

# 文化×バイオ×コンピュータでの解析-言語類型論に基づく言語データベースとゲノムデータの統合的解析の提案-

松前ひろみ<sup>1</sup> 長谷武志<sup>2</sup> 清水健太郎<sup>3,4</sup>

**概要**：ヒトのゲノム情報からは、人類の系統や混血などの歴史を詳細かつ統計的に推定することができる。そうしてゲノムから推定される民族集団史の系統関係と、文化、とりわけ言語の分類には、一定の関連があると考えられてきた。しかし言語の分類のうち、言語族という語彙レベルで近縁な言語間の関係を除くと、遠い言語同士の関係（例えば日本語、アイヌ語、韓国語）を定量的に分析することは、これまで困難であった。私たちは文法の比較法である言語類型論の研究者とともに、文法のデータベースから定量的に語族を超えた言語の特徴の類似性を抽出し、ゲノムに基づく民族集団の関係との関係性を分析することに成功した。データベースに蓄積された文化と生物学の情報を統合的に解析する方法を提案する。

**キーワード**：言語類型論, 人類集団史, ゲノム, 文化進化

## 1. 背景

ゲノムとは、その生物が持つ全ての DNA 情報のことである。2001 年にヒトのドラフトゲノムが解読されて以来、その解析技術の飛躍的な進歩により、世界中の様々な民族集団に由来する個人のゲノム情報の解読とデータベース上での蓄積がなされてきた。加えて遺跡から発掘された数千～数万年前の人々の遺骨から DNA を分析し、現代の人々のゲノムと直接比較をすることも可能となっている。その結果、集団遺伝学と呼ばれる分野ではゲノムを用いて、文字を持たない社会や時代も含めた人類の歴史（系統関係、分岐年代推定、混血などの移住史、人口動態の変遷など）について、統計的に高い精度での解明が進んでいる[1]。

民族集団間の文化の類似性もそうした遺伝的背景と関連があるのではないかというアイデアは、(現在から見ると極めて)断片的な DNA 情報が入手可能になった 1990 年前後には存在していた。遺伝学者の Cavalli-Sforza らは、DNA に基づく世界の民族の系統関係に、言語学者による言語の分類をマッピングするという、先駆的な研究を行った[2]。しかしこうした研究には批判も根強かった。遺伝学は理論が先行して確立し、次にデータの生産・規格化・蓄積が進行し、理論を支持するパラメータも実験・経験的に得られてきている。DNA は A,T,G,C の 4 つの塩基、すなわち物質的要素で構成されるが、集団遺伝学では主に世代交代で生じた DNA 上のこれらの変異が集団中に蓄積していく過程や結果を見ることで、統計的に歴史を遡る（それを集団史と呼ぶ）。一方、言語を始めとする文化の伝達や進化の様式については、そもそもこういった要素をデータ化すべきなのか、またどのように伝達・進化していくのか不明な点が多く、議論の余地が多い[3]。特に言語はゲノムに比べ進化速度が速いと考えられており、同一祖先に由来する語彙(祖語)を共有する”言語族”内の言語同士の関係(例えばインド・ヨーロッパ語族等)を除くと、語族を超えた遠い言語同士の関係性を推定することは難しかった。例えば日本語と

韓国語はそれぞれ別の語族に由来するため、借用語を除くと語彙の共有も極めて少ない。世界の人類集団全体を遺伝的に見れば、韓国人と本州の日本人は遺伝的に互いに最も近い民族集団の一つであるにも関わらず、言語学的に見れば語族という壁が存在する。世界には現在知られているだけで約 7000 語が存在し、それらは約 400 の語族に集約される[4]。韓国語と日本語に限らず、これら 400 語族が歴史的に互いにどのような関係をもつのか、Cavalli-Sforza らのアイデアを定量的に検証するには、言語はデータや解析方法の不足で実証が長らく困難であった。

