Chapter 7
Genes: Interactions with Language on Three Levels—Inter-Individual Variation, Historical Correlations and Genetic Biasing

Dan Dediu

Abstract The complex inter-relationships between genetics and linguistics encompass all four scales highlighted by the contributions to this book and, together with cultural transmission, the genetics of language holds the promise to offer a unitary understanding of this fascinating phenomenon. There are inter-individual differences in genetic makeup which contribute to the obvious fact that we are not identical in the way we understand and use language and, by studying them, we will be able to both better treat and enhance ourselves. There are correlations between the genetic configuration of human groups and their languages, reflecting the historical processes shaping them, and there also seem to exist genes which can influence some characteristics of language, biasing it towards or against certain states by altering the way language is transmitted across generations. Besides the joys of pure knowledge, the understanding of these three aspects of genetics relevant to language will potentially trigger advances in medicine, linguistics, psychology or the understanding of our own past and, last but not least, a profound change in the way we regard one of the emblems of being human: our capacity for language.

7.1 Introduction

The phenomenon of language can be explored at many scales, as so eloquently illustrated by the present volume, ranging, in time, from the human life extending over tens of years up to evolutionary timescales of (tens of) millions of years and, in social complexity, from the single individual up to interconnected societies at the continental scale. However, these levels cannot and must not be treated separately

D. Dediu (✉)
Language and Genetics Department, Max Planck Institute for Psycholinguistics,
Wundtlaan 1, 6525 XD, Nijmegen, The Netherlands
e-mail: Dan.Dediu@mpi.nl

139

except for very clearly set pedagogical goals, as they interact in complex ways, each shaping all the others and being shaped by them.

As clearly shown by the previous contributions, the two main factors conferring unity to language across all these diverse scales are represented by the *cultural transmission of language across generations* (see also Croft, Kirby and Oudeyer, this volume) and the *genetic bases* of the capacity for language (Zuidema, this volume). As expected, these two factors interact at all these levels to produce the bewildering complexity and diversity of language and languages.

This chapter will explore the main aspects of these interactions and will introduce some fundamental concepts, methods and findings by focusing both on well-known and accepted cases as well as recent and provocative hypotheses.

7.2 Being Different

It is obvious that we are different in many respects, including, for example, height, hair color, memory, mathematical abilities or language and speech. As some of us are shorter and some taller, some of us, despite normal or above-average intelligence, struggle with words while others have an almost artistic way of speaking and writing. Some of us speak clearly, articulating in a distinctive manner while others mumble at the limit of intelligibility. But why are we different in the first place?

One well-known reason is the experience of *different environments* by different people: eating well will allow you to grow taller than starving during childhood, being encouraged to read early by your parents will increase your chances of succeeding at school and exercising regularly will make your muscles stronger and bigger.

But it is equally well-known that some differences are down to innate factors, to *genetic differences* between individuals: those unfortunate enough to have a supplementary chromosome 21 will have the many problems associated with Down's syndrome (Plomin et al. 2001) while being born to taller parents will probably make you taller (Weedon et al. 2007).

However, the distinction between these two causes is not clear-cut and they are not somehow opposed, as the old question of "nature versus nurture" would imply. In fact, most inter-individual variability results from a combination of genetic and environmental differences, usually interacting in complex ways. A fascinating illustration of this is given by the recent study of the relationship between IQ and breastfeeding (Caspi et al. 2007), where it was found that breastfeeding tended to increase children's IQ but only if they had a certain allele of the *FADS2* gene. Alleles are possible variants of a given gene (Jobling et al. 2004) and this particular one allows its carriers to be able to process the human-specific fatty acids in mothers' milk, which are seemingly involved in brain development (Caspi et al. 2007).

What this shows is that if there happens to be the right nurture (breast feeding) but not the right nature (*FADS2* allele), or the other way around, there is no increase in the IQ. *Both* nature *and* nurture must be there and work together to bring about the inter-individual phenotypic variation (Ladd et al. 2008).

Understanding these inter-individual differences will allow us not only to better conceptualize human nature but also to both treat those characteristics perceived as pathological and enhance those seen as desirable.

7.2.1 Heritability

A first step is trying to somehow disentangle the relative contributions of genetic and environmental differences to inter-individual phenotypic variation in a given population. A popular way of doing this is represented by *heritability*, defined as the proportion of the phenotypic variation accounted for by genetic variation between individuals in a certain population (Plomin et al. 2001; Halliburton 2004) and reflecting the expected degree of similarity between parents and offspring on genetic grounds.

Heritability can be estimated in many ways through, for example, adoption, family aggregation and pedigree studies, and most commonly, by comparing *monozygotic* (MZ) and *dizygotic* (DZ) twins (Stromswold 2001; Plomin et al. 2001). While DZ (or fraternal) twins form from two separate ova fertilized by two different spermatozoa and are genetically as related as any two regular siblings, MZ (or identical) twins form from a single ovum fertilized by a single spermatozoon which then divides into two embryos, providing them with identical genomes (barring new mutations). By comparing the similarities between MZ (r_{MZ}) and DZ (r_{DZ}) twins, it is possible to estimate the heritability of the considered phenotypic trait in the target population as $h^2 = 2(r_{MZ} - r_{DZ})$, because presumably any increase in similarity between MZ relative to DZ twins is due to their greater (on average, two times bigger) genetic similarity (Stromswold 2001; Plomin et al. 2001).

However, it must be noted that there are a number of caveats concerning heritability estimates (see, for example, Charney 2012) including the fact that they are meaningful only for phenotypes which show variation in the target population, they are higher in homogeneous compared to variable environments, they differ between populations and can vary with age (Bishop 2003; Plomin et al. 2001). Therefore, heritability is not an absolute measure of some kind of intrinsic genetic contribution to the trait but a *relative* estimate fundamentally reflecting the specific population and environmental conditions.

Heritability estimates for aspects of speech and language generally point to the involvement of genes in explaining the inter-individual variation.[1] For language and speech disorders—by far the best studied to date—the heritabilities of their various aspects seems to be high (typically $h^2 > 0.50$) with differing heritabilities for different aspects of language (Stromswold 2001; Bonneau et al. 2004; Fisher et al. 2003; Felsenfeld 2002; Plomin et al. 2002; Plomin and Kovas 2005). For example, liability to stuttering is highly heritable ($h^2 \approx 0.70$; Felsenfeld 2002) and the heritability of

---

[1] Probably the best review of the heritability of aspects of speech and language remains Stromswold (2001).

SLI (*Specific Language Impairment*, a complex and controversial language pathology; Bishop 2003, OMIM[2] 602081) is also high (Bishop 2003, cites $h^2 \approx 0.76$, 0.48 and 0.54, while Bonneau et al. (2004) give $h^2 \approx 0.70$ and 0.73, depending on the study).

Likewise, normal inter-individual variation in aspects of speech and language seems to also have a genetic component, ranging from a very low $h^2 \approx 0.02$ for expressive vocabulary at 14 months to $h^2 \approx 0.38$ at 24 months and $h^2 \approx 0.72$ for WISC-R vocabulary (Stromswold 2001). As a comparison (Halliburton, 2004, p. 540), the heritability of height in humans is $h^2 \approx 0.65$, of schizophrenia $h^2 \approx 0.70$ and of fingerprint ridge count $h^2 \approx 0.92$.

7.2.2 Linkage Studies and the FOXP2 Gene

Having established that a phenotypic trait is heritable is not, of course, the end of the story and the next interesting question concerns the nature of the gene(s) involved and the specific mechanisms bridging the causal gap between genetic and phenotypic variation. A first step in this direction is knowing where the gene(s) might be, on which of the 23 pairs of chromosomes (Jobling et al. 2004) and in which specific position.

The main idea is to use the fact that genes on the same chromosome are not transmitted independently, and that this *linkage* between them generally increases the closer together they are, offering a way to build *genetic maps*. A *linkage study* exploits the association between the trait under consideration and various *genetic markers* across generations in a family tree, on the assumption that the genes(s) responsible and the linked markers will tend to be inherited together (Jobling et al. 2004; Halliburton 2004; Gibson and Gruen 2008; for more details see Box 1). This design was instrumental to the discovery of *FOXP2*, probably the most discussed gene when it comes to speech and language.

> **Box 1: Genetic linkage and linkage studies**
>
> Two *loci* or "positions" on different chromosomes or on the same chromosome but far apart are transmitted independently to the offspring, but two loci near to one another on the same chromosome are not. Chromosomes are of two types: *autosomes* come in identical pairs while the two *sex chromosomes* come in an identical pair in women (XX) but are different in men (XY) (Jobling et al. 2004; Halliburton 2004). Let's consider two loci on an autosome, *1* and *2*; therefore, each individual will have two alleles for each locus. Let's consider an example individual *heterozygous* for both loci, carrying two different alleles for each of them (as opposed to *homozygous* loci for which (s)he would carry the same

---

[2] *Online Mendelian Inheritance in Man*, an online comprehensive database of genetic disorders freely browsable at http://www.ncbi.nlm.nih.gov/omim.

allele): *A* at locus *1* and *B* at locus *2* on one chromosome and *a* at locus *1* and *b* at locus *2* on the other chromosome of the pair.

