

# Using Genetic Data to Make Linguistic Arguments: Are clicks evidence of late externalization?

Remo Nitschke\*

In this paper<sup>1</sup> I will show that using data from other fields, such as genetics may be very valuable to linguistics, but that we must take care to ensure that the data we use from other fields are suitable for the kinds of linguistic arguments we want to make. Huybregts (2017) brings forth a particular argument regarding the nature of click-phonemes and what they can tell us about human language evolution, based on genetic data on Khoe-San, Hadza and Sandawe language communities. I will show that the data they use is unsuitable for that specific argument. I will further provide different genetic evidence that calls theirs into question and provide some additional linguistic considerations that support a different view of the matter. Finally, I will propose an alternative story on the historical nature of click phonemes in these languages, suggesting that clicks might be a more recent linguistic innovation.

## 1 Huybregts 2017

Huybregts (2017) provides a welcome example of how linguistics can utilize insights from other sciences, such as genetics. It also provides a useful example for how we need to be careful with data from other sciences. This section will briefly discuss the main points of Huybregts (2017), as they are the starting point of further investigations.

Huybregts (2017) attempts to use genetic evidence of the earliest known split in the anatomically modern human population to date the latest possible point of the emergence of a language faculty (FL). It should be noted, however, that their arguments only fully apply if one assumes FL to be some sort of (uniquely human) genetic endowment. For the purpose of this paper the author will do just that.

The earliest known divergence of the human population can be dated anywhere between 300kya<sup>2</sup> (Scally and Durbin 2012:478), over 180kya (Rito et al. 2013:4) to 100kya (Schleebusch et al. 2012:376). The different ranges here largely depend on the type of DNA sampling that is done (cf. Scally and Durbin 2012). Huybregts (2017:28) assume around 125kya. During this divergence event the Khoisan group, here understood as an ancestry group, the first node in a series of proliferations, separated from the main human population. Based on this early split, Huybregts (2017) forms two hypotheses:

**Hypothesis 1** Assuming that the human language ability is largely due to some kind of genetic endowment, the time of this divergence constitutes the latest possible point in time at which such a mutation must have surfaced in the population. In Chomskyan terms: the Khoisan divergence marks the latest possible point in time for the emergence of an I-language. We can thus reasonably date the

---

\*I would like to acknowledge the help of, and/or useful discussions with, the following persons: Riny Huijbregts, Shigeru Miyagawa, Robert Berwick, Mark Lipson, David Reich, Peter Norquest, Antonia Pock and Aaron Golish

<sup>1</sup>As this paper will deal with a blend of linguistic and genetic terms, I wish to clarify some of these right away: If I refer to 'genealogy' or 'ancestry' in this paper, I take these to be understood in genetic terms, not in linguistic terms, unless otherwise specified. If I talk about 'Khoisan (peoples)' I take these to be a bio-genetic affiliation which may be taken to be a proposed *ancestor*-group to (some) modern peoples living in present day South Africa, including speakers of Khoe-Khwadi languages, Tuu languages, and Ju-ǀHoan languages. However, I wish to make a clear distinction between this proposed ancestor group, and the actual modern day peoples, about who I do not wish to make any kind of argument. When using the term 'Khoe-San', I take this to be a proposed umbrella term for the above language groups, without making a statement about any linguistic relationships between those three families. Finally, I use the terms 'Hadza' and 'Sandawe' as a linguistic and bio-genetic affiliations of modern day peoples living in present day Tanzania.

<sup>2</sup>300,000 years ago

lower bound of the possible time-range of language evolution in the biological sense.<sup>3</sup>

The basic argument behind this hypothesis is as follows: Assuming the modern human population had not yet developed a genetic endowment for FL at the time of the Khoisan divergence, then we would expect to see modern descendant populations from that split that do not have a capacity for language today. As we do not find these populations, humans must have had the necessary genetic mutation at the time of the split.