近年、異なる言語の文法を比較する言語類型論の研究者である Bickel らは世界の 2,000 以上の言語について文法と音素の特徴を集約した AUTOTYP データベースを構築した[5], [6]。これにより、言語でもゲノムに対応させられるような、データに基づく研究が可能となった。そこで我々は、Bickel らのデータベースを用いて、ゲノムに基づく人類の歴史と文化、とりわけ語族を超えた言語の類似性には相関はあるのか、という人類学上の課題へ挑戦することにした。特に我々はこれまで、東アジアの民族集団史の解明に取り組み、例えば約 2400 年前の縄文人や、地理的に隔離されやすい島嶼部やシベリアの先住民族と、現代日本列島の集団の間における関係性を掘り下げてきた[7], [8]。興味深いことに、東アジアは、研究が進んでいる西ヨーロッパに比べると、文法レベルでの言語の多様性が高いことが、以前より指摘されている[9]。また音楽(歌)の多様性からは言語やゲノムとは異なる民族集団間の関係性が示唆されていた[10]。そこで、東アジアにおける集団史と、言語・音楽の多様性の関係についてケーススタディとして取り上げ、実データでの解析を行った[11]。そこでゲノムと文化的データという本質的に異なるデータを統合的に解析するための枠組みを構築した。本報告では、既にデータの解析方法が確立している生物学的データに対応させつつ、人文科学のデータを活用するための解析方法に焦点を当てて報告する。

## 2. 各要素のデータ化

異なる性質のデータを直接比較するには、データがもつ特性に配慮したデータ化と解析が必要である。前述との通り、ゲノムデータから歴史的情報を引き出すには集団遺伝学という分野が確立されているため、それらの標準的な方法論に従った[12]。本研究で利用した文化的データは言語（文法及び音素）と音楽（歌）であり、それらは既に要素化されてデータベースや論文として公開されているものを利用した。ゲノムと文化の各データの解析にあたり、本研究では、集団遺伝学のアナロジーや、分野を超えて一般的によく利用される多変量解析の手法を組み合わせ文化データを解析した。実は集団遺伝学のアナロジーを文化の解析へ適用すること自体は、それほど珍しいことではないが、後述の通り、データの構造や特性をよく理解して使い分けの必要がある。

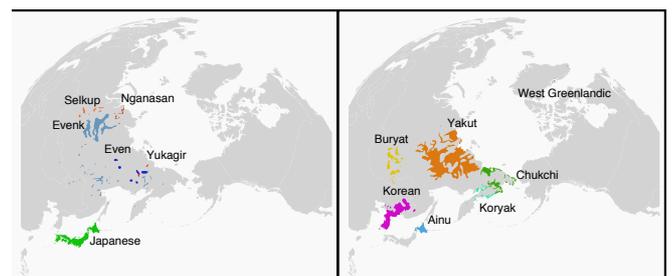
### 2.1 対象となる民族集団の選択

ゲノム、言語、文化の情報が、各々の分野で議論するのに十分な解像度で集約的にアノテーションされているデータベースはほぼないと言って良い。そこで、広い意味での東アジア（シベリアや北極圏を含む）から、ゲノム、言語、音楽（歌）の三つのデータが揃っている民族集団を13集団・言語を、既に論文で公開されているデータを中心にアノテーションし、選択する作業から行った(表1, 図1)。