If loci *1* and *2* are transmitted independently, then this individual's offspring would have the same probability of inheriting allele *A* (locus *1*) and allele *B* (locus *2*)—*AB* for short—, *Ab*, *aB* or *ab* from this parent and something else (and irrelevant for now) from the other parent. However, if loci *1* and *2* are on the same chromosome and linked, then the probability of the offspring inheriting *AB* or *ab* will be greater than that of inheriting the *recombinants Ab* or *aB*. The strength of the linkage is given by this difference in probabilities and is measured in *centiMorgans*, cM, representing a 1% probability of recombination between two loci in a generation (Jobling et al. 2004), with independence for separations bigger than 50 cM. Very roughly, there is a linear relationship between strength of linkage and physical distance between the loci, with 1 cM ≈ 1 Mb at the scale of the entire human genome, but, as always, things are really not that simple (Jobling et al. 2004; Speed and Zhao 2007). Physical distances are measured in units of the fundamental building blocks of the DNA molecule, the *base pair*, bp, and multiples, like millions of base pairs, Mb (Jobling et al. 2004).

The figure below illustrates the concept of linkage disequilibrium: the two *homologous* chromosomes contain alleles *A* and *a* at locus *1* and alleles *B* and *b* at locus *2* in the parental generation. If these two loci are independent (light gray bars), then it is expected that all four possible combinations of alleles to be present in the same proportion (25% each) in the gametes (sex cells, ova and spermatozoa) producing the next generation. However, if there is linkage between them (dark gray bars), the *parental* combinations (*AB* and *ab*) are expected more frequently than the *recombined* ones (*Ab* and *aB*).
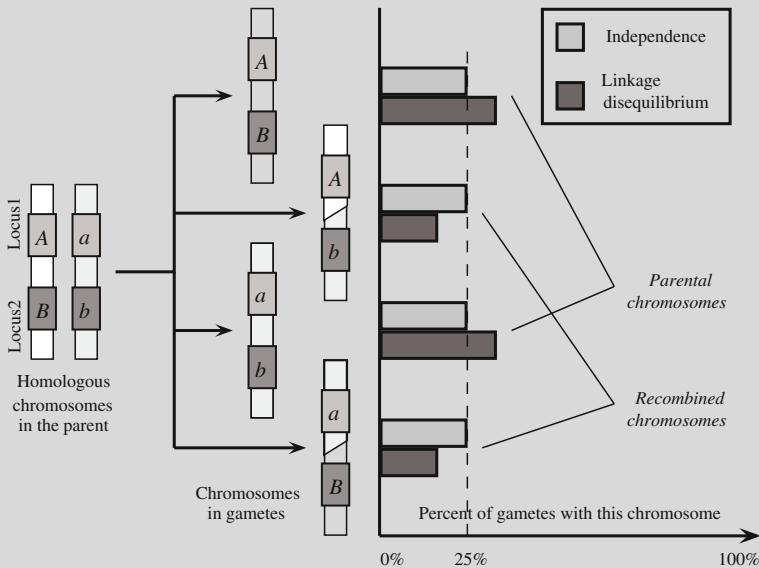
The British "KE" family has a very interesting three-generations pedigree, with half the members (15 out of 31) affected by a complex pathology, involving speech and language (Hurst et al. 1990; Gopnik and Crago 1991), classified as *developmental verbal dyspraxia* (OMIM 602081) and included in the SLI category. The exact phenotype is very complex (Fisher et al. 2003; Lai et al. 2001; Vargha-Khadem et al. 1998; Marcus and Fisher 2003) and includes *articulatory problems*, due to troubles with coordinating complex oro-facial movements, *cognitive impairments* and *language impairments*, affecting spoken (expressive and receptive) and written language, as well as the comprehension and production components of grammar (understanding complex sentences, inflectional and derivational morphology).

This pathology follows a *dominant autosomal* mode of inheritance and is *fully penetrant*, meaning that the gene (named *SPCH1*, OMIM 602081) is located on an autosome (chromosome 7; Fisher et al. 1998) and an individual carrying even a single mutation of the gene will manifest the pathology (Hurst et al. 1990; Bishop 2003; Lai et al. 2001). The gene was finally identified as *FOXP2* (Lai et al. 2000, 2001), a member of the numerous *Forkhead box* (Fox) family of genes which act as *transcription regulators*, controlling the expression of other genes (Scharff and White 2004). Recently, Vernes et al. (2008) have shown that *FOXP2* downregulates the gene *CNTNAP2* which was previously involved in language delay in autism Alarcón et al. (2008) and is strongly associated with non-word repetition (Vernes et al. 2008). The specific mutation affecting the "KE" family[3] disrupts the regulatory function of this gene, leading to a cascade of events resulting in the pathological phenotype.

It is not entirely clear what *FOXP2* does in humans, but it is certainly *not* "the language gene", as sometimes boasted in the media. As always, the story is much more fascinating, as it turns out that *FOXP2* has complex roles in neural (and not only) development, with many targets in the developing brain (Spiteri et al. 2007) and affecting brain structure and functioning (Vargha-Khadem et al. 1998; Watkins et al. 2002; Liégeois et al. 2003). This picture is complicated by studies involving birds with learned and non-learned song, echolocating bats and vocal-learning non-human mammals (Li et al. 2007; Webb and Zhang 2005; Teramitsu et al. 2004; Scharff and Haesler 2005; Haesler et al. 2004; Shu et al. 2005), including the recent engineering of mice having the human-specific gene (Enard et al. 2009), suggesting a complex role for *FoxP2*.

Evolutionary studies seem to imply that, overall, *FoxP2* is very conserved across taxa (Enard et al. 2002) but that there seem to exist two "human-specific" mutations, fixed in the human population and under strong selection (Enard et al. 2002; Zhang et al. 2002). It is, however, unclear what exactly these mutations have been selected for, nor when this has happened, with earlier estimates (Enard et al. 2002; Zhang et al. 2002) around 100–200 thousand years ago, apparently coinciding with the emergence of anatomically modern *Homo sapiens*. Recently, Krause et al. (2007) found that these "human-specific" mutations are shared with Neandertals and conclude that the human *FOXP2* originated and was selected before the split between us and them,

---

[3] This is not the only *deleterious* mutation affecting *FOXP2*: for example, two different mutations in two unrelated individuals produce similar pathologies (Lai et al. 2001; MacDermot et al. 2005).

more than 300 thousand years ago, but Coop et al. (2008) read the data as suggesting low rates of gene flow between modern humans and Neadertals. Nevertheless, the publication of the Neandertal draft genome (Green et al. 2010) confirmed that we share this variant of *FOXP2*, most probably present in the last common ancestor.

In conclusion, *FOXP2*, the first gene specifically identified for its role in language and speech, vividly shows how complex the questions concerning the genetic influences on this human-specific phenotype are (Fisher and Scharff 2009). It is to be expected that there is no "language gene" which would suddenly have allowed a lucky mutant somewhere down the human line to speak, nor even a small set of such genes. However, the unravelling of these complex interactions between genes and environmental factors will undoubtedly help us better understand what language really is.

7.2.3 Association Studies and Future Directions

In *association studies*, a group of unrelated individuals displaying the trait of interest and a group of individuals without it are compared, looking for those genetic markers which correlate with the presence of the trait (Jobling et al. 2004; Plomin et al. 2001). If, based on previous research, one or more *candidate genes* are hypothesized to be involved in the target phenotype, the search will be constrained to variants of these genes. In contrast, the "hypothesis-free" approach of *genome-wide association studies* (GWAS) consists of looking for correlations between the phenotype of interest and genetic markers (such as SNPs[4]) covering the entire genome in the hope that very strong correlations are due to real genetic influences (Pearson and Manolio 2008; McCarthy et al. 2008).

The GWAS are very promising and currently intensively used in topics ranging from the genetic bases of cancer or diabetes (The Wellcome Trust Case Control Consortium 2007) to schizophrenia and bipolar disorder (Maher et al. 2008; Lewis et al. 2003; Segurado et al. 2003). However, there are a number of issues mainly concerning the fact that when hundreds of thousands of genetic markers are used the chance of "discovering" non-existent correlations with the trait (*false positives*) is huge and techniques for *multiple testing correction* (which lower the threshold of statistical significance from the "standard" 0.05 to much smaller values, like 10^{-8}) must be employed (Speed and Zhao 2007) together with caution in interpreting the results of the study (Pearson and Manolio 2008; McCarthy et al. 2008). These and other difficulties have resulted in a low degree of replication of the early findings and the recent publication of stringent guidelines (Chanock et al. 2007). Concerning specifically language and speech, it is to be expected that large GWAS comparing both pathological to normal phenotypes as well as the range of normal variation will

---

[4] Single Nucleotide Polymorphisms, one of the simplest types of genetic variation where a single DNA "letter" (nucleotide) varies between individuals and is present at relatively high frequencies.

be areas of intense research in the near future, but they will require better definition and measurement of interesting phenotypes.

Another very promising avenue is represented by *copy-number variation* (CNVs), which represent inter-individual variations in the number of times a certain region of the genome is present (McCarroll and Altshuler 2007; Redon et al. 2006). Due to the fact that, as opposed to SNPs, which have only two alleles, CNVs can have a number of possible repeats, they can alter the *dosage* of gene products, a plausible mechanism behind complex phenotypic variation (McCarroll and Altshuler 2007) and very promising for the study of language and speech.