There are of course two other possibilities that would allow for a genetic endowment to arise at a later point. Both branches of the human population could have developed the same, or a similar endowment after the divergence. We can disregard this possibility, as the chances of the same genetic mutation occurring twice independently seem astronomically low, and even if the two branches had developed just *similar* endowments, we should be able to see the differences in ability or externalization of those endowments in present day humans, which we do not find. A second possibility would be that one branch develops a genetic endowment after the split, and transfers it to the other branch in later admixture events. This second option is not as easily disregarded as the first, but we will put it aside for this discussion.

**Hypothesis 2** The externalization of the aforementioned genetic endowment for language (FL) occurred after the separation of the Khoisan branch (cf. Huybregts 2017:287-288). Furthermore: “clicks emerged in the oldest human lineage [Khoisan] after the first divergence in human population history” (Huybregts 2017:288) and that is why “[within] Africa clicks are [...] uniquely associated with Khoisan [here: Khoe-San]” (Huybregts 2017:281).

Huybregts (2017) labels the modern descendants of the Khoisan divergence event as the communities that are encompassed by the, formerly proposed but no longer accepted (cf. Güldemann and Sands 2009), “Southern Khoisan” (South Africa) language family and the isolates of the Hadza and Sandawe communities (East Africa), which constitute all of the language communities that have native click phonemes.<sup>4</sup> They primarily base this argument on evidence by Pickrell et al. (2012) who established distant shared ancestry between the Sandawe and Hadza groups of Eastern Africa and the speaker communities of Khoe-Khwadi languages, Tuu languages, and Ju-ǀHoan languages, which they assume to be the modern descendants of the first split in modern human populations.

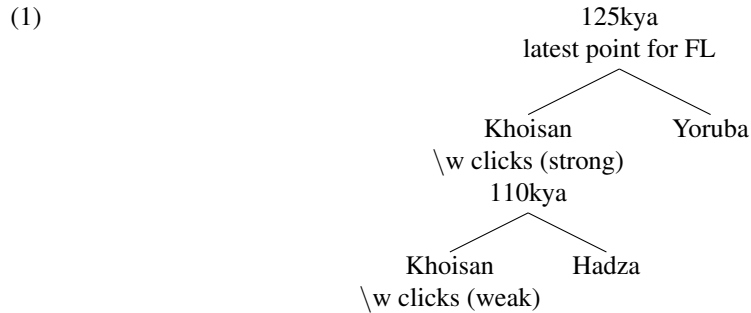
The Hadza and Sandawe languages were once considered to be part of the “Khoisan” languages,<sup>5</sup> but they have since been re-analyzed as isolates of a larger “Khoisan” family, or, more likely, not related at all (cf. Güldemann and Sands 2009:205) (for Sandawe, cf. du Plessis 2014:583). Huybregts (2017) therefore needs to establish a genetic relationship between these groups and the Southern Khoisan language groups, to be able to make the argument that clicks developed only in that part of the human population that split 300-100kya. The descendant communities must therefore have retained clicks throughout their history up to this day. According to Huybregts (2017), language externalization happened after the Khoisan split from the general human population. This allowed the “Khoisan” group to develop click phonemes, which were then retained within the ancestral languages of the Khoe-Khwadi, Tuu, Ju-ǀHoan, the Hadza and the Sandawe language communities for up to 110k years<sup>6</sup> (Huybregts 2017:287), which explains why clicks are not natively found in any other language community. The general idea is visualized in example (1) which is an abridged version of Huybregts (2017:287) Fig.2.

<sup>3</sup>It should be noted, that we still have very little to say about the higher bound.

<sup>4</sup>This requires us to disregard Dahalo, and possibly Damin.

<sup>5</sup>A Greenberg (1963) classification which encompassed Khoe-Khwadi, Tuu, Ju-ǀHoan, Sandawe, and Hadza, largely on the account that they all have click-phonemes.

