム〕の関係を可視化

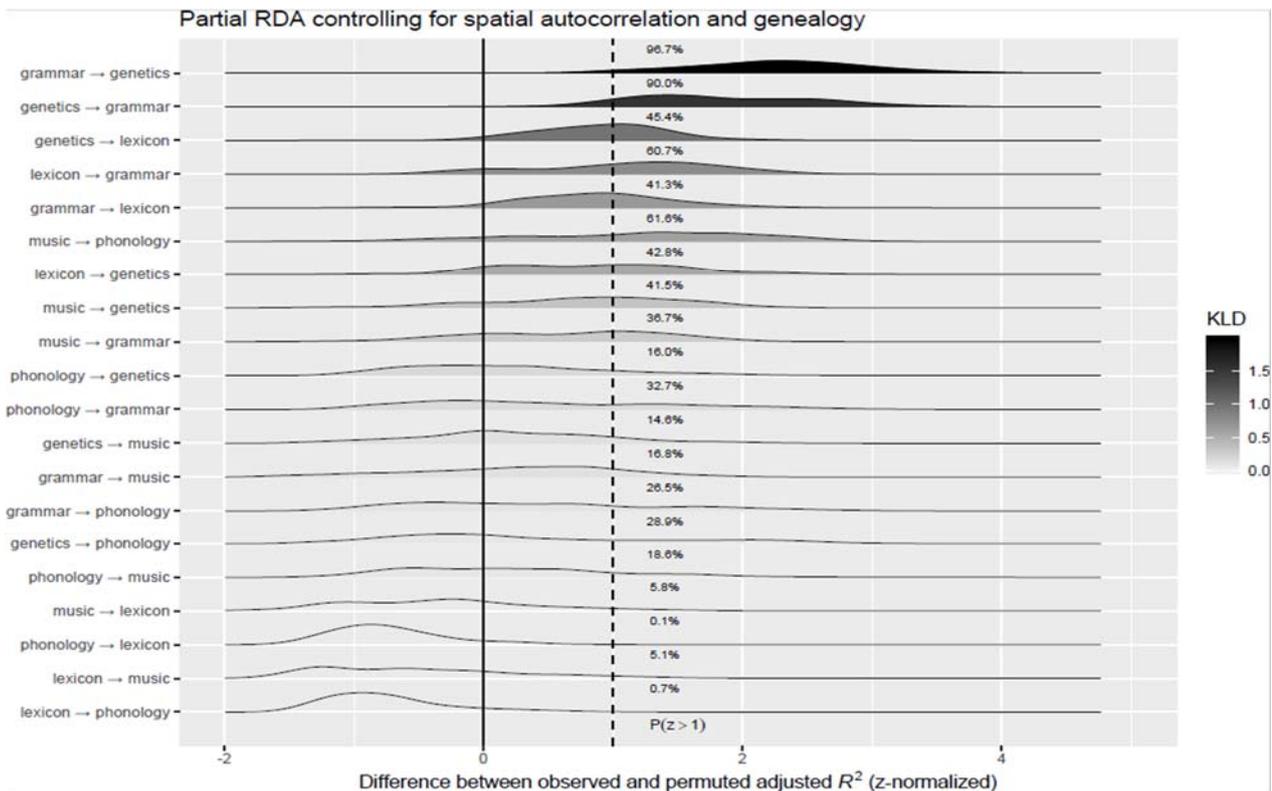


【図3】5つの要素〔言語（文法、音韻、語彙）、音楽、ゲノム〕を冗長性解析で分析した。
*は $p \leq 0.05$ を示す。

では、なぜ言語のうち文法と遺伝的な歴史の間に関連が見られるのか、その説明として以下の3つの仮説を立てました。

- (1) 最近の文化（言語）接触により、隣り合う言語同士は似てしまった可能性（地理的な自己相関で説明できる場合）
- (2) データセットに含まれる言語の中には、同一語族に属する言語が複数あるため、そうした系統的に近い言語のバイアスが反映された可能性
- (3) 上記2つでは説明できない場合、語族間の古い類似性を示している可能性

上記(1)の仮説については、言語が話されている地域の地理情報を用いて、空間的自己相関の影響を除去するような解析を行いました。上記(2)の仮説については、同一言語族に属する言語はランダムサンプリングすることで、言語間のバイアスを抑える解析を行いました。その結果、ゲノムと文法の類似性の相関は、最近の言語同士の接触や、同一言語族によるバイアスではこのシグナルを説明できず、現在の語族が形成される前に生じた関係を示唆していました【図4】。