Epigenetic processes (Jiang et al. 2008), whereby non-genetic marks (chemical changes of the nucleotides or the proteins associated with the DNA) alter the expression of genes, seem to play an important role not only in development and disease (Hirst and Marra 2009) but also in normal neuronal plasticity (Fagiolini et al. 2009). These mechanisms also seem to have evolutionary relevance because they are transmitted across multiple generations (Youngson and Whitelaw 2008; Jablonka and Raz 2009). They subtend *parent of origin effects* where the same allele has different effects depending on the parent it has been inherited from, as has been suggested for the gene *LRRTM1* involved in handedness and schizophrenia (Francks et al. 2007). In the future, epigenetics will probably play an important role in understanding language development and evolution.

7.3 Changing Together

As briefly reviewed above, it is clear that human individuals differ from each other at the genetic level (with the possible exception of identical twins) by carrying different alleles of the same gene. But do populations also differ in what concerns their genes and, if so, in what ways?

It is generally agreed that humans are genetically very uniform when compared with other mammals (Jobling et al. 2004; Relethford 2001; Templeton 1998), and that the existing genetic variation is apportioned mostly (approximately 85%) *within* populations (between the individuals from the same group), but there is some variation (approximately 15%) due to *inter-population* genetic differences (*between* populations) (Jobling et al. 2004; Lewontin 1972). These inter-group differences are in great majority due to different *allele frequencies* in one group as compared to another and not to group-specific genetic variants.

When considering together many genetic loci, the variation in the frequencies of their alleles between populations combine and allows the study of the genetic structure of humans at the population level, the reconstruction of demographic history and even the inference of the population of origin for a given individual (Rosenberg et al. 2002; Bamshad et al. 2003; Falush et al. 2003). However, it is still debated if this variation between populations is gradual or if there are some continental-level boundaries, where the changes are abrupt (Barbujani and Belle 2006). Nevertheless,

it must be highlighted that the existence of such genetic structure offers no support to racist theories (Jobling et al. 2004; Banton 1998).

Having established that there are (small but detectable) differences between human populations, the next question concerns their origin and dynamics. Some are caused by *genes under selection*, whereby variants of a gene are (dis-)favored in certain environments, leading to differences between populations experiencing these different environmental pressures (Halliburton 2004; Jobling et al. 2004; Skelton 1993). Well-known examples are represented by *Sickle cell disease*, where a recessive autosomal mutation affecting haemoglobin confers resistance to heterozygotes against malaria but inflicts sickle cell anemia to homozygotes (*heterozygote advantage*; see Box 2), and *lactase persistence* where the ability to digest milk is retained into adulthood in dairying populations (Jobling et al. 2004; see Box 3).

Box 2: Sickle cell anemia: natural selection and heterozygote advantage

The red blood cells' function is to transport oxygen from the lungs throughout the body and, to this end, they contain vast quantities of *haemoglobin*, a complex iron-containing protein. In some people, a *point mutation* in the *HBB* gene on chromosome 11 (this mutation is known as *HbS*), results in the substitution of one *amino acid* for another and the production of a malfunctioning haemoglobin molecule. When an individual is homozygous for this mutation, it will produce only malfunctioning haemoglobin, manifested as *sickle cell anemia*, a complex disease which usually kills before puberty (Guégan et al. 2007, see OMIM 603903 for details).

However, a heterozygous *carrier* will have a more or less normal phenotype, as the normal (*wild-type*, denoted *HbA*) copy of the gene will produce enough functioning red blood cells to compensate for the mutated ones. Interestingly, these heterozygous individuals actually suffer less from *malaria*, a deadly infectious disease caused by a parasite of the red blood cells (genus *Plasmodium*, the most dangerous being the *falciparum* species), transmitted through mosquito (genus *Anopheles*) bites.

Thus, the *HbS* mutation, on one hand, is very deleterious in homozygous form (*HbS-HbS*) by causing sickle cell anemia, but is better than the normal homozygous (*HbA-HbA*) when in heterozygous form (*HbS-HbA*) *in environments* where malaria is *endemic* (Guégan et al. 2007; Cavalli-Sforza et al. 1994). This phenomenon is called *heterozygote advantage* and explains why some alleles deleterious in homozygous form are, nevertheless, maintained at relatively high frequencies, in the population (Cavalli-Sforza et al. 1994; Skelton 1993): in a sense, this is a cost some have to pay for the benefit of many.

Box 3: Lactose tolerance: culture impacting on genes

One of the defining characteristics of mammals is represented by the production of milk to feed their young. *Lactose* is one of the most important sugars in milk and digesting it is ensured by the enzyme *lactase*. After weaning, lactase production gradually decreases, resulting in a total inability or greatly reduced capacity to digest lactose in adults. However, in some human populations, the proportion of individuals *retaining* lactose tolerance is very high and the molecular mechanism seems to involve changes in a non-coding region of the *MCM6* gene, which upregulates the *LCT* gene responsible for the production of the lactase enzyme (Ingram and Swallow 2008, OMIM 223100).

Interestingly, the capacity to digest lactose is retained in populations with a history of milking, reaching high frequencies in North–West Europe and some parts of Africa (see map in Ingram and Swallow 2008, p. 1128) and it seems that there were several independent origins of this capacity. Therefore, the mutations allowing adults to digest milk were selected in those groups with a culture favoring milking, clearly showing that culture can lead to new selective pressures on genes (Ingram and Swallow 2008; Cavalli-Sforza et al. 1994).

However, this dramatic example is just the simplest possible type of feed-back from culture onto genes, with more complex cases being discussed in the *niche construction* (Odling-Smee et al. 2003) or *cultural evolution* (Boyd and Richerson 1985; Cavalli-Sforza and Feldman 1981) literatures.

But most are due to *selectively neutral alleles*, whose frequencies change through *random drift*, going up or down in a given population as chance dictates (Jobling et al. 2004; Halliburton 2004). When this is the case, the genetic differences between populations reflect their history of *splits* and *admixture*, as genetically similar populations share a more recent common ancestor and/or have recently been in intense contact, while more dissimilar populations have diverged further in the past (Cavalli-Sforza et al. 1994; Jobling et al. 2004).

When populations split, they carry with them not only their genes, but also their languages, so that the two will diverge in parallel ways, albeit at very different rates. Also, when populations meet, they not only exchange genes through inter-marriages but their languages converge through borrowing. Therefore, it looks as if it would be possible to find *correlations between genetic and linguistic diversities* across populations due to common processes making them change together (Cavalli-Sforza et al. 1994; Jobling et al. 2004). However, such correlations are far from perfect as languages and genes can become decoupled through *language shift* (or language replacement, where one population replaces its original language with a new one) or *elite dominance* (a small elite, military, religious, etc., manages to impose its language and culture but not its genes), for example (Cavalli-Sforza et al. 1994).

The seminal work of Cavalli-Sforza et al. (1994) introduced the methodology used to find and investigate such correlations, including the building of population trees—

based on the *genetic distances* between such populations—and their comparison with *historical linguistic* trees in the attempt to capture the historical processes of descent with modification at the demographic and linguistic levels. However, this approach encounters a number of issues, like the usage of contentious linguistic classifications, the problematic concept of "population" and population trees, and even the method of comparing linguistic and population trees.

This early literature tended to relay on contentious linguistic classifications, like Ruhlen's (Ruhlen 1991, 1994) or Greenberg's (Greenberg 1971, 1987, 1998), which transcend the accepted *historical linguistic comparative method* (McMahon and McMahon 2005; Matisoff 1990; Sims-Williams 1998), producing such debated constructs like "Indo-Pacific" (Dixon 1997), "Australian" (Dixon 2001; Dench 2001) and especially "Amerind" (Bateman et al. 1990; Bolnick et al. 2004; Matisoff 1990; Sims-Williams 1998). Moreover, this work uses even more contentious "macrofamilies" like "Nostratic" and "Eurasiatic", which purport to represent a level beyond linguistic families, but which seem to lack any linguistic validity (Campbell 1999; McMahon and McMahon 2005; Trask 1999; Appleyard 1999). The main drive behind this systematic and voluntary error, which still persists despite vehement critique by linguistics (see, for a recent example, Belle and Barbujani 2007), is represented by the perceived need to have a fully hierarchical classification of languages to feed into the comparison, no matter how complex the linguistic reality actually is (see, for example, the clear statement in Cavalli-Sforza 2000, p. 139).

Another critique is represented by the "populations" used, which are generally sampling units based on language criteria and cultural significance (McMahon 2004) and which are not equivalent in the sense that, for example, the Hadza of Tanzania (population of ≈1000), South Chinese (population ≈500 million) and French (population ≈ 60 million) cannot be treated as comparable (MacEachern 2000). Moreover, it is assumed that there is a one-to-one relationship between tribal and language names, leading to a the concept of "language as a powerful ethnic guidebook" (Cavalli-Sforza 1994, p. 23), but this is totally misleading (Sims-Williams 1998; McMahon 2004; MacEachern 2000). However, this sampling procedure is here to stay and the best we can do is try to minimize the impact it has on studies of correlations between languages and genes.