<sup>6</sup>At least according to the numbers that Huybregts (2017) uses. If we consider Skoglund et al. (2017) this would mean retention of those phonemes over 200-300,000 years.



Huybregts (2017) offers two possible takes on Hypothesis 2: Externalization happened in the Khoisan<sup>7</sup> branch before the ancestors of modern Hadza and possibly Sandawe split from the Khoisan branch (strong hypothesis). Or, externalization happened in the Khoisan branch after the ancestors of modern Hadza split from the Khoisan branch. Hadza and Sandawe acquired clicks through later contact situations (weak hypothesis). In the following sections I will show why both versions of this hypothesis are problematic.

## 2 Mismatches in the Genetic Evidence

I would like to preface this section with a warning:<sup>8</sup> The genetic structure of early humans within the time frame we are discussing is still not well understood. While genetics may seem like an exact science from a linguist's point of view, there are a lot of issues at play, and neither the data nor the conclusions we can draw from it are as clear cut as we would like.

The first issue that we need to address, is that different studies use different methods, which may produce wildly different results. The most common methods used in these studies are based on mtDNA, Y-chromosomes, genome wide analysis, genome wide shotgun sequencing, or a combination of the above. Mitochondrial DNA (mtDNA) is popular for these purposes, because it has a “higher copy number per cell than nuclear DNA and hence is more readily sampled” (Scally and Durbin 2012:746). But, mtDNA is only passed down through the maternal line, which may be problematic if there was sex-biased gene flow (cf. Shriner et al. 2018:880). It also has a high mutation rate and different regions (of mtDNA) seem to mutate at different speeds (cf. Scally and Durbin 2012:746). This is a problem, as it muddies the data. Y-chromosome based methods also suffer from the sex-bias problem, as these are passed down through the paternal line. Finally, genome wide analyses and shotgun analyses<sup>9</sup> tend to be the most accurate.<sup>10</sup>

Exacerbating these inaccuracies is the fact that these models have to rely on estimated mutation rates and estimated generation times to derive their phylogenetic trees (cf. Scally and Durbin 2012). If any of these are slightly off, this results in wildly different predictions for time frames. For example, a lot of recent studies use *de novo* mutation rates (Fig. 1 Scally and Durbin 2012:476), which means they count mutations per generation. So, assuming we are looking at data that suggests 6000 generations according to our parameters, then we are looking at 150,000 years, with an assumed 25 years per generation. Assuming 30 years per generation, this number jumps to 180,000 years. For a more in depth-excursion into these issues, I recommend Scally and Durbin (2012).

Let us briefly revisit the main idea behind Hypothesis 2: At the first split in human population, we had not externalized our capacity for language. After the Khoisan branch split, externalization occurred in both branches individually. Only the Khoisan branch developed click phonemes. Speakers of Khoe-Khwadi, Tuu, Ju-ǀHoan, Sandawe, and Hadza are all descendant populations of the Khoisan branch. That is why they are the only language communities that *natively* have click

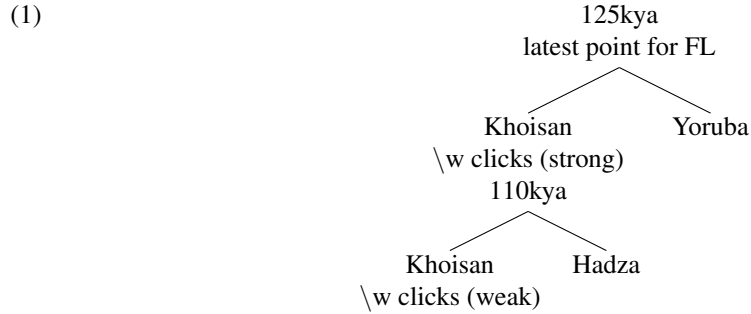
<sup>7</sup>Khoisan is used as a genetic affiliation here, not as a linguistic affiliation.