表1 本研究で用いた民族集団と対応する言語族

集団/言語名 (英語)	言語族	言語に対応する ID [ISO][AUTOTYP][Glottolog]
本州日本人 Mainland Japanese	ジャポニック	[i-jpn][a-118][g-nucl1643]
韓国人 Korean	コリアニック	[i-kor][a-141][g-kore1280]
北海道アイヌ Hokkaido Ainu	アイヌ	[i-ain][a-12][g-ainu1240]
エヴェン Even	ツングース	[i-eve][a-738][g-even1260]
エヴェンキ Evenki	ツングース	[i-evn][a-527][g-even1259]
西グリーンランド・ イヌイト West Greenlandic Inuit	エスキモー・ アリュート	[i-kal][a-511][g-Greenlandic kala1399]
コリヤーク Koryak	チュクト・ カムチャツカ	[i-kpy][a-1808][g-kory1246]
チュクチ Chukchi	チュクト・ カムチャツカ	[i-ckt][a-56][g-chuk1273]
ブリヤート Buryat	モンゴル	[i-bua][a-1095][g-buri1258]
ガナサン Nganasan	フィン・ウラル	[i-nio][a-2172][g-ngan1291]
セルクープ Selkup	フィン・ウラル	[i-scl][a-2393][g-selk1253]
ヤクート Yakut	テュルク	[i-sah][a-2662][g-yaku1245]
ユカギール Yukagir	ユカギール	[i-yux][a-2797][g-sout2750]

図1 本研究で利用した集団・言語の地理情報



13集団の話す言語は、語族としては10語族に相当する。それぞれの言語には既存の3つの言語学のデータベース（ISO, AUTOTYP, Glottolog）と相互参照可能なIDを対応付けた。また話者の情報から、各言語が話されている地理情報を取得した(図1)。当然ながら、集団・言語・文化の選択には、地域史についての一定の知見や仮説がないと難しいため、遺伝学者、言語学者、民族音楽学者の間で、仮説検証に最適な集団・言語の選択について議論しながら作業を進めた。

### 2.2 ゲノムデータ

ゲノム情報の取得には様々な技術があるが、今回はSNPアレイと呼ばれる手法で得られたデータを論文から収集して用いた[11]。統計的な誤差を減らす目的から、各集団につき6人以上の近親ではない個人のデータについて、ゲノム上の68,658箇所の塩基を解析に用いた(表2)。

表2 ゲノムデータのイメージ例。この例では、本州日本人は3人、韓国人は2人分の行で表されている。この表の5人全てが68,658列からなる遺伝情報の変数をもつ。

個人ID	集団・言語	ゲノム上の位置						68,658
		1	2	3	4	5	...	
JPT_1	本州日本人	A	G	C	T	G		C
JPT_2	本州日本人	A	G	A	C	T		A
JPT_3	本州日本人	T	G	A	C	G		A
Korean_1	韓国人	T	G	A	C	T		A
Korean_2	韓国人	A	G	A	C	G		C

このデータに対して、集団遺伝学で用いられる代表的な指標の中である $\phi_{st}$ を適用し、集団間の遺伝的距離を求めた。これにより、1集団につき複数個体が含まれるデータから、集団ペア間の差を表す距離行列が生成された(図2)。

図2. 実際に得られた民族集団間の遺伝的距離。数値がゼロに近いほど遺伝的な差が少なく、大きな値ほど民族間の差が大きい。後述の過程で、このような集団間・言語間の距離行列を言語・音楽でも算出した。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Chukchi	0												
2	Nganasan	0.26375	0											
3	Evenk	0.04385	0.02315	0										
4	Koryak	0.02301	0.06195	0.04017	0									
5	Yakut	0.02428	0.22486	0.21058	0.03182	0								
6	Sekup	0.06505	0.04774	0.03059	0.06177	0.04134	0							
7	Korean	0.064	0.05744	0.02786	0.06072	0.02729	0.05634	0						
8	Alu	0.1048	0.1032	0.07088	0.1002	0.03849	0.1002	0.0628	0					
9	Buryat	0.05058	0.03721	0.03307	0.04762	0.03566	0.03871	0.01755	0.07073	0				
10	Evenk	0.0513	0.0289	0.00784	0.04932	0.003573	0.03843	0.02773	0.08032	0.01508	0			
11	Japanese	0.06329	0.05797	0.03183	0.06197	0.03125	0.0608	0.002494	0.0404	0.02204	0.03229	0		
12	westGreenland	0.0469	0.07163	0.0463	0.06129	0.08864	0.05346	0.06695	0.1109	0.05227	0.06682	0.07138	0	
13	Yakug	0.0589	0.04819	0.02105	0.05457	0.03117	0.03777	0.04805	0.09511	0.0313	0.02937	0.05229	0.0472	0