The third, and probably the most criticized aspect, concerns the methods for comparing linguistic and population classifications. Tree classifications of populations were produced from genetic distances, resulting in *phenetic* (Skelton 1993) populations trees, which cluster populations based on their similarity, but assumed (contentiously; Bateman et al. 1990; MacEachern 2000) to represent the actual genealogical history of those populations. These trees were then compared to the linguistic classifications, apparently supporting a high concordance between the two (Cavalli-Sforza et al. 1994), but this appearance seems totally due to the method used (Bateman et al. 1990), as it exploited the ability of branches in a tree to freely rotate in order to artificially increase the visual similarity between the two compared trees.

In recent work, however, there is a tendency to move away from tree-based methods and controversial linguistic classifications, using, for example, *Mantel correlations* (correlations adapted for the case of distance matrices where the "observations" are not independent) between genetic and linguistic distances (e.g., Poloni et al. 1997; Rosser et al. 2000), *boundary detection* (e.g., Manni et al. 2004) or *AMOVA* (Analysis of Molecular Variance; Excoffier et al. 1992).

Probably the best-known example of language-genes correlation hypothesis is represented by the *language/farming co-dispersal* theory, which tries to explain the modern distribution of a major component of genetic and linguistic diversity as being the result of the spread of farming from its primary origins, carrying together the genes and languages of the original farmers (Jobling et al. 2004; Cavalli-Sforza et al. 1994; Mithen 2003; Diamond 1998). To focus on the *Indo-European* family of languages (Fortson 2004), it is postulated that its spread started around 10,000 years ago when agriculturalists from somewhere around Anatolia expanded westward and eastward (Diamond 1997, 1998; Diamond and Bellwood 2003; Cavalli-Sforza et al. 1994; Renfrew 2002). This theory is highly contentious on linguistic (Mallory 1991) as well as genetic and archaeological (Jobling et al. 2004; Sims-Williams 1998) grounds, and the SE to NW genetic gradients discovered across Europe (Cavalli-Sforza et al. 1994), initially taken to support such a demic diffusion, seem to reflect a far more complex history (Jobling et al. 2004; Sims-Williams 1998). However, recent (and no less controversial) applications of *phylogenetics* to Indo-European *cognate sets*[5] seem to support an origin of this family around the time of agricultural expansion from Anatolia (Gray and Atkinson 2003; Atkinson and Gray 2006).

However, in general, it seems that the bulk of the correlations between linguistic historical classifications and genetics is due to geography, with great distances and major obstacles being the best predictors of large differences in languages and genes (Belle and Barbujani 2007; Rosser et al. 2000), but that there is also a residual correlation between languages and genes after controlling for geography (Cavalli-Sforza et al. 1994; Belle and Barbujani 2007), due to shared history and *linguistic assortative mating*, whereby people tend to marry within their own linguistic community (Cavalli-Sforza et al. 1994).

7.4 Pushing and Pulling at Language

If the previous two sections tried to review in a more or less objective manner well-established aspects of the interaction between genes and language, the present one will concern a very new and controversial development in which the author is directly involved, so that some subjectivity is bound to remain despite all efforts to the contrary. The differences between individuals (Sect. 7.2) are limited in both time

---

[5] Roughly, two words in two different languages are cognate if they descend from the same original proto-word.

and scale to the ontogenetic and personal, while the inter-population correlations between genes and languages (Sect. 7.3) concern historical (cultural) and continental (macroregional) scales.

The correlations between linguistic classifications and genetic differences between populations discussed above consider explicitly the *phylogenetic diversity*, characterized by the number of different *language lineages* (language families, subfamilies, etc.) and reflecting historical processes (Nettle 1999). There are around 100 language families more or less accepted to date (Gordon 2005), vastly unequal with respect to their geographical distribution and size (measured either as number of speakers or number of languages; Nettle 1999; Gordon 2005), but, given the difficulties involved, these data are very approximate (Trask 1996; Campbell 2004; Lass 1997).

But another dimension along which languages differ is represented by their *structure*, allowing the definition of *structural diversity* (Nettle 1999). Languages can be described (to a good approximation) using a set of abstract categories (variously named *features*, *variables* or *parameters*) which can have different values for different languages and covering every aspect of language, from phonetics and phonology, through morpho-syntax to semantics and pragmatics (studied by *linguistic typology*; Croft 1990; Comrie 1981; Haspelmath et al. 2005). One example of such a feature is "number of vowels" (Maddieson 2008): some languages have a lot (like English or German) while some very few (like Berber or many Australian aboriginal languages). Another well-known example is represented by the order of Subject, Object and Verb in declarative clauses (Dryer 2008): many languages (like Japanese) prefer SOV, others (like English) prefer SVO, while the others go for VSO, VOS, OVS, OSV or show no particular preference.

7.4.1 Linguistic Tone

Another linguistic feature is represented by *tone*. *Non-tone* languages (like English or Arabic) use *voice pitch* to conveys certain meanings at the sentence level (*intonation*) like marking an utterance as a question or an exclamation. In addition, *tone languages*, like Chinese or Yoruba, use voice pitch to distinguish words or grammatical structures in very much the same way all languages use vowels and consonants (Yip 2002; Dediu and Ladd 2007): for example, "nian" in Mandarin Chinese means "year" when pronounced with rising pitch and "read" when pronounced with falling pitch. The number of tones varies between tone languages, but usually is between two and seven (Yip 2002).

The classification of languages as tonal or not is not straightforward, with many borderline cases (like Swedish or Norwegian), but a bit more than half the world's 6000 or so languages (Gordon 2005) seem to be tonal (Haspelmath et al. 2005). Their geographic distribution is not random, with clusters predominating in sub-Saharan Africa, continental and insular South–East Asia, and Central America and

Amazonia[6] (Haspelmath et al. 2005; Dediu and Ladd 2007) and a very interesting question concerns the causes of this distribution: is it purely "random", following the laws of cultural transmission and the vagaries of history and geography, or are there some other causes, as well?

It is known that languages sharing a common ancestor tend to inherit its value for tone (in a way, tonal begets tonal and non-tonal, non-tonal, but this is highly simplifying) and it is also known that tone (or non-tone, for that matter) tends to spread to neighboring languages through regular language contact (Yip 2002; Dediu and Ladd 2007). Also, we know that *tonogenesis* can happen through regular historical linguistic processes (Hyman 1978; Yip 2002) and that the reverse process, of tone loss, appears, for example, in situations of usage as *lingua franca* (like Swahili).

But the hypothesis D. Robert Ladd and I have recently put forward (Dediu and Ladd 2007) suggests that there might be a slight *genetic biasing* (Dediu 2011) at work, contributing to the dynamics and distribution of tone across the world's languages.

7.4.2 Two Brain-Related Genes

This genetic biasing is proposed to be due to the so-called *derived* or *adaptive*[7] *haplogroups* of two human genes, *ASPM* and *Microcephalin* (in this chapter, we will denote these two haplogroups as *ASPM-D* and *MCPH-D*, respectively). Both *ASPM* and *Microcephalin* are clearly involved in brain growth and development because people carrying deleterious mutations of any of these two genes develop *microcephaly*, having heads much smaller than the average (but these are not the only genes causing microcephaly; Gilbert et al. 2005; Cox et al. 2006; Woods 2004). The exact mechanism by which the deleterious mutations of both *ASPM* and *Microcephalin* induce this pathology are not fully understood yet, but they very probably affect the formation of neural cells (Dediu and Ladd 2007; Caviness et al. 1995).

However, the derived haplogroups *ASPM-D* and *MCPH-D* do not cause microcephaly, or, for that matter, any other obvious phenotypic effects. Previous research failed to find any correlations with, for example, variation in intelligence or head size in the normal population[8] (Mekel-Bobrov et al. 2007; Woods et al. 2006) or any association with schizophrenia (Rivero et al. 2006). But they are fascinating because they seem to be under strong natural selection in humans,[9] have a skewed geographical

---

[6] For a map of tone, see the World Atlas of Language Structures (Haspelmath et al. 2005) Online, the chapter "Tone" by Ian Maddieson, at http://wals.info/chapter/13.

[7] The names "derived" and "adaptive" come from the original work by Bruce Lahn's group (Mekel-Bobrov et al. 2005; Evans et al. 2005) which, in fact, identified them. They are called "derived" as opposed to the original (ancient) form of the genes and "adaptive" because of the selection pressures acting on them, that the authors claim to have detected. For our purposes here, these are just labels.

[8] But see the recent claim that another SNP of *Microcephalin* is associated with cranial volume in normal Chinese males (Wang et al. 2008).

[9] However, the methodology used to infer this is probably not adequate and the claimed recent selection has not been replicated.

distribution and have relatively recent origins (Mekel-Bobrov et al. 2005; Evans et al. 2005; Dediu and Ladd 2007).