<sup>8</sup>As pointed out to me by Mark Lipson, 3/19/2019 (pers. comm.).

<sup>9</sup>Shotgun sequencing describes a method where DNA sequences are randomly broken up into smaller bits for easier computation and later reassembled.

<sup>10</sup>Although, for example, a genome wide analysis of only mtDNA would suffer the same problems as described for just an mtDNA analysis.

phonemes. The presence of these click phonemes is therefore supporting evidence for late externalization (cf. Huybregts 2017). Such is the basic argument for the strong version of Hypothesis 2. In the weak version clicks are developed *after* the Hadza and possibly Sandawe ancestors split from the Khoisan branch and are later transferred through some contact event. However, since the whole point of this argument is the late externalization, this would entail that externalization for the Khoisan branch had to happen after the Hadza split. This would delay the date for externalization in that branch by multiples of thousands of years, 15,000 according to Huybregts (2017) own numbers. Let us revisit figure (1) for visualization:



Huybregts (2017) needs strong genetic evidence linking the Hadza and Sandawe communities living in East Africa to the other click speaking language communities in South Africa to support this hypothesis. Since there is no linguistic evidence that these languages are related (cf. Güldemann and Sands 2009), they need a genetic link. Huybregts (2017) largely relies on Pickrell et al. (2012) for this. Pickrell et al. (2012:4) attest that their analysis “provides strong evidence for a shared origin for the Khoisan-related genetic material in the Hadza and Sandawe”. This would support Hypothesis 2, but there are a series of other studies pointing in different directions. Lachance et al. (2012) for example, using whole genome sequencing of Hadza, Sandawe, and Cameroon Pygmy individuals, come to a different result. They propose that ancient Sandawe and Hadza populations may have diverged from ancient Pygmy populations (Lachance et al. 2012:458, F), which puts into question the idea that clicks developed in the ancient Khoisan divergence and have since been retained by the Hadza and Sandawe for 110k years (Huybregts 2017:287). Even if, in this scenario, ancient pygmies can be derived as branching of the Khoisan group (cf. Duda and Zrzavý 2016), one would have to explain why the other descendants of ancient pygmies lost their clicks,<sup>11</sup> while Hadza and Sandawe retained them.

Skoglund et al. (2017) used genome-wide sampling, Y-chromosome sampling, and mtDNA sampling of 16 individuals that lived in Africa up to 8.1k years BP<sup>12</sup> and compared them to genetic data of modern day Africans to gain more fine grained insight into early divergence events. They find that “individuals from the Hadza group in Tanzania can be modeled as deriving all their ancestry from a lineage related deeply to ancient eastern Africans such as the Ethiopia 4500BP individual” (Skoglund et al. 2017:63), which they do not model as branching of the divergence that constitutes the Khoisan split (cf. Skoglund et al. 2017:65, Fig. 3, A). They also model Sandawe “as having ancestry similar to the Hadza but also admixture related to that of neighboring populations” (Skoglund et al. 2017:64).

Shriner et al. (2018), in a genome-wide study, suggest that “Hadza ancestry is closer to Omotic<sup>13</sup> ancestry than to Khoisan ancestry [...] [and] the core of the Sandawe sample is predominantly Omotic” (Shriner et al. 2018:880). They suggest that the admixture of Khoisan ancestry found by Pickrell et al. (2012) is a result of a more recent migration event (cf. Shriner et al. 2018:879). This could be potentially squared with a weak version of Hypothesis 2, where clicks are introduced to Hadza and Sandawe in a later contact situation.

Tishkoff et al. (2007) has similar findings to Shriner et al. (2018), arguing that Hadza and Sandawe are “genetically more similar to their Nilotic-, Cushitic-, and Bantu-speaking neighbors

<sup>11</sup>Branching off the Khoisan tree, they would have to have click phonemes, in this scenario.