## 2.3 言語データ

AUTOTYP データベースにある言語類型論のデータは遺伝情報と異なり個人差がデータベース化されているわけではなく、各言語の文法的特徴を1つの代表値で縮約できる点が大きく異なる。表3に言語が持ちうる文法の要素を単純化した例を挙げた。例えば、言語の語順は、主語(S)-動詞(V)-述語(O)で表すことができる。多くの日本語を母国語とする者が英語の文法を習ったときに、日本語はSOV型で、英語はSVO型、と教わるのと同じである。このように語順はS,V,Oの組み合わせでできるため、語順を示す組み合わせは、SVO,SOV,VSO,VOS,OSV,OVSの6つ、それに特定の語順に定まらないパターンを含めると7つである(世界にはVやOで始まる言語もある)。DNAがA,T,G,Cの4変数を取る一方で、語順は7変数を取る、という数字だけの違いだけでデータの構造は似ている。また他の文法要素の例としては、動詞が過去形を取るか否か、というものがある。その場合、過去時制を持たない言語は少なくない上に、逆に近い過去と遠い過去を区別するような言語も存在するため、3パターン以上が存在する。このように一概に文法の変数といっても多様な要素を包含する。なおAUTOTYPにおける実際のデータ構造は、この例よりも更に複雑であり、その理解には言語類型論の知識が必要である(詳細はAUTOTYPデータベースを参照のこと)。このように、言語のデータは、そのままではゲノムデータとは構造が異なるため、直接比較ができない。そこで13言語で共有する12個の文法的特徴について主成分分析を行った。主成分は距離行列に変換できることが数学的に証明されており、主成分を言語ペアの距離行列に変換することで、ゲノムデータとの互換性を得た。

表 3. 言語データの簡易的なイメージ

言語	文法要素 1 (語順)	文法要素 2 (過去形)	文法要素 3...
本州日本	SVO	あり	
北海道アイヌ	SVO	なし	
チュクチ	決まった語順なし	なし	

また言語の音素については、文法よりはデータの定量的な解析が進んでいる。しかし、音素は、言語の本来の”系統的な関係”よりも、最近の言語接触による影響を強く受けると言われている。そこで文法と比較する形で、AUTOTYP上の音素のデータも同様の手段で解析した。

## 2.4 音楽(歌)データ

音楽(歌)の選択と、音楽要素の抽出と定量化は、民族音楽の専門家であるSavageとBrownらがこれまで構築してきた手法を利用した[10], [13]。音楽のデータは、ゲノムのデータと同様に、1集団につき複数の歌を含み(表4)、また同じような(近縁的すぎる)歌は含まないように事前に整理されている。各歌の要素は、歌詞は含まず、音楽の構造やメロディについて、リズムや日本民謡のこぶしの有無まで、民族音楽学に基づく指標から、41の要素として数値化されている(表4)。そこで遺伝情報のアナロジーからφstに基づく集団間の距離を求め、ゲノムデータとの直接比較を可能にした。

表 4 音楽データの簡易的なイメージ。データの構造がゲノムデータと同じであることに着目。

歌	集団・言語	41の音楽要素			
		1	2	3	...
盆踊り1	本州日本人	n/a	iii	1	
子守歌1	本州日本人	a	i	4	
茶摘み歌1	本州日本人	n/a	ii	3	
子守歌_1	韓国人	a	ii	2	
漁の歌_1	韓国人	n/a	i	1	