It was found (Dediu and Ladd 2007; Ladd et al. 2008) that the geographic distribution of *ASPM-D* and *MCPH-D* correlates very well with the distribution of tone languages, even after controlling for the main two sources of such *spurious* correlations (see Sect. 7.3): the historical relatedness between languages and language contact mediated by spatial closeness (this was done by computing the *partial Mantel correlations* between linguistic, genetic, historical linguistic and geographical distances between populations—see Dediu and Ladd 2007 for technical details). Moreover, when comparing the correlations between tone, *ASPM-D* and *MCPH-D* with all the possible correlations between 25 or so other linguistic features and 1000 genes, it was found that it was well in the upper tail of the empirical distribution of this type of correlation, suggesting that this association is very "special", in a statistical sense (Dediu and Ladd 2007). Therefore, it was concluded that the relationship between *ASPM-D* and *MCPH-D* and tone is stronger than expected by chance and not fully explained by classical genes-languages co-dispersal processes (Dediu and Ladd 2007; Ladd et al. 2008).

7.4.3 Biasing Language Transmission

Populations having *low* frequencies of both derived haplogroups tend to speak *tone* languages, while populations with *high* frequencies tend to speak *non-tone* languages, with populations with *low ASPM-D* and *high MCPH-D* showing no preference[10] (see also the figure in Dediu and Ladd 2007). Therefore, we suggested that the derived haplogroups of these two genes might be able to bias language *towards non-tonality* or *against tonality* (Dediu and Ladd 2007; Ladd et al. 2008).

It is obvious that such a bias, if it indeed exists, does not have any major effects at the individual level, as any normal child can acquire the language(s) of his/her community irrespective of their genetic background (Ladd et al. 2008). What such a bias does is to have an extremely small, almost invisible impact at the individual level, but it will be amplified and made manifest by the cultural transmission of language across generations, by "pushing" or "pulling" language towards its preferred state (see also various other contributions in this volume). Again, such a bias will emphatically *not* determine the fate of language, as many other factors play a major role in language change and evolution (contact, history, sheer accident), but it will statistically bias the distribution of structural diversity (see Ladd et al. 2008 for a thorough discussion and relevant examples). Moreover, such a bias probably does not represent the reason why *ASPM-D* and *MCPH-D* are under natural selection, but is simply a byproduct of their effects on brain development (Dediu and Ladd 2007; Ladd et al. 2008).

---

[10] There are no *high ASPM-D* and *low MCPH-D* in the sample (Dediu and Ladd 2007; Ladd et al. 2008).

Confirmation or falsification of this hypothesis rests with experimental approaches of the type discussed above (Sect. 7.2), by trying to define the effects of this bias at the individual level, design appropriate measures and test if the variation in such measures correlates with the presence or absence of the derived haplogroups. Interestingly, it was reported very recently (Wong et al. 2012) that *ASPM-D* seems to be involved in pitch processing at the individual level, but more work is needed in order to draw any clear conclusions. However, the methodology introduced by Dediu and Ladd (2007) holds the potential to detect more candidate genetic biases and their effects in shaping linguistic diversity, being fundamentally a *hypothesis-generating* mechanism (Nettle 2007).

7.5 Conclusions

It should come as no surprise that our genes impact on our language and speech but what might be unexpected is the complexity of the mechanisms involved. It is clear that the old question of "nature vs nurture" is profoundly misguided, as genes and environment interact in subtle ways and require each other in order to produce their phenotypic effects (see, for example, Ladd et al. 2008 for a detailed discussion of this issue in the context of language).

It is also such inter-individual differences which, if structured in appropriate ways, can influence the trajectory of language change. People with the derived haplogroups of *ASPM* and *Microcephalin* might indeed turn out to be different from those without them in ways relevant to learning, processing or producing linguistic tone, but if only one such individual appears each generation in a population, it will not be enough to bias language change towards non-tone languages. However, intuition and computer models (Dediu 2008b) seem to suggest that when enough such individuals exist in a population for long enough time, language will indeed be influenced by their combined biases.

On the other hand, inter-individual differences and genetic biases are embedded in the larger context of language dispersal, birth, change and death (Nettle 1999) and these three levels interact in complex ways to produce the patterns of language and genetic diversity. And, finally, such small genetic biases could form the basis of language evolution, viewed as the co-evolution between our languages and our brains and bodies (Dediu 2008a; Christiansen and Chater 2008).

But understanding this interaction—and the multifaceted *phenomenon of language*—requires understanding all the intervening levels: brain, hearing, speech and gesturing, discourse now and here, and language change on the historical and evolutionary timescales. It requires, thus, a truly integrative approach.

References

Alarcón, M., Abrahams, B.S., Stone, J.L., Duvall, J.A., Perederiy, J.V., Bomar, J.M., Sebat, J., Wigler, M., Martin, C.L., Ledbetter, D.H., Nelson, S.F., Cantor, R.M., Geschwind, D.H.: Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. Am. J. Human Genet. **82**(1), 150–159 (2008). doi:10.1016/j.ajhg.2007.09.005

Appleyard, D.: Afroasiatic and the Nostratic hypothesis. In: Renfrew, C., Nettle, D. (eds.) Nostratic: Examining a Linguistic Macrofamily, pp. 289–325. The McDonald Institute for Archaeological Research, Oxbow Books, Oxford (1999)

Atkinson, Q.D., Gray, R.D.: How old is the Indo-European language family? Illumination or more moths to the flame? In: Clackson, J., Forster, P., Renfrew, C. (eds.), Phylogenetic Methods and the Prehistory of Languages, pp. 91–109. MacDonald Institute, Cambridge (2006)

Bamshad, M., Wooding, S., Watkins, W., Ostler, C., Batzer, M., Jorde, L.: Human population genetic structure and inference of group membership. Am. J. Hum. Genet. **72**, 578–589 (2003)

Banton, M.: Racial Theories, 2nd edn. Cambridge University Press, Cambridge (1998)

Barbujani, G., Belle, E.: Genomic boundaries between human populations. Hum. Hered. **61**(1), 15–21 (2006)

Bateman, R., Goddard, I., O'Grady, R., Funk, V., Mooi, R., Kress, W., Cannell, P.: Speaking of forked tongues: the feasibility of reconciling human phylogeny and the history of language. Curr. Anthropol. **31**, 1–13 (1990)

Belle, E., Barbujani, G.: Worldwide analysis of multiple microsatellites: language diversity has a detectable influence on DNA diversity. Am. J. Phys. Anthropol. **133**(4), 1137–1146 (2007)

Bishop, D.V.M.: Genetic and environmental risks for specific language impairment in children. Int. J. Pediatr. Otorhinolaryngol. **6751**, S143–S157 (2003)

Bolnick, D., Shook, B., Campbell, L., Goddard, I.: Problematic use of Greenberg's linguistic classification of the Americas in studies of native American genetic variation. Am. J. Hum. Genet. **75**, 519–523 (2004)

Bonneau, D., Verny, C., Uzé, J.: Les facteurs génétiques dans les troubles spécifiques du langage oral. Archives de pédiatrie **10**, 1213–1216 (2004)

Boyd, R., Richerson, P.J.: Culture and the Evolutionary Process. University of Chicago Press, Chicago (1985)

Campbell, L.: Nostratic and linguistic palaeontology in methodological perspective. In: Renfrew, C., Nettle, D. (eds.) Nostratic: Examining a Linguistic Macrofamily, pp. 179–230. The McDonald Institute for Archaeological Research, Oxbow Books, Oxford (1999)

Campbell, L.: Historical Linguistics: An Introduction. Edinburgh University Press, Edinburgh (2004)

Caspi, A., Williams, B., Kim-Cohen, J., Craig, I.W., Milne, B.J., Poulton, R., Schalkwyk, L.C., Taylor, A., Werts, H., Moffitt, T.E.: Moderation of breast feeding effects on the IQ by genetic variation in fatty acid metabolism. Proc. Natl. Acad. Sci. USA **104**(47), 18860–18865 (2007)

Cavalli-Sforza, L.: Genes, Peoples, and Languages. Allen Lane: The Penguin Press, London (2000)

Cavalli-Sforza, L., Menozzi, P., Piazza, A.: The History and Geography of Human Genes. Princeton University Press, Princeton (1994). (abridged paperback edition)

Cavalli-Sforza, L.L., Feldman, M.W.: Cultural Transmission and Evolution: A Quantitative Approach. Princeton University Press, Princeton (1981)

Caviness, V.J., Takahashi, T., Nowakowski, R.: Numbers, time and neocortical neuronogenesis: a general developmental and evolutionary model. Trends Neurosci. **18**(9), 379–383 (1995)

Chanock, S., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D., Thomas, G., Hirschhorn, J., Abecasis, G., Altshuler, D., Bailey-Wilson, J., Brooks, L., Cardon, L., Daly, M., Donnelly, P., Fraumeni, J., Freimer, N., Gerhard, D., Gunter, C., Guttmacher, A., Guyer, M., Harris, E., Hoh, J., Hoover, R., Kong, C., Merikangas, K., Morton, C., Palmer, L., Phimister, E., Rice, J., Roberts, J., Rotimi, C., Tucker, M., Vogan, K., Wacholder, S., Wijsman, E., Winn, D., Collins, F.: Replicating genotype-phenotype associations. Nature **447**(7145), 655–660 (2007). doi:10.1038/447655

Charney, E.: Behavior genetics and post genomics. Behav. Brain Sci. **35**(6), 1–80 (2012)

Christiansen, M., Chater, N.: Language as shaped by the brain. Behav. Brain Sci. **31**(5), 489–508 (2008). (discussion 509–58)

Comrie, B.: Language Universals and Linguistic Typology. Basil Blackwell Publisher Limited, Oxford (1981)