【図4】空間的自己相関と言語間のバイアスを除去した partial RDA 解析の結果。縦軸は2つの因子間の関係（説明変数→応答変数）。横軸は、本研究で得られた相関関係が偶然（ランダムサンプリング）で説明する場合からどのくらい離れているかをZスコアという指標で示しており、1000回繰り返した結果を分布（カーブ）で示している。分布が黒線（y=0）近くにある場合は偶然と区別が付きませんが、右側に行けば行くほど、偶然では説明しにくいことを意味する。また1000回の繰り返し実験を行った場合、そのZスコアが1より高い値になった割合を%で示している。更にその差をカルバックライブラー情報量（KLD）という指標でグレースケールで示している。例えば一行目の grammar → genetics は96.7%の再現性をもって文法間の類似性で遺伝的距離を説明できることを意味しており、genetics → grammar は遺伝的距離で文法間の類似性を説明した場合に90.0%の再現性があったことを示している。

■本研究の意義と今後の展望

本研究成果は、文法が遺伝的歴史の文化的な指標である可能性を示唆していると同時に、文化的な関係と遺伝的な関係の違いを示し、人類の文化や歴史の複雑さを浮き彫りにしていると考えられます。民族間の文化と遺伝的な歴史の関係は西ユーラシアでは研究が進んでいましたが、言語学的複雑性が高い日本を含めた北東アジアでは、方法論的な限界により関係性が分かっていませんでした。本研究により、1980年代にカヴァツリ＝スフォルツァが課題とした、定量的に文化を測る手法が新たに確立しました。そして、北東アジアにおいても言語の文法の類似性が、遺伝的關係に遡る情報を維持している可能性が示されました。このことは文化が入り組んでいる北東アジアの歴史の解明を更に推し進める上で重要になると考えられます。

また、デジタル化された文化をデータ解析することにより、北東アジアのように複雑な文化を持った地域の研究に展開できるほか、文化データを有効活用することに繋がられることを示しています。こうした言語や文化のバリエーションと進化の過程を明らかにしていくことは、ヒト特有の行動と社会からの影響について分析する上でも重要だと考えています。

本研究は、日本、スイス、ドイツ、カナダ、英国にまたがる国際的共同研究によって進められ、生物学、言語学、音楽学、統計学の学際的な研究により成果を得ました。なお、本研究成果は、以下の外部資金（抜粋）等によるものです。

- ・ 科研費・新学術・共創言語進化 JP18H05080、JP20H05013
- ・ 科研費16H06469
- ・ 科研費19KK0064
- ・ チューリヒ大学「URPP Evolution in Action」 「URPP Language and Space」
- ・ スイス National Center of Competence in Research「Evolving language」

また本研究を行うにあたり、本研究ではこれまで世界的にほとんど存在しなかった、サハリン先住民のニブフのゲノムデータをSNPアレイを用いて分析しました。このDNAサンプルは、1990年代に速水正憲博士（京都大学ウイルス研究所教授・当時）らが収集し、ヒトT細胞白血病ウイルス（HTLV-1）やヒト白血球抗原（HLA）、mtDNAの分析等に利用したもので（Gurtsevitch V, et al. 1995. *Int. J. Cancer*; Lou et al, 2008 *Tissue Antigens*; Tajima et al, 2004, *J Hum Genet*）、故・宝来聡博士（総合研究大学院大学教授・当時）、田辺秀之博士（総合研究大学院大学准教授）らおよびAsia DNA Repository Consortiumによって維持・管理されているサンプルの提供を受けました。

【掲載論文】

雑誌名 : 『Science Advances』（2021年8月18日掲載）
タイトル : Exploring correlations in genetic and cultural variation across language families in Northeast Asia
DOI : <https://doi.org/10.1126/sciadv.abd9223>

【筆者】

松前ひろみ^{a,b,†,‡,*}、Peter Ranacher^{‡,c,d,*}、Patrick E. Savage^{e,f,*}、Damián E. Blasi^{g,h,i}、Thomas E. Currie^j、
小金淵佳江^k、西田奈央^l、佐藤丈寛^m、田辺秀之ⁿ、田嶋敦^m、Steven Brown^o、Mark Stoneking^p、
清水健太郎^{a,b,q}、太田博樹^{k,r,s,*}、Balthasar Bickel^{g,q,*}

^a Department of Evolutionary Biology and Environmental Studies, University of Zurich

^b 横浜市立大学木原生物学研究所

^c Department of Geography, University of Zurich

^d URPP Language and Space, University of Zurich

^e 慶應義塾大学環境情報学部

^f 東京藝術大学

^g Department of Comparative Language Science, University of Zurich

^h Department of Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History

ⁱ Human Relations Area Files

^j Human Behaviour & Cultural Evolution Group, Centre for Ecology & Conservation, Department of Biosciences, University of Exeter

^k 北里大学大学院医療系研究科

^l 国立国際医療研究センター

^m 金沢大学 医薬保健研究域医学系 革新ゲノム情報学分野

ⁿ 総合研究大学院大学 先導科学研究科

^o Department of Psychology, Neuroscience & Behaviour, McMaster University

^p Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology

^q Center for the Interdisciplinary Study of Language Evolution (ISLE)

^r 北里大学医学部

^s 東京大学大学院理学系研究科

[†] Current Address: 東海大学医学部

[‡] **These authors contributed equally to this work.**

*** Corresponding authors**