## 3. 可視化と統合的解析

### 3.1 関係性の可視化

データの可視化は、未知のデータ構造やその学術的意義を議論する上で重要である。遺伝的距離に対して最もよく利用される可視化の方法は、木構造で表現された進化系統樹である。進化的に同一祖先をもつ種の間を木構造で表現したものを進化系統樹という。木構造は人間にとって直観的に全体像を把握しやすい。その反面、進化生物学において系統樹と木構造は厳密に区別されるものであるが、それ以外の分野では両者が混同されて理解されている例が少なくない。また木構造を超えた関係性を理解するには適していない。集団遺伝学においては、仮定した系統関係に対して、その系統関係を超えた関係性(例えば、系統的には異なる遠い民族集団同士が移住により接触して混血した場合)を検出する指標や方法が複数提案されている。このようにゲノムでは木構造の限界を踏まえた上で、系統樹の情報を解釈する。文化では、進化系統樹の前提である「同一祖先から枝分かれして発生した」かどうか明らかではないケースが多い(そもそも独立に進化した他人のそら似かもしれない場合には同一祖先を共有するという仮定が成立しないし、同一祖先を共有していたとしても進化が早すぎて過去の情報が要素間で共有されていない場合などが考えられる)。そこであくまでも類似性を見る方法として木構造を利用することをお勧めする。なお系統樹作成法については複数の方法があるので、教科書などを参考にされたい

[12]. 参考までに、遺伝的距離(図2)と文法の距離行列を、一般的に用いられる系統樹作成法の NJ 法と、混血を表現する Neighbornet 法で表現した例を掲載する[14]. 日本列島の集団の遺伝的背景は、単系統ではなく、縄文人の子孫であるアイヌと、渡来系弥生人に近い集団である韓国人や中国の漢民族の混血で説明できることが分かっている[15]. 従って、遺伝的距離に基づく集団の系統樹を二つの方法で作製すると、図3のように、Japanese は Ainu と Korean の中間に位置するようになる. さらに同じ方法で文法の距離を比較してみると、図4のようになる. ここでは、コリヤーク-チュクチ、エヴェン-エヴェンキのように同じ語族に属する言語がクラスターするだけでなく、異なる語族の韓国語と日本語が常にクラスターすることが見て取れた. ただし、文法の進化メカニズムがよく分かっていない現時点では、あくまでも言語の系統ではなく類似性を反映しているに過ぎないことに留意する.

図3. 遺伝的距離に基づく集団の系統樹. 左側が NJ 法による無根系統樹, 右側が Neighbornet による系統樹.

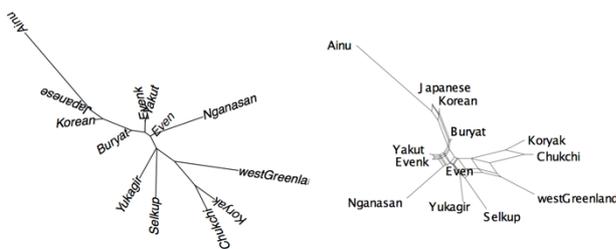
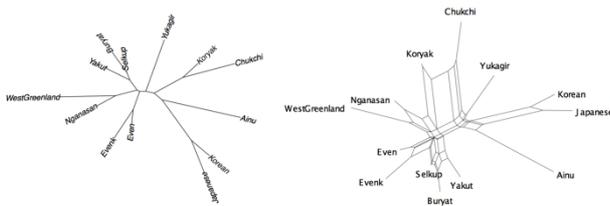


図4. 遺伝的距離と同じ方法で文法の距離行列を系統樹作成法で表現したもの. 左側は NJ 法無根系統樹, 右側は Neighbornet.



また階層的クラスタリングは似たデータをまとめるのに適しているが、二分木による階層構造がやはり直線的な関係性への印象を与えることがある. そこで、我々は木構造を避けてデータを可視化する方法を模索している. 例えば、レーダーチャート(スパイダーチャート)やネットワーク解析などが代替手段となりうると考えている.

### 3.2 統合的解析

ゲノム、言語、音楽という全く異なる性質のデータを、距離行列や主成分に落とし込むことで、同じレベルでその関係性を解析できるようになった. 本研究では冗長性解析という手法を用いて二つの変数間の関係性を分析した. さ

らに、統計地理学の観点から、言語や民族集団に対応する地理情報を、単一の点を(例えば本州の日本人なら東京を代表点として)取得するのではなく、図1で示したようにその言語がカバーするエリアから 1000 回、地点をランダムサンプリングすることで、統計的な説明力を強化することにした.