Coop, G., Bullaughey, K., Luca, F., Przeworski, M.: The timing of selection at the human FOXP2 gene. Mol. Biol. Evol. **25**(7), 1257–1259 (2008). doi:10.1093/molbev/msn091

Cox, J., Jackson, A.P., Bond, J., Woods, C.G.: What primary microcephaly can tell us about brain growth. Trends Mol. Med. **12**(8), 358–366 (2006)

Croft, W.: Typology and Universals. Cambridge University Press, Cambridge (1990)

Dediu, D.: Causal correlations between genes and linguistic features - the mechanism of gradual language evolution. In: Smith, A.D.M., Smith, K., Ferrer, R., Cancho, I. (eds.) The Evolution of Language: Proceedings of the 7th International Conference (EVOLANG7), pp. 83–90. World Scientific Press, Singapore (2008a)

Dediu, D.: The role of genetic biases in shaping language-genes correlations. J. Theor. Biol. 254, 400–407 (2008b)

Dediu, D.: Are languages really independent from genes? If not, what would a genetic bias affecting language diversity look like? Hum. Biol. **83**(2), 279–296 (2011). doi:10.3378/027.083.0208

Dediu, D., Ladd, D.: Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and microcephalin. Proc. Natl. Acad. Sci. USA **104**(26), 10944–10949 (2007)

Dench, A.: Descent and diffusion: the complexity of the pilbara situation. In: Aikhenvald, A., Dixon, R. (eds.) Areal Diffusion and Genetic Inheritance, pp. 105–133. Oxford University Press, Oxford (2001)

Diamond, J.: The language steamrollers. Nature **389**, 544–546 (1997)

Diamond, J.: Guns, Germs and Steel: A Short History of Everybody for the Last 13,000 Years. Vintage, London (1998)

Diamond, J., Bellwood, P.: Farmers and their languages: the first expansions. Science **300**, 597–603 (2003)

Dixon, R.: The Rise and Fall of Languages. Cambridge University Press, Cambridge (1997)

Dixon, R.: The Australian linguistic area. In: Aikhenvald, A., Dixon, R. (eds.) Areal Diffusion and Genetic Inheritance, pp. 64–104. Oxford University Press, Oxford (2001)

Dryer, M.S.: Order of subject, object and verb. In: Haspelmath, M., Dryer, M.S., Gil, D., Comrie, B. (eds.) The World Atlas of Language Structures Online. Max Planck Digital Library, Munich (2008)

Enard, W., Gehre, S., Hammerschmidt, K., Hölter, S.M., Blass, T., Somel, M., Brückner, M.K., Schreiweis, C., Winter, C., Sohr, R., Becker, L., Wiebe, V., Nickel, B., Giger, T., Müller, U., Groszer, M., Adler, T., Aguilar, A., Bolle, I., Calzada-Wack, J., Dalke, C., Ehrhardt, N., Favor, J., Fuchs, H., Gailus-Durner, V., Hans, W., Hölzlwimmer, G., Javaheri, A., Kalaydjiev, S., Kallnik, M., Kling, E., Kunder, S., Mossbrugger, I., Naton, B., Racz, I., Rathkolb, B., Rozman, J., Schrewe, A., Busch, D.H., Graw, J., Ivandic, B., Klingenspor, M., Klopstock, T., Ollert, M., Quintanilla-Martinez, L., Schulz, H., Wolf, E., Wurst, W., Zimmer, A., Fisher, S. E., Morgenstern, R., Arendt, T., de Angelis, M.H., Fischer, J., Schwarz, J., Pääbo, S.: A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. Cell **137**(5), 961–971 (2009). doi:10.1016/j.cell.2009.03.041

Enard, W., Przeworski, M., Fisher, S., Lai, C., Wiebe, V., Kitano, T., Monaco, A., Pääbo, S.: Molecular evolution of FOXP2, a gene involved in speech and language. Nature **418**, 869–872 (2002)

Evans, P., Gilbert, S., Mekel-Bobrov, N., Vallender, E., Anderson, J., Vaez-Azizi, L., Tishkoff, S., Hudson, R., Lahn, B.: Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. Science **309**(5741), 1717–1720 (2005)

Excoffier, L., Smouse, P., Quattro, J.: Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics **131**, 479–491 (1992)

Fagiolini, M., Jensen, C.L., Champagne, F.A.: Epigenetic influences on brain development and plasticity. Curr. Opin. Neurobiol. **19**(2), 207–212 (2009). doi:10.1016/j.conb.2009.05.009

Falush, D., Stephens, M., Pritchard, J.K.: Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics **164**(4), 1567–1587 (2003)

Felsenfeld, S.: Finding susceptibility genes for developmental disorders of speech: the long and winding road. J. Commun. Disord. **35**, 329–345 (2002)

Fisher, S., Lai, C., Monaco, A.: Deciphering the genetic basis of speech and language disorders. Annu. Rev. Neurosci. **26**, 57–80 (2003)

Fisher, S., Vargha-Khadem, K., Monaco, A., Pembery, M.: Localization of a gene implicated in a severe speech and language disorder. Nat. Genet. **18**, 168–170 (1998)

Fisher, S.E., Scharff, C.: FOXP2 as a molecular window into speech and language. Trends Genet. **25**(4), 166–177 (2009). doi:10.1016/j.tig.2009.03.002

Fortson, B.: Indo-European Language and Culture: An Introduction. Blackwell Publishing, Oxford (2004)

Francks, C., Maegawa, S., Laurén, J., Abrahams, B.S., Velayos-Baeza, A., Medland, S.E., Colella, S., Groszer, M., McAuley, E.Z., Caffrey, T.M., Timmusk, T., Pruunsild, P., Koppel, I., Lind, P.A., Matsumoto-Itaba, N., Nicod, J., Xiong, L., Joober, R., Enard, W., Krinsky, B., Nanba, E., Richardson, A.J., Riley, B.P., Martin, N.G., Strittmatter, S.M., Möller, H.J., Rujescu, D., Muglia, P., Roos, J.L., Fisher, S.E., Wade-Martins, R., Rouleau, G.A., Stein, J.F., Karayiorgou, M., Geschwind, D.H., Ragoussis, J., Kendler, K.S., Airaksinen, M.S., Oshimura, M., DeLisi, L.E., Monaco, A.P.: LRRTM1 on chromosome 2p12 is a maternally suppressed gene that is associated paternally with handedness and schizophrenia. Mol. Psychiatry **12**, 1129–1139 (2007)

Gibson, C.J., Gruen, J.R.: The human lexinome: genes of language and reading. J. Commun. Disord. **41**(5), 409–420 (2008)

Gilbert, S.L., Dobyns, W.B., Lahn, B.T.: Genetic links between brain development and brain evolution. Nat. Rev. Genet. **6**(7), 581–590 (2005)

Gopnik, M., Crago, M.: Familial aggregation of a developmental language disorder. Cognition **39**, 1–50 (1991)

Gordon, R.G. (ed.): Ethnologue: Languages of the World, 15th edn. SIL International, Dallas (2005)

Gray, R.D., Atkinson, Q.D.: Language-tree divergence times support the anatolian theory of Indo-European origin. Nature **426**, 435–439 (2003)

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.-Y., Hansen, N.F., Durand, E.Y., Malaspinas, A.-S., Jensen, J.D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H.A., Good, J.M., Schultz, R., Aximu-Petri, A., Butthof, A.,Höber, B., Höffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E.S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Doronichev, V.B., Golovanova, L.V., Lalueza-Fox, C., de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R.W., Johnson, P.L.F., Eichler, E.E., Falush, D., Birney, E., Mullikin, J.C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D., Pääbo, S.: A draft sequence of the Neandertal genome. Science **328**(5979), 710–722 (2010). doi:10.1126/science.1188021

Greenberg, J.: The Indo-Pacific hypothesis. In: Sebeok, T. (ed.) Current Trends in Linguistics, Linguistics in Oceania, vol. 8, pp. 807–871. Mouton, The Hague (1971)

Greenberg, J.: Language in the Americas. Stanford University Press, Stanford (1987)

Greenberg, J.: The convergence of Eurasiatic and Nostratic. In: Salmons, J., Joseph, B. (eds.) Nostratic: Sifting the Evidence, pp. 51–60. John Benjamins Publishing Co., Amsterdam (1998)

Guégan, J.-F., Prugnolle, F., Thomas, F.: Global spatial patterns of infectious diseases and human evolution. In: Stearns, S.C., Koella, J.C. (eds.) Evolution in Health and Disease, pp. 19–29. Oxford University Press, Oxford (2007)

Haesler, S., Wada, K., Nshdejan, A., Morrisey, E., Lints, T., Jarvis, E., Scharff, C.: FoxP2 expression in avian vocal learners and non-learners. J. Neurosci. **24**, 3164–3175 (2004)

Halliburton, R.: Introduction to Population Genetics. Pearson Eduction Inc., Upper Saddle River (2004)

Haspelmath, M., Dryer, M. S., Gil, D., Comrie, B. (eds.): The World Atlas of Language Structures. Oxford University Press, Oxford (2005)

Hirst, M., Marra, M.A.: Epigenetics and human disease. Int. J. Biochem. Cell Biol. **41**(1), 136–146 (2009). doi:10.1016/j.biocel.2008.09.011