<sup>12</sup>Before Present: Before 1950

<sup>13</sup>Modern speakers of Afro-Asiatic languages

than they are to the SAK-speaking [Khoe-San] population” (Tishkoff et al. 2007:2192). They also point out that modern day Hadza are estimated at a population size of around 1,000 individuals (cf. Tishkoff et al. 2007:2181). Such small and isolated populations<sup>14</sup> are at a big risk of random genetic drift, making exact readings more difficult.

## 2.1 The Dahalo Case

One final issue, is the case of Dahalo. Huybregts (2017) needs to establish that the Khoe-San group-ing, Hadza, and Sandawe constitute the only click speaking language communities for his argument, as they are the only ones they can argue to derive from the Khoisan split.<sup>15</sup> This requires them to disregard Dahalo, a Cushitic language that uses clicks.<sup>16</sup> They offer an explanation where Dahalo clicks may be a result of a language shift as argued in Güldemann and Sands (2009). However, Güldemann and Sands (cf. 2009:215) assume that Dahalo speakers may originally have spoken a language similar, or related to, Sandawe, before shifting to a different language family (Cushitic). Therefore, if Hypothesis 2 is to be assumed, one would have to be able to link Dahalo speakers genetically to the Khoisan split as well. As far as I am aware, no work has been done on this yet. In any case, in trying to discount Dahalo, Huybregts (2017) inadvertently links it to the Khoisan divergence, only complicating their case further.

All these conflicting genetic data put a big dent in the likelihood of the strong version of Hypothesis 2. As we see conflicting results from a number of studies, we cannot establish a clear genetic link with the Hadza and Sandawe groups, and the Khoisan split. Perhaps the future will bring more sophisticated technology and more precise data, but as of now genetics cannot deliver the kind of fine-grained and clear results we would need to sufficiently argue for a strong Hypothesis 2. In the next section I will bring forth some linguistic considerations that call in question both the strong and the weak version of Hypothesis 2.

## 3 Linguistic Considerations

In this section I will bring forth some linguistic considerations that call into question Hypothesis 2. I also aim to show that we can be equally justified in thinking about clicks as recent innovation that is not language family dependent, but instead can be considered an areal feature.

Assuming that clicks have been part of the Khoisan branch of human ancestry for at least 100k years,<sup>17</sup> is highly questionable from a linguistic point of view. We know that clicks are not uniquely acquirable by peoples with Khoisan ancestry, as many Bantu languages have extensively borrowed clicks from the Southern Khoe-San languages through the Bantu expansion (Huybregts 2017). Clicks, like any other type of phoneme, are not restricted to genetic inheritance and may therefore be innovated or borrowed and transferred through language contact situations.

Under this light, if we want to hold on to any version of Hypothesis 2, we need to assume two things: Clicks are somehow extremely retentive and once acquired will stay in a language community for 100k years. There has been no significant language contact situations with click speaking populations and non-click speaking populations before the Bantu expansion. I will discuss the prior first.

Huybregts (2017:280) states that “click-consonants systems seem very stable”, but they do not divulge exactly how they are stable or in any way more stable than other consonant systems. In fact, clicks seem if anything, unexceptional in terms of their stability and learnability. Güldemann and Sands (2009:215) cite multiple instances of click-loss in Khoe-Khwadi languages, such as the loss of /ʔ, / in ||Xwégi. They argue, based on the various losses, borrowings, and re-borrowings of clicks among the Khoe-San language group and Bantu languages that: “Clicks appear to undergo sound

<sup>14</sup>Hadza are a fairly isolated hunter and gatherer society.

<sup>15</sup>Although I have already shown that this relationship cannot be established firmly.

<sup>16</sup>I will disregard Damin here, but come back to it later, as it never had any native speakers.