冗長性解析の実施に際し、今回のデータセットの特徴を踏まえ、二つの点に留意した. 第一に、多くの生物学的・環境的因子は空間的距離と相関することが経験的によく知られているため、地理情報の制御を行った. 有名な例では、ヨーロッパ内の民族集団の遺伝的な近さは、地理的な近さと関連があることが分かっている[16]. 加えて、文化に目を向ければ、地理的に隣り合う文化ほど近い関係にありそうなことは直観的に理解できるだろう. 地理的な近さと言語の類似性の近さが、系統的な関係に由来するのか、それとも異なる系統関係に由来する言語が移住などの関係で地理的に隣接することで生じる言語接触で説明されるのか、区別することは本研究の本来の目的には重要である. 従って、地理的距離による自己相関を除去する操作を行った.

第二に、表1に示したとおり、同一語族に属する言語が含まれているため、近い者同士の関係が過大評価される可能性があった. そこで、同一語族の言語の場合、ランダムに片方を選択するような処理を行い、言語の系統的な影響を抑えた.

その上で、冗長性解析によって、ゲノム、言語(文法と音素)、音楽、という4つの因子の関係を解析した結果、ゲノムと文法の類似性間に一定の関係性があり、その他の因子ペアにはそうした関係性が見出せないことを統計的に提示することができた[11]. この結果は、これまで人類学や言語学で直観的に考えられてきた、人類の系統関係と言語分類の体系の間に、歴史的な関係性があることを示す初めての直接的かつ定量的な証拠となった. 加えて、この結果は、文化独自の進化についての研究を進展させる可能性を秘めている. 文化を遡ってその分岐年代を推定することは現在の知見では極めて困難であるため、文化の類似性がいつごろ生じたものに由来するのか、そのままでは解釈が難しい. しかし、ゲノムからは集団間の分岐年代について一定の情報を取り出すことができるという利点が存在する. そこで、ゲノムに基づく分岐年代の推定が分かれば、それが文法においていつ言語が分かれたのか、間接的な知見を得ることが可能になると考えている.

### 4. まとめ

人文科学のデータは、データの規格化と蓄積という課題だけでなく、文化進化という観点ではデータの解析手法も発展途上である. 人文科学のデータを活かす一つの方法として、我々の研究は、「人類の民族集団史と言語の類似性の間に関連はあるのだろうか」という自然科学と人文科学を

包含する人類学上の課題を解決しようとする中で、生物学で培われてきたゲノム解析の手法やアナロジーを人文科学のデータの解析に応用できる新たな可能性を示した。また生物学と人文科学のデータを統合的に解析するには、データを共通の統計解析に落とし込めるようなデータ変換が必要となる。それには元データや解析手法の特性をよく踏まえて解析を進めていくことが重要になる。これはデータベース上にどのようにデータを規格化するか、データベースの設計にも繋がる問題である。従って、人文科学データから統計的に何かを見出すには、データベースの設計が鍵を握ると言っても過言ではない。今回使用した文法のデータは言語類型論に基づいているが、言語類型論とはまさに文法の形式的な特徴を規格化していく言語学の一分野であり、他の文化データに比べるとデータベースが整備されており、データ解析に入りやすいという利点があった。他の文化的データで解析を進めていく場合には、音楽で Savage が行ったように、データの形式化から丹念に進めていくことが求められる。また生物学を始めとする自然科学では、経験則的なデータ解析のノウハウが蓄積している。例えば、複数の異なるデータ解析手法でも概ね似たような結果得られるかどうか、再現性を確認することは一般的である。解析手法間で再現性が得られない場合は、関係性が弱いからではなく、それらの手法がデータの異なる側面を抽出している場合もあるので（例えば上述の系統関係から外れた混血の関係など）、解析手法の特性を理解しておくことが重要である。加えて繰り返し述べたように、文化進化的なアプローチに、進化生物学や集団遺伝学のアナロジーを利用する場合には、注意が必要である。