Hurst, J., Baraitser, M., Auger, E., Graham, F., Norell, S.: An extended family with a dominantly inherited speech disorder. Dev. Med. Child Neurol. **32**, 352–355 (1990)

Hyman, L.: Historical tonology. In: Fromkin, V. (ed.) Tone: A Linguistic Survey, pp. 257–269. Academic, London (1978)

Ingram, C.J.E., Swallow, D.M.: Population genetics of lactase persistence and lactose intolerance. In: Cooper, D.N., Kehrer-Sawatzki, H. (eds.) Handbook of Human Molecular Evolution, vol. 2, pp. 1127–1132. John Wiley and Sons Ltd, Chichester (2008)

Jablonka, E., Raz, G.: Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. Q. Rev. Biol. **84**(2), 131–176 (2009)

Jiang, Y., Langley, B., Lubin, F.D., Renthal, W., Wood, M.A., Yasui, D.H., Kumar, A., Nestler, E.J., Akbarian, S., Beckel-Mitchener, A.C.: Epigenetics in the nervous system. J. Neurosci. **28**(46), 11753–11759 (2008). doi:10.1523/JNEUROSCI.3797-08.2008

Jobling, M., Tyler-Smith, C., Hurles, M.: Human Evolutionary Genetics: Origins, Peoples and Disease. Garland Science, New York (2004)

Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R., Burbano, H., Hublin, J.-J., Hänni, C., Fortea, J., de la Rasilla, M., Bertranpetit, J., Rosas, A., Pääbo, S.: The derived FOXP2 variant of modern humans was shared with neandertals. Curr. Biol. **17**(21), 1908–1912 (2007)

Ladd, D., Dediu, D., Kinsella, A.: Languages and genes: reflections on biolinguistics and the nature-nurture question. Biolinguistics **2**(1), 114–126 (2008)

Lai, C., Fisher, S., Hurst, J., Levy, E., Hodgson, S., Fox, M., Jeremiah, S., Povey, S., Jamison, D., Green, E., Vargha-Khadem, F., Monaco, A.: The SPCH1 region on human 7q31: genomic characterization of the critical interval and localization of translocations associated with speech and language disorder. Am. J. Hum. Genet. **67**, 357–368 (2000)

Lai, C., Fisher, S., Hurst, J., Vargha-Khadem, F., Monaco, P.: A forkhead-domain gene is mutated in a severe speech and language disorder. Nature **413**, 519–523 (2001)

Lass, R.: Historical Linguistics and Language Change. Cambridge University Press, Cambridge (1997)

Lewis, C., Levinson, D., Wise, L., DeLisi, L., Straub, R., Hovatta, I., Williams, N., Schwab, S., Pulver, A., Faraone, S., Brzustowicz, L., Kaufmann, C., Garver, D., Gurling, H., Lindholm, E., Coon, H., Moises, H., Byerley, W., Shaw, S., Mesen, A., Sherrington, R., O'Neill, F., Walsh, D., Kendler, K., Ekelund, J., Paunio, T., Lönnqvist, J., Peltonen, L., O'Donovan, M., Owen, M., Wildenauer, D., Maier, W., Nestadt, G., Blouin, J.-L., Antonarakis, S., Mowry, B., Silverman, J., Crowe, R., Cloninger, C., Tsuang, M., Malaspina, D., Harkavy-Friedman, J., Svrakic, D., Bassett, A., Holcomb, J., Kalsi, G., McQuillin, A., Brynjolfson, J., Sigmundsson, T., Petursson, H., Jazin, E., Zoëga, T., Helgason, T.: Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: schizophrenia. Am. J. Hum. Genet. **73**(1), 34–48 (2003)

Lewontin, R.: The apportionment of human diversity. In: Dobzhansky, T., Hecht, M.K., Steere, W.C. (eds.) Evolutionary Biology, vol. 6, pp. 381–398. Appleton-Century-Crofts, New York (1972)

Li, G., Wang, J., Rossiter, S., Jones, G., Zhang, S.: Accelerated FoxP2 evolution in echolocating bats. PLoS ONE **2**(9), e900 (2007)

Liégeois, F., Baldeweg, T., Connelly, A., Gadian, D., Mishkin, M., Vargha-Khadem, F.: Language fMRI abnormalities associated with FOXP2 gene mutation. Nat. Neurosci. **6**, 1230–1237 (2003)

MacDermot, K., Bonora, E., Sykes, N., Coupe, A., Lai, C., Vernes, S., Vargha-Khadem, F., McKenzie, F., Smith, R., Monaco, A., Fisher, S.: Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. Am. J. Hum. Genet. **76**, 1074–1080 (2005)

MacEachern, S.: Genes, tribes, and african history. Curr. Anthropol. **41**, 357–384 (2000)

Maddieson, I.: Vowel quality inventories. In: Haspelmath, M., Dryer, M.S., Gil, D., Comrie, B. (eds.) The World Atlas of Language Structures Online. Max Planck Digital Library, Munich (2008)

Maher, B., Riley, B., Kendler, K.: Psychiatric genetics gets a boost. Nat. Genet. **40**, 1042–1044 (2008). doi:10.1038/ng0908-1042

Mallory, J.: In Search of the Indo-Europeans: Language, Archaeology and Myth. Thames and Hudson, London (1991)

Manni, F., Guérard, E., Heyer, E.: Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using monmonier's algorithm. Hum. Biol. **76**, 173–190 (2004)

Marcus, G., Fisher, S.: FOXP2 in focus: what can genes tell us about speech and language? Trends Cogn. Sci. **7**, 257–262 (2003)

Matisoff, J.: On megalocomparison. Language **66**, 106–120 (1990)

McCarroll, S.A., Altshuler, D.M.: Copy-number variation and association studies of human disease. Nat. Gen. **39**(7 Suppl), S37–S42 (2007)

McCarthy, M., Abecasis, G., Cardon, L., Goldstein, D., Little, J., Ioannidis, J., Hirschhorn, J.: Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat. Rev. Genet., **9**, 356–369 (2008). doi:10.1038/nrg2344

McMahon, A., McMahon, R.: Language Classification by Numbers. Oxford University Press, Oxford (2005)

McMahon, R.: Genes and languages. Commun. Genet. **7**, 2–13 (2004)

Mekel-Bobrov, N., Gilbert, S., Evans, P., Vallender, E., Anderson, J., Hudson, R., Tishkoff, S., Lahn, B.: Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens. Science **309**(5741), 1720–1722 (2005)

Mekel-Bobrov, N., Posthuma, D., Gilbert, S., Lind, P., Gosso, M., Luciano, M., Harris, S., Bates, T., Polderman, T., Whalley, L., Fox, H., Starr, J., Evans, P., Montgomery, G., Fernandes, C., Heutink, P., Martin, N., Boomsma, D., Deary, I., Wright, M., de Geus, E., Lahn, B.: The ongoing adaptive evolution of ASPM and Microcephalin is not explained by increased intelligence. Hum. Mol. Genet. **16**(6), 600–608 (2007)

Mithen, S.: After the ice: a global human history. Phoenix: Orion Books Ltd, London (2003)

Nettle, D.: Linguistic Diversity. Oxford University Press, Oxford (1999)

Nettle, D.: Language and genes: a new perspective on the origins of human cultural diversity. Proc. Natl. Acad. Sci. USA **104**(26), 10755–10756 (2007)

Odling-Smee, F.J., Laland, K.N., Feldman, M.W.: Niche Construction: The Neglected Process in Evolution. Princeton University Press, Princeton (2003)

Pearson, T., Manolio, T.: How to interpret a genome-wide association study. J. Am. Med. Assoc. **299**(11), 1335–1344 (2008)

Plomin, R., Colledge, E., Dale, P.: Genetics and the development of language disabilities and abilities. Curr. Paediatr. **12**, 419–424 (2002)

Plomin, R., Defries, J.C., McClearn, G.E., McGuffin, P.: Behav. Genet., 4th edn. Worth Publishers, New York (2001)

Plomin, R., Kovas, Y.: Generalist genes and learning disabilities. Psychol. Bull. **131**, 592–617 (2005)

Poloni, E., Semino, O., Passarino, G., Santachiara-Benerecetti, A., Dupanloup, I., Langaney, A., Excoffier, L.: Human genetic affinities for Y-chromosome P49a, f/TaqI haplotypes show strong correspondence with linguistics. Am. J. Hum. Genet. **61**, 1015–1035 (1997)

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., González, J.R., Gratacòs, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., Hurles, M.E.: Global variation in copy number in the human genome. Nature **444**(7118), 444–454 (2006)

Relethford, J.: Genetics and the Search for Modern Human Origins. Wiley-Liss, New York (2001)

Renfrew, C.: 'The Emerging Synthesis': the archaeogenetics of farming/language dispersals and other spread zones. In: Bellwood, P., Renfrew, C. (eds.) Examining the Farming/Language Dis-

persal Hypothesis, pp. 3–16. The McDonald Institute for Archaeological Research, Cambridge (2002)

Rivero, O., Sanjuán, J., Moltó, M.-D., Aguilar, E.-J., Gonzalez, J.-C., de Frutos, R., Nájera, C.: The microcephaly ASPM gene and schizophrenia: a preliminary study. Schizophr. Res. **84**(2–3), 427–429 (2006)