<sup>17</sup>If we believe some of the newer studies like Skoglund et al. (2017), then we are looking at an even longer time.

change in the same way as do other sounds; there is no evidence to view them as being particularly prone either to loss or to fossilization.” (Güldemann and Sands 2009:216)

The complete lack of clicks outside of the Khoe-San group, Dahalo, Bantu, Hadza, and Sandawe, would force a believer of Hypothesis 2 to assume, that, before the Bantu expansion none of the “Khoisan” languages encountered any language contact situations that would facilitate the transfer of clicks to another community. Especially, since we would also need to believe that clicks are unusually ‘stable’, so once acquired in a population, we would expect to still see them today.<sup>18</sup> This does not only seem unlikely, but is also called into question through the same genetic studies cited by Huybregts (2017). For example, Pickrell et al. (2012) attest admixture in the Hadza ancestry with “a population most closely related to the Dinka” (Pickrell et al. 2012:4). More importantly, Skoglund et al. (2017:65) shows that: “ancient southern Africans who lived ~2,000 BP were not strictly an outgroup to extant lineages in other parts of sub-Saharan Africa.” Sikora et al. (2011) show that the ancestral Khoisan group is strongly related to ancient Pygmies, a signal that Shrinier et al. (2018) also find. It is therefore unlikely, that the ancestors of modern Khoe-San speakers did not come in any significant contact situations before the Bantu expansion. If that is the case, then why are clicks not more widespread?

### 3.1 The Myth of the “Ancient” Click

I would like to discuss the seemingly accepted idea, that clicks are somehow ancient. There seems to be an unspoken consensus in some linguistic communities, that linguistic phenomena are somehow ‘older’, or ‘ancient’, and that African language communities in general, and hunter-gatherer societies in particular, are somehow more conservative or “a window to the past”. As far as I can see, nothing we know about linguistics supports this belief. And nothing about the Khoe-San language group supports the belief that they are somehow ‘ancient’. As du Plessis (2014:573) points out: “In reality, none of the Khoisan [Khoe-San] language groupings shows the kind of extensive proliferation that would be commensurate with a great antiquity.”

Africa is not a linguistically conservative ecosystem. If anything, it is the other way around. The continent boasts the second highest Linguistic Diversity Index, according to Harmon and Loh (2010), second only to Eurasia (cf. Harmon and Loh 2010:109). Africa also has the highest phonemic diversity among the world’s regions, as pointed out by Perreault and Matthew (2012).<sup>19</sup> Considering these facts, we should view Africa as a highly diverse and *innovative* region.

### 3.2 Clicks as linguistic Innovations

The case of Damin, a ritual register of Lardil,<sup>20</sup> shows us that clicks can be innovated by language communities regardless of genetic heritage. Damin is often dismissed, as it was purely a ritual language and thus never had any native speakers. But, it serves as excellent evidence that clicks *can* make the jump from paralinguistic use to full phonemes and that they do not *need* to be borrowed but may be *innovated*.

If we believe in the idea of a “Khoisan language family” and disregard Dahalo and Damin, then an idea like Hypothesis 2 seems not so unlikely and clicks seem to be a rare and uniquely “Khoisan” phenomenon. However, there is no such thing as a Khoisan language family and the cases of Dahalo and Damin do exist. Neither are hunter-gatherer societies particularly linguistically conservative. Considering the linguistic and phonemic diversity of Africa, it does not seem particularly strange to me that clicks may just constitute a *recent linguistic innovation*. Perhaps clicks were innovated by different communities at different, more recent, times in the past and have since existed as an areal feature in Southern Africa and Eastern Africa independently. A recent innovation hypothesis has the

<sup>18</sup>Essentially, if we believe that clicks were retained over 100k years in Khoe-San, then we have to believe that they would be retained anywhere else.

<sup>19</sup>However, they take this to be indicative of some kind of Proto-World idea, where diversity is lost with distance to Africa. This seems an interpretation of the data that is very much in the spirit that I criticize in the prior paragraph.