最後に、本報告で説明しきれなかった本研究の詳細な結果や、利用した専門的なソフトウェア、コードについては、生物系のプレプリントサーバである bioRxiv に、本文及び Supplementary Information として掲載しているので[11]、そちらを参照して頂きたい。

**謝辞** 本研究は JSPS 科研費(新学術・共創言語進化)JP18H05080 及びチューリヒ大学重点化プログラム URPP Evolution in Action の助成を受けたものです。

## 参考文献

- [1] R. Nielsen, J. M. Akey, M. Jakobsson, J. K. Pritchard, S. Tishkoff と E. Willerslev, 「Tracing the peopling of the world through genomics」, *Nature*, vol. 541, no. 7637, 2017, doi: 10.1038/nature21347.
- [2] L. L. Cavalli-Sforza, E. Minch と J. L. Mountain, 「Coevolution of genes and languages revisited.」, *Proc. Natl. Acad. Sci. U. S. A.*, vol. 89, no. 12, pp. 5620–4, 6 月 1992.
- [3] メスーディアレックス, 文化進化論—ダーウィン進化論は文化を説明できるか. NTT 出版, 2016.
- [4] 「Ethnologue: Languages of the World」, *Ethnologue*. [Online]. Available at: <https://www.ethnologue.com/>. [参照: 07-1 月-2020].
- [5] B. Bickel と J. Nichols, 「Oceania, the Pacific Rim, and the theory of linguistic areas」, *Proc. Annu. Meet. Berkeley Linguist. Soc.*,

vol. 32S, pp. 3–15, 2006.

- [6] B. Bickel ほか, *The AUTOTYP typological databases. Version 0.1.0*. 2017.
- [7] H. McColl ほか, 「The prehistoric peopling of Southeast Asia」, *Science*, vol. 361, no. 6397, pp. 88–92, 2018.
- [8] T. Gakuhari ほか, 「Jomon genome sheds light on East Asian population history」, *bioRxiv*, p. 579177, 3 月 2019, doi: 10.1101/579177.
- [9] Johanna. Nichols, *Linguistic diversity in space and time*. University of Chicago Press, 1999.
- [10] P. E. Savage ほか, 「How ‘circumpolar’ is Ainu music? Musical and genetic perspectives on the history of the Japanese archipelago」, *Ethnomusicol. Forum*, 2015, doi: 10.1080/17411912.2015.1084236.
- [11] H. Matsumae ほか, 「Exploring deep-time relationships between cultural and genetic evolution in Northeast Asia」, *bioRxiv*, p. 513929, 1 月 2019, doi: 10.1101/513929.
- [12] 長田直樹, 進化で読み解く バイオインフォマティクス入門. 森北出版株式会社, 2019.
- [13] P. Savage, E. Merritt, T. Rzeszutek と S. Brown, 「CantoCore: a new cross-cultural song classification scheme」, *Anal. Approaches World Music*, 2012.
- [14] D. Bryant と V. Moulton, 「Neighbor-net: an agglomerative method for the construction of phylogenetic networks.」, *Mol. Biol. Evol.*, vol. 21, no. 2, pp. 255–65, 2 月 2004, doi: 10.1093/molbev/msh018.
- [15] T. Jinam ほか, 「The history of human populations in the Japanese Archipelago inferred from genome-wide SNP data with a special reference to the Ainu and the Ryukyuan populations.」, *J. Hum. Genet.*, vol. 57, no. 12, pp. 787–95, 2012, doi: 10.1038/jhg.2012.114.
- [16] J. Novembre ほか, 「Genes mirror geography within Europe.」, *Nature*, vol. 456, no. 7218, pp. 98–101, 11 月 2008, doi: 10.1038/nature07331.