Rosenberg, N., Pritchard, J., Weber, J., Cann, H., Kidd, K., Zhivotovsky, L., Feldman, M.W.: Genetic structure of human populations. Science **298**, 2381–2385 (2002)

Rosser, Z.H., Zerjal, T., Hurles, M.E., Adojaan, M., Alavantic, D., Amorim, A., Amos, W., Armenteros, M., Arroyo, E., Barbujani, G., Beckman, G., Beckman, L., Bertranpetit, J., Bosch, E., Bradley, D.G., Brede, G., Cooper, G., Côrte-Real, H.B., de Knijff, P., Decorte, R., Dubrova, Y.E., Evgrafov, O., Gilissen, A., Glisic, S., Gölge, M., Hill, E.W., Jeziorowska, A., Kalaydjieva, L., Kayser, M., Kivisild, T., Kravchenko, S.A., Krumina, A., Kucinskas, V., Lavinha, J., Livshits, L.A., Malaspina, P., Maria, S., McElreavey, K., Meitinger, T.A., Mikelsaar, A.V., Mitchell, R.J., Nafa, K., Nicholson, J., Norby, S., Pandya, A., Parik, J., Patsalis, P.C., Pereira, L., Peterlin, B., Pielberg, G., Prata, M.J., Previderé, C., Roewer, L., Rootsi, S., Rubinsztein, D.C., Saillard, J., Santos, F.R., Stefanescu, G., Sykes, B.C., Tolun, A., Villems, R., Tyler-Smith, C., Jobling, M.A.: Y-chromosomal diversity in europe is clinal and influenced primarily by geography, rather than by language. Am. J. Hum. Genet. **67**, 1526–1543 (2000)

Ruhlen, M.: A Guide to the World's Languages, vol. I. Classification (with a Postscript on Recent Developments). Arnold, London (1991)

Ruhlen, M.: On the Origin of Languages: Studies in Linguistic Taxonomy. Stanford University Press, Stanford (1994)

Scharff, C., Haesler, S.: An evolutionary perspective on FoxP2: strictly for the birds? Curr. Opin. Neurobiol. **15**, 694–703 (2005)

Scharff, C., White, S.: Genetic components of vocal learning. Ann. N. Y. Acad. Sci. **1016**, 325–347 (2004)

Segurado, R., Detera-Wadleigh, S., Levinson, D., Lewis, C., Gill, M., Nurnberger, J., Craddock, N., DePaulo, J., Baron, M., Gershon, E., Ekholm, J., Cichon, S., Turecki, G., Claes, S., Kelsoe, J., Schofield, P., Badenhop, R., Morissette, J., Coon, H., Blackwood, D., McInnes, L., Foroud, T., Edenberg, H., Reich, T., Rice, J., Goate, A., McInnis, M., McMahon, F., Badner, J., Goldin, L., Bennett, P., Willour, V., Zandi, P., Liu, J., Gilliam, C., Juo, S.-H., Berrettini, W., Yoshikawa, T., Peltonen, L., Lönnqvist, J., Nöthen, M., Schumacher, J., Windemuth, C., Rietschel, M., Propping, P., Maier, W., Alda, M., Grof, P., Rouleau, G., Del-Favero, J., Broeckhoven, C., Mendlewicz, J., Adolfsson, R., Spence, M., Luebbert, H., Adams, L., Donald, J., Mitchell, P., Barden, N., Shink, E., Byerley, W., Muir, W., Visscher, P., Macgregor, S., Gurling, H., Kalsi, G., McQuillin, A., Escamilla, M., Reus, V., Leon, P., Freimer, N., Ewald, H., Kruse, T. A., Mors, O., Radhakrishna, U., Blouin, J.-L., Antonarakis, S., Akarsu, N.: Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: bipolar disorder. Am. J. Hum. Genet. **73**(1), 49–62 (2003)

Shu, W., Cho, J., Jiang, Y., Zhang, M., Weisz, D., Elder, G., Schmeidler, J., De Gasperi, R., Sosa, M., Rabidou, D., Santucci, A., Perl, D., Morrisey, E., Buxbaum, J.: Altered ultrasonic vocalization in mice with a disruption in the foxp2 gene. Proc. Natl. Acad. Sci. USA **102**, 9643–9648 (2005)

Sims-Williams, P.: Genetics, linguistics, and prehistory: thinking big and thinking straight. Antiquity **72**, 505–527 (1998)

Skelton, P.: Evolution: a biological and palaeoantological approach. The Open University, Milton Keynes (1993)

Speed, T., Zhao, H.: Chromosome maps. In: Balding, D., Bishop, M., Cannings, C. (eds.) Handbook of Statistical Genetics, vol. 1, 3rd edn. Chapter 1, pp. 3–39. John Wiley and Sons., Ltd., Chichester (2007)

Spiteri, E., Konopka, G., Coppola, G., Bomar, J., Oldham, M., Ou, J., Vernes, S., Fisher, S., Ren, B., Geschwind, D.: Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. Am. J. Hum. Genet. **81**(6), 1144–1157 (2007)

Stromswold, K.: The heritability of language: a review and metaanalysis of twin, adoption, and linkage studies. Language **77**, 647–723 (2001)

Templeton, A.: Human races: a genetic and evolutionary perspective. Am. Anthropol. **100**, 632–650 (1998)

Teramitsu, I., Kudo, L., London, S., Geschwind, D., White, S.: Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction. J. Neurosci. **24**, 3152–3163 (2004)

The Wellcome Trust Case Control Consortium.: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature **447**, 661 (2007). doi:10.1038/nature05911

Trask, R.: Why should a language have any relatives? In: Renfrew, C., Nettle, D. (eds.) Nostratic: Examining a Linguistic Macrofamily, pp. 157–178. The McDonald Institute for Archaeological Research, Oxbow Books, Oxford (1999)

Trask, R.L.: Historical Linguistics. Arnold, London (1996)

Vargha-Khadem, F., Watkins, K., Price, C., Ashburner, J., Alcock, K., Connelly, A., Frackowiak, R., Friston, K., Pembrey, M., Mishkin, M., Passingham, R.: Neural basis of an inherited speech and language disorder. Proc. Natl. Acad. Sci. USA **95**, 12695–12700 (1998)

Vernes, S.C., Newbury, D.F., Abrahams, B.S., Winchester, L., Nicod, J., Groszer, M., Alarcón, M., Oliver, P.L., Davies, K.E., Geschwind, D.H., Monaco, A.P., Fisher, S.E.: A functional genetic link between distinct developmental language disorders. New Eng. J. Med. **359**(22), 2337–2345 (2008)

Wang, J.-K., Li, Y., Su, B.: A common SNP of MCPH1 is associated with cranial volume variation in Chinese population. Hum. Mol. Genet. **17**(9), 1329–1335 (2008)

Watkins, K., Vargha-Khadem, F., Ashburner, J., Passingham, R., Connelly, A., Friston, K., Frackowiak, R., Mishkin, M., Gadian, D.: MRI analysis of an inherited speech and language disorder: structural brain abnormalities. Brain **125**, 465–478 (2002)

Webb, D., Zhang, J.: FoxP2 in song-learning birds and vocal-learning mammals. J. Hered. **96**, 212–216 (2005)

Weedon, M.N., Lettre, G., Freathy, R.M., Lindgren, C.M., Voight, B.F., Perry, J.R.B., Elliott, K.S., Hackett, R., Guiducci, C., Shields, B., Zeggini, E., Lango, H., Lyssenko, V., Timpson, N.J., Burtt, N.P., Rayner, N.W., Saxena, R., Ardlie, K., Tobias, J.H., Ness, A.R., Ring, S.M., Palmer, C.N.A., Morris, A.D., Peltonen, L., Salomaa, V., Initiative, D.G., Consortium, W.T.C.C., Smith, G.D., Groop, L.C., Hattersley, A.T., McCarthy, M.I., Hirschhorn, J.N., Frayling, T.M.: A common variant of HMGA2 is associated with adult and childhood height in the general population. Nat. Genet. **39**(10), 1245–1250 (2007). doi:10.1038/ng2121

Wong, P.C.M., Chandrasekaran, B., Zheng, J.: The derived allele of aspm is associated with lexical tone perception. PLoS One **7**(4), e34243 (2012). doi:10.1371/journal.pone.0034243

Woods, C.G.: Human microcephaly. Curr. Opin. Neurobiol. **14**(1), 112–117 (2004)

Woods, R., Freimer, N., Young, J.D., Fears, S., Sicotte, N., Service, S., Valentino, D., Toga, A., Mazziotta, J.: Normal variants of Microcephalin and ASPM do not account for brain size variability. Hum. Mol. Genet. **15**(12), 2025–2029 (2006)

Yip, M.: Tone. Cambridge University Press, Cambridge (2002)

Youngson, N.A. Whitelaw, E.: Transgenerational epigenetic effects. Annu. Rev. Genomics Hum. Genet. **9**, 233–257 (2008). doi:10.1146/annurev.genom.9.081307.164445

Zhang, J., Webb, D., Podlaha, O.: Accelerated protein evolution and origins of human-specific features: Foxp2 as an example. Genetics **162**, 1825–1835 (2002)