<sup>20</sup>Lardil is a Tangkic language spoken in the Wellesley Islands of Australia.

advantage that it also does away with the problem of why we do not see clicks in populations where we can see old genetic admixtures from common ancestors of modern click-speakers, such as the pygmy case. These populations were simply in contact with each other before those communities innovated clicks. There were no clicks to borrow, yet. I agree with Güldemann and Sands (2009:218) when they argue that: “The possibility is very real that the emergence of clicks as phonemes in Africa represents a far later episode in the diversification of human speech”.

## 4 Discussions

Huybregts (2017) shows both the opportunities and the pitfalls of using genetic data in linguistics. Hypothesis 1 is a great contribution to the study of the evolution of language capacity. Even if the geneticist cannot agree on a clear date yet, the argument is still sound. The insight, that at the point of the earliest diversification event in human history we must have had, at latest, a genetic endowment for language, is a very useful one, and a good example of how insights from other sciences can inform linguistics.

Hypothesis 2 is ambitious in its scope, trying to show that externalization happened after this split on the basis of clicks. Unfortunately, the evidence from genetics is not conclusive in any way. Since the late externalization argument, which is the main point of Hypothesis 2, hinges largely on the proposition that clicks are a result of late externalization, the hypothesis becomes untenable once the genetic evidence becomes inconclusive.

The kind of attitudes I discuss in section 3.1 can distort our view of the data and bring us to conclusions that do not match up with our knowledge about linguistics. In the worst cases, we create feedback loops. A major reason why there are so many genealogical studies of the Khoe-San language communities and their relations to the Hadza and Sandawe is the old Greenberg (1963) classification of the “Khoisan language family”. As geneticists try to inform their decisions with insights from linguistics, they reinforce the old idea of an ancient “Khoisan language family”.<sup>21</sup>

Using evidence from other fields is something linguistics can greatly benefit from, as Hypothesis 1 shows. We just need to ensure that the arguments we construct rely on carefully vetted data and that the data we use actually tells the kind of story we are aiming to tell. To that end, we will have to work more closely with the people from those fields.

## References

- Duda, Pavel, and Jan Zrzavý. 2016. Human population history revealed by a supertree approach. *Scientific Reports* 6:1–10.
- Greenberg, Joseph. 1963. *The Languages of Africa*. Bloomington: Indiana University Press.
- Güldemann, Tom, and Bonny Sands. 2009. What click languages can and can’t tell us about language origins. In *The Cradle of Language*, ed. Rudolf Botha and Chris Knight, 204–218. Oxford: OUP.
- Harmon, David, and Jonathan Loh. 2010. The index of linguistic diversity: A new quantitative measure of trends in the status of the world’s languages. *Language Documentation & Conservation* 4:97–151.
- Huybregts, M.A.C. (Riny). 2017. Phonemic clicks and the mapping asymmetry: How language emerged and speech developed. *Neuroscience and Biobehavioral Reviews* 81:280–294.
- Lachance, Joseph, Benjamin Vernot, Clara C. Elbers, Bart Ferwerda, Alain Froment, Jean-Marie Bodo, Godfrey Lema, Wenqing Fu, Thomas B. Nyambo, Timothy R. Rebbeck, Kun Zhang, Joshua M. Akey, , and Sarah A. Tishkoff. 2012. Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse african hunter-gatherers. *Cell* 150:457–469.
- Perreault, Charles, and Sarah Matthew. 2012. Dating the origin of language using phonemic diversity. *PLOS One* 7:e35289.

<sup>21</sup>For example: Veeramah et al. (2012) speak of “Khoisan” languages throughout their paper.

- Pickrell, Joseph K., Nick Patterson, Chiara Barbieri, Falco Berthold, Linda Gerlach, Tom Gldemann, Blesswell Kure, Sununguko Wata Mpoloka, Hiroshi Nakagawa, Christfried Naumann, Mark Lipson, Po-Ru Loh, Joseph Lachance, Joanna Mountain, Carlos D. Bustamante, Bonnie Berger, Sarah A. Tishkoff, Brenna M. Henn, Mark Stoneking, David Reich, and Brigitte Pakendorf. 2012. The genetic prehistory of southern africa. *Nature Communications* 3.
- du Plessis, Men'an. 2014. The damaging effects of romantic mythopoeia on khoesan linguistics. *Critical Arts* 28:569–592.
- Rito, Teresa, Martin B. Richards, Veronica Fernandes, Farida Alshamali, Viktor Cerny, Luisa Pereira, and Pedro Soares. 2013. The first modern human dispersals across africa. *PLOS one* 8:1–16.
- Scally, Aylwyn, and Richard Durbin. 2012. Revising the human mutation rate: implications for understanding human evolution. *Nature Reviews Genetics* 13:745–753.
- Schleebusch, Carina M., Pontus Skoglund, Per Sjdin, Lucie M. Gattepaille, Dena Hernandez, Flora Jay, Sen Li, Michael De Jongh, Andrew Singleton, Micheal G. B. Blum, Himla Soodyall, and Mattias Jakobsson. 2012. Genomic variation in seven khoe-san groups reveals adaption and complex african history. *Science* 338:374–379.
- Shriner, Daniel, Fasil Tekola-Ayele, Adebawale Adeyemo, and Charles N. Rotimi. 2018. Genetic ancestry of hadza and sandawe peoples reveals ancient population structure in africa. *Genome Biol. Evol.* 10:875–882.
- Sikora, Martin, Hafid Laayouni, Francesc Calafell, David Comas, and Jaume Bertranpetit. 2011. A genomic analysis identifies a novel component in the genetic structure of sub-saharan african populations. *European Journal of Human Genetics* 19:84–88.
- Skoglund, Pontus, Jessica C. Thompson, Mary E. Prendergast, Alissa Mittnik, Kendra Sirak, Mateja Hajdinjak, Tasneem Salie, Nadin Rohland, Swapan Mallick, Alexander Peltzer, Anja Heinze, Inigo Olalde, Matthew Ferry, Eadaoin Harney, Megan Michel, Kristin Stewardson, Jessica I. Cerezo-Roman, Chrissy Chiumia, Alison Crowther, Elizabeth Gomani-Chindebvu, Agness O. Gidna, Katherine M. Grillo, I. Taneli Helenius, Garrett Hellenthal, Richard Helm, Mark Horton, Saioa Lopez, Audax Z.P. Mabulla, John Parkington, Ceri Shipton, Mark G. Thomas, Ruth Tibesasa, Menno Welling, Vanessa M. Hayes, Douglas J. Kennett, Raj Ramesar, Matthias Meyer, Svante Pbo, Nick Patterson, Alan G. Morris, Nicole Boivin, Ron Pinhasi, Johannes Krause, and David Reich. 2017. Reconstructing prehistoric african population structure. *Cell* 171:59–71.
- Tishkoff, Sarah A., Mary Katherine Gonder, Brenna M. Henn, Holly Mortensen, Alec Knight, Christopher Gignoux, Neil Fernandopulle, Godfrey Lema, Thomas B. Nyambo, Uma Ramakrishna, Floyd A. Reed, and Joanna L. Mountain. 2007. History of click-speaking populations of africa inferred from mtDNA and y chromosome genetic variation. *Mol. Biol. Evol.* 24:2180–2195.
- Veeramah, Krishna, Daniel Wegmann, August Woerner, Fernando Mendez, Joseph Watkins, Destro-Bisol Giovanni, Himla Soodyall, Leslie Louie, and Micheal Hammer. 2012. An early divergence of khoesan ancestors from those of other modern humans is supported by an abc-based analysis of autosomal resequencing data. *Mol. Biol. Evol.* 29:617–630.

Department of Linguistics  
P.O. Box 210025  
The University of Arizona  
Tucson AZ 85721  
nitschke@email.arizona